In-group Favoritism and Peer Effects in Wrongful Acquittals: NBA Referees as Judges

Naci Mocan
Louisiana State University and NBER
mocan@lsu.edu

Eric Osborne-Christenson
Pace University
eosbornechristenson@pace.edu

May 2022

We provide the first analysis of racial in-group bias in Type-I and Type-II errors. Using player-referee matched data from NBA games we show that there is no overall racial bias or in-group bias in foul calls made by referees. Similarly, there is no racial bias or in-group bias in Type-I errors (incorrect foul calls). On the other hand, there is significant in-group favoritism in Type-II errors. These are wrongful acquittals where the referee did not blow the whistle although a foul was committed. Although higher error rates during the season lower referees' probability to be selected to officiate games in the playoffs, month-to-month adjustment in behavior takes place only in more visible and consequential Type-I errors. Each game is officiated by a crew of 3 referees. When we analyze peer effects, we find a pattern consistent with the critical mass hypothesis. Black referees' proclivity to make Type-II errors in favor of black players exists unless black referees have two white peers with them on the court. In-group favoritism of white referees emerges when they have two black peers. We provide evidence showing that the results are not attributable to skill differences between referees.

We thank Richard Boylan, Philip Marx, Dan Hamermesh and Murat Iyigun for helpful comments.

In-group Favoritism and Peer Effects in Wrongful Acquittals: NBA Referees as Judges

I. Introduction

Decisionmakers evaluate others based their current, past, or expected future performance, and make decisions based on these evaluations, ranging from hiring, firing and promotion choices to approvals or rejections of loan and credit applications, to decisions to rent an apartment to a prospective tenant. An enormous body of research has focused on the extent to which the race of the individual who is being evaluated has an impact on the judgement of the evaluator, and the outcome. A large segment of this work identified racial disparities which may be attributed to racial bias, although the particular source of the bias is difficult to pin down (statistical discrimination, racial animus etc.).[1]

The analysis of racial bias, however, is incomplete without the accompanying investigation of in-group bias, which is the examination of the extent to which individuals treat those who are members of their in-group differently in comparison to those who are members of the out-group. In-group bias can exist even in the absence of overall racial bias because in-group bias can mask the overall bias. For example, black evaluators may be biased in favor of blacks in comparison to whites; and white evaluators may exhibit favorable in-group bias for whites. In this circumstance, depending on number of black and white evaluators, and depending on the magnitude of the in-group bias in each group, in-group biases may wash out the overall bias

[1] Among the many examples of research regarding the impact of race on the decisionmaker's behavior, see Munnell et al. (1996) on mortgage lending, Alesina and Ferrera (2014) on the administration of the death penalty, Fisman et al. (2020) on credit lending, Bertrand and Mullainathan (2004) on interview callbacks, Giuliano et al. (2009) on hiring decision, Arnold, Dobbie and Yang (2018) on bail decisions, Ayres, Banaji and Jolls (2005) on online auctions.

(Bielen, Marneffe and Mocan 2021, Mocan 2020).[2] Conversely, detection of overall bias in favor of the first group does not imply lack of in-group bias among the members of the second group.[3] Finally, in-group bias does not have to be in favor of the in-group. Put differently, in-group bias can be negative, indicating that decisionmakers are harsher towards members of their in-group (Bielen, Marnefee and Mocan 2021; Bar and Zussman 2020; Depew, Eren and Mocan 2017). Apart from experimental work in laboratory settings (Mussweiler and Ockenfles 2013; Bernhard, Fishbacher and Fehr, 2006; Goette, Huffman and Meier, 2006) there is limited research on in-group bias because it is difficult to find suitable data sets that include detailed information including race, on both the decisionmakers and the individuals who are being evaluated.[4]

A related question is how *errors* in these decisions are related to race. For example, it is important to analyze whether judges make *systematic errors* in their conviction and acquittal decisions and whether these errors are related to the attributes of the defendants. These types of analyses are essential because these Type-I and Type-II errors could be related to the race of

[2] As a stylized example, assume that there are 5 black and 5 white loan officers and each one evaluates 10 white and 10 black applicants who are identical in all aspects other that race. Assume that white loan officers approve white applicants 60 percent of the time, while they approve black applicants with 40 percent probability, and suppose that black loan officers approve black applications with 60 percent probability while they approve white applicants 40 percent of the time. In this case, there is no overall racial bias because of the 100 black applicants 50 are approved, and similarly 50 of the 100 Blacks are approved.

[3] For example, suppose there are 5 black and 5 white loan officers and each one evaluates 10 white and 10 Black applicants who are identical in all aspects other that race. Assume that white loan officers approve white applicants 70 percent of the time, while they approve black applicants with 40 percent probability, and suppose that black loan officers approve black applications with 60 percent probability while they approve white applicants 50 percent of the time. In this case, overall racial bias would emerge because there would be 70 white applicants who are approved in comparison 50 black approvals, even though both black and white loan officers exhibit in-group favoritism.

[4] Empirical studies of in-group bias include research on the decisions of U.S. baseball umpires (Parsons et al. 2010), Israeli judges (Shayo and Zussman2011), Israeli driving test evaluators (Bar and Zussman 2020), juvenile court judges in Louisiana (Depew, Eren and Mocan 2017).

person on which the decision has been made. Furthermore, and more importantly, as in the case of in-group bias, racial bias may exist in Type-I and Type-II errors even though it may not register in the data as overall racial bias. That is, racial bias may fail to reveal itself in visible decisions such as hiring a person, convicting a defendant, approving a loan, letting someone pass a test, and so on, but bias may be hidden in the propensity of the correctness or incorrectness of these decisions.

Despite the importance of analyzing the existence of Type-I and Type-II errors, there has been virtually no empirical research on the subject because of unavailability of data, and there is no work on racial bias in these errors.[5] For example, wrongful convictions in judicial decisions are not possible to analyze systematically because the accuracy of these decisions are unknown (unless some outside body, such as an appellate court, evaluates them to determine errors). Similarly, once a defendant is found not guilty of a crime, and once he is acquitted, it is not possible to determine whether the decision was correct or incorrect.[6]

In this paper we leverage a data set which is ideally suited to investigate these questions. We analyze the decisions of National Basketball Association (NBA) referees, who adjudicate incidents during basketball games. The incidents involve two players: the player who is under scrutiny (the alleged perpetrator who might have committed a foul), and the second player who was the target of the alleged foul. The referee, who observes the incident, makes a decision: he

[5] Chan et al. (2022) analyze pneumonia diagnoses of radiologists. Their data include information on false negative diagnoses (Type II errors), but not false positives. The data do not permit an analysis of in-group bias in any attribute (race, gender, etc.) of radiologists either.

[6] The same problem exists in all other domains where the decision of the evaluators are not securitized. For example, Bar and Zussman (2020) detect in-group ethnic (Arab vs. Jewish) bias and gender bias in the assessments of driving test evaluators, but the authors are not able to analyze Type-I or Type-II errors, let alone whether such errors are related to race/ethnicity or gender-matching between the drives and the evaluator. This is because once a driver passes or fails the driving test, there is no procedure to determine whether the decision of the evaluator was correct or incorrect.

either blows his whistle and declares the action of the first player as a violation of the rules of the game (a foul), or he allows the play to continue, which means that the referee decided that there was no foul in the incident, and that the player was innocent. We have information on the race of the referees and the race of the players, along with a wealth of player, referee and game attributes. This allows us to analyze the existence of racial bias in referee decisions. More specifically, we investigate whether black players are more or less likely to be "convicted" by referees, all else the same. Because we know the race of both the players and the referees, and because we know which referee made a foul call on which player, we also investigate the existence of in-group bias in foul calls.

A novelty of our paper lies in the fact that we can also analyze acquittals. These are situations where the referee, who observed the incident, decided that the action of the "perpetrating" player was legal. As a result, the referee did not blow the whistle. Here too, we investigate the overall racial bias, and the in-group bias. That is, we analyze whether the race of the player has an impact on his propensity of having been acquitted, and whether race-matching between the referee and the accused offending player has an impact on the player being declared not guilty.

Another key innovation of the paper is the ability to investigate Type-I and Type-II errors in these decisions. This is possible because every single incident in the data (during the last two minutes of each game) has been analyzed and evaluated by NBA's League Operations Senior Management Team, which determined whether the referees were correct or incorrect in their decisions. This means that we know if each foul call and each non-call was correct or incorrect. This information provides a rare insight into the race effects in Type-I and Type-II errors.

There is a large literature analyzing how exposure to peers with certain attributes influences an individual's behavior. This research covered a wide range of subjects from the

analysis of peer effects on students (e.g., Sacerdote 2001; Whitmore 2005; Carrell et al. 2013) to worker productivity (Mas and Moretti 2009; Cornelissen et al. 2017), to misconduct in the army (Murphy 2019). Peer effects on judicial decisions have also been analyzed by focusing on how the race or gender composition of a panel of jurors or judges impact the outcomes (Anwar, Bayer and Hjalmarsson 2012, Kastellec 2013, Grossman et al. 2016), or by analyzing how individual judges are impacted by the variation in their peer composition (Eren and Mocan 2021).

Each NBA game is officiated by a crew of three referees. Thus, each referee has two peer referees on the basketball court. Because the race of each referee is known, we are able to investigate whether the decisions of referees and the impact of player race in these decisions are affected by the racial composition of peers. More specifically, we analyze how the race effect in Type-I and Type-II errors changes as the racial composition of the peers of a referee varies.

Previous Research and the Contribution of this Paper

The question of racial in-group bias has been investigated in the context of the NBA games (Price and Wolfers 2010). The authors identified in-group bias in fouls calls using data from the 1991 to 2003 seasons. They, however, did not have access to data on calls made by specific referees. Consequently, the authors employed game-level observations, and analyzed the foul rate of each player in a game as a function of the proportion of white referees on the court. They reported that black players received more foul calls relative to white players when the ratio of white referees officiating the game went up. Subsequent work, analyzing more recent data with the same empirical design, found that this racial in-group bias did not exist in the 2007-2010 seasons (Pope, Price, and Wolfers 2018). As the authors postulated, a possible reason for the latter finding is that the results of the former study have received wide-spread attention in the media. In response, the NBA referees could have adjusted their behavior, which in turn could

have eliminated the in-group bias identified in earlier seasons. Because we have data on each specific call, we conduct the analysis of in-group bias in foul calls using player-referee matched data. Consistent with Pope, Price and Wolfers (2018) we find no evidence of in-group bias or overall bias of foul calls made by referees.

The innovation of our paper is the analysis of non-calls (in addition to foul calls), and the errors in both foul calls and non-calls. This analysis reveals no racial bias in incorrect foul calls. That is, we find no evidence of racial bias in Type-I errors. However, there is significant racial favoritism in incorrect non-calls. These decisions are wrongful acquittals of players who are of the same race as the referee. Following the framework of Chan, Gentzkow and Yu (2021) we provide evidence showing that the results are not driven by skill differences between referees.

When we analyze peer effects, we find that the positive racial bias in wrongful acquittals of black referees persists unless they have two white peers with them on the court is black. Although white referees, as a group, do not exhibit in-group favoritism in these Type-II errors, the bias pops up when white referees have two black peers on the court. This means that being in the numerical minority of the group (being the member of a referee crew, two other members of which are of the other race) alters the behavior of the referees. In Section VIII we provide an explanation for this pattern, which is also reported by previous research.

It is noteworthy that in-group bias does not exist in foul calls; nor does it exist in incorrect foul calls (wrongful convictions), although in-group bias is revealed in non-calls (wrongful acquittals). This finding is illuminating to the extent that incorrect foul calls are more visible and potentially costlier to make for a referee in comparison to incorrect non-calls. These results are important because they demonstrate the nuanced nature of racial bias and they underline the importance of deeper analyses. More specifically, the results show that the lack of racial bias in

one layer of decision-making does not imply the non-existence of bias in other layers.

Referees receive feedback from the NBA about their performance both after each game, and in the form of bi-weekly evaluations. They also receive performance evaluations in mid-season, and at the end of the season. To investigate the extent to which referees adjust their behavior and their error rate during the season in light of the feedback from the NBA, we analyze the error dynamics during the season. We find that the incorrect call rate is negatively impacted by the previous month's incorrect call rate, indicating that referees adjust their rate of Type-I errors. On the other hand, the rate Type-II errors in a given month of the season is independent of the error rate in the previous month, indicating no adjustment dynamics. Relatedly, the rate of Type-II errors is twice as high compared to Type-I errors. These results are meaningful to the extent that incorrect non-calls are arguably less visible to the fans and less consequential in a particular game.

To shed light into potential disincentives of making decision errors and recognizing that officiating a playoff game generates financial benefits to referees, we analyze whether the referees' propensity to obtain an assignment in the playoffs is related to their error rate in foul calls as well as the error rate in non-calls during the season. We find that an increase in the rate of incorrect foul calls (wrongful convictions) during the season reduces the likelihood of having been assigned to a playoff game, whereas an increase in the error rate of incorrect non-calls (wrongful acquittals) has no statistically significant impact, although the point estimates are similar between the two types of incorrect decisions.

In section II we describe the data used in the paper. Section III discusses the process of referee assignment to NBA games. Sections IV and V present the analyses of overall racial bias, and in-group bias, respectively. Section VI unbundles the in-group bias in Type-II errors.

Section VII discusses skill vs. preferences as drivers of the results, and Section VIII presents the results of the peer effects analyses. Section IX contains the analyses of Referees Rewards and Learning, and Section X is the conclusion.

II. Data

The data are compiled from two different sources. The primary data on calls and non-calls of referees are obtained from the National Basketball Association's *Last Two Minute Reports* (LTMR). The NBA began releasing these reports for purposes of transparency starting midway through the 2014/2015 season. Every game, where teams are within three points of each other at any point in the last two minutes of the game, is reviewed by NBA's League Operations Senior Management Team and their findings are released for public scrutiny the following day.[7]

All foul calls as well as non-calls that are "material plays directly related to the outcome of a possession" are evaluated, including defensive fouls, offensive fouls, flagrant fouls and technical fouls.[8] In all cases, the NBA determines the committing player, who was the "perpetrator" and the disadvantaged player who was the target.[9] Importantly, the NBA also determines whether each of these referee decisions was correct or incorrect.

[7] It is worth noting that this office is not related to the Office of Referee Development and Training nor is it related to the Referee Union. Thus, there is no evident incentive to misrepresent the accuracy of referee decisions to make the referees look good.

[8] Other types of rule violations are excluded because they do not involve two players, and usually these violations are too obvious for the referee to make a judgement call. These include travels, double-dribbles, out of bounds, five-second inbound violations, eight-second back court violations, ten-second free-throw violations, twenty-four second shot clock violations, jump balls, kick balls, offensive lane violations, defensive lane violations, offensive goal tending, and defensive goal tending.

[9] The standard for correct and incorrect is defined as follows: "Similar to the NBA's instant replay standards, there must be clear and conclusive video evidence in order to make a determination that a play was incorrectly officiated."

The LTMRs are released by the NBA without the names of the specific referee(s) responsible for each decision.[10] However, the names of the referees who officiated each particular game were provided by the NBA in these LTMRs. The LTMRs also contain the links to video clips of each play (i.e. each incident). Using the videos of the plays, research assistants determined the referees who made the decisions (making a foul call, or not making the call in each incident.)[11] More specifically, all plays on the NBA web site were watched by two different groups of research assistants, who made a note of the referee responsible for each call/no-call. In circumstances where two referees simultaneously made a call on the same incident, they are considered as separate decisions by two referees.

Research assistants were provided with two resources. First, they were given a Referee Picture Book, which comes from NBA.com and contains headshots and jersey numbers of all active referees.[12] Second, each video has an associated Play Comment, which is released by the NBA along with the LTMRs. These comments include a detailed description of each play, the names of the players involved, and a timestamp for when the incident has occurred during the game. The NBA labeled the player who committed the act as the "*Committing Player*" and the player who was the target is named the "*Disadvantaged Player*." For accuracy, the entries of each research assistance group are compared with one another. Only those cases where both groups agreed on the identity of the referee who made the call/non-call are kept as part of the final dataset.

[10] The LTMRs have been employed by other researchers as well (e.g., Gong 2022; Deutscher 2015).

[11] The Research Assistant group consisted of individuals who were students or recent graduates of Louisiana State University.

[12] The Referee Picture Book for the 2021-22 season can be found at the link: 2021-22 NBA Referee Headshots.xlsx

Referee decisions, for which no evaluation was provided by the NBA, cannot be used in the analyses, because it is impossible to determine if these calls/non-calls were correct or incorrect. In addition, because these reports consist exclusively of incidents occurring during the last two minutes, there are a significant amount of (2,911) intentional fouls which are not relevant for our analysis. There are also 1,487 team level infractions, in which there is no identifiable offending player.[13] Some other plays are excluded because either the videos were missing on the NBA web site, or the incidents involved only one player (goal tending, travels, double-dribbles, lane violations, and so on). A few cases are also excluded because there was no unanimous agreement among our research team regarding the referee who made the decision. The final sample includes 15,978 are referee decisions in 1,482 games between the 2015-2019 seasons.

