Toward an Understanding of Discrimination: The Case of Parsing Multiple Sources *

Majid Ahmadi[1], Gwen-Jirō Clochard[2,3,*], Jeffrey Lachman[4], and John A. List[3,4,5]

[1]Georgia Institute of Technology
[2]Institute of Social and Economic Research, Osaka University
[3]Joint Initiative for Latin American Experimental Economics
[4]Walmart
[5]University of Chicago
[*]Corresponding author: gjclochard@iser.osaka-u.ac.jp

February 2024

Abstract

When multiple forces potentially underlie discriminatory behavior it is difficult to parse the sources of discrimination, rendering proposed policy solutions as speculative. This study advances an empirical approach to parse two specific channels of discrimination: customer side and manager side bias. To showcase our general idea, we combine proprietary data and several publicly available data sets to identify channels of discrimination within the Major League Baseball draft. In doing so, we show that customer preferences are importantly linked to the players drafted at the top end of the draft–players who are most likely to receive immediate public attention and end up playing for the club. Alternatively, we find manager homophily in the latter parts of the draft, when players who receive little attention and have a scant chance to ever play with the club are drafted. The opportunity cost of expressing such preferences is considerable foregone success of the club. Our results have general implications for future work measuring discrimination and how to tackle the multiple channel problem.

Keywords: racial bias, customer discrimination, sports

JEL Classification: J70, J71, M51

*We are grateful to Charlie Brobst and Charles Shi for excellent research assistance.

1 Introduction

One would be hard-pressed to find an issue as important and timely for a nation as equitable treatment of its citizens. Yet our understanding of the sources of discrimination within the marketplace remains speculative. This attribution problem is especially difficult when multiple sources of discriminatory behavior exist. In his seminal work, Becker (1971) highlighted this difficulty when identifying three potential sources of bias that can lead to discrimination: managers (in hiring decisions), co-workers, and consumers. The literature has recognized each of these factors in the past five decades, both theoretically and empirically (see, e.g., Arrow (1972); Phelps (1972); Akerlof (1997); Riach and Rich (2002).

While there has been empirical headway measuring and detecting discriminatory patterns (see, e.g., Jowell and Prescott-Clarke (1970); Goldin and Rouse (2000); Riach and Rich (2002); List (2004); Bohren et al. (2019, 2022), an important lesson learned from the vast literature on discrimination is that data availability places severe constraints on efforts to understand the nature of discrimination, forcing researchers to present aggregate measures of discrimination. Such aggregation, by necessity, estimates the force of all potential channels. Yet, there are crucial areas of modern economies wherein multiple sources of discrimination can influence behaviors. For example, consider the restaurant that hires "front room" workers (e.g., bartenders, hosts, and waiters) as well as "back room" workers (i.e., dishwashers, cooks, food prep, and managers). In this manner, any organization that has both publicly facing employees and "behind the scenes" employees provides a fertile testing ground to parse the source and nature of discriminatory preferences (Combes et al., 2016). Such an exploration permits, in theory, the ability for the researcher to disentangle potentially complementary and competing preferences.

We showcase our general idea by combining rich proprietary data with several publicly available data sources that we scraped from the internet to examine choices made in the Major League baseball (MLB) draft from 2000-19. Over this time period, approximately 12,000 players were drafted by MLB clubs. Our proprietary data comes from a partner MLB team, and contains scouting evaluations, as well as information about each scout, including race. Our publicly available data are drawn from five sources. First, we scraped from the website Perfect Game USA,[1] which provides a list of approximately 60,000 players names and whether they were drafted in the MLB. Our second public data source is a database of high school and college prospects who were drafted in the MLB. The data includes over 10,000 prospects, including individual characteristics and performance. In addition to the information of drafted players, we also scraped all players who played at least one game in the MLB for the years 2008-2019.[2] We then categorized their racial affiliation using the racial identification data from Rosenman et al. (2022).

Our third public data source is the universe of scouting directors for all teams in the MLB, including their race. This database is used to evaluate whether the race of scouting directors influences teams' draft strategy and whether scouting directors tend to recruit players of their own race. Our fourth publicly available data source is a metric of fan bias. To create this metric, we take advantage of a naturally occurring event: the Black Lives Matter movement in June and July 2020. Following demonstrations, every team in the MLB posted messages on social media about Black Lives Matter. These posts received much attention (both negative and positive) from fans. The measure of fan bias is a combination of several elements. We collected reactions from the teams' posts on Facebook and Twitter. We use the data of Kumar and Pranesh (2021) to analyze the content of the messages. We use their algorithm based on approximately 9,000 tweets related to Black Lives Matter, which the authors manually labeled, to create a dummy variable for each message (equal to 1 if the message is flagged as negative). Our fifth data source relates to the attendance of stadiums for every game played over the period 2008-2019. We use these data as an alternative measure of fan bias.

Exploring these data within standard Mincer-type models, we generate several insights. First, there

[1] Accessible here.
[2] Gathered from the ESPN website.

is no link between race and the probability of a player being drafted. If anything, African-American players are slightly more likely than White players to be drafted. Second, Using data on signing bonuses, conditional on being drafted, player payment is similar across race. Indeed, controlling for player quality, there is some evidence that Asian and White players are paid lower amounts than their peers.

Yet, this aggregation masks a key insight in our data, leading to our third result: the data are consonant with discrimination when individual channels are explored while alternative channels are muted. For example, empirical results indicate that team draft of African-American players and fan bias are strongly correlated in the early rounds of the draft. Yet, this relationship does not exist in later rounds of the draft. Together, these results are consistent with the MLB club accounting for customers preferences when drafting players who will receive more scrutiny and public attention. [3]

Our fourth major result is that when we examine the race of scouting directors across the MLB clubs, we find evidence consistent with homophily in selections made in latter rounds of the MLB draft. For example, conditional on player quality, we find that scouting directors have a bias toward players of their own race in that players from their racial group are approximately 10 percent more likely to get drafted in rounds 16-40. This finding is consistent with the notion that when the stake are decreased, and the public scrutiny is diminished, scouting directors are more likely to express their own preferences. This follows from the fact that such players have a very low probability of ever reaching a MLB club.

Our final result is that expression of such biased preferences, either due to customers or managers, is associated with certain economic costs. For example, our estimates indicates that teams are willing to draft players who are worth approximately $500k less when fan bias increases by one standard deviation. Yet, when we examine scouting biases, the effect is financially less important because the players they draft, while less likely to make the MLB club than their peers, still are longshots to make the majors.

[3] We caution the reader not to invoke causality here. In fact, the causality could very well run in the opposite direction, with fans being biased due to the lack of representation of certain races on their favorite team.

Even though our results are drawn from a specific corner of the world, we view the major lessons learned as having broad import. Following List (2020) in considering the generalizability of our insights, considering the nature of preferences, beliefs, and constraints we observe much similarities between our environment and economically more important ones. For example, our setting is one in which multiple sources of discrimination operate simultaneously. We cannot imagine any important industry, whether business to business or business to consumer, that does not have similar trade-offs. As such, the broad lesson learned is that scholars who examine data for discrimination using establishment level data run the risk when measuring aggregates. If there is a tension in biased preferences between management and customers, then key aggregates can underestimate, or mask, key biases.

A second implication of our results is that relative prices matter when it comes to discrimination. For example, an organization might express preferences to cater to their customers, but when the rewards for expressing such bias decrease, the organizational preferences begin to take greater import. Viewed from the other side of the coin, homophily is important, but not important enough to express when the stakes grow too large.

Third, from a policy perspective, given that scalable prescriptions to combat bias necessarily must pinpoint correct channels to advance efficient policies, our study provides a framework for modeling and estimating relevant sources of discrimination. In addition, methodologically, our study provides what is necessary to generate insights that can be applied in other settings. Most importantly, it highlights that standard economic models provides useful first steps into what markets and settings are most ripe to explore.

The remainder of the paper is organized as follows. In Section 2, we present a short survey of the literature on consumer-side vs supplier-side discrimination. In Section 3, we present the data used in this paper. We present our empirical strategy in Section 4 and the results in Section 5. Finally, Section 6 concludes.

2 Literature review

In his seminal book, Becker (1971) identifies three sources of bias that can lead to discrimination: managers (in hiring decisions), co-workers and consumers. This paper speaks to discrimination on the part of managers (here, the scouting team) and consumers (here, the fans). This paper also speaks to the literature that specifically applies these principles in the context of sports. In this Section, we briefly summarize these strands of the literature and discuss how the present research relates to them.

2.1 Biased managers and co-workers

The literature on biased managers has primarily focused on documenting discrimination in hiring, with the most common method being the correspondence study (Bertrand and Mullainathan, 2004; Kline et al., 2022). Evidence suggests not only that racial discrimination remains pervasive in the US labor market, but that it is also not decreasing over time (Quillian et al., 2017). Evidence of discrimination has also been identified in other fields such as housing (Auspurg et al., 2019), transportation (Liebe and Beyer, 2021), but also online dating (Potârcă and Mills, 2015), clothing stores (Bourabain and Verhaeghe, 2019), childcare (Boyd-Swan and Herbst, 2019), speeding tickets in police encounters (Aggarwal et al., 2022) or charitable giving (Landry et al., 2006).

This literature is closely linked to the literature of homophily in the labor market. This strand of literature has found that people tend to gather in groups that are homogeneous with respect to age, gender, religion, ethnicity, and even behavior and ability (Bell, 1981; McPherson et al., 2001). In the labor market, Giuliano et al. (2009) found that within a given supermarket chain, White managers tend to recruit fewer African-Americans than African-American managers, without being able to distinguish whether the difference comes from supply or demand. This paper shows that homophily can persist even in a highly competitive environment, in particular when scrutiny is not too present.

The present paper also speaks to the literature on biased managers and performance. Glover et al. (2017) show that under biased managers, the productivity of minority workers is decreased. Bandiera et al. (2009) further show that when managers' pay is fixed, workers are more productive if they are connected to the manager. In the medical literature, FitzGerald and Hurst (2017) review the literature on doctors' bias towards patients of their own groups, and found correlational evidence that doctors' implicit bias and quality of care are linked. In this paper, we find that players who are drafted by a scouting director of their racial group are not of lower productivity (measured in terms of performance).

In the literature on biased managers and performance, Lippens et al. (2021) also show that penalties for discriminating are negatively correlated with discriminatory behavior. In this paper, we find that scrutiny of actions (measured by draft round) is negatively correlated with biased decisions on the part of scouting teams, with the race of the scouting director playing no role in draft decisions for early rounds of the draft.

The literature on racial bias has also investigated the link between biased coworkers and productivity and retention. For instance, Giuliano et al. (2009) show that workers are less likely to quit when more coworkers have the same race, in particular for White and Asian workers, while Bygren (2010) finds that workers are *less* likely to leave if the share of the opposite gender is larger. Hedegaard and Tyran (2018) show that workers are willing to forego 8% of their earnings to work with someone with the same ethnicity. Hjort (2014) and Hamilton et al. (2012) further show that team diversity in Kenya and in California can decrease team productivity. In this paper, however, due to a lack of data on coworker (here, other players) bias, we cannot investigate the link between coworker bias and labor market outcomes, nor can we investigate whether having biased players already in the team influences draft strategies.

2.2 Biased consumers

The present paper also speaks to the literature on the effects of consumer bias. Similarly to the presence of biased managers, there exists ample evidence that some consumers are biased against certain groups of workers (Lang and Spitzer, 2020). For instance, (Doleac and Stein, 2013) find that Black online sellers receive fewer and lower offers than White sellers, Zussman (2013) finds evidence of discrimination against Arab sellers of used cars in Israel, Parrett (2011) and Kelley et al. (2024) find evidence of discrimination against waitresses and female online sales agents, respectively. In this paper, we do not focus on absolute levels of consumer bias, but on the relative level of fan bias, and its link with teams' draft strategies.

There also exist evidence that consumer bias is linked with firms' hiring strategies. Bar and Zussman (2017) show that a significant share of customers discriminate against Arab workers in Israel, and that this bias is then passed on to firms, who reduce their employment of Arab workers, and those who hire Arab workers charge significantly lower service prices. Boyd-Swan and Herbst (2019) also find evidence of a strong link between the share of minority children attending a childcare center and the probability that a minority applicant would be selected. Interestingly, Kuppuswamy and Younkin (2020) find that employers in the film industry discriminate against minority actors because of a perceived bias of the public, when movies with a diverse cast actually raise more money than movies with homogeneous casts. In this paper, we show that teams whose fans are more biased against African-Americans tend to recruit fewer of these players.