Player and Referee Information

Control variables are obtained from Basketball-Reference.com for the entire sample period. These include player characteristics (for both the *Committing Player* and the *Disadvantaged Player*), such as all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center); player statistics (both for the committing player and the disadvantaged player), including assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). Game attributes include indicators for home game status for the committing player and for whether the game was a playoff contest. Team characteristics include a dummy variable indicating whether a team made the playoffs that season (for both

[13] Team infractions are situation such as the ball going out of bounds after a shot is taken, illegal defense calls, backcourt violations, and so on.

committing and disadvantaged players), and an indicator variable which takes the value of one if the coach of the committing player is Black. Similarly, information on referee age, experience, number of games officiated are also obtained from the Basketball-Reference.com and NBA.com.[14]

The race of the players and the referees are determined by inspection of media sources and the NBA's Referee Picture Book. This was done by the authors and research assistants. There are four female referees in the sample. These four referees made a total of 205 of the 15,978 decisions analyzed in the paper. They are included in the analyses, but dropping them did not alter the results.

Descriptive Statistics

Our analyses are based on 1,482 games played during the 2015-2019 seasons, including the playoff games. Dropping playoff games does not alter the results. As discussed earlier, the NBA's assessment of the accuracy of referee decisions during the last two minutes of each game and the resultant *Last Two Minutes Reports* started in the middle of the 2015 season. These reports cover close games where the point difference between the teams was three or fewer sometime during the last two minutes of the regulation time. There are 15,978 decisions made by 81 unique referees, 43 of whom are black (53 percent). These decisions involved 590 unique players, 460 of whom are black (78 percent).

Figure 1 presents the structure of the data analyzed in the paper. Analyses are conducted at different levels, the most standard of which is the investigation of the determinants of a foul call being made (the box that separates the decisions into "Foul Calls" vs. "Non-Calls"). The

[14] One referee's age was not posted on either site, and was found using his publicly available Florida voting records.

novelty of our paper, however, is the analyses of the actions at the next level. This is the investigation of the determinants of making a wrong foul call, which is equivalent to a wrongful conviction because in this case the player is found guilty of an act that he did not commit. Similarly, we analyze the determinants of the referee being wrong, conditional on the decision of not making a foul call. This is the investigation of incorrect non-calls, or wrongful acquittals, where the player committed a foul but the referee decided that the act was legal.

Tables 1 and 2 present the descriptive statistics of selected variables. Table 1 summarizes the referee decisions by the race of the *Committing Players*, whose actions were evaluated by referees. Eighty percent of the decisions are made on Black players (12,891 decisions out of a total of 15,978). This is consistent with the fact that 78 percent of players are black. As shown in the middle section of Panel A of Table 3,116 of these decisions were foul calls (i.e. convictions). The remaining 12,817 decisions were no-calls (acquittals), where the referee has decided that there was no infraction on the part of the committing player.

The middle section of Panel A of Table 1 shows that of the 2,618 fouls calls received by black players, 5 percent were incorrect. These are Type-I errors made on black players. Similarly, about 6 percent of the fouls called on white players are incorrect. The bottom section of Panel A reveals that referees "made" 12,817 no-calls. These are the cases where the committing player is found "not guilty" and no whistle is blown. About 8 percent of these "acquittal" decisions were incorrect for both black and white players. Figure 1 presents these different layers that will be analyzed in the paper.

Panel B of Table 1 presents information pertaining to the committing players who were the subjects of these referee decisions. *Win for Committing Player* is a dummy that takes the value of 1 if the team of the committing player won that game. *Playoff Game* signifies if the

game in which the decision took place was a playoff contest. The last two variable in Panel B are indictors for whether the coach of the committing player was black and whether the committing player's team was played on their home court.

Table 2 presents selected player attributes. This information, as it was employed in the regressions, pertains to season averages of players. That is, each committing player contributes one observation per season. There are differences between white and black players in terms of their season performance, some of which are related to the positions they are playing. For example, white players have higher averages in both offensive and defensive fouls, but they score fewer points and make fewer attempts at two-point and three-point shots.[15] As Panel B demonstrates, white players are also more likely to play Center, while black players are more likely to play Guard.

We have the same information about player attributes for "disadvantaged players." These are players who were involved in the incident with "committing players." In case of a foul call, a disadvantaged player is the one who was fouled. In case of a no-call, the disadvantaged player was involved in the play as the potential "victim," but the referee decided that he was not fouled by the committing player. Descriptive statistics on Disadvantaged Players are presented in Appendix Table A.1.

There are 81 referees in the sample, 43 of whom are black. The mean age for black referees is 47.85 and average white referee is 44 years old. The average experience for black and white referees is 13.5 years, and 12.6 years, respectively. On average black referees officiate 59 games per season, while white referees officiate 61 games. The age difference between black and white referees is addresses later in the paper.

[15] These statistics are adjusted to reflect 48 minute performance.

III. Referee Assignment to Games

Since the 1988-1989 season, each NBA game has been officiated by three referees, called the Crew Chief, Referee, and Umpire. The Crew Chief is the leader, and is responsible for making final determinations on any disputed call. S/he, along with the Referee, handles communication between other officials and the teams. The Umpire has no responsibilities outside accurately making calls. However, in terms of performance directly related to game play, the responsibilities and expectations of each official are the same.

Referee crews are chosen from a pool of sixty to seventy; and once selected, the crew stays together for a few games before each member being re-assigned. The exact position to which each referee is assigned is based on experience and performance. To reduce any potential bias, the NBA limits each referee to nine games per team and prohibits more than one assignment per city within the same two-week window. NBA also limits each referee to 75 total games and tries to balance experience within each three-person team. These restrictions are not particularly problematic for the purposes of this study, as they are not related to the race of the referees or to the racial composition of the teams. Thus, while referee assignment to the games is not random, it is random with respect to race.

Nevertheless, we investigate whether referee race is related to teams' racial composition. There are 1,482 games in the data. Table 3A displays the distribution of the percentage of minutes played by black players with zero white referees, and with 0, 1, 2, or 3 white referees in these games. The unit of observation behind Table 3A is a team-game. The play time of black players on the basketball court is uncorrelated with the racial composition of the refereeing crew. For example, in 2019, when games were officiated with three black referees, about 82 percent of minutes were played by black players. Games with 1 white or 2 black referees had 84 percent of

the game time played by black players. The percentage of minutes played by black players was 82 and 83 when there were 2 white referees and 3 white referees on the court, respectively. As displayed in Columns (VI) and (VII), regressions of average minutes played by black players on the number of white referees confirm the lack of a relationship between the two variables.

Table 3B presents the results of a similar exercise where the average number of black starting players per team-and-game is displayed by season and by the racial composition of the refereeing crew. For example, in 2019 there were about 4.2 starting black players on each team when there the refereeing crew was all black; and as shown in Columns (II) and (III), the same was true when the crew included one or two white referees. When the crew consisted of three white referees, the average number of starting black players was about 4.1. The number of black players starting a game is not related to the racial composition of the refereeing crew in other seasons either.

The make-up of the refereeing crew is known to teams by 9:00 pm EST the night before each game. Thus, teams may adjust the racial composition of their players for the game, based on the racial composition of the refereeing crew. As Table 3B reveals, however, the number of starting players is uncorrelated with the racial composition of the referee crew, indicating that teams do not adjust their racial make-up (at least in the beginning of the game) according to the racial make-up of the referee crew.

Panel A of Table 4 presents information on the distribution of decisions made on black and white players by different members of the refereeing crew. About 34 percent of the 12,891 decisions (calls and no-calls) on black players were made by the Chiefs of the refereeing crews, 33 percent are made by the umpires, and 33 percent are made by the referees. The distribution is similar for the decisions made on white players. Thus, there is no evidence that a particular

member of the refereeing crew takes more initiative or is more active in making decisions during the game.

Panel A of Table 4 also shows that 47 percent of decisions on black players are made when the refereeing crew had a black Chief, and similarly, 49 percent of the decisions on white players were made when the Chief of the crew was black. Finally, the last row of Panel A shows that when a decision was made (a crew member blowing his whistle, or letting the play continue) on black players, the average number of black members of the crew was 1.37. When a decision was made on white players, there was, on average, 1.41 black officials on the basketball court.

Panel B of Table 4 repeats this exercise using the sample of foul calls made on black and white players. These are decisions where the committing player was found guilty and a whistle has been blown to stop the game. Panel C displays information about no-calls where the official, who was watching the incident, did not make a call (declared the committing player not guilty). Similar to Panel A, in none of these panels is there evidence that a particular member of the officiating crew (Chief, Umpire or Referee) is more likely to have made a decision; and there is no evidence that calls and no-calls are matched by player and referee race.

In summary, even though the official titles of the three decision-makers on the basketball court are different (the Crew Chief, the Referee and the Umpire), there is no difference in their behavior related to their titles. Thus, we refer to them as referees in the rest of the paper.

Even though neither the racial composition of the starting lineup of the teams nor the average number of minutes played by black players are related to the racial make-up of the refereeing crew, it could be the case that black or white referees seek back or white players during the game to make calls. To investigate this possibility, we ran regressions where the race of the referee in each decision is regressed on the race of the committing player. This exercise

investigates the randomness between the race of the referee who made the decision and the race of the player who was the subject of the decision. Table 5 displays the results. Regardless of the controls included in the specification, there is no association between the race of the referee who made the call/no-call, and the race of the committing player. If we run the same analysis using the sample of foul calls (N=3,161) we obtain the same result. Even in the model with no control variables the coefficient of *Committing Player Black* is very small and statistically insignificant (-0.001, se=0.024). The same is true in the sample of no-calls (N=12,817). When we run the simple regression displayed in Column (1) of Table 5 in the sample of no-calls, the coefficient of *Committing Player Black* is -0.016 (se=0.014).[16] Thus, these analyses indicate that there is no systematic matching between the race of the referee who makes a decision on players, and the race of the players whose actions were evaluated.

IV. Analysis of Overall Racial Bias

To investigate the existence of overall racial bias in foul calls, we estimate models as depicted by Equation (1) below. Using the same framework, we also investigate racial bias in referees' propensity to make an incorrect foul call (Type-1 error), and their propensity to make an incorrect no-call (Type-II error).

$$(1)\quad Call_{prgs} = \alpha + \beta\, \text{Committing Player Black}_{p} + \text{Referee FE}\ \Gamma + \mathbf{C}_{ps}\boldsymbol{\Omega} + \mathbf{S}_{pgs}\boldsymbol{\Psi} + \mathbf{G}_{gs}\Phi + \lambda_g + \xi_s + \varepsilon_{prgs}$$

where *Committed Player Black* is an indicator variable which takes the value of 1 if player *(p),* who committed the act, is black. $Call_{pr}$ represents the conviction-acquittal decision

[16] In the interest of space, these regressions are not reported.

made on committing player p. Specifically, *Call*=1 if the referee (*r)* blew the whistle and called a foul on player *p*, and it is zero if the action of the player was determined by the referee to be a permissible act according to the rules of the game.

Each player *p* can have multiple interactions with referee *r* in a particular game (*g*) of season (*s*). The model adjusts for a large set of player attributes (for both the offending player and the disadvantaged player) as well as game and team attributes, which are combined under four categories: *Player Characteristics*, represented by vector **S** in Equation (1), which vary between seasons. Examples include the percentage of games started in that season, all-star status, average minutes played per game, position (guard, center etc.), and so on. Vector **S** also includes 11 specific performance statistics of player *p* in game *g*, and season s, ranging from defensive rebounds and offensive rebounds to three point shot attempts and turnovers. The full list of these variables is provided in Panel A of Table 2.

Note that each incident involves the committing player, as well as a player from the opposite team who is involved in the incident. Following the terminology of the NBA, we call this player "Disadvantaged Player." We also know the identity of the disadvantaged player. Thus, vector S includes attributes of both the "Committing Player (whose action has been evaluated by the referee), and also the "Disadvantaged Player" (who faced the alleged infraction of the committing player.)

G stands for game characteristics such as the home game status of the committing player, whether the game was a playoff contest, if the team of the committing player won, and the total points scoring in the game. Team controls, **T**, include information about whether the team of the committing player made the playoffs that season, and an indicator which takes the value of 1 if the coach of the committing player is black. Specifications also include Referee fixed-effects, as

well as game fixed effects (λ_g) or season fixed effects (ξ_s). Standard errors are clustered by referee. The coefficient β signifies the magnitude and the direction of the racial bias.

Recall that the NBA has assessed whether the decisions made by the referees were correct or incorrect. This information allows us to analyze potential racial bias in Type I vs. Type II errors. These specifications use all foul calls made by the referees (i.e. all convictions) and analyze whether the foul call was correct or incorrect. Similarly, we analyze all instances in which there was an actionable incident between two players (as determined by the NBA), but where the referee decided that the action of the committing player was not a foul. These models use the same specification as depicted by Equation (1), but the dependent variables are indicators of an incorrect call being made (*Incorrect Call*), or an indicator for an *Incorrect No-Call*, respectively. These specifications allow us to analyze the extent to which the race of the committing player has an impact on the probability of wrongful conviction by the referee when the action was not a foul, and on the probability of a wrongful acquittal where the committing player committed a foul, which was not called by the referee.

The results of these analyses are displayed in Table 6. We present the estimated coefficients of a few selected variables. Other control variables included in the models are described at the bottom of the table and in its footnote. Columns (I) and (II) present the results of the analysis of foul calls. There are 15,978 observations in this sample, indicating that the referees made 15,978 decisions on whether to incident was a foul or it was a permissible act within the rules of the game. The mean of the dependent variable is 0.20, indicating that 20 percent of these incidents ended up being judged as a foul.

Columns (III) and (IV) display regression results where the dependent variable takes the value of 1 if the foul decision was incorrect, and zero if the decision was correct. Put differently,

these models investigate whether the race of the committing player has an impact on the probability of the referee making an incorrect foul call (i.e. a wrongful conviction.) In contrast, columns (V) and (VI) present the results of the analysis of Type II errors. Here, the sample includes all acquittal decisions, where the referee decided not to make a foul call. Eighty percent of all incidents were judged to be as legal, leading to 12,817 no-calls (See Figure 1). The dependent variable in columns (V) and (VI) takes the value of 1 if the non-call was incorrect.

Columns (I) and (II) of Table 6 reveal that the coefficient of *Committing Player Black* is small and never statistically different from zero in any regression. This means that the probability of receiving a foul call does not depend on the race of the player who is evaluated by the referee. Similarly, columns (III) to (VI) show that the race of the committing player has no impact on the accuracy of the foul call, or the accuracy of the non-call.[17] Thus, Table 6 reveals no overall racial bias in foul calls and no bias in Type-I or Type-II errors.

V. In-Group Bias

The lack of overall racial bias in referees' decisions, analyzed using Equation (1), does not rule out more subtle or more multi-layered racial biases. For example, assume that black referees exhibit positive bias towards black players. That is, presume that black referees are more lenient towards black players (or harsher towards white players). Also assume that white referees are favorably biased towards white players. In this scenario, depending upon the magnitude of these in-group biases and depending on the distribution of decisions between white and black referees, the overall racial bias could be zero even if there is in-group bias in both

[17] The estimated coefficient of *Committing Player is Black* is borderline significant at the 10 percent level in column (V), but the coefficient in model with game fixed effects in column (VI) is not different from zero.

races. Put differently, such in-group bias may mask an overall bias.

To investigate in-group bias in referee decisions we estimate models in the same form as previous work (Bielen, Marneffe and Mocan, 2021; Depew, Eren and Mocan 2017; Gazal-Ayal, and Sulizeanu-Kenan 2010; Shayo and and Zussman 2011) as depicted by Equation (2) below

(2) $\text{Call}_{prgs} = \gamma_1 + \delta_1 \text{Black Referee}_r \times \text{Committing Player Black}_p +$

δ_2 Committing Player Black_p + Referee FEs $\boldsymbol{\beta} + + \mathbf{S}_{ps}\boldsymbol{\Theta} + \mathbf{G}_{gs}\Lambda + (\kappa_g + \pi_s) + u_{prgs}$

Equation (2) is also estimated with the inclusion of committing player fixed effects (instead of *Committing Player Black* dummy), which does not alter the inference. In Equation (2) δ_1 is the in-group bias coefficient. It signifies the differential decisions of white versus black referees in their evaluation of black versus white committing players. The link between in-group bias and overall bias is evident as the overall bias (β in Equation 1) can be calculated from the estimates of Equation (2) as $\delta_2+(1-\rho)\,\delta_1$, where ρ stands for the proportion of decisions made by white referees.

Table 7A displays the results of estimating the in-group bias regression where the outcome is the foul call. The estimated coefficient of the interaction term between *Black Referee* and *Committing Player Black* reveals whether referees of a given race (black or white) are more or less lenient towards players of their own race when making a foul call. The results indicate no in-group bias in foul calls.

To put this result in perspective, note that Price and Wolfers (2010) identified in-group bias in fouls calls in NBA games analyzed for the 1991 to 2003 seasons. The authors did not have access to data on calls made by specific referees as we do in this paper, but they analyzed

the foul rate of each player as a function of the proportion of white referees of the referee crew (0, 1/3, 2/3, or 1). They reported that black players received more fouls relative to white players when the ratio of white referees officiating the game went up (Price and Wolfer 2010). Subsequent work, analyzing more recent data, found that this racial in-group did not exist in the 2007-2010 seasons (Pope, Price, and Wolfers 2018), possibly because of the media attention on the former prompted the referees to adjust their behavior. The result of Table 7A confirm the finding of Pope, Price and Wolfers (2018), and indicate no in-group bias in fouls calls.