Holzer and Ihlanfeldt (1998) show that the racial composition of an establishment's customer base is closely related to the race of hired workers, particularly in jobs that require close contact with customers. Combes et al. (2016) also find that more French residents widens more the wage gap between French and African workers with customer contact than in sectors without such contact. Laouenan (2017) further show evidence of discrimination from both employers and customers against African-American entry-level workers in the US.

In terms of the effects of performance of firms who tailor to biased consumers, Leonard et al. (2010) show that in areas with larger shares of Whites, having more African-American employees is correlated with smaller sales. They further show that matching employee demographics to those of potential customers has little payoff, except when the customers do not speak English. Avery et al. (2012) also identify a positive correlation between racial and ethnic representativeness of the workforce and productivity, particularly in stores with larger minority customer bases.

In this paper, we do not find that the teams draft strategies are driven by their local demographic composition, but we do find that teams with biased are willing to recruit lower-quality White players to cater to their fans' bias.

2.3 Racial bias in sports

In this section we present the literature on racial bias in the context of sports. Because of the amount of easily accessible data, sports have long been used to test for theories (Kahn, 2000). In particular, there is ample evidence of bias against racial minorities and immigrants in sports, from broadcast commentary in the NFL (Schmidt and Coe, 2014) or in European soccer (McLoughlin, 2020; MacLeod and Newall, 2022; Principe and van Ours, 2022), to amateur clubs recruiting fewer immigrants (Gomez-Gonzalez et al., 2021), and even bias in the eligibility of international players in the MLB (Hauptman, 2009; Ross and James Jr, 2015). This paper contributes to this literature by disentangling two potential sources of bias in the decision to draft minority players: biased management (scouts and scouting directors) and biased fans.

There also exists evidence of consumer bias in sports. For instance, Kahn and Sherer (1988) show evidence of an attendance against African-American players in the NBA, with White players significantly raising attendance by 5700-13000 per year per White player in the NBA in the 1980s, although not confirmed by Dey (1997); McCormick and Tollison (2001); Kahn and Shah (2005). Similarly, Brown and Jewell (1995) show that White college basketball players generate substantially more revenues that African-American players, "providing a strong incentive for colleges to discriminate against recruiting black student-athletes", and Kanazawa and Funk (2001) show that White players significantly increase Nielsen television ratings for the team.

To identify evidence of consumer discrimination, a strand of literature has specifically focused on situations where teams play no decisions, such as the market for memorabilia or online evaluations of players. For instance, Broyles and Keen (2010); Bryson and Chevalier (2015) show no aggregate racial bias from fans in NBA trading cards and in the Fantasy Premier League. Sur and Sasaki (2020) also show no evidence of discrimination in player evaluation in Indian cricket. However, there exists evidence that racial discrimination exists in the baseball cards market, with Nardinelli and Simon (1990) showing that cards for non-White players have a lower price than those of White players, and List (2004) showing that minority sellers (of baseball cards) get offered significantly lower initial and final offers than White sellers. On the contrary, Depken II and Ford (2006) show that African-American and Hispanic players were *favored* by voters for the All-Star Game ballot, relative to White players.

Similarly to the literature on consumer bias in general, several papers have documented a positive correlation between the racial composition of basketball teams and that of their respective metropolitan areas (Brown et al., 1991; Burdekin and Idson, 1991; Hoang and Rascher, 1999; Burdekin et al., 2005). However, Kahn and Sherer (1988); Kahn and Shah (2005) find no evidence of racial bias in the NBA draft, controlling for college performance. In this paper, we find opposite results, with no evidence of links between the metropolitan composition and team draft strategy, but evidence of racial bias in recruiting, controlling for prospect quality.

A few papers have also examined the evolution of fan bias in the context of sports over time. Kahn (2009) found evidence that racial discrimination from fans in the NBA decreased since the 1980s, while Maennig and Mueller (2021) shows that fan bias in attendance tended to grow in the 2000s relative to the 1980s. In this paper, we document that the link between scouting directors' race and draft decisions only exists for early seasons in our data (2008-2015).

3 Data

We use several sources of data to analyze sources of racial bias in the baseball draft. Descriptive statistics for the players are shown in Table A.1.

3.1 Data on players for the extensive margin analysis

The first data set is scraped from the website Perfect Game USA,[4] which provides a list of approximately 60k players names and whether they were drafted at all to the Major League Baseball. We merge the players' last names and the data from Rosenman et al. (2022) to compute the most likely race of players. This measure of race is not very precise, which is why we refine it below for a subset of players who were drafted (and were therefore easier to identify). The website also provides an evaluation of the players (*PG Grade*). Because this prospect quality metric is not perfect (as shown below, scout evaluation of players could depend on their race), when possible, we use our own measure of player quality (see below).

3.2 Data on players

The second data source is a database of high school and college prospects who were drafted in the MLB. The data includes over 10k prospects with their characteristics and performance. Of these, we

[4] Accessible here.

collected available racial information for approximately 2,000 players. We supplement the data with a measure of prospect quality. This metric is the result of a machine learning algorithm developed in a previous paper (Ahmadi et al., 2022), which takes as input the performance of players in high school or college and is trained using the output of players who made it to the MLB.

Although this database does not cover the universe of drafted players, it is a good sample of the overall population of baseball players drafted between 2000 and 2019. From 2012-2017, 41 of the 204 players drafted in the first round of the MLB draft (20.1%) were African-American.[5] In our data for the same period, 34 of 225 players were African-American. Our data is most accurate in the early rounds because the more prominent players are easier to verify and collect racial data, while racial data is more difficult to collect in later rounds of the draft. We were also able to obtain data on signing bonuses for a subset of players.

In addition to the information of drafted players, we also scraped all players who played at least one game in the MLB for the years 2008-2019.[6] We then categorized their racial affiliation using the racial identification data from Rosenman et al. (2022). It should be noted that this data also includes players who never participated in the draft but who were recruited as free agents. This includes a lot of foreign-born players, making the share of Hispanic and Asian players higher than in the draft data.

3.3 Data on scouting teams

The third data source is the universe of scouting directors for all teams in the MLB, including their race. This database is used to evaluate whether the race of scouting directors influences teams' draft strategy and whether scouting directors tend to recruit players of their own race.

The fourth data source comes from a partner MLB team, and contains scouting evaluations, as well

[5]The discrepancy in the number of drafted players is likely due to some variation in what constitutes a "Round 1" of the MLB draft, where short compensation rounds are often squeezed in between full rounds.

[6]Gathered from the ESPN website.

as information about each scout, including their race. We use this data to investigate whether scouts evaluate more favorably players of their own race, controlling for player quality.

3.4 Data on fan bias from Black Lives Matter

The fifth data source is a metric of fan bias. To create this metric, we take advantage of a naturally occurring event: the Black Lives Matter movement in June and July 2020. Following demonstrations, every team in the MLB posted messages on social media about Black Lives Matter. These posts received a lot of (negative and positive) reactions from fans. The measure of fan bias is a combination of several elements. We collected reactions from the teams' posts on Facebook and Twitter. We use the data of Kumar and Pranesh (2021) to analyze the content of the messages. We use their algorithm based on approximately 9,000 tweets related to Black Lives Matter, which the authors manually labeled, to create a dummy variable for each message (equal to 1 if the message is flagged as negative). Examples of tweets classified as biased and non-biased are shown in Table B.1. We defined bias in comments as the proportion of negative comments for the team's post. We also collect the ratio of "angry" reactions to "likes" on Facebook posts. Restricting replies to the tweets to people who follow the team's Twitter page yields similar metrics (see Table B.2).

To capture the overall bias from the three metrics described above (tweets, comments on Facebook posts, and angry reactions), we calculated the sum of the three normalized metrics, which we then normalized to create a composite Index of fan bias. The measures for each team are shown in Table B.3.

3.5 Data on bias in stadium attendance

The fifth data source relates to the attendance of stadiums for every game played over the period 2008-2019. We use this data to get an alternative measure of fan bias. To do so, we first normalize the attendance at the team level. Second, we take the average attendance at the monthly level. Third, we regress the normalized attendance on the share of African-American players who played at least one game over the month, controlling for the share of wins in this month and year-fixed effects. For simplicity and comparability with the index variable defined above, we normalize the value of the variable and take the negative of the coefficient. The variable "Bias attendance" is therefore positive if, for a given team, increasing the share of African-American players is associated with fewer people attending the games. The attendance bias is presented in Table B.4.

3.6 Other data sources

For the external side analysis, we complement the analysis by including the following controls: racial and political composition of fans ("Fans Black" and "Fans Republicans"), using the data from a Morning Consult Brand Intelligence survey on racial and political composition of each major U.S. sports teams (Silverman, 2020); racial and political composition at the local level, with the share of African-Americans ("Local Blacks") and the county-level vote share for Donald Trump in the 2020 presidential election ("Local Republicans").[7] Because we showed that the race of the scouting directors is related to teams' draft strategy, we included two additional regressions, controlling for: a dummy variable on whether the team has had an African-American scouting director at any point during the 2000-2020 period; and the years that the team was under an African-American scouting director during the same period.

4 Empirical strategy

In this section we present the estimation strategy for our analysis of racial bias in baseball.

[7]For the New York Mets and New York Yankees, we took the local shares of Republicans and African-Americans for all of New York City. For the Washington Nationals, we took the values for Washington, D.C. There is no value for the vote for Donald Trump for the Toronto Blue Jays, the team being located in Canada.

4.1 Extensive margin analysis

We first analyze the probability for players of being drafted at all in the MLB, as a function of their race and their quality (as defined by their best ranking on the website Perfect game). The players' race is defined using the most likely race from their last name (matched with data from Rosenman et al. (2022)). We estimate the following linear probability model:

$$\mathbb{P}(\text{Drafted}_i = 1) = \beta_0 + \beta_R \text{ Race}_i + \gamma \text{ PG Grade}_i + \epsilon_i \tag{1}$$

where Drafted is a dummy variable for whether the player was drafted to the MLB, Race is a categorical variable representing the player's race and PG Grade is the evaluation given to players by the Perfect Game website (proxy for player quality).

Once we analyzed the link between race and the probability of being drafted, we analyze the amount of the signing bonus, conditional on the player quality and for the players we have the information (approximately 1,100 players), estimating the following linear equation:

$$\text{Bonus}_i = \beta_0 + \beta_R \text{ Race}_i + \gamma \text{ ML quality prediction}_i + \epsilon_i \tag{2}$$

4.2 Internal bias analysis

To evaluate the bias on the side of teams' scouting teams, we perform two analyses. First, we evaluate whether players are more likely to be drafted into a particular team when they are of the same racial group as the scouting director from that team. To do so, we use our data set of drafted players to create a data set with 30 observations per player (one for each team), and two dummy variables: one for whether the player shares the same race as the scouting director, and one for whether the player was drafted by the team. We then estimate the following linear probability model:

$$\mathbb{P}(\text{Drafted}_{i,j} = 1) = \beta_0 + \beta_R \text{ Race}_i + \gamma \text{ X}_i + \delta \text{ Same race}_{i,j} + \epsilon_{i,j} \tag{3}$$

In Equation 3 we control for the quality of the prospect as well as the distance to the team and the scouting director, which was shown to be linked to the probability of being drafted by a specific team (Ahmadi et al., 2022). The coefficient of interest is δ We estimate the model for the entire sample, and then divide the sample depending on their draft rounds: rounds 1-5, rounds 6-15 and rounds 16+.

The second dimension for the internal bias analysis is the evaluation of players by scouts. Using the data from a partner MLB team, we want to analyze the link between player's race and the scout's race. We estimate the following linear equation, with η being the coefficient of interest:

$$Evaluation_{i,j} = \beta_0 + \beta_R \text{ Race}_i + \gamma_R \text{ Race}_j + \delta \text{ ML quality prediction}_i + \eta \text{ Same race}_{i,j} + \epsilon_{i,j} \tag{4}$$

4.3 External bias analysis

For the external bias analysis, we replicate the methodology used for the internal bias by creating a data set with 30 observations per player and combining it with the level of bias from the teams. Because our measure of fan bias (see above) is defined as a bias against African-Americans, we focus primarily on the link between fan bias and the draft of African-American players. We then estimate the following equation, with η being the main coefficient of interest.

$$\mathbb{P}(\text{Drafted}_{i,j} = 1) = \beta_0 + \beta_R \text{ Race}_i + \gamma \text{ X}_i + \delta \text{ Fan bias}_j + \eta \text{ Black}_i \times \text{ Fan bias}_j + \epsilon_{i,j} \tag{5}$$

As for the internal side analysis, we estimate Equation 5 on the entire sample as well as disaggregated by draft round.

We then compute the share of African-American players drafted by teams and estimate its correlation with the level of fan bias using a linear framework.