Type-I and Type-II Errors

Bias, if detected, would reveal that members of one group are treated differently in comparison to the members of another group. For example, all else the same, if wage offers are lower for people in group A in comparison to wages offered to group B, this wage gap may reflect bias, but it does not tell us whether group A is underpaid or group B is overpaid. As another example, assume that judges assign longer sentences to the members of group A in comparison to those of group B. This relative harshness in sentence length does not provide information as to whether group A receives the appropriate sentence and group B is over-punished, or whether group B receives the appropriate sentence, but group A is being treated leniently. This is because in these cases it is not possible to determine what the correct or optimal benchmark is to which individual outcomes can be compared. In contrast, in our study we have information about whether each decision of each referee was correct or incorrect, which allows us to analyze the *errors* in these decisions.

Table 7B reports the results of the in-group bias analyses where the outcome is referees' propensity to make an incorrect foul call (Type-I error). The estimated coefficients are never

statistically different from zero regardless of the specification, indicating that referees' propensity to make incorrect foul calls (wrongful convictions) do not depend on race-matching between the referee and the players who received the foul call.

The regression results displayed in Table 7C pertain to in-group bias in Type-II errors. The analysis sample is all no-calls, and the dependent variable that takes the value of 1 if the no-call was incorrect. These results portray a starkly different picture, and reveal in-group bias in these "wrongful acquittals." Referees are about 2.2 percentage points more likely not to call a foul on a player who committed the foul if that player is of the same race of the referee. This is a 27 percent impact from the baseline of 8.5 percent.[18]

The models control for referee fixed effects which account for different styles of refereeing, and the difference between referees regarding their tolerance of rough play. Nevertheless, to account for the possibility that the leniency/harshness of black referees might differ from the leniency/harshness of white referees, which in turn may impact players with different attributes differentially, we interact the indicator variable for black referee with all player characteristics and player statistics displayed in Table 2. The results, presented in Appendix Table A.3A, A.3B, and A.3C, remained the same.

These results have potentially important implications as they demonstrate that analyzing one dimension of the decisionmaker's judgment may not reveal the full picture. NBA referees exhibit no overall racial bias in how they make foul calls. Similarly, there is no in-group bias in their foul calls or in their Type-I errors. But significant in-group favoritism exists in Type-II errors, which reflects a more subtle and difficult-to-observe bias in the decision-making process.

[18] The results also reveal that players are more likely to get away with a foul if they are playing on their home court, and the same is true for a player who became All Star in that season.

VI. Unbundling the In-group Bias in Type-II Errors

In-group bias identified in Type-II errors may stem from the behavior of both white and black referees, or it could be driven by one of these groups. It is also possible for one group of referees (white or black) to exhibit strongly positive in-group bias (favoring players of the same race and therefore being less likely to call a foul on them), and the other group to exhibit negative in-group bias (disfavoring the players of the same race and being more likely to call a foul on them). All of these scenarios would lead to the detection of in-group bias in no-calls (Depew, Eren and Mocan, 2007; Bielen, Marneffe and Mocan, 2021).

To see this point more clearly, consider estimating Equation (3) using the no-call decisions of white referees

(3) Incorrect No-Call $_{prgs} = \tau + \mu_1$ Committing Player White$_r$ + All Controls $+\upsilon_{prgs}$

And consider Equation (4), which is estimated using the non-call decisions made by black referees

(4) Incorrect No-Call $_{prgs} = \tau + \mu_2$ Committing Player Black$_r$ + All Controls $+\omega_{prgs}$

Equation (3) is estimated using the decisions of white referees. Thus, μ_1 identifies the differential treatment of white players by white referees, which is equivalent to $(-\delta_2)$ from Equation (2). Equation (4) uses the sample of decisions made by black referees. Therefore, μ_2 in Equation (4) represents the differential treatment of black players by black referees, which is captured by $(\delta_1+\delta_2)$ in Equation (2). Thus $\mu_1 + \mu_2 = (-\delta_2) + (\delta_1 + \delta_2) = \delta_1$. This means that the in-group bias coefficient δ_1, which is estimated from Equation (2), can be decomposed as $\mu_1 + \mu_2$.

It is appropriate to divide the sample of non-calls by referee race (as shown in Equations 3 and 4) under the assumption of no selection by referee race, which is supported conceptually in our context, and confirmed empirically.

The result of estimating Equations (3) and (4) are displayed in columns (I)-(II) and (III)-(IV) of Table 8, respectively. Because the sample is divided by the race of the referees and because we have 38 black and 41 white referees, we report the p-values associated with bootstrapped standard errors in {curly brackets}. The estimated coefficient of *Committing Player White* is very small, and it is statistically indistinguishable from zero in the sample of white referee decisions in columns (I) and (II). This indicates that white referees' propensity to make an incorrect non-call (Type II error) is not impacted by the race of the player. On the other hand, columns (III) and (IV) reveal that black referees are 2.4 percentage points more likely not to call a foul (wrongfully acquitting the committing player) if the committing player is black. Thus, Table 8 indicates that in-group bias in incorrect non-calls is pronounced in the decisions of black referees, although the analysis of peer effects (presented below), reveals that this in-group bias also shows up in white referees' decisions.[19]

The theoretical framework that produces decision errors needs to consider the costs and benefits of these errors to the decision maker. The cost of call inaccuracy (Type-I errors) and the cost of missing a foul (Type-II errors) would be reflected in performance evaluation of the referees by the NBA. As mentioned earlier, the NBA evaluates referees regularly (after each game, by bi-weekly reports, in mid-season, and so on). Type-I and Type-II errors of referees would be associated with financial costs as they impact referee assignments to future games and to the playoffs. We provide evidence on this point in Section IX below. On the other hand, if

[19] As displayed in Table 7B, there is no in-group bias in Type-I errors (incorrect calls). To make sure that potentially nuanced race-specific biases in Type-I errors are not masked by the analysis of overall Type-I errors reported in Table 7B, we performed that analysis by the race of the referees. Put differently, we conducted the analysis reported in Table 8 for Type-I errors. The results, reported In Appendix Table A.4, show that there is no in-group bias in Incorrect Calls within white or black referees, confirming that there is no racial bias in Type-I errors.

Type-I and Type-II errors provide an advantage to a group of players with whom the referee associates himself, these errors would provide psychic benefits to the referee.[20]

VII. Referee Skills vs. Preferences

Chan, Gentzkow and Yu (2022) develop a framework where differential decisions of agents reflect variations in both their preferences and skills. They employ data on radiologists' pneumonia diagnoses, where false negatives (miss rates) are recorded, but false positives (where the patient is incorrectly diagnosed with the disease) are not observable. The authors propose a methodology which allows for the determination of whether the variation in diagnosis rates are attributable to variation in doctors' skills. They find that radiologist who diagnose at higher rates have higher miss rates (false negatives). This indicates variation in skills between the radiologists.[21] The authors also show that the true positive rate and the false positive rate across radiologists are negatively correlated, which is also an indication of radiologists differing in their skills, but not preferences.

We apply the same tools to our data. We find that the call rates of referee (the rate at which referees blow a whistle, given the incidents they observe) is negatively but not significantly related to the rate of Type-II errors (the false negative rate). Specifically, a regression of the Type-II error rate of referees on their rate of foul calls reveals a negative relationship (coefficient=-0.10, se=0.04). This means that more trigger-happy referees, who

[20] For a detailed discussion of a framework regarding the decisionmaker's optimization process see Canay et al. (2020) and Arnold et al. (2018).

[21] This is because it reflects the existence of (i) radiologists who have high rates of a positive diagnosis, and when they provide negative diagnoses, their error rate is high. (ii) radiologists who are less likely to diagnose a patient with the disease and when they provide a negative diagnosis, their error rate is low. Clearly group (i) is less skilled than group (ii).

make fouls calls at a higher rate, are not more likely to make errors in their non-calls. This provides evidence against skill variation across referees.

Second, we calculate for each referee their true positive rate (TPR) and the false positive rate (FPR)[22] Both the TPR and FPR range from 0 to 1, and the concave curve in this space represents the referee's "technology" frontier as it depicts the tradeoff between TPR and (1-FPR)[23]. Put differently, TPR and (1-TPR) can be thought of as the two elements of a production possibilities frontier; and the referee chooses an optimal point on this frontier based on his preferences. (See Figures 1 and 2, and the related discussion of Chan, Gentzkow and Yu 2022). If referees varied in their preferences but not skills, they would operate on different locations of the same "technology frontier." In equilibrium this would be revealed by a positive correlation between TPR and FPR pairs across referees. If, on the other hand, referees differ in their skills, less skilled referees' "production frontiers" would live beneath of those with higher skills, and this would be reflected by a negative correlation between TPR and FPR across referees.

Because the components that are needed to compute TPR and FPR for each referee are available in our data, the correlation between TPR and FPR can be calculated. A regression of the former on the latter produces a coefficient of 0.195 (se=0.364). We obtain the same conclusions in both exercises (in the call rate - Type-II error rate space; or in the TPR-FPR space) regardless of whether we aggregate all decisions of each referee in all seasons to one point, or when we analyze referee-by season observations. The same inference is obtained when

[22] TPR=TP/(TP+FN), FPR=FP/(FP+TN) and TP: True Positive rate (the rate of correct calls), FP: False Positive rate (the rate of incorrect calls- Type I error), FN: False Negative rate (the rate of incorrect non-calls- Type-II errors), and TN: True negative rate (the rate of current non-calls). FP+TP+FP+FN=1

[23] The "ROC" (Receiver Operating Characteristics) curve in the terminology of Chan, Gentzkow and Yu (2022)

we conduct these analyses for black or for white referees separately. The details are provided in Appendix Table A.2 and Appendix Figures A.F1-A.F6 These results are consistent with the hypothesis that referee skills are homogenous, and that referees differ in their preferences and not skills. Thus, the results of our paper are not attributable to skill differences between referees.

VIII. Peer Effects

In this section we analyze the extent to which the racial in-group bias in Type-II errors of black referees is impacted by the racial composition of these referees' peers. Each referee has two peers on the court. These peers could be both white (W, W), both black (B, B), or one white and one black (W, B). We run the same regressions displayed in Table 8, but we partition the sample by the peer composition of the referee who made the no-call decision. Specifically, we divide this sample into four groups: (i) Type-II errors of referees when they had two white peers on the court, (W,W); (ii) Type-II errors when the referee who is responsible for the error had one white and one black peer (W,B); (iii) Type-II errors when the referee had at least one black peer (W,B) or (B,B); (iv) Type-II errors when the referee had two black peers (B,B).

The results for black referees are presented in Table 9A. Bootstrapped p-values are reported in curly brackets. Column (I) shows that the propensity of black referees to make a Type-II error does not depend on the race of the player who committed the foul when the decision-making black referee has two white peers (W, W) with him on the court. On the other hand, Column (II) reveals that black referees are 3.5 percentage points more likely to wrongfully acquit a player (i.e. not to call the foul) if that player is black and if the black referee has one black peer on the court. The mean of the dependent variable is 0.097, indicating that 9.7 percent of the no-calls made by black referees when they had one white and one black peer on the court

was incorrect. This in turn implies that in these cases black players are about 32 percent more likely to get wrongfully acquitted by black referees.

Column (III) uses the sample of incorrect no-calls by black referees when these black referees had *at least one* black peer on the court. In this sample, too, the coefficient of *Committing Player Black* is positive and highly statistically significant. That is, if black referees have at least one black peer with them on the court, their propensity to incorrectly acquit black players is 3.1 percentage points higher in comparison to the same propensity when the committing player is white. Column (IV) displays the results of the incorrect no-calls regression of black referees when they have two black peers. The sample size goes down substantially, but the point estimate remains the same and it is still borderline significant with a p-value of 0.11. Thus, the results of Tables 8 and 9A reveal that black referees exhibit positive in-group bias in Type-II errors, and that the bias is eliminated only if black referees have two white peers on the court.

Columns (I) and (II) of Table 8 revealed that white referees, on average, exhibit no racial bias in their Type-II errors. That is, the race of the committing player has no impact of a white referee's propensity to make an incorrect no-call. Nevertheless, we also analyze peer effects in white referees' incorrect non-call decisions. The results are reported in Table 9B. If white referees have two white peers (column I), or if they have mixed-race peers (column II), their propensity not to blow the whistle is not impacted by the race of the player who committed the foul. Similarly, there is no impact on white referees' proclivity for Type-II errors as long as one of their peers is white (column III). On the other hand, as shown in column (IV), if white referees have two black peers on the court, they are 5 percentage points more likely not to call a foul on a white committing player.

That we did not identify a statistically significant overall in-group bias in Type-II errors of white referees (columns I and II of Table 8) is consistent with the results of Table 9B. This is because the in-group bias in white referees' Type-II errors appears if white referees have two black peers on the court (column IV of Table 9B). A white referee has two black peers during only 21.7 percent of his decisions (1,496 of 6,891 decisions). Thus, the in-group bias, triggered by the existence of two black peers, is not large enough to register in the entire set of white referee decisions.

The upshot of Tables 9A and 9B is that black referees exhibit positive in-group bias in incorrect no-calls (by not blowing the whistle on a black player when the player commits a foul) *unless* both of the peer referees on the court are white. White referees do the same for a white player (incorrectly acquitting a white player when he committed a foul) *if* white referees have two black peers on the court. This pattern consistent with the "critical mass" hypothesis, which postulates that people who are in the numerical minority of a group adopt the behavioral norms of the majority, but when they are in numerical majority they start making decisions that reflect their true tendencies. This behavioral pattern has been detected for females on corporate boards, in politics and in science and in judicial decisions (Eren and Mocan 2021; Etzkowitz et al. 1994; Dahlerup 1988; Joecks et al. 2013; Kanter 1977).

In our case, when black referees are in the numerical majority on the court (which is the case when a black referee has one or two black peers) they are more likely not to call a foul when the offending player is also black. When black referees are in numerical minority on the court (when the other two members of the crew are both white) in-group favoritism disappears. The pattern exists for white referees as well. When white referees are in the numerical minority (with two black peers) they conform to the behavior of black referees and exhibit in-group

favoritism in Type-II errors. When white referees are in the majority on the court, their in-group favoritism disappears.

Although the results are robust regardless of how the models are estimated, it can still be argued that black and white referees differ in some dimensions which may be the driving the results. For example, as mentioned early in the paper, black referees are about four years older and they have about one more year of experience. To address this concern, we interacted the age and experience of the referee with the *Committing Player Black* indicator. We also followed Price and Wolfers (2010) and obtained information on each referees' state of birth, created an indicator which takes the value of 1 if they were born in the South, and interacted it with the *Committing Player Black* dummy.[24] Adding these additional variables to the models did not alter the result meaningfully.[25] These results are provided in Appendix Tables A.5- A.8B

Visibility

Parsons et al. (2010) detect in-group bias in the decisions Major League Baseball home-plate umpires, and find that umpires favor pitchers of their own race/ethnicity. Specifically, umpires are more likely to call strikes (a successful play for the pitcher) if the pitcher belongs to their race/ethnicity in-group of the umpire. The authors also find that the bias disappears if the decisions of umpires are scrutinized electronically or watched by a large crowd of spectators. Archsmith et al. (2021) also analyze Major League Baseball umpires and report that they apply

[24] Thirty-five of the 81 referees were both in the South. Southern states are Alabama, Arkansas, Delaware, the District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. These are states south of the Mason and Dixon Line, the Ohio River, and the 36°30′ parallel.

[25] In-group favoritism for white referees in the presence of two black peers became not distinguishable from zero in table A.8B.

greater effort in higher-stake decisions.[26]

To investigate this dimension of decision-making, we can, in principle, estimate the same models reported in the paper using the sample of playoff games, under the assumption that those games are watched more intensively. This is, however, not feasible because doing so reduces our sample sizes dramatically, which prohibits reliable inference. When we drop the playoff games from the analysis sample and run the models using the games of the regular season, the obtained very similar point estimates to those reported in Tables 7A-9. This is likely because the non-playoff games in our analysis are also high-stakes contests. Specifically, the data analyzed in the paper pertain to incidents during the last two minutes of games where the teams were within 3-point from each other during any point in this time interval.

IX. Reward for Performance, and Learning

As mentioned earlier, the NBA evaluates referees' performance after each game. These game day observations are given to referees in bi-weekly performance evaluations. Referees also receive a longer mid-season evaluation from coaches and an even longer end of season review from the league's referee operations staff, using input from coaches and data on call accuracy. At the end of the season 36 referees are selected to officiate games in the playoffs. These referees receive additional compensation.[27]

In this section we analyze whether Type-I and Type-II errors made by the referees during

[26] Other examples of analyses of decisionmakers and the factors that impact their decisions in professional sports include Garicano, Palacios-Huerta and Prendergast (2005), and Rickman and Witt (2008).

[27] The referee collective bargaining agreement is not publicly available, but it's speculated that referees receive an additional $800 to $5,000 per game, on top of their salary, depending on rank and the round of the playoff game. https://www.888sport.com/blog/nba-referee-salary#:~:text=These%20referees%20are%20then%20eligible,expenses%2C%20insurance%20and%20retirement%20plans.

the season influence their chances for being selected for the playoff games at the end of the season. If referee performance during the season is a key determinant of whether they are assigned to playoff games, their error rate during the season should negatively impact their propensity to be rewarded with playoff games at the end of that season.

To investigate the link between the rate of wrong referee decisions during the season and referees' propensity to be assigned to the playoffs, we run following regression

(5) $$\textit{Officiated in Playoffs}_{rs} = \omega + \psi_1 \textit{ Incorrect Call Rate}_{rs} + \psi_2 \textit{ Incorrect No-Call Rate}_{rs} + \\ + \text{Controls} + \nu_{rs}$$

where *Officiated Playoffs* takes the value of one if referee *r* was assigned by the NBA to the playoffs after season *s*. *Incorrect Call Rate* is the percentage of incorrect foul calls made by the referee during the last two minutes of close games of the season (i.e. not including the playoff games), and *Incorrect Non-Call Rate* is the percentage of incorrect non-calls of the referee during the season. Depending on specification, *Controls* include season fixed effects as well as referee characteristics such as years of experience, race, or referee fixed effects. Standard errors are clustered at the referee level. Clustering them by referee-season did not alter the inference.

The first column of Table 10 presents the results of the specification where the probability of officiating a playoff game is explained by referee's race and experience. There are 309 referee-season observations and the mean of the dependent variable is 0.57. Black referees are no more or no less likely to officiate in the playoffs. Referee experience has a strong positive impact on the probability of having been assigned to the playoffs. Column (II) reports the result of the model which includes the percentage of incorrect calls (Type-I errors) as well as the percentage of incorrect non-calls (Type-II errors) made by the referees, and shows that both types of errors negatively impact the propensity to have as assignment in the playoffs.

To investigate the potential path-dependence in referee assignment to playoffs, we report a model where the probability of refereeing in the playoffs in a given season is explained by having been refereed in previous year's playoffs. As shown in column (III) there is strong path dependence. If a referee participated in the playoffs in a given season, he is 69 percentage points more likely to participate next season as well.

It is plausible to assume that there are unobservable referee attributes which would impact their chances of assignment to the playoffs. Temperament, the ability to deal with star players on the basketball court, the ability to stay calm under pressure could be some of these attributes which would impact the NBA's decision to assign a referee to a playoff game. To account for these unobservables, the model in column (IV) includes referee fixed effects in addition to lagged playoff participation. Doing so reduces the coefficient of lagged playoff participation and eliminates its statistical significance.

Column (V) presents our preferred specification which adds the percentage of incorrect calls as well as the percentage of incorrect no-calls to the model, along with referee fixed effects and lagged playoff participation. The coefficient of the incorrect call rate is negative and statistically significant, indicating that a higher Type-I error rate during the season lowers referees' chances of being selected for playoff games. This result is meaningful to the extent that incorrect foul calls are visible both because the game stops after a foul call which provides more time for the audience to scrutinize the call, and also because an incorrect foul call has potentially significant implications as they unjustly contribute to the foul count of players who have to leave the game after committing six fouls. Because of these visibility reasons, referees are penalized for their Type-I errors. It is interesting to note that even though the coefficient of the rate of incorrect non-call rate is smaller than that of the incorrect call rate, and although it is not

statistically different from zero, the magnitude (-0.502) is non-trivial, suggesting that a higher rate of Type-II errors may also be detrimental to referees' selection for playoff games. Still, referees have significantly higher Type-II error rates, perhaps because incorrect non-calls may receive less attention from the public and from the media compared to scrutiny of incorrect calls. The results in Table 10 indicate that if a referee increased his error rate in incorrect fouls calls to the level of his error rate of incorrect non-calls (i.e. from 0.055 to 0.100), his probability of participating in the playoffs would go down by about 3 percentage points, or by 5 percent.

Do Referees Adjust their Behavior?

Referee crews are highly trained, but they area also subject to considerable supervision and oversight. Before each game, they meet with an observer from the referee operations staff and discuss potential issues related to the upcoming match. After the game concludes, they reconvene, review video, and discuss performance. The observer then reports to the director of officiating, who issues bi-weekly performance evaluations for each referee. Coaches also provide feedback each game and more comprehensively in a mid-season report. Finally, the league's referee operations staff, using input from coaches and data on call accuracy, issue end-of-season report that ranks the pool of referees in terms of overall performance. These reports determine the 36 officials who are chosen for playoff games, thereby providing considerable incentive to be as accurate as possible.

In this last section of the paper we analyze whether referees modify their behavior over time. Specifically, we investigate whether a rise in the error rate in one month leads to an adjustment in the next month's games. For this analysis we create average monthly rate of incorrect calls (Type-I error rate) and the rate of incorrect no-calls (Type-II error rate) for each

referee, yielding an unbalanced panel of monthly observations. We estimate the models of the following form:

(6) $\textit{Referee Error Rate}_{r,t} = \theta + \lambda\ \textit{Referee Error Rate}_{r,t-1} + \text{Controls} + u_{r,t.}$

where *Referee Error Rate* represents either the percentage of incorrect foul calls of the referee in a given month, or the percentage of incorrect non-calls. Standard errors are clustered at the referee level. The coefficient of the autoregressive term, λ, estimates the extent to which the error rate in a particular month has an impact on the error rate of the following month. The results are displayed in Table 11. Holding constant referee fixed effects, incorrect foul calls in Panel A are negatively autocorrelated, indicating that an increase in the percentage of incorrect foul calls in a given month leads to a decrease in the rate of incorrect foul calls the next month. Column (II) shows that this result does not change when we also hold constant the lagged rate of incorrect non-calls. Column (III) indicates that the same inference is obtained when we exclude the playoff games from the analysis. Thus, columns (I) to (III) of Table 11 reveal that an increase in referees' rate of incorrect foul calls in a month leads to a reduction in that error rate in the following month, and that incorrect call rate exhibits mean reversion. Columns (IV) and (V) show that both white and black referees adjust their behavior during the season regarding the rate of their Type-I errors.[28] Thus, the results of Table 11 are consistent with that of Table 10, which showed that the probability of the referees participating in the playoffs is negatively impacted by the rate of their incorrect calls during the season.

Panel B of Table 11 displays the analysis of incorrect non-calls and shows that the error rates in incorrect non-calls are serially uncorrelated. Put differently, an increase in referees' rate

[28] Adding month-by-year fixed effects, instead of a time trend provided the same inference. For example, in that specification the coefficient of the *Incorrect Call Rate* was -0.113 (se=0.037) in column (III). It was -0.102 (se=0.049) in column (IV), and -0.142 (se=0.056) in column (V).

of Type-II errors does not lead to an adjustment in that error rate in the following month, and that monthly rate of Type-II errors fluctuates randomly around their long-run mean value. This implies, for example, that a significant jump in the rate of incorrect calls in a given month does not motivate the referees to adjust their behavior in the following month.

The results of Tables 10 and 11, taken together, suggest that because Type-I errors significantly impact referee assignment to playoffs, referees make fewer Type-I errors compared to Type-II, they adjust their Type-I error rate down when it has risen in the previous month, and they show no in-group bias in Type-I errors. Type-II errors, which are less visible, are committed by the referees at twice the rate as Type-I errors (10 percent vs. 5.5 percent). Type-II error rates exhibit no time dependence, and in-group bias exhibits itself in these wrongful acquittals.

X. Summary and Discussion

The accuracy of performance evaluations, and any biases that may exist in these evaluations are important to investigate both to better understand human behavior and for concerns about economic efficiency and fairness. Examples of such evaluations are numerous, ranging from hiring and promotion decisions to passing judgements on others in such domains as test grading, and judicial decisions. Although it is important to investigate the existence of biases based on race, gender or some other attribute of the person being evaluated, these investigations are incomplete, and they could be misleading without the analysis of the second-layer bias, which is the study of in-group favoritism.[29]

[29] Also important is a theoretical framework to bridge the concept of the optimization process of the decision maker and the empirical implementation (Canay et al. 2020; Arnold et al. 2018)

Equally crucial is the determinants of the *errors* in these decisions, and whether these errors are related to the attributes of the evaluators and those who are being evaluated. These analyses, however, are difficult to carry out because of two reasons. First, they require a benchmark for the "correct decision" to which each individual evaluation can be compared. Second, in many cases the accuracy of the evaluations may not be observable. For example, assume that a manager hires a new employee based the evaluation of job candidates' credentials. After the new hire has spent some time on the job, it can be determined whether the hiring decision was correct or incorrect. On the other hand, it is impossible to ascertain the correctness of "not-hiring" decisions on the applicants who were rejected. Similarly, consider the decision of a judge to convict or acquit defendants. If the defendants are found guilty, it is theoretically possible to analyze Type-I errors (incorrect convictions) if the cases are re-evaluated by an upper court. On the other hand, it is impossible to analyze Type-II errors (incorrect acquittals) because once a defendant is found not guilty, the case is closed.[30] The upshot is that in most cases it is not possible to investigate *biases in evaluation errors*, although they contain important information.

In this paper we use a novel data source that allows us to investigate racial bias and in-group bias in performance evaluations and errors in these evaluations. We analyze the decisions of NBA referees who observe incidents that involve potential violations of the rules of the game.

[30] At the risk of belaboring the point, consider as a final example, the evaluation of patients by cardiologists regrading whether the patients need a stent to be inserted into a cardiac vein to prevent a heart attack. If the cardiologist evaluates the patient negatively and concludes that a stent is not necessary, the subsequent outcome will determine if the diagnosis was correct or incorrect. If the patient lives, the diagnosis would be a correct negative. If the patient dies because of a heart attack, the diagnosis would be a false negative. If, on the other hand, the patient is diagnosed positively and if a stent is inserted, the patient will live regardless of whether the evaluation was correct or incorrect. Thus, it is impossible to determine the extent of false positive diagnoses.

Referees make foul/no-foul determinations on the players who may have committed fouls. These are consequential decisions because they are evaluations of incidents during the last two minutes of close games, where the outcome of the game is at stake.[31]

This setting provides significant advantages over other environments. First, we have a wealth of information on both the presumed perpetrator (the committing player) and the victim (the disadvantaged player) who were involved in the incident, ranging from their performance statistics to their race. Similarly, we know the identity of the referees who made each specific decision. As mentioned above, these decisions involve either a conviction (blowing the whistle and declaring the committing player guilty of a foul) or an acquittal (letting the play continue). Importantly, we also know whether each of these decisions was correct or incorrect. This is possible because the NBA has evaluated the accuracy of each decision for its own purposes, which allows us to analyze both Type-I and Type-II errors. We consider the assessment of the NBA as the correct verdict in each incident because NBA's incentive is to evaluate the referee performance correctly. Furthermore, NBA's evaluations are transparent as they are revealed to the public via the videos of these referee decisions along with the assessment of their correctness.[32] As described in the previous section, NBA has a well-defined and detailed procedure for referee assessment and feedback, including evaluations after the games, bi-weekly reports, and mid- and post-season evaluations involving multiple parties including referees, NBA officials, and coaches.

[31] The decisions we analyze are made during the last two minutes of games, where the teams were within 3 points of each other anytime during those two minutes.

[32] Even if NBA were biased in its evaluation of the decisions made by one particular group of referee, such an act would not be the driver of the intricate pattern of results identified in the paper.

We detect no overall racial bias in foul calls; that is, black players are no more or no less likely to receive a foul call from referees in comparison to white players. Similarly, there is no racial in-group bias in foul calls: referees do not treat players of their own race differently when they make foul calls. The same is also true for Type-I errors, meaning that there is no evidence for racial bias or in-group bias in referees' propensity to make incorrect foul calls.

On the other hand, we document substantial in-group bias in Type-II errors. Referees are more likely to let a player get away with a foul if that player is of the same race as the referee. NBA referees are highly trained individuals, and they operate within well-established rules of the game. Nevertheless, their decisions reveal in-group favoritism, albeit in a relatively covert domain of Type-II errors. That we find no evidence of in-group bias in Type-I errors is interesting because Type-I errors are more visible than Type-II errors. If an incorrect foul is called on a player, the implications for both the teams and the player (who has a 6 foul limit to stay in the game) could be substantial. The repercussions of a missed foul call (an incorrect no-call) is arguably lower.

In-group favoritism detected in Type-II errors suggest the likely (psychic) benefit for referees to making these decisions. On the other hand, errors have consequences. Consistent with institutional structure of the NBA regarding referee evaluation, we show in the paper that the error rates of referees during the season influence their assignment to playoff games, highlighting the cost of decision errors to referees. More specifically, Type-I error rate during the season has a significant negative impact on referees' participation in the playoff games of that season. Relatedly, an increase in the Type-I error rate of referees in a given month during the season leads to a reduction in the same error rate in the following month. On the other hand, referees' monthly Type-II error rates are uncorrelated over time, indicating that they fluctuate around a

mean value, although Type-II errors in a season have also a negative but statistically insignificant influence on whether referees are selected to officiate a playoff game.

Bertrand, Chugh and Mullainathan (2005) argue that the results of a number of decisions analyzed in previous studies, ranging from screening resumés in Bertrand and Mullainathan (2004) to tipping taxi drivers in Ayres, Vars and Zakariya (2004), can be attributed to implicit bias, rather than being an indication of taste-based or statistical discrimination. The racial in-group bias identified in this paper could be another example of implicit bias to the extent that it is a reflection of spontaneous reactions of referees. It should be recognized, however, that, regardless of whether it is deliberate or spontaneous, in-group favoritism does not imply out-group hatred, and that in group-favoritism may have an evolutionary component[33]

Importantly, we show that in-group bias reacts to the environment. We find that referees' in-group bias is impacted by the racial composition of their peers on the court. In-group favoritism in Type-II errors of black referees disappears when they have two white peers on the court, and white referees' in-group favoritism appears if they have two black peers.
This pattern is consistent with the "critical mass" hypothesis, which suggests that people who are in the numerical minority of a group adopt the behavioral norms of the majority, but when they become a member of the numerical majority, their decisions reflect their true tendencies.[34] In our case, when a black referee is in the numerical minority on the court (when he has two white peers) he exhibits no in-group favoritism. When a black referee is among the numerical majority of the refereeing crew (which is the case when the 3-member crew includes no more than one

[33] See Bloom (2013), Mahajan and Wynn (2012), and Efferson, Lalive and Fehr (2008). Along the same lines, in a field experiment Feld, Salamanca and Hamermesh (2015) find that the difference in outcomes (grades assigned to students) by nationality is attributable to endophilia and not to exophobia.

[34] Empirical examples of this pattern are provided, among others, by (Eren and Mocan 2021; Etzkowitz et al. 1994; Joecks et al. 2013; Dahlerup 1988; Kanter 1977).

white referee), he is more likely to let a player get away with a foul if that player is also black. This pattern also exists for white referees. When a white referee is the numerical minority (with two black peers on the court) he conforms to the behavior of black referees and exhibits in-group favoritism in Type-II errors. When the white referee is in the majority on the court (having zero or one black peer), in-group favoritism disappears.

These results indicate that analyses of overall bias in decision-making may be incomplete without the analysis of underling layers of those decisions. Such analyses, of course, necessitate additional data on the details of these decisions, and a reasonable benchmark to which these decisions can be compared.