We then perform the same analysis using our alternative measure of fan bias, the link between the share of African-American players and stadium attendance.

4.4 Cost of racial bias

To assess the cost of expressing racial bias, we use the performance metric of WAR (Win Above Replacement), a standard measure of performance used in professional baseball. Because many players never reach the majors and therefore do not have a WAR metric, we assign them a value of 0. (In the appendix, we also perform the analysis on players who did reach the majors and find similar results.)

We estimate the following linear regression for player i drafted by team j:

$$\begin{aligned} WAR_i &= \beta_0 + \beta_R \text{ Race}_i + \gamma \text{ Fan bias}_j + \delta \text{ White}_i \times \text{ Fan bias}_j + \eta \text{ Same race}_{i,j} \\ &\quad + \kappa \text{ ML quality prediction}_i + \epsilon_{i,j} \end{aligned} \tag{6}$$

where Same race refers to whether the player is of the same race as the scouting director. The main coefficients of interest are δ, which measures whether White players drafted by teams with biased fans are of worse quality (if δ is negative), and η which measures if players drafted by a team whose scouting director is of the same race are less good.

5 Results

5.1 Extensive margin analysis: probability of reaching the MLB and signing bonus

The first step of analysis is the aggregate analysis of the extensive margin, i.e., whether minority players are less likely to be drafted in the MLB than White players.

Table 1: Extensive margin analysis

	Drafted (1)
Asian	-0.109
	(0.139)
Black	0.073
	(0.051)
Hispanic	-0.081
	(0.057)
Other races	-0.020
	(0.111)
PG Grade	1.803***
	(0.026)
Constant	-18.271***
	(0.237)
N	61,362
Pseudo-R^2	0.476

Note: The dependent variable is a dummy variable for whether the player was drafted. The variable PG Grade corresponds to the evaluation of players on the Perfect Game website. Racial identity was obtained from players' last names. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

Result 1a: There is no statistical difference between races in the probability of being drafted on the extensive margin, controlling for quality.

Results from the estimation of Equation 1 are presented in Table 1. They clearly show that, conditional on player quality, there is no link between race and the probability of being drafted. If anything, African-American players are slightly more likely than White players to be drafted into the MLB.

Result 1b: Controlling for prospect quality, there is no statistical difference in signing bonus amounts based on race.

Using data on signing bonuses, conditional on being drafted, we show in Table 2 that controlling for player quality, no racial group is paid significantly differently than White players. On the contrary, controlling for player quality, Asian and White players appear to be paid the least relative to all other races (as the point estimates for other races are positive).

Table 2: Analysis of signing bonuses

	Log(Signing bonus) (1)
2 or more races	0.321
	(0.196)
Asian	−0.660
	(0.434)
Black	0.112
	(0.118)
Hispanic	0.183
	(0.137)
Native American	0.210
	(0.368)
ML quality prediction	7.244***
	(0.256)
Constant	10.659***
	(0.079)
N	1,162
R^2	0.415

Note: The dependent variable is the log of the signing bonus obtained by players. The omitted category is White players. The variable "ML quality prediction" is the output of a Machine Learning algorithm computing the probability of a prospect to reach the MLB, as a function of their high-school / college performance. * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

We now turn to the analysis of bias on the side of teams, first focusing on internal bias, i.e., bias in favor of the scouts' own racial group.

5.2 Internal Bias: Bias in Scouting

In this section, we present the results of the evaluation of racial bias on the part of teams' scouting departments. First, we examine whether players are more likely to be drafted by teams whose scouting directors are of the same race. Second, we examine whether individual scouts evaluate players of their own race more favorably.

5.2.1 Bias in favor of the scouting director's racial group

First, we analyze teams' draft strategies as a function of the race of their scouting directors. To do so, we created a synthetic entry for each team for all players considered by teams between 2000 and 2019. We then estimate a logistic regression to assess the probability of a given team drafting each player, conditional on player quality and a dummy variable equal to one if the player and scouting director share the same race. We control for player race and distance to teams and scouting directors (which has been shown to be correlated with the probability of being drafted, Ahmadi et al. (2022)).

Result 2: Players are more likely to be drafted by teams whose scouting director is of the same race. The correlation is only statistically significant for late rounds of the draft (16+).

The results shown in Table 3 (column 1) illustrate that there is a bias in favor of players of the same race as the scouting director. Furthermore, we show in columns 2 to 4 of Table 3 that the bias in favor of players of the same race as the scouting directors only appears for players drafted in late rounds of the draft (rounds 16+). This result corroborates the findings of a previous paper (Ahmadi et al., 2022), where the propinquity bias (drafting players who are spatially close) of scouting directors plays a stronger role for lower picks of the draft, where they have more decision-making power. Disaggregating by time (Table C.1), we find that the results hold only for the early years available in our data (2008-2015), but not for later years. This result may indicate that the bias of scouting directors becomes less important as statistics and models become more heavily weighted.

These results suggest that scouting directors have a bias toward players of their own race, and this bias is evident when scouting directors have a lot of decision-making power, for players who have a very lower chance of ever reaching the MLB and are less visible.

Table 3: Link between being the same race as the scouting director and the probability of being drafted by a given team

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-15 (3)	Rounds 16+ (4)
Same race as scouting director	0.142**	0.141	−0.037	0.300**
	(0.071)	(0.135)	(0.128)	(0.125)
2 or more races	0.248*	0.243	0.111	0.420
	(0.136)	(0.221)	(0.238)	(0.345)
Asian	0.347	0.173	0.420	1.200*
	(0.272)	(0.434)	(0.604)	(0.616)
Black	0.097	0.116	0.009	0.149
	(0.082)	(0.165)	(0.150)	(0.140)
Hispanic	0.039	0.130	−0.238	0.187
	(0.099)	(0.191)	(0.183)	(0.170)
Native American	0.127	0.150	0.042	0.286
	(0.292)	(0.407)	(0.475)	(1.040)
ML predictor	0.014	−0.110	0.572	0.020
	(0.137)	(0.233)	(0.434)	(0.289)
Constant	−3.483***	−3.441***	−3.426***	−3.567***
	(0.083)	(0.165)	(0.159)	(0.143)
Performance metrics and distance controls	Yes	Yes	Yes	Yes
N	77,310	21,782	21,433	26,489
Pseudo R^2	0.0003	0.0002	0.0009	0.0014
Mean dep. variable	0.033	0.033	0.033	0.033

Note: The dependent variable is the probability of being drafted, conditional on player quality. The omitted category is White players. * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

5.2.2 Bias in scouting evaluations

After examining the relationship between scouting director race and player draft prospects, we turned to examining the internal evaluations of a Major League Baseball team. We combined a partner MLB team's internal reports on players with the race of their scouts. The outcome variable considered here is group rating, a variable that takes values from 1 (most skilled) to 7 (least skilled).