Figure 1

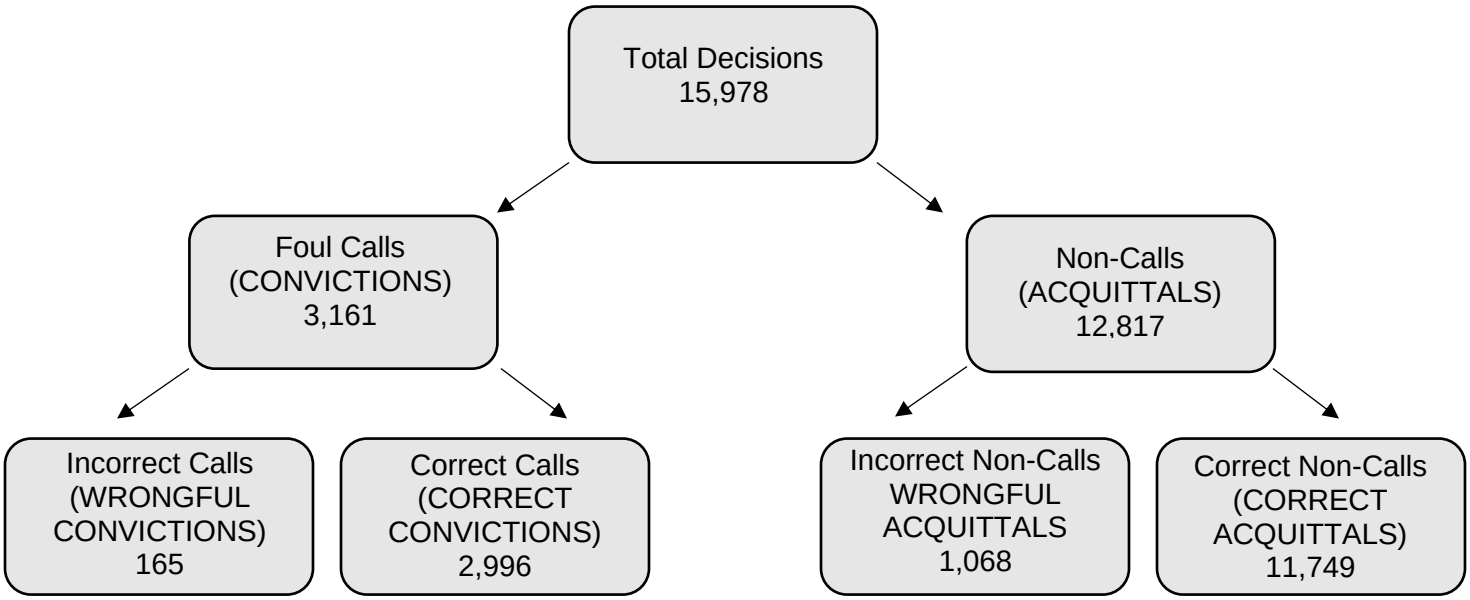

Total Decisions
15,978
Foul Calls
(CONVICTIONS)
3,161
Non-Calls
(ACQUITTALS)
12,817
Incorrect Calls
(WRONGFUL
CONVICTIONS)
165
Correct Calls
(CORRECT
CONVICTIONS)
2,996
Incorrect Non-Calls
WRONGFUL
ACQUITTALS
1,068
Correct Non-Calls
(CORRECT
ACQUITTALS)
11,749

Table 1
Descriptive Statistics

Panel A: Referee Decisions by Player Race				
	Black Players	White Players	Diff in Mean	p-value
	Decisions (N=15,978)			
Foul Call	0.203	0.176	0.027	0.001
	(0.402)	(0.381)		
N	12,891	3,087		
	Foul Calls (N= 3,161)			
Incorrect Foul Call	0.051	0.059	-0.008	0.438
(Type-I Error)	(0.220)	(0.236)		
N	2,618	543		
	No-Calls (N= 12,817)			
Incorrect No-Call	0.084	0.082	0.002	0.750
(Type-II Error)	(0.277)	(0.274)		
N	10,273	2,544		
Panel B: Game Characteristics by Player Race (N=8,743)				
	Black Players	White Players	Diff in Means	p-value
Win for	0.511	0.480	0.031	0.011
Committing Player	(0.500)	(0.500)		
Playoff Game	0.097	0.075	0.022	0.001
	(0.296)	(0.264)		
Black Coach for	0.277	0.227	0.050	0.000
Committing Player	(0.448)	(0.419)		
Home Game for	0.476	0.490	-0.014	0.254
Committing Player	(0.499)	(0.500)		
N	7,173	1,570		

Notes: In Panel A displays the data source is the LTMR. The data source in Panel B is Basketball-Refence.com. For both panels standard deviations are in parentheses. The p-values in the final column are for the null hypotheses that the variables in the left-hand column are not related to player race.

Table 2
Descriptive Statistics

Panel A: Committing Player Season Statistics (Averages for 48 Minutes)				
	Black Player	White Player	Diff in Mean	p-value
Personal Fouls	4.127	4.570	-0.444	0.000
	(1.284)	(1.422)		
Points	20.400	19.392	1.008	0.007
	(6.368)	(5.301)		
Offensive Rebounds	1.993	2.739	-0.746	0.000
	(1.576)	(1.939)		
Defensive Rebounds	6.433	7.713	-1.280	0.000
	(2.646)	(2.821)		
Assists	4.514	3.896	0.618	0.000
	(2.815)	(2.429)		
Blocks	0.987	1.085	-0.098	0.071
	(0.900)	(0.848)		
Steals	1.580	1.357	0.223	0.000
	(0.589)	(1.352)		
Turnovers	2.668	2.589	0.079	0.203
	(1.059)	(0.828)		
Two-Point Attempts	11.540	10.850	0.690	0.012
	(4.510)	(4.444)		
Three-Point Attempts	5.272	4.738	0.534	0.011
	(3.392)	(3.673)		
Free-Throw Attempts	4.425	4.124	0.301	0.035
	(2.405)	(2.066)		
Panel B: Committing Player Attributes				
	Black Player	White Player	Diff in Mean	p-value
Minutes Per Game	24.620	21.682	2.938	0.000
	(7.481)	(6.893)		
Games Played	62.125	61.338	0.787	0.500
	(19.089)	(19.116)		
Games Started	34.942	29.771	5.171	0.004
	(29.464)	(28.648)		
Center	0.148	0.379	-0.231	0.000
	(0.355)	(0.485)		
Forward	0.373	0.396	-0.023	0.442
	(0.483)	(0.489)		
Guard	0.479	0.225	0.254	0.000
	(0.499)	(0.418)		
All-Star	0.088	0.038	0.050	0.002
	(0.284)	(0.192)		

Notes: Data are from Basketball-Reference.com. Player statistics (Panel A) and player attributes (Panel B) are measured at the season-level. There are 1,611 player-season observations, 1,271 of which are from Black players and 340 are from white players. The p-values in the final column are for the null hypotheses that the variables in the left-hand column are not related to player race.

Table 3A
The Percentage of Minutes Played by Black Players by Refereeing Crew Racial Composition

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
	0 White Referees	1 White Referee	2 White Referees	3 White Referees	*N*	t-test (p-value)	t-test (p-value)
Overall	0.799	0.807	0.810	0.807			
2015	0.745	0.784	0.808	0.770	268	0.331	0.332
2016	0.800	0.798	0.803	0.807	854	0.439	0.176
2017	0.784	0.797	0.793	0.811	750	0.260	0.268
2018	0.830	0.812	0.832	0.802	604	0.657	0.659
2019	0.817	0.841	0.822	0.827	488	0.504	0.505
N	270	1,020	1,202	472	2,964		

Notes: Data are from Basketball-Reference.com. The unit of observation is team-by-game. Each cell represents the percent of minutes played by Black players in games officiated by zero, one, two, or three white referees. The p-values in Column (VI) are based on regressions where the average number of minutes played by Black players are regressed on the number of white referees in that game. Each observation pertains to one team (2 observations per game). The p-values are for the null hypotheses that the number of white referees on the court are not related to the percentage of minutes played by Black players. Column (VII) reports the same p-values from regressions that also include game controls. The *N* represents the number of team-games represented in each column or row.

Table 3B
The Number of Black Starters by Refereeing Crew Racial Composition

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
	0 White Referees	1 White Referee	2 White Referees	3 White Referees	*N*	t-test (p-value)	t-test (p-value)
Overall	4.004	4.013	4.046	4.015			
2015	3.667	3.790	3.983	3.735	268	0.379	0.368
2016	3.932	3.966	3.987	4.018	854	0.395	0.322
2017	3.989	3.983	3.963	4.030	750	0.862	0.629
2018	4.146	4.042	4.153	4.042	604	0.925	0.970
2019	4.158	4.219	4.177	4.081	488	0.286	0.408
N	270	1,020	1,202	472	2,964		

Notes: Data are from Basketball-Reference.com. The unit of observation is team-by-game. Each cell represents the number of Black starters in games officiated by zero, one, two, or three white referees. The p-values in Column (VI) are based on regressions where the total number of Black starters is regressed on the number of white referees in that game. Each observation pertains to one team (2 observations per game). The p-values are for the null hypotheses that the number of white referees on the court are not related to the number of black starters. Column (VII) reports the p-values from regressions that also include game controls. The *N* represents the number of team-games represented in each column or row.

Table 4

The Relationship between Player Race and the Referees

Panel A: All Decisions on Black and White Committing Players				
	Black Players	White Players	Diff in Means	p-value
Call Made by Crew Chief	0.337	0.321	0.015	0.101
	(0.473)	(0.467)		
Call Made by Umpire	0.333	0.330	0.002	0.814
	(0.471)	(0.470)		
Call Made by Referee	0.326	0.344	-0.018	0.053
	(0.469)	(0.475)		
Crew Chief Black	0.474	0.492	-0.018	0.067
	(0.499)	(0.500)		
Total Black Referees	1.367	1.410	-0.043	0.011
	(0.854)	(0.860)		
N	12,891	3,087		
Panel B: Foul Calls on Committing Players				
	Black Players	White Players	Diff in Means	p-value
Call Made by Crew Chief	0.337	0.324	0.015	0.101
	(0.473)	(0.468)		
Call Made by Umpire	0.325	0.346	0.002	0.814
	(0.469)	(0.476)		
Call made by Referee	0.332	0.326	-0.018	0.053
	(0.471)	(0.469)		
Crew Chief Black	0.485	0.499	-0.018	0.067
	(0.500)	(0.500)		
Total Black Referees	1.368	1.390	-0.043	0.011
	(0.870)	(0.873)		
N	2,618	543		
Panel C: No-Calls on Committing Players				
	Black Players	White Players	Diff in Means	p-value
Call Made by Crew Chief	0.337	0.321	0.015	0.101
	(0.473)	(0.467)		
Call Made by Umpire	0.334	0.327	0.002	0.814
	(0.472)	(0.469)		
Call Made by Referee	0.324	0.348	-0.018	0.053
	(0.468)	(0.476)		
Crew Chief Black	0.471	0.491	-0.018	0.067
	(0.499)	(0.500)		
Total Black Referees	1.366	1.414	-0.043	0.011
	(0.850)	(0.858)		
N	10,273	2,544		

Table 5
Relationship between Player Race and Referee Race

	Dependent Variable: Black Referee				
	(I)	(II)	(III)	(IV)	(V)
Committing Player Black	-0.014	-0.011	0.009	-0.011	0.012
	(0.013)	(0.013)	(0.009)	(0.014)	(0.009)
Game FE	NO	NO	YES	NO	YES
Season FE	YES	YES	NO	YES	NO
Player FE	NO	NO	NO	NO	NO
Referee FE	NO	NO	NO	NO	NO
Player Characteristics	NO	YES	YES	YES	YES
Player Statistics	NO	NO	NO	YES	YES
Game Controls	NO	YES	YES	YES	YES
Team Controls	NO	YES	YES	YES	YES
N	15,978	15,978	15,978	15,978	15,978

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. ***, **, and * signify statistical significance at the 1%, 5%, and 10%, respectively.

Table 6
Overall Racial Bias in Foul Calls, Type-I and Type-II Errors

	Foul Call		Incorrect Foul Call		Incorrect No-Call	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Committing Player Black	0.011	0.000	0.002	0.011	0.011*	0.008
	(0.009)	(0.010)	(0.011)	(0.017)	(0.007)	(0.007)
Disadvantaged Player Black	-0.017	-0.015	0.000	0.024	-0.012	-0.010
	(0.010)	(0.011)	(0.013)	(0.020)	(0.009)	(0.010)
Black Coach for Committing Player	-0.015**	-0.017*	-0.010	-0.017	-0.003	0.000
	(0.007)	(0.009)	(0.010)	(0.019)	(0.006)	(0.009)
Home Game for Committing Player	-0.000	-0.007	-0.000	-0.006	0.010**	0.013**
	(0.006)	(0.007)	(0.008)	(0.013)	(0.005)	(0.005)
Playoff Game	-0.060***		0.016		-0.007	
	(0.015)		(0.016)		(0.011)	
Committing Player All Star in that Season	-0.008	-0.006	0.011	0.015	-0.025***	-0.020*
	(0.011)	(0.012)	(0.019)	(0.028)	(0.009)	(0.010)
Committing Player Starter (% Games Of Season)	-0.026*	-0.037**	0.009	-0.009	0.006	0.000
	(0.015)	(0.017)	(0.017)	(0.024)	(0.014)	(0.017)
Game FE	NO	YES	NO	YES	NO	YES
Season FE	YES	NO	YES	NO	YES	NO
Committing Player FE	NO	NO	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES	YES	YES
N	15,978	15,978	3,161	3,161	12,817	12,817

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. ***, **, and * signify statistical significance at the 1%, 5%, and 10%, respectively.

Table 7A
In-Group Bias in Foul Calls

	Dependent Variable: Foul Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	-0.002	-0.005	-0.008	-0.010
	(0.014)	(0.015)	(0.017)	(0.018)
Committing Player Black	0.012		0.004	
	(0.011)		(0.014)	
Disadvantaged Player Black	-0.017	-0.015	-0.015	-0.011
	(0.010)	(0.011)	(0.011)	(0.012)
Black Coach for Committing Player	-0.015**	-0.044***	-0.017*	-0.054***
	(0.007)	(0.010)	(0.009)	(0.013)
Home Game for Committing Player	-0.000	-0.003	-0.007	-0.008
	(0.006)	(0.006)	(0.007)	(0.007)
Playoff Game	-0.060***	-0.061***		
	(0.015)	(0.015)		
Committing Player All Star in that Season	-0.008	-0.004	-0.006	-0.004
	(0.011)	(0.013)	(0.012)	(0.017)
Committing Player Starter (% Games of Season)	-0.026*	-0.019	-0.037**	-0.020
	(0.015)	(0.022)	(0.017)	(0.025)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	15,978	15,978	15,978	15,978

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level.

***, **, and * signify statistical significance at the 1%, 5%, and 10%, respectively.

Table 7B
In-Group Bias in Type-I Errors

	Dependent Variable: Incorrect Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.012	0.000	-0.006	-0.033
	(0.023)	(0.025)	(0.035)	(0.039)
Committing Player Black	-0.004		0.013	
	(0.012)		(0.022)	
Disadvantaged Player Black	-0.001	-0.008	0.024	0.009
	(0.013)	(0.015)	(0.020)	(0.024)
Black Coach for Committing Player	-0.010	-0.002	-0.017	-0.010
	(0.010)	(0.012)	(0.019)	(0.034)
Home Game for Committing Player	0.000	-0.003	-0.005	-0.009
	(0.008)	(0.010)	(0.014)	(0.016)
Playoff Game	0.016	0.011		
	(0.016)	(0.017)		
Committing Player All Star in that Season	0.009	0.003	0.014	0.038
	(0.019)	(0.027)	(0.028)	(0.045)
Committing Player Starter (% Games of Season)	0.008	0.052	-0.009	0.025
	(0.016)	(0.037)	(0.024)	(0.054)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	3,161	3,161	3,161	3,161

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level.

***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Table 7C
In-Group Bias in Type-II Errors

	Dependent Variable: Incorrect No Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.024**	0.023*	0.027**	0.022*
	(0.013)	(0.013)	(0.011)	(0.012)
Committing Player Black	0.000		-0.005	
	(0.009)		(0.008)	
Disadvantaged Player Black	-0.012	-0.011	-0.010	-0.008
	(0.009)	(0.009)	(0.010)	(0.011)
Black Coach for Committing Player	-0.003	0.004	0.000	0.007
	(0.006)	(0.010)	(0.009)	(0.012)
Home Game for Committing Player	0.010**	0.011**	0.013**	0.013**
	(0.005)	(0.005)	(0.005)	(0.006)
Playoff Game	-0.007	-0.010		
	(0.011)	(0.011)		
Committing Player All Star in that Season	-0.025***	-0.043***	-0.020*	-0.037***
	(0.009)	(0.012)	(0.010)	(0.014)
Committing Player Starter (% Games of Season)	0.006	-0.002	-0.000	-0.020
	(0.014)	(0.020)	(0.017)	(0.023)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	12,817	12,817	12,817	12,817

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level.

***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Table 8
Bias in Type-II Errors Conditional on Referee Race

	Dependent Variable: Incorrect No-Call			
	White Referees		Black Referees	
	(I)	(II)	(III)	(IV)
Committing Player Black			0.023**	0.024**
			{0.029}	{0.022}
Committing Player White	-0.000	0.004		
	{0.961}	{0.611}		
Disadvantaged Player Black	-0.009	-0.011	-0.017	-0.012
	(0.011)	(0.014)	(0.015)	(0.018)
Black Coach for Committing Player	-0.006	0.001	-0.001	-0.003
	(0.007)	(0.012)	(0.009)	(0.014)
Home Game for Committing Player	0.005	0.008	0.018**	0.023***
	(0.007)	(0.008)	(0.007)	(0.008)
Playoff Game	-0.009		-0.004	
	(0.016)		(0.012)	
Committing Player All Star in that Season	-0.028**	-0.022	-0.021	-0.021
	(0.012)	(0.016)	(0.013)	(0.014)
Committing Player Starter (% Games of Season)	-0.019	-0.027	0.033	0.023
	(0.021)	(0.030)	(0.020)	(0.023)
Game FE	NO	YES	NO	YES
Season FE	YES	NO	YES	NO
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	6,891	6,891	5,926	5,926

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}. ***, **, and * signify statistical significance at 1%, 5%, and 10% level, respectively.

Table 9A
Type-II Errors of Black Referees with Peer Effects

	Dependent Variable: Incorrect No-Call			
	(W,W)	(W,B)	(W,B)/(B,B)	(B,B)
	(I)	(II)	(III)	(IV)
Committing Player Black	0.005	0.035**	0.031**	0.033
	{0.786}	{0.029}	{0.029}	{0.109}
Disadvantaged Player Black	-0.023	-0.020	-0.015	-0.002
	(0.028)	(0.016)	(0.014)	(0.025)
Black Coach for Committing Player	-0.008	0.006	0.003	0.001
	(0.017)	(0.013)	(0.011)	(0.021)
Home Game for Committing Player	-0.004	0.028***	0.029***	0.042**
	(0.011)	(0.010)	(0.008)	(0.021)
Playoff Game	-0.033	0.004	0.009	0.067*
	(0.027)	(0.025)	(0.021)	(0.033)
Committing Player All Star in that Season	-0.023	-0.022	-0.016	-0.008
	(0.022)	(0.022)	(0.019)	(0.031)
Committing Player Starter (% Games of Season)	0.006	0.043	0.044*	0.058
	(0.026)	(0.025)	(0.023)	(0.048)
Game FE	NO	NO	NO	NO
Season FE	YES	YES	YES	YES
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	1,792	2,969	4,134	1,165

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}. ***, **, and * signify statistical significance at 1%, 5%, and 10% level, respectively.

Table 9B
Type-II Errors of White Referees with Peer Effects

	Dependent Variable: Incorrect No-Call			
	(W,W)	(W,B)	(W,B)/(B,B)	(B,B)
	(I)	(II)	(III)	(IV)
Committing Player White	0.008	-0.025	-0.002	0.050**
	{0.605}	{0.145}	{0.839}	{0.049}
Disadvantaged Player Black	0.001	-0.002	-0.009	-0.018
	(0.022)	(0.013)	(0.011)	(0.024)
Black Coach for Committing Player	0.001	-0.011	-0.007	-0.002
	(0.012)	(0.010)	(0.009)	(0.023)
Home Game for Committing Player	0.007	0.006	0.005	0.009
	(0.013)	(0.010)	(0.007)	(0.013)
Playoff Game	-0.008	-0.002	-0.011	-0.018
	(0.030)	(0.018)	(0.017)	(0.029)
Committing Player All Star in that Season	-0.048**	-0.032	-0.017	0.016
	(0.021)	(0.019)	(0.015)	(0.028)
Committing Player Starter (% Games of Season)	-0.041	-0.016	-0.008	0.006
	(0.040)	(0.026)	(0.021)	(0.035)
Game FE	NO	NO	NO	NO
Season FE	YES	YES	YES	YES
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	1,978	3,417	4,913	1,496

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}. ***, **, and * signify statistical significance at 1%, 5%, and 10% level, respectively.

Table 10
Referees' Probability of Officiating in the Playoffs

	Dependent Variable: Referee Officiated the Playoffs				
	(I)	(II)	(III)	(IV)	(V)
Incorrect Call Rate		-0.629**			-0.803*
		(0.244)			(0.461)
Incorrect Non-Call Rate		-0.575**			-0.502
		(0.265)			(0.525)
Referee was in Playoffs Last Season			0.692***	0.067	0.040
			(0.059)	(0.147)	(0.151)
Black Referee	-0.028	-0.004	-0.003		
	(0.082)	(0.083)	(0.037)		
Referee Experience	0.030***	0.030***	0.002		
	(0.004)	(0.004)	(0.004)		
Season FE	YES	YES	YES	YES	YES
Referee FE	NO	NO	NO	YES	YES
N	309	309	241	241	241

Notes: The dependent variable is a dummy for refereeing in the playoffs in given season. The unit of observation referee-by-season.. Standard errors are clustered by referee and are in parentheses. In column (2) the sample mean of *Incorrect Call Rate* is 0.055, and it is 0.099 for the *Incorrect Non-Call Rate*. The mean of the dependent variable is 0.573. In column (5) the sample mean of *Incorrect Call Rate* is 0.051, and it is 0.082 for the *Incorrect Non-Call Rate*. The mean of the dependent variable is 0.593. ***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Table 11

Referee Learning

	Panel A: Dependent Variable: Incorrect Call Rate (t)				
	All	All	All	Black Refs	White Refs
	(I)	(II)	(III)	(IV)	(V)
Incorrect Call	-0.077***	-0.080**	-0.107***	-0.084**	-0.129**
Rate (t-1)	(0.028)	(0.030)	(0.033)	(0.041)	(0.053)
Incorrect Non-		0.069	0.053	0.000	0.091
Call Rate (t-1)		(0.088)	(0.100)	(0.094)	(0.153)
Month FE	YES	YES	YES	YES	YES
Referee FE	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES
Including Playoffs	YES	YES	NO	NO	NO
N	640	640	618	279	339
	Panel B: Dependent Variable: Incorrect No-Call Rate (t)				
	All	All	All	Black Refs	White Refs
	(I)	(II)	(III)	(IV)	(V)
Incorrect Call		0.024	0.020	0.016	0.025
Rate (t-1)		(0.041)	(0.042)	(0.036)	(0.081)
Incorrect Non-	-0.009	-0.012	-0.032	-0.051	-0.012
Call Rate (t-1)	(0.042)	(0.041)	(0.041)	(0.061)	(0.057)
Month FE	YES	YES	YES	YES	YES
Referee FE	YES	YES	YES	YES	YES
Time Trend	YES	YES	YES	YES	YES
Including Playoffs	YES	YES	NO	NO	NO
N	818	818	794	355	439

Notes: The unit of observation is referee-by-month for the months in which referee made fouls calls or non-calls. Standard errors, lustered at the referee level, are reported in parentheses. In Panel A the mean value of *Incorrect Call Rate* is 0.050 and the mean of *Incorrect Non-Call Rate* is 0.094 in column (2). In panel B, A the mean value of *Incorrect Call Rate* is 0.048 and the mean of *Incorrect Non-Call Rate* is 0.097 in column (2). ***, **, and * represent statistical significance at 1%, 5%, and 10%, respectively.

References

Alesina, A. & Ferrara, E. L. (2014). A Test of Racial Bias in Capital Sentencing. *The American Economic Review*, *104*(11), 3397-3433.

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, *127*(2), 1017-1055.

Archsmith, J., & Heyes, A., Neidell, M. & Sampat, B. (2021). The Dynamics of Inattention in the (Baseball) Field. *IZA Discussion Papers*, *14440*.

Arnold, D., Dobbie, W. & Yang, C. (2018). Racial Bias in Bail Decision. *The Quarterly Journal of Economics, 133*(4), 1885-1932.

Ayres, I., Banaji, M., Jolls, C. (2015). Race Effects on eBay. *Rand Journal of Economics*, *46*, 891-917.

Ayres, I., Vars, F. E., & Zakariya, N. (2004). To insure prejudice: Racial Disparities in Taxicab Tipping. *Yale Law Journal*, *114*, 1613.

Bar, R. & Zussman, A. (2020). Identity and Bias: Insights from Driving Tests. *Economic Journal*, *130*(625), 1-23.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial Altruism in Humans. *Nature, 442*, 912–15.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit Discrimination. *The American Economic Review*, *95*(2), 94–98.

Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991-1013.

Bielen, S., Mocan, N., & Marneffe, W. (2021). Racial Bias and In-group Bias: Evidence from Virtual Reality Courtrooms. *Journal of Law and Economics, 64*(2).

Bloom, P. (2013). *Just Babies.* New York: Crown Publishers.

Canay, I.A., Mogstad, M., & Mountjoy, J. (2020). On the Use of Outcomes Tests for Detecting Bias in Decision Making. *NBER Working Paper*, *No:27802*.

Carrell, S.E., Sacerdote, B.I., & West, J.E. (2013). From Natural Variation to 0ptimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica, 81*(3), 855-882.

Chan, D.C., Gentzkow, M., Yu, C. (2022). Selection with Variation in Diagnostic Skill: Evidence from Radiologists, *The Quarterly Journal of Economics*, *137*(2), 729-783.

Correll, J., Park, B., Judd, C.M., & Wittenbrink, B. (2002). The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. *Journal of Personality and Social Psychology, 83*(6), 1314–29.

Correll, J., Park, B., Judd, C.M., Wittenbrink, B., Sadler, M.S., & Keesee, T. (2007). Across the thin blue line: Police Officers and Racial Bias in the Decision to Shoot. *Journal of Personality and Social Psychology 92*(6), 1006–23.

Cornelissen, T., Dustmann, C., & Schonberg, U. (2017). Peer Effects in the Workplace. *American Economic Review, 107*(2), 425-456.

Dahlerup, Drude (1988). Fram a Small to a Large Minority: Women in Scandinavian Politics. Scandinavian Political Studies, 11(4), 275-298.

Depew, B., Eren, O., & Mocan, N. (2017). Judges, Juveniles, and In-group Bias. *The Journal of Law and Economics, 60*(2), 209-239.

Deutscher, C. (2015). No Referee Bias in the NBA: New Evidence with Leagues' Assessment Data. *Journal of Sports Analytics, 1*(2), 91-96.

Efferson, C., Lalive, R., & Fehr, E. (2008). The Coevolution of Cultural Groups and Ingroup Favoritism. *Science, 321*(5897), 1844-9.

Eren, O. & Mocan, N. (2021). Judge Peer Effects in the Courthouse. *NBER Working Paper No:27713*.

Etzkowitz, H., C. Kemelgor, M. Neuschatz, B. Uzzi, and J. Alonzo (1994). The Paradox of Critical Mass for Women in Science. *Science* 266 (5182), 51- 54.

Feld, J., Salamanca, N., & Hamermesh, D. (2015). Endophilia or Exophobia: Beyond Discrimination. *The Economics Journal, 126*(594), 1503-1527

Fisman, R., Sarkar, A., Skrastins, J. & Vig, V. (2020). Experience of Communal Conflicts and Intergroup Lending. *Journal of Political Economy*, *128*(9), 3346–75.

Garicano, L., Palacios-Huerta, I. & Prendergast, C. (2005). Favoritism Under Social Pressure., *The Review of Economics and Statistics*, *87*(2), 208-216.

Gazal-Ayal, O. & Sulitzeanu-Kenan, R. (2010). Let My People Go: Ethnic In-Group Bias in Judicial Decisions- Evidence from a Randomized Natural Experiment. *Journal of Empirical Legal Studies*, *7*(3), 403-428.

Goette, L., Huffman, D., & Meier, S. (2006). The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups. *American Economic Review*, *96*(2), 212-216.

Gong, H. (2022). The Effect of the Crowd on Home Bias: Evidence from NBA Games During the COVID-19 Pandemic. *Journal of Sports Economics*, 1-26.

Grossman, G., Gazal-Ayal, O., Pimentel, S. & Weinstein, J. (2016). Descriptive Representation and Judicial Outcomes in Multiethnic Societies. *American Journal of Political Science, 60*(1), 44-69.

Giuliano, L., Levine, D. I., & Leonard, J. (2009). Manager Race and the Race of New Hires. *Journal of Labor Economics*, *27*(4), 589–631.

Joecks, J., Pull K. & Vetter K. (2013) Gender Diversity in the Boardroom and Firm Performance: What Exactly Constitutes a "Critical Mass?" *Journal of Business Ethics*. 118(1) 61-72.

Kanter, R. M. (1977). Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *American Journal of Sociology* 82 (5), 965-990.

Kastellec, J.P. (2013). Racial Diversity and Judicial Influence on Appellate Courts. *American Journal of Political Science*, *57*, 167-183.

Mas, A. & Moretti, E. (2009). Peers at Work. *American Economic Review, 99*(1), 112-145.

Mahajan, N. & Wynn, K. (2012). Origins of "us" versus "Them": Prelinguistic Infants Prefer Similar Others, *Cognition 124*(2), 227-33.

Mocan, N. (2020). Biases in Judicial Decision-Making. *Bias in the Law, Joseph Avery and Joel Cooper (eds.)*, London: Lexington Books, 97-114

Murphy, F.X. (2019). Does Increased Exposure to Peers with Adverse Characteristics Reduce Workplace Performance? Evidence from a Natural Experiment in the US Army. *Journal of Labor Economics*, *37*(2), 435-466.

Munnell, A., Tootell, G., Browne, L.E. & McEneaney, J. (1996). Mortgage Lending in Boston: Interpreting HMDA Data. *American Economic Review, 86*(1), 25-53.

Mussweiler, T. & Ockenfels, A. (2013). Similarity Increases Altruistic Punishment in Humans. *Psychological and Cognitive Scisnces, 110*(48).

Parsons, C.A., Sulaeman, J., Yates, M.C., & Hamermesh, D.S. (2011). Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review*, *101*(4), 1410-35.

Pope, D.G., Price, J., & Wolfers, J. (2018). Awareness Reduces Racial Bias. *Management Science*, *64*, 4988-4995.

Price, J. & Wolfers, J. (2010). Racial Discrimination Among NBA Referees. *Quarterly Journal of Economics, 125*(4), 1859-1887.

Rickman, N. & Witt, R. (2008). Favoritism and Financial Incentives: A Natural Experiment. *Economica*, *75*(298), 296-309.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics, 116*(2), 681-704.

Shayo, M. & Zussman, A. (2011). Judicial Ingroup Bias in the Shadow of Terrorims. *The Quarterly Journal of Economics*, *126*(3), 1447-1484.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review*, *95*(2), 199-203.

Appendix Table A.1
Descriptive Statistics

Panel A: Disadvantaged Player Season Statistics (Averages for 48 Minutes)				
	Black Player	White Player	Diff in Mean	p-value
Personal Fouls	4.052	4.471	-0.418	0.000
	(1.226)	(1.399)		
Points	20.592	19.797	0.795	0.040
	(6.370)	(5.284)		
Offensive Rebounds	1.899	2.645	-0.745	0.000
	(1.518)	(1.920)		
Defensive Rebounds	6.342	7.615	-1.272	0.000
	(2.614)	(2.904)		
Assists	4.625	4.022	0.603	0.000
	(2.825)	(2.470)		
Blocks	0.938	1.050	-0.112	0.037
	(0.856)	(0.855)		
Steals	1.585	1.363	0.222	0.000
	(0.595)	(1.379)		
Turnovers	2.687	2.612	0.075	0.243
	(1.068)	(0.851)		
Two-Point Attempts	11.556	10.897	0.659	0.020
	(4.539)	(4.390)		
Three-Point Attempts	5.425	4.969	0.455	0.035
	(3.380)	(3.714)		
Free-Throw Attempts	4.466	4.172	0.294	0.047
	(2.416)	(2.123)		
Panel B: Disadvantaged Player Attributes				
	Black Player	White Player	Diff in Mean	p-value
Minutes Per Game	24.936	22.321	2.616	0.000
	(7.233)	(6.738)		
Games Played	61.851	62.890	-1.039	0.394
	(19.767)	(18.130)		
Games Started	35.729	31.511	4.218	0.022
	(29.499)	(28.711)		
Center	0.134	0.382	-0.249	0.000
	(0.341)	(0.487)		
Forward	0.359	0.362	-0.003	0.931
	(0.480)	(0.481)		
Guard	0.507	0.255	0.251	0.000
	(0.500)	(0.436)		
All-Star	0.089	0.041	0.048	0.004
	(0.285)	(0.198)		

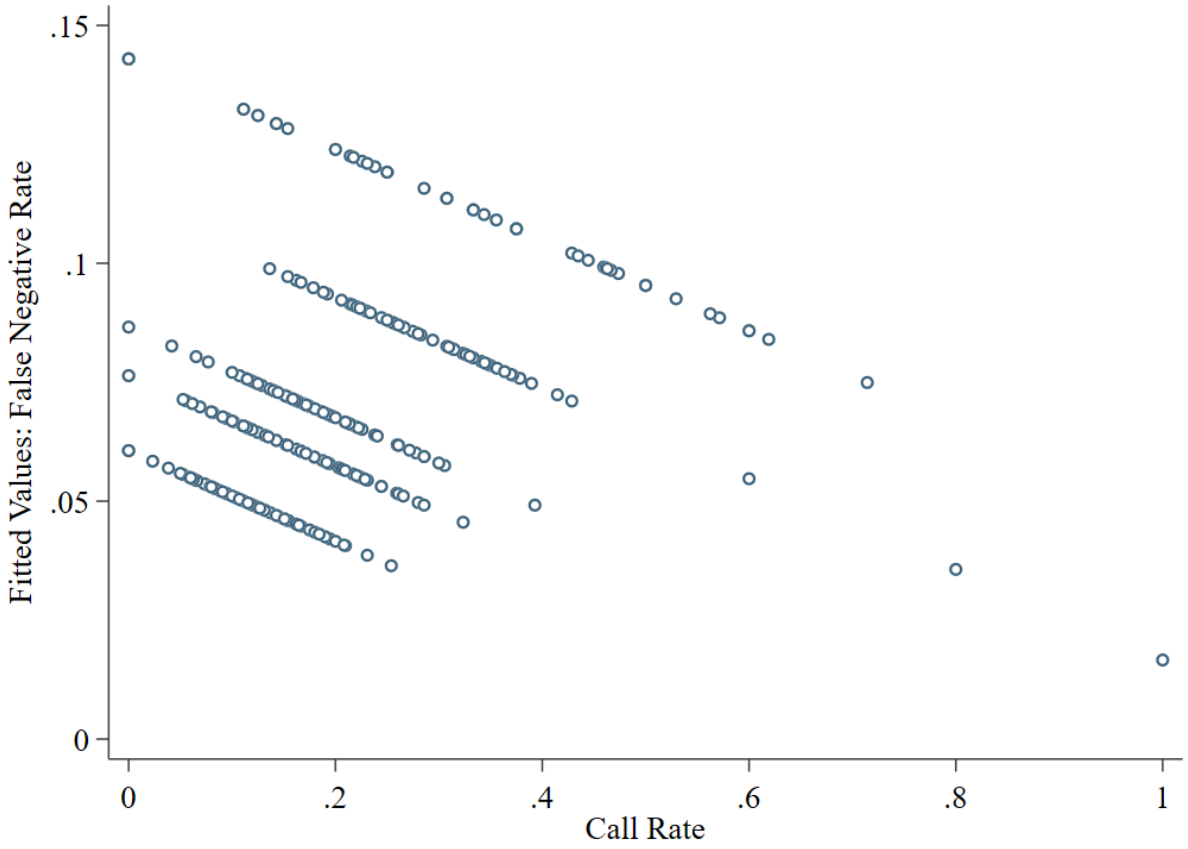
Notes: Data is from Basketball-Reference.com. Player statistics and player role variables are measured at the season level and are restricted to the players that appear in the LTMR dataset. There are 1,590 player-season observations. 1,271 are from black players and 319 are from white players. The p-values in the final column are for the null hypothesis that the variables in the left-hand column are not related to player race.