Result 3: Scouts rate players of their own race more favorably, controlling for prospect quality.

As can be seen in Table 4, controlling for the predicted player quality (which is very strongly correlated with player's rating), we observe a very strong bias in favor of players who are the same race as the scout (a negative coefficient means a higher rating). Quantitatively, being the same race as the scout is associated with moving 1/5 of a category higher in the evaluation. Combined with the fact that African-American and Hispanic scouts tend to be sent to observe players of lower quality (Table C.2), this creates a strong bias in favor of White players.

Disaggregating by rounds in which the players are drafted (Table C.3), we observe that the bias is stronger for players in rounds 6+ (columns 3 and 4). We can interpret this finding as an indication that scouts' bias in favor of players of their own race is stronger for players whose quality is less "exceptional", i.e., for players who do not stand out as future stars. Disaggregating by time (Table C.4), we see that the scout's bias in favor of players of their own race has increased over the period studied.

Interestingly, we also observe that Hispanic, African-American and mixed-race players tend to receive a "premium" above their skill level in group ratings, controlling for prospect quality, relative to White players.

Table 4: Link between race and internal scouting evaluation

	Same race as scout (1)
Same race as scout	-0.211^{***}
	(0.065)
ML quality prediction	-4.558^{***}
	(0.144)
Player race	
2 or more races	-0.255^{*}
	(0.139)
Asian	-0.291
	(0.306)
Black	-0.215^{***}
	(0.074)
Hispanic	-0.483^{***}
	(0.082)
Native American	-0.393
	(0.287)
Scout race	
Black	0.201^{***}
	(0.077)
Hispanic	-0.089
	(0.074)
Constant	4.747^{***}
	(0.074)
N	2,493
R^2	0.304

Note: The dependent variable is the group scoring from the internal reports for a partner MLB team. A lower value means a better evaluation. The omitted categories are White players and White scouts. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

5.3 External Bias: Fan Bias and Draft Strategy

In this section, we examine another potential source of racial bias in baseball: fans. To do so, we use the naturally occurring event of all teams posting a message of support for the Black Lives Matter movement in June or July 2020. We collected all comments on the related posts on Twitter and Facebook. Using the dataset and methodology of Kumar and Pranesh (2021), we then computed a measure of bias for each team as the average of negative comments for each team. We then computed a composite index based on the different measures of negative fan reactions (replies on Twitter, comments on Facebook, and the ratio of "angry" to "like" reactions on Facebook).

We also complement the analysis by using an alternative measure of fan bias related to the link between attendance and racial composition of the team. Using this alternative measure does not change the results.

5.3.1 Fan bias and Team Draft Strategy

Result 4: African-American players are less likely to be drafted by teams with more biased fans, and the effect is statistically significant only for the early rounds of the draft (1-5).

Results from the estimation of Equation 5 are displayed in Table 5. Results indicate that African-American players are less likely to be drafted into a team if the team's fans are biased. Contrary to what was found in the internal bias, the correlation is stronger for *early* rounds of the draft (Rounds 1-5, Column 2), where scrutiny from the fans is higher. For later rounds of the draft, the point estimate is still negative, but the magnitude is much lower and the correlation is no longer significant (Columns 3 and 4).

Table 5: Fan bias and probability of being drafted

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-15 (3)	Rounds 16+ (4)
2 or more races	0.070	0.036	0.114	0.148
	(0.124)	(0.194)	(0.213)	(0.326)
Asian	0.246	0.066	0.503	0.895
	(0.265)	(0.418)	(0.595)	(0.605)
Black	−0.041	−0.010	−0.011	−0.110
	(0.057)	(0.123)	(0.110)	(0.092)
Hispanic	−0.074	−0.035	−0.176	−0.053
	(0.080)	(0.156)	(0.151)	(0.135)
Native American	−0.070	−0.122	0.070	−0.037
	(0.294)	(0.417)	(0.459)	(1.019)
ML quality prediction	−0.028	−0.129	0.607	0.062
	(0.139)	(0.235)	(0.439)	(0.290)
Index of fan bias	0.029	0.020	0.024	0.022
	(0.023)	(0.042)	(0.044)	(0.040)
Black × Index of fan bias	−0.176***	−0.289**	−0.048	−0.141
	(0.059)	(0.130)	(0.111)	(0.094)
Constant	−3.339***	−3.326***	−3.469***	−3.303***
	(0.040)	(0.089)	(0.092)	(0.067)
N	71,511	20,676	19,784	23,869
Pseudo R^2	0.0005	0.0001	0.0009	0.0008
Mean dep. variable	0.033	0.033	0.033	0.033

Note: The dependent variable is a dummy of whether the player was drafted by a given team. The dependent variable Index of fan bias is defined above. The other dependent variables are player's race (White is the omitted category) and prospect quality. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

5.3.2 Aggregate effect of fan bias

Result 5: On the aggregate, teams with more biased fans draft fewer African-American players, and the result is statistically significant in the early rounds of the draft.

After showing the effect of fan bias on the probability of being drafted, we here present aggregate results. To do so, we regressed the total share of African-American players on the index of fan bias from the reactions to Black Lives Matter posts. Results are presented in Table 6 and in Figure 1.

Overall, there is a negative correlation between the index of team bias and the share of African-American players drafted by the team. In terms of magnitude, a one standard deviation increase in fan bias is associated with a 2.8 p.p. decrease in the share of African-American players drafted by the team. This corresponds to approximately 17% of the league average, or 0.34 standard deviations.