Appendix Table A.2
The Relationship between Referees' False Negative Rates and Call Rates

	Panel A: Dependent Variable: False Negative Rate				
	All	All	Black Ref	White Ref	All
	(I)	(II)	(III)	(IV)	(V)
Call Rate	-0.095**	-0.113**	-0.075	-0.112*	-0.022
	(0.037)	(0.049)	(0.051)	(0.057)	(0.037)
Season FE	YES	YES	YES	YES	NO
Referee FE	NO	YES	NO	NO	NO
N	328	328	154	174	81
	Panel B: Dependent Variable: True Positive Rate				
	All	All	Black Ref	White Ref	All
	(I)	(II)	(III)	(IV)	(V)
False Positive Rate	0.051	0.195	0.174	0.229	0.176
	(0.299)	(0.364)	(0.384)	(0.453)	(1.443)
Season FE	YES	YES	YES	YES	NO
Referee FE	NO	YES	NO	NO	NO
N	320	320	148	172	81

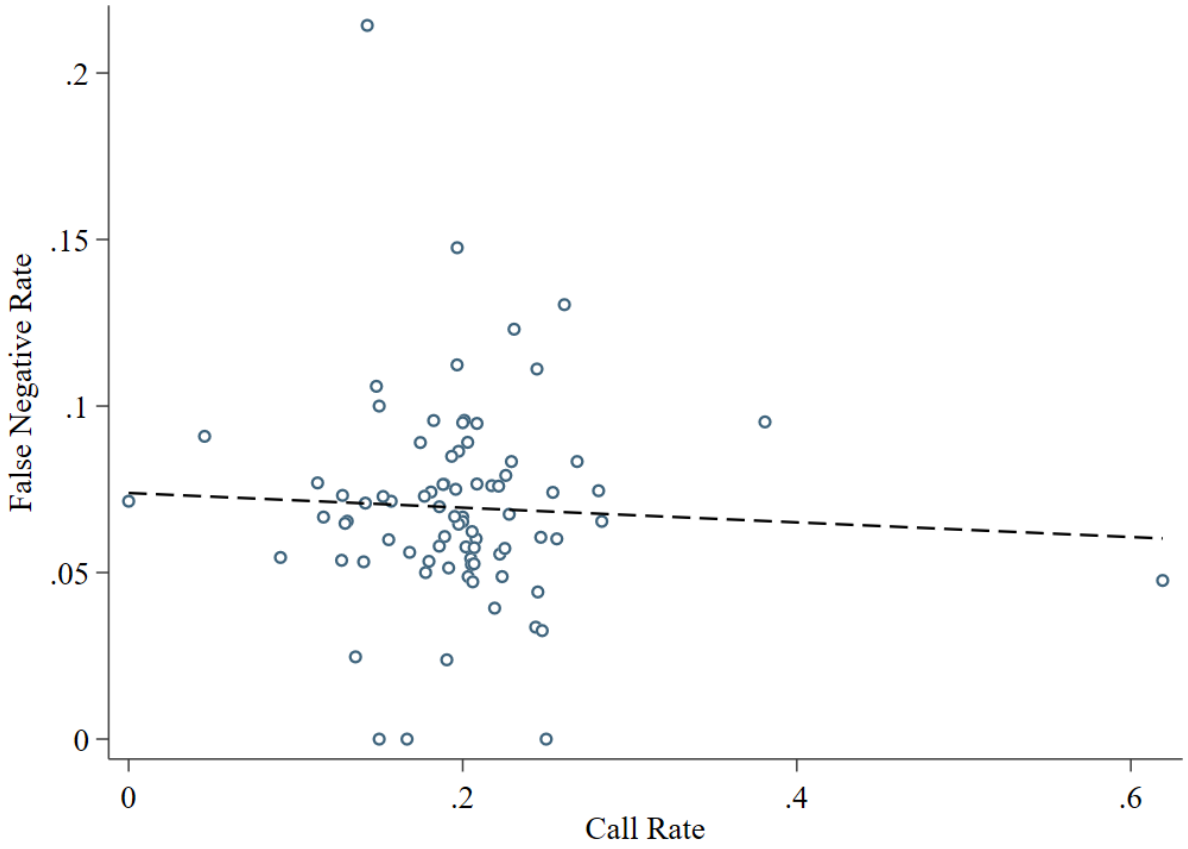
Notes: The unit of observation is referee-by-season for the first four columns and the referee for the final column.
False Negative Rate is the rate of incorrect non-calls.
Call rate is the ratio of foul calls to total decisions (foul calls + no-calls)
Standard errors are clustered at the referee level. ***, **, and * represent statistical significance at 1%, 5%, and 10% level

Appendix Figure A.F1



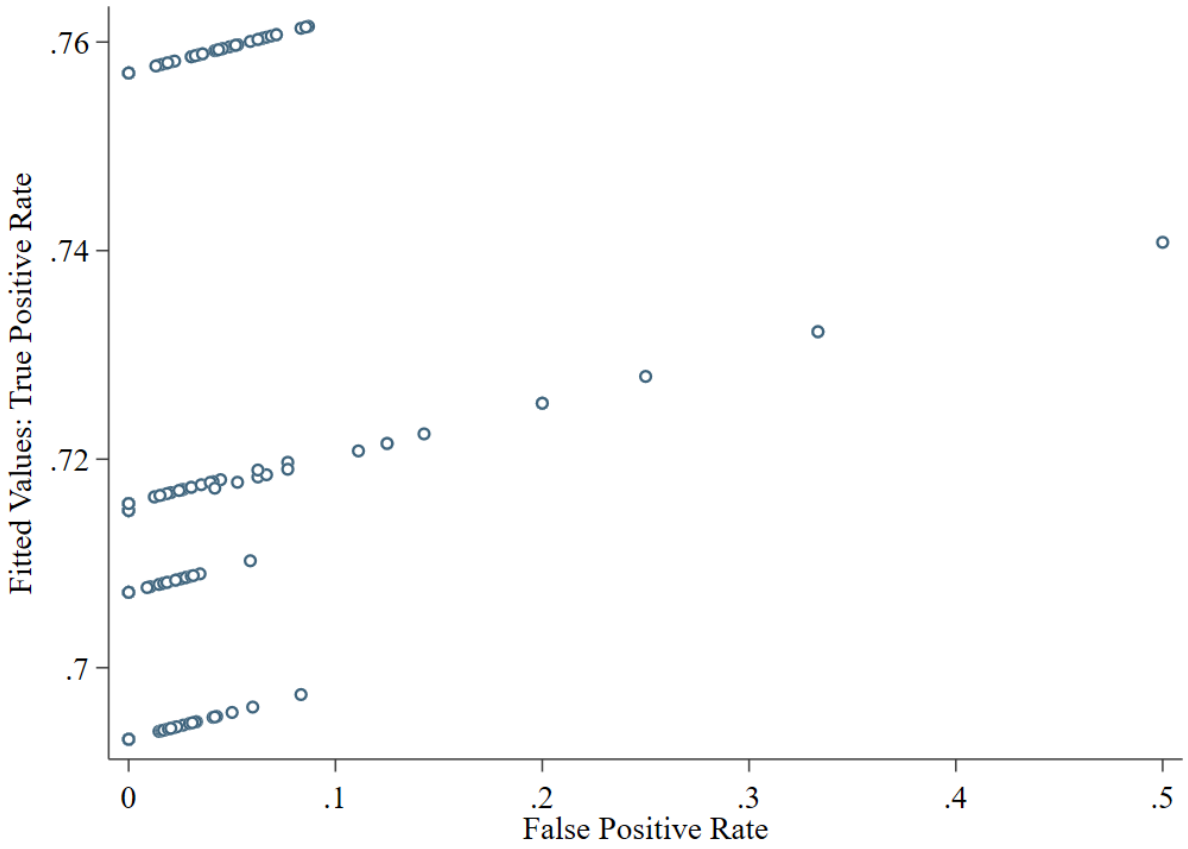
Notes: Fitted values of False Negative Rates are obtained from the model in Column (1) of Panel A in Appendix Table 2. Different intercepts reflect season fixed-effects. The fitted lines from the top pertain to seasons 2015, 2016, 2017, 2019, and 2018, respectively.

Appendix Figure A.F2



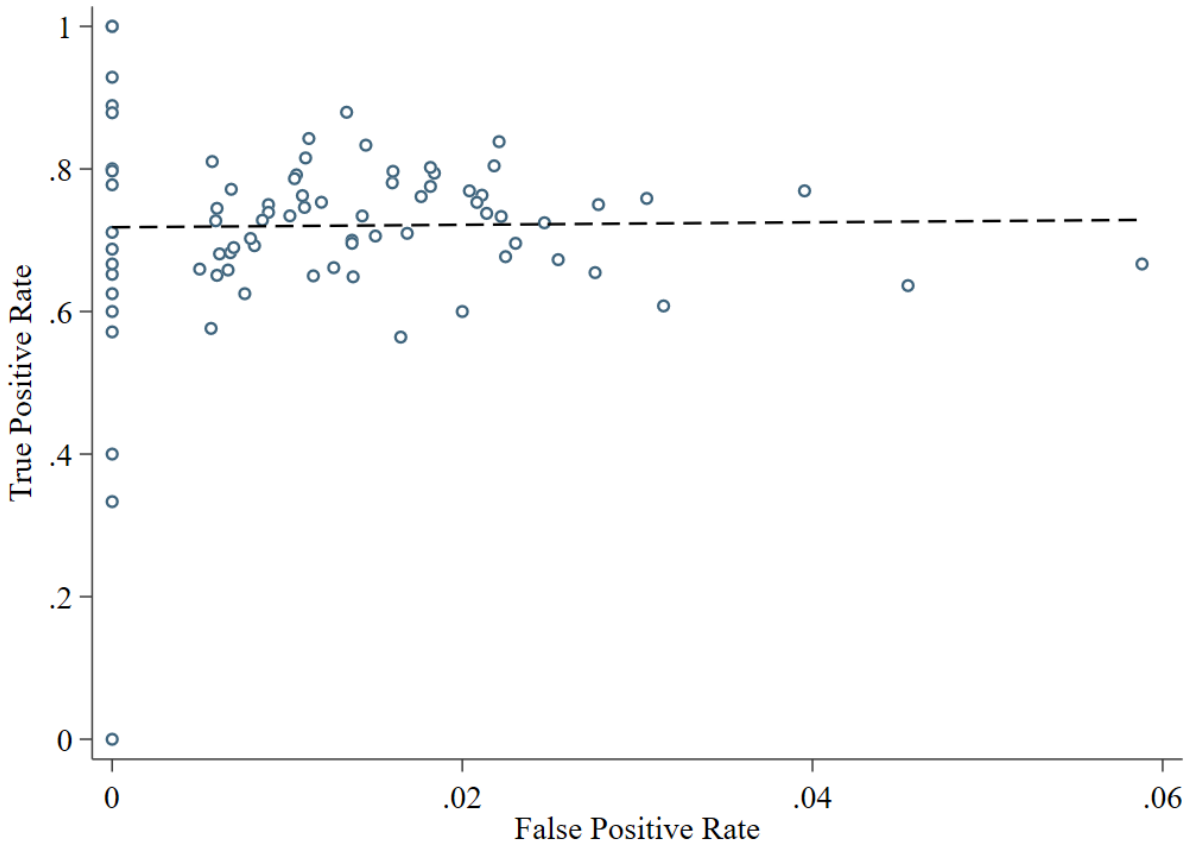
Notes: This graph displays 81 observations in column (V) of Panel A in Appendix Table A2.

Appendix Figure A.F3
All Referees



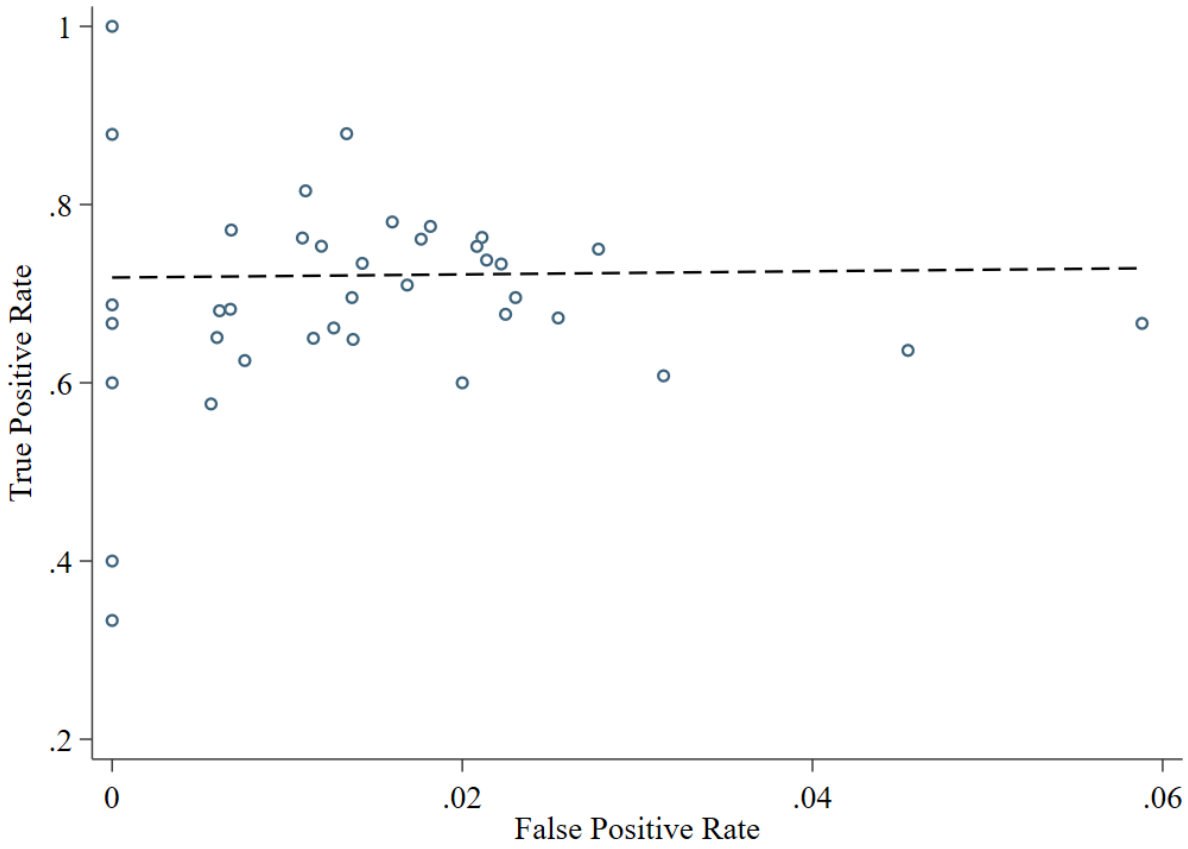
Notes: Fitted values of True Positive Rates (TPR) are obtained from the model in Column (1) of Panel B in Appendix Table 2. Different intercepts reflect season fixed-effects. TPR=TP/(TP+FN), False Positive Rate (FPR)=FP/(FP+TN) and TP: True Positive rate (the rate of correct calls), FP: False Positive rate (the rate of incorrect calls- Type I error), FN: False Negative rate (the rate of incorrect non-calls- Type-II errors), and TN: True negative rate (the rate of current non-calls). FP+TP+FP+FN=1