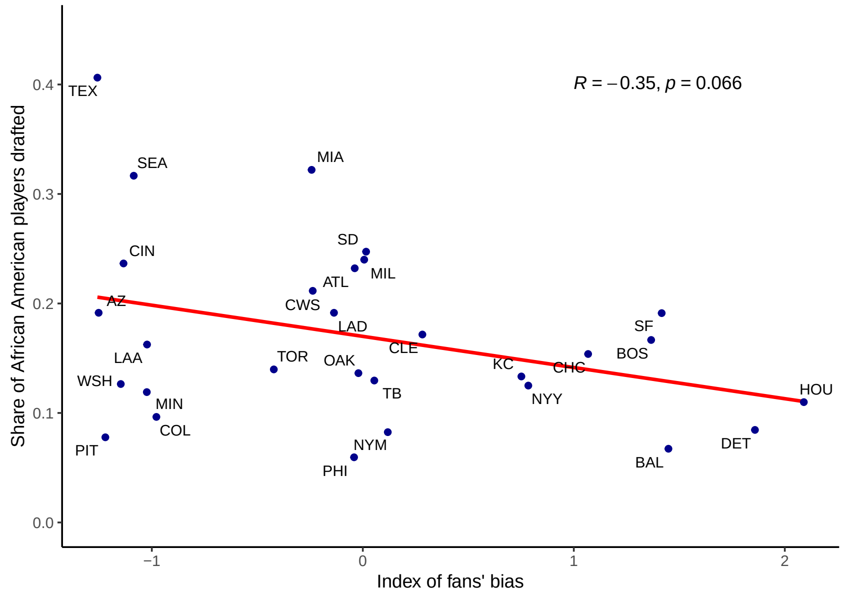
Table 6: Correlation fans bias and draft of Black players by round of draft

	All rounds (1)	Rounds 1-5 (2)	Rounds 6+ (3)
Index	-0.028^{*}	-0.049^{**}	-0.014
	(0.015)	(0.022)	(0.019)
Constant	0.170^{***}	0.139^{***}	0.176^{***}
	(0.015)	(0.022)	(0.018)
N	29	29	29
R^2	0.119	0.154	0.019

Note: The dependent variable in Column 1 is the share of African American players drafted in any round of the draft. In Column 2, it is the share of African American players drafted in the first five rounds. In Column 3, it is the share of African American players drafted in later rounds (6+). The independent variable is the index of fans' bias, defined above. No Facebook post related to Black Lives Matter was found for the St Louis Cardinals, explaining why the sample size is 29. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

After examining the overall relationship between fan bias and team draft strategy, we differentiated whether the relationship is stronger for earlier or later picks in the draft. The rationale behind this analysis is that players drafted early in the draft (especially in the first 5 rounds) are more likely to make it to the MLB, and therefore receive more media exposure and are more financially important to teams.

Figure 1: Index of fan bias and share of African-American players drafted



Note: The X-axis corresponds to the composite index of fan bias computed with reactions to posts related to the Black Lives Matter movement. The Y-axis corresponds to the share of African-American players drafted during the study period (2008-20). The red line is the regression line.

The results of the estimations are shown in Table 6. The results indicate that team draft of African-American players and fan bias are strongly correlated in the early rounds of the draft (column 2), while it does not seem to be significant in later rounds of the draft (although the coefficient is still negative, column 3). These results suggest that that team owners may take fan bias into account when planning to draft players who are more highly publicized. The results for early rounds of the draft are robust to the inclusion of controls such as fan base composition, local characteristics, and the presence or absence of African-American scouting directors (Table D.1).

An important point to note is that we cannot make a causal claim that biased fans cause fewer African-American players to be drafted. In fact, the causality could very well run in the opposite direction, with fans being biased due to the lack of representation of talented African-American players by their favorite team.

5.3.3 Robustness check using fan bias in attendance

Result 6: On the aggregate, using an alternative measure of fan bias based on stadium attendance, teams with more biased fans draft fewer African-American players.

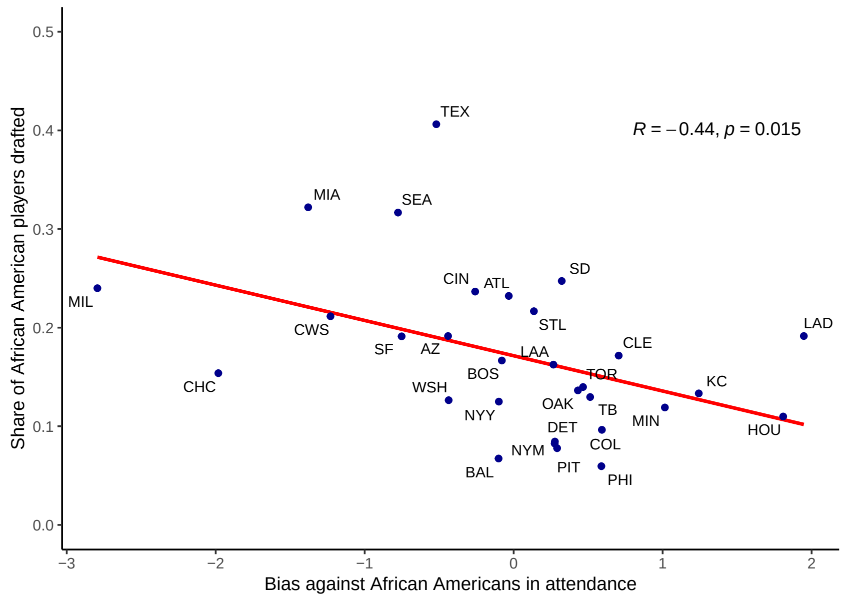
To complement our analysis of the link between fan bias and drafting strategies, we perform similar estimations using an alternative measure of fan bias, based on attendance in the stadium. The measure, defined in Section 2, represents the correlation between stadium attendance and the share of African-American players in the team. For comparability, the coefficient is normalized and inverted, so that a positive coefficient corresponds to more biased fans.

The two measures of fans bias (the index from the BLM posts) and the bias in attendance are positively correlated, although insignificantly (Table E.1).

The attendance bias is also strongly correlated with the share of African-American players drafted, even when controlling for local factors (Figure 2, Table E.2, columns 1 and 2). Interestingly, the two measures of indices are still significantly correlated with the share of African-American players drafted (Table E.2, column 3). Disaggregating by draft round (Table E.3), the correlations between draft strategy and attendance bias are no longer statistically significant, but the magnitude of the correlation follows the pattern observed in the previous section, with a stronger link in first rounds of the draft.

The use of an alternative measure of fan bias therefore confirms the link between fan bias and draft strategies. It therefore appears that team executives could take into account fans preferences when making draft decisions.

Figure 2: Attendance bias and share of African-American players drafted



Note: The X-axis corresponds to the composite index of fan bias. The Y-axis corresponds to the share of African-American players drafted during the study period (2008-20). The red line is the regression line.