Appendix Figure A.F4
All Referees



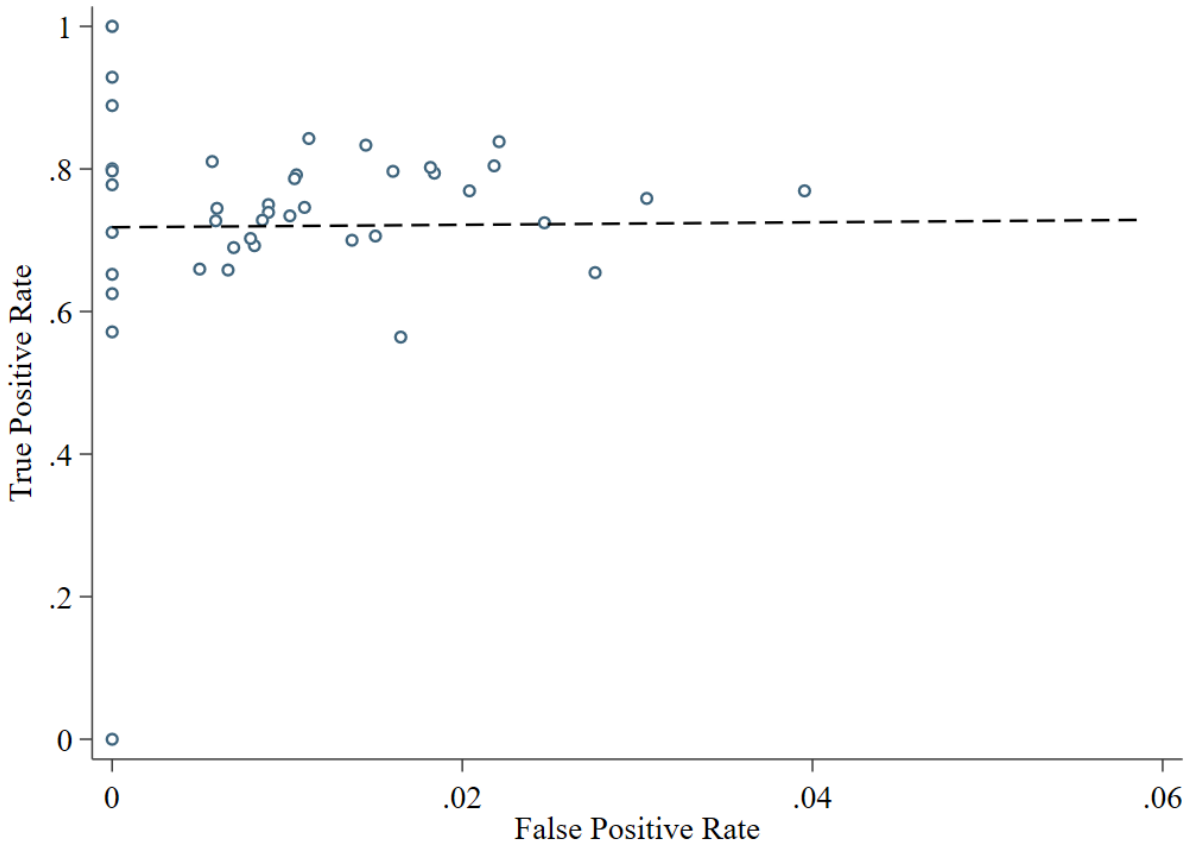
Notes: The graph is based on 81 observations used in column (5) of Panel B in Appendix Table A.2. Data points reflect averages for each referee for the entire sample period (2015-2019 season)

Appendix Figure A.F5
Black Referees



Notes: The graph is based on 38 Black referees (also used in column (5) of Panel B in Appendix Table A.2. Data points reflect averages for each referee for the entire sample period (2015-2019 season)

Appendix Figure A.F6
White Referees



Notes: The graph is based on 53 white referees (also used in column (5) of Panel B in Appendix Table A.2. Data points reflect averages for each referee for the entire sample period (2015-2019 season)

Appendix Table A.3A
In-Group Bias

	Dependent Variable: Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	-0.006	-0.005	-0.015	-0.021
	(0.018)	(0.015)	(0.020)	(0.021)
Committing Player Black	0.013		0.007	
	(0.012)		(0.015)	
Disadvantaged Player Black	-0.017*	-0.015	-0.015	-0.011
	(0.010)	(0.011)	(0.011)	(0.012)
Black Coach for Committing Player	-0.015**	-0.044***	-0.017*	-0.054***
	(0.007)	(0.010)	(0.009)	(0.013)
Home Game for Committing Player	-0.000	-0.003	-0.007	-0.008
	(0.006)	(0.006)	(0.007)	(0.007)
Playoff Game	-0.060***	-0.061***		
	(0.015)	(0.015)		
Committing Player All Star in that Season	-0.001	-0.004	0.001	0.002
	(0.013)	(0.013)	(0.016)	(0.020)
Committing Player Starter (% Games of Season)	-0.020	-0.019	-0.038	-0.015
	(0.023)	(0.022)	(0.026)	(0.032)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	15,978	15,978	15,978	15,978

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. This model also includes an interaction term between the Black Referee dummy and all committing player characteristics and statistics. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level.

***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.3B
In-Group Bias

	Dependent Variable: Incorrect Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.035	0.000	0.019	-0.014
	(0.024)	(0.025)	(0.036)	(0.042)
Committing Player Black	-0.015		0.004	
	(0.011)		(0.021)	
Disadvantaged Player Black	0.000	-0.008	0.024	0.010
	(0.013)	(0.015)	(0.021)	(0.025)
Black Coach for Committing Player	-0.010	-0.002	-0.018	-0.011
	(0.010)	(0.012)	(0.019)	(0.035)
Home Game for Committing Player	0.002	-0.003	-0.003	-0.006
	(0.008)	(0.010)	(0.014)	(0.015)
Playoff Game	0.013	0.011		
	(0.016)	(0.017)		
Committing Player All Star in that Season	0.007	0.003	0.019	0.047
	(0.023)	(0.027)	(0.032)	(0.046)
Committing Player Starter (% Games of Season)	-0.022	0.052	-0.032	0.037
	(0.018)	(0.037)	(0.024)	(0.064)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	3,161	3,161	3,161	3,161

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. This model also includes an interaction term between the Black Referee dummy and all committing player characteristics and statistics. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level.

∗∗∗, ∗∗, and ∗ signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.3C
In-Group Bias

	Dependent Variable: Incorrect No-Calls			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.028**	0.023*	0.035***	0.026**
	(0.014)	(0.013)	(0.012)	(0.012)
Committing Player Black	-0.002		-0.008	
	(0.009)		(0.008)	
Disadvantaged Player Black	-0.013	-0.011	-0.010	-0.008
	(0.009)	(0.009)	(0.010)	(0.011)
Black Coach for Committing Player	-0.004	0.004	-0.000	0.007
	(0.006)	(0.010)	(0.009)	(0.012)
Home Game for Committing Player	0.010**	0.011**	0.013**	0.013**
	(0.005)	(0.005)	(0.005)	(0.006)
Playoff Game	-0.007	-0.010		
	(0.010)	(0.011)		
Committing Player All Star in that	-0.024*	-0.043***	-0.014	-0.027
Season	(0.012)	(0.012)	(0.015)	(0.018)
Committing Player Starter	-0.017	-0.002	-0.020	-0.041
(% Games of Season)	(0.021)	(0.020)	(0.025)	(0.032)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	12,817	12,817	12,817	12,817

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. This model also includes an interaction term between the Black Referee dummy and all committing player characteristics and statistics. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. ***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.4
Bias in Type-I Errors Conditional on Referee Race

	Dependent Variable: Incorrect Call			
	White Referees		Black Referees	
	(I)	(II)	(III)	(IV)
Committing Player Black			0.021	0.020
			{0.365}	{0.432}
Committing Player White	0.011	-0.018		
	{0.334}	{0.261}		
Disadvantaged Player Black	-0.020	0.018	0.027	0.054
	(0.016)	(0.026)	(0.018)	(0.054)
Black Coach for Committing Player	-0.015	-0.020	-0.002	-0.026
	(0.014)	(0.033)	(0.015)	(0.041)
Home Game for Committing Player	0.002	0.017	-0.002	-0.026
	(0.010)	(0.022)	(0.015)	(0.041)
Playoff Game	0.002		0.028	
	(0.016)		(0.036)	
Committing Player All Star in that Season	0.009	-0.002	0.005	0.016
	(0.023)	(0.032)	(0.033)	(0.071)
Committing Player Starter (% Games of Season)	-0.022	-0.010	0.044	0.029
	(0.018)	(0.027)	(0.029)	(0.066)
Game FE	NO	YES	NO	YES
Season FE	YES	NO	YES	NO
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	1,777	1,777	1,384	1,384

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}. ∗∗∗, ∗∗, and ∗ signify statistical significance at 1%, 5%, and 10% level, respectively.

Appendix Table A.5
Overall Bias

	Call		Incorrect Call		Incorrect No-Call	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Com Player Black	-0.051	-0.056	0.026	0.113	0.042	0.023
	(0.057)	(0.068)	(0.079)	(0.131)	(0.051)	(0.049)
(Ref Age)x	0.003*	0.003	-0.001	-0.004	-0.000	-0.000
(Com Player Black)	(0.001)	(0.002)	(0.002)	(0.004)	(0.001)	(0.001)
(Ref Experience)x	-0.004***	-0.004**	0.001	0.006	-0.001	-0.000
(Com Player Black)	(0.001)	(0.002)	(0.003)	(0.004)	(0.001)	(0.001)
(Ref Southern)x	-0.010	-0.012	0.006	0.026	-0.006	-0.005
(Com Player Black)	(0.013)	(0.015)	(0.022)	(0.034)	(0.013)	(0.012)
Game FE	NO	YES	NO	YES	NO	YES
Season FE	YES	NO	YES	NO	YES	NO
Com Player FE	NO	NO	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES	YES	YES
N	15,978	15,978	3,161	3,161	12,817	12,817

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parenthesis and are clustered at the referee level.

$***$, $**$, and $*$ signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.6A
In-Group Bias in Calls

	Dependent Variable: Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	-0.007	-0.010	-0.014	-0.017
	(0.015)	(0.016)	(0.017)	(0.018)
(Ref Age)x(Com Player Black)	0.003*	0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
(Ref Exp)x(Com Player Black)	-0.004***	-0.004**	-0.004***	-0.004**
	(0.001)	(0.002)	(0.002)	(0.002)
(Ref Southern)x(Com Player Black)	-0.009	-0.012	-0.009	-0.011
	(0.013)	(0.013)	(0.014)	(0.015)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	15,978	15,978	15,978	15,978

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. Additionally, specification 1 and 3 control for committing player race, while all four control for disadvantaged player race. All observations are at the player-event level. Standard errors are in parenthesis and are clustered at the referee level.

***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.6B
In-Group Bias in Type-I Errors

	Dependent Variable: Incorrect Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.017	0.004	-0.002	-0.036
	(0.024)	(0.027)	(0.032)	(0.039)
(Ref Age)x(Com Player Black)	-0.002	-0.001	-0.004	-0.002
	(0.003)	(0.003)	(0.004)	(0.005)
(Ref Exp)x(Com Player Black)	0.002	0.001	0.006	0.002
	(0.003)	(0.003)	(0.004)	(0.005)
(Ref Southern)x(Com Player Black)	0.003	-0.002	0.026	0.031
	(0.020)	(0.025)	(0.034)	(0.038)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	3,161	3,161	3,161	3,161

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. Additionally, specification 1 and 3 control for committing player race, while all four control for disadvantaged player race. All observations are at the player-event level. Standard errors are in parenthesis and are clustered at the referee level.

$***$, $**$, and $*$ signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.6C
In-Group Bias in Type-II Errors

	Dependent Variable: Incorrect No Call			
	(I)	(II)	(III)	(IV)
(Black Ref)x(Com Player Black)	0.035***	0.034***	0.037***	0.031**
	(0.013)	(0.013)	(0.011)	(0.012)
(Ref Age)x(Com Player Black)	-0.002	-0.002	-0.002	-0.002
	(0.001)	(0.002)	(0.001)	(0.002)
(Ref Exp)x(Com Player Black)	0.001	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
(Ref Southern)x(Com Player Black)	-0.012	-0.013	-0.011	-0.012
	(0.013)	(0.013)	(0.011)	(0.012)
Game FE	NO	NO	YES	YES
Season FE	YES	YES	NO	NO
Player FE	NO	YES	NO	YES
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	12,817	12,817	12,817	12,817

Notes: ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. Additionally, specification 1 and 3 control for committing player race, while all four control for disadvantaged player race. All observations are at the player-event level. Standard errors are in parenthesis and are clustered at the referee level.

***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.7
Bias in Type-II Errors Conditional on Referee Race

	Dependent Variable: Incorrect No-Call			
	White Referees		Black Referees	
	(I)	(II)	(III)	(IV)
Com Player Black			0.134	0.117
			{0.118}	{0.179}
(Ref Age)x			-0.002	-0.002
(Com Player Black)			(0.002)	(0.002)
(Ref Experience)x			0.000	0.001
(Com Player Black)			(0.002)	(0.002)
(Ref Southern)x			-0.020	-0.012
(Com Player Black)			(0.019)	(0.019)
Com Player White	-0.071	-0.060		
	{0.400}	{0.491}		
(Ref Age)x	0.002	0.001		
(Com Player White)	(0.003)	(0.002)		
(Ref Experience)x	-0.000	0.000		
(Com Player White)	(0.002)	(0.002)		
(Ref Southern)x	0.001	0.002		
(Com Player White)	(0.018)	(0.017)		
Game FE	NO	YES	NO	YES
Season FE	YES	NO	YES	NO
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	6,891	6,891	5,926	5,926

Notes: Means for black referee age, experience, and being southern-born are 49.376, 15.249, and 0.521, respectively. Means for white referee age, experience, and being southern-born are 44.504, 13.430, and 0.359, respectively. ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}

$***$, $**$, and $*$ signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.8A
Type-II Errors of Black Referees with Peer Effects

	Dependent Variable: Incorrect No-Call			
	(W,W)	(W,B)	(W,B)/(B,B)	(B,B)
	(I)	(II)	(III)	(IV)
Committing Player Black	0.059	0.228*	0.171*	0.014
	{0.726}	{0.056}	{0.104}	{0.941}
(Ref Age)x(Com Player Black)	0.000	-0.004	-0.003	-0.001
	(0.004)	(0.003)	(0.002)	(0.004)
(Ref Experience)x(Com Player Black)	-0.004	0.002	0.002	0.005
	(0.003)	(0.002)	(0.002)	(0.005)
(Ref Southern)x(Com Player Black)	-0.029	-0.022	-0.016	0.009
	(0.034)	(0.024)	(0.021)	(0.031)
Game FE	NO	NO	NO	NO
Season FE	YES	YES	YES	YES
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	1,792	2,969	4,134	1,165

Notes: With two white peers, means for black referee age, experience, and being southern-born are 49.651, 15.729, and 0.505, respectively; for mixed-race peers, the means are 49.137, 14.999, and 0.517; for mixed race or two black peers, the means are 49.260, 15.046, and 0.528; with two black peers the means are 49.571, 15.164, and 0.556. ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets} ***, **, and * signify statistical significance at 1%, 5%, and 10%, respectively.

Appendix Table A.8B
Type-II Errors of White Referees with Peer Effects

	Dependent Variable: Incorrect No-Call			
	(W,W)	(W,B)	(W,B)/(B,B)	(B,B)
	(I)	(II)	(III)	(IV)
Committing Player White	-0.090	-0.069	-0.079	-0.048
	{0.911}	{0.601}	{0.816}	{0.787}
(Ref Age)x(Com Player White)	0.003	0.001	0.002	0.003
	(0.004)	(0.004)	(0.003)	(0.005)
(Ref Experience)x(Com Player White)	-0.001	0.000	-0.000	-0.003
	(0.004)	(0.004)	(0.003)	(0.005)
(Ref Southern)x(Com Player White)	-0.003	-0.009	0.001	0.034
	(0.033)	(0.031)	(0.022)	(0.039)
Game FE	NO	NO	NO	NO
Season FE	YES	YES	YES	YES
Player FE	NO	NO	NO	NO
Referee FE	YES	YES	YES	YES
Player Characteristics	YES	YES	YES	YES
Player Statistics	YES	YES	YES	YES
Game Controls	YES	YES	YES	YES
Team Controls	YES	YES	YES	YES
N	1,978	3,417	4,913	1,496

Notes: With two white peers, means for white referee age, experience, and being southern-born are 44.484, 13.259, and 0.344, respectively; for mixed-race peers, the means are 44.272, 13.120, and 0.362; for mixed race or two black peers, the means are 44.513, 13.500, and 0.365; with two black peers the means are 45.042, 14.161, and 0.371. ***Player Characteristics*** include (for both committing and disadvantaged player) all-star status in a season, percentage of games started in a season, average minutes played, number of games played, and position (guard, forward, center). ***Player Statistics*** include (both for the committing player and the disadvantaged player) assists, blocks, defensive rebounds, offensive rebounds, personal fouls, turnovers, steals, free throw attempts, two point shot attempts, three point shot attempts, and points per game (all measured on a per 48 minutes played basis). ***Game Controls*** include points scored and indicators for home game status for the committing player, for whether the game was won by the committing team, and whether the game was a playoff game. ***Team Controls*** include a dummy variable indicating whether a team made the playoffs that season (for both committing and disadvantaged players) and a dummy variable indicating if the committing player's coach is black. All observations are at the player-event level. Standard errors are in parentheses and are clustered at the referee level. Wild bootstrapped p-values are {in brackets}. ∗∗∗, ∗∗, and ∗ signify statistical significance at 1%, 5%, and 10%, respectively.