5.4 Cost of racial bias

To measure the cost of racial bias, we estimate the value that players bring to the team in terms of performance, as measured by WAR (Win Above Replacement) for their first year in the majors. For the main specification, we assign a value of 0 to all players who never made it to the majors (or have not yet made it to the majors).

Result 7a: Teams with biased fans draft less qualified White players in the early rounds of the draft, resulting in significant sporting losses.

Result 7b: Homophily of the scouting director does not lead to less qualified players, mainly because the bias manifests itself on players with a very low probability of ever reaching the majors.

The results are presented in Table 7. As can be seen in Column 1, White players drafted by teams with more biased fans tend to perform less well than those drafted by less biased teams, and conversely, non-White players drafted by teams with more biased fans tend to perform better, although the results are nos statistically significant.

Disaggregating by draft round, we find that in the early rounds of the draft, where fan bias matters for draft strategy (see above), White players drafted by teams with biased fans tend to perform worse than those drafted by teams with less biased fans. In terms of magnitude, a one standard deviation increase in fan bias for white players is associated with a decrease of approximately 0.1 WAR. Given the approximate value of $5 million per WAR (Steinberg, 2022), this estimate suggests that teams are willing to draft players who are worth approximately $500k less when fan bias increases by one standard deviation.

Regarding internal bias, we see in Table 7 that there is no correlation between players' WAR and being of the same race as the scouting director. For the late rounds of the draft (column 4), where the internal bias was identified, we observe a negative coefficient, indicating that players drafted late who belong to the same racial group as the scouting director perform slightly worse, but the correlation is clearly not statistically significant ($p = 0.76$).

In summary, the results indicate that teams with biased fans are willing to forgo significant amounts of money by drafting inferior white players, while their internal bias (bias in favor of the scouting director's race) does not appear to be financially important.

As a robustness check, we perform the same analysis for the subset of players who reached the MLB and find (Table F.1) similarly that fan bias has a cost in terms of performance, with white players drafted by teams with biased fans performing worse on average.

Table 7: Cost of racial bias in performance of drafted players

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-15 (3)	Rounds 16+ (4)
Index of fan bias	0.014	0.080*	0.014	−0.010
	(0.017)	(0.048)	(0.022)	(0.016)
White players × Index of fan bias	−0.024	−0.095*	−0.034	0.017
	(0.020)	(0.054)	(0.026)	(0.019)
Same race as scouting director	0.035	0.053	0.053	−0.008
	(0.031)	(0.084)	(0.038)	(0.028)
Constant	−0.042	0.032	−0.035	−0.023
	(0.032)	(0.093)	(0.042)	(0.029)
Race-fixed effects	Yes	Yes	Yes	Yes
Prospect quality metric	Yes	Yes	Yes	Yes
N	2,417	689	657	838
R^2	0.037	0.037	0.026	0.068

Note: The dependent variable is the WAR (Win Above Replacement) value. For players who did not reach the MLB (or have not done it yet), the value 0 is set. The index of fan bias is defined in Section 3. Player's race-fixed effects are included (White is the omitted category). * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

6 Conclusion

In this paper, we empirically test two channels of racial bias in the context of the Major League Baseball draft using a set of publicly available and proprietary data. We find that consumer bias against African Americans is associated with the drafting of African American players, but only in the early rounds of the draft. This comes at a substantial cost to teams, as teams with more biased fans are willing to draft lower quality white players. On the other hand, we find that scouting directors' homophily - drafting players from their own racial group - occurs in later rounds of the draft, when scrutiny is low, and at no significant financial cost to teams.

This paper has several implications. First, while our results are drawn from the specific context of baseball, we believe that our findings are applicable to many important economic settings where multiple sources of bias interact simultaneously. For example, this work echoes previous findings (Combes et al., 2016) that the consequences of consumer bias vary with exposure to customers. Looking at bias at the aggregate level of bias may mask different and potentially conflicting sources of bias.

Our second implication is that firms respond to incentives to discriminate. In our data, we find that homophily on the part of scouting directors does not matter when stakes are high, but does when incentives are low. On the other hand, we find that teams cater to their fans' biases in the early rounds of the draft when stakes are high, but not when stakes and scrutiny are low.

Understanding and disaggregating the sources of bias also has important policy implications. For example, methods to reduce scouting team bias, such as contact interventions (Paluck et al., 2019; Clochard, 2024), could reduce the influence of homophily in the late rounds of the draft, but would have little effect on racial bias in the early rounds of the draft, where decisions are driven by fan bias.

References

Aggarwal, P., A. Brandon, A. Goldszmidt, J. Holz, J. A. List, I. Muir, G. Sun, and T. Yu (2022). Using high-frequency location data to evaluate racial bias in policing. Technical report, NBER Working Paper.

Ahmadi, M., N. Durst, J. Lachman, J. A. List, M. List, N. List, and A. T. Vayalinkal (2022). Nothing propinks like propinquity: Using machine learning to estimate the effects of spatial proximity in the major league baseball draft. Technical report, National Bureau of Economic Research.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society 65*(5), 1005–1027.

Arrow, K. J. (1972). Models of job discrimination. In *Racial discrimination in economic life*. Lexington, Massachusetts.

Auspurg, K., A. Schneck, and T. Hinz (2019). Closed doors everywhere? a meta-analysis of field experiments on ethnic discrimination in rental housing markets. *Journal of Ethnic and Migration Studies 45*(1), 95–114.

Avery, D. R., P. F. McKay, S. Tonidandel, S. D. Volpone, and M. A. Morris (2012). Is there method to the madness? examining how racioethnic matching influences retail store productivity. *Personnel Psychology 65*(1), 167–199.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica 77*(4), 1047–1094.

Bar, R. and A. Zussman (2017). Customer discrimination: evidence from israel. *Journal of Labor Economics 35*(4), 1031–1059.

Becker, G. S. (1971). *The economics of discrimination*. University of Chicago press.

Bell, R. R. (1981). Friendships of women and of men. *Psychology of Women Quarterly 5*(3), 402–417.

Bertrand, M. and S. Mullainathan (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.

Bohren, J. A., P. Hull, and A. Imas (2022). Systemic discrimination: Theory and measurement. Technical report, National Bureau of Economic Research.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American economic review 109*(10), 3395–3436.

Bourabain, D. and P.-P. Verhaeghe (2019). Could you help me, please? intersectional field experiments on everyday discrimination in clothing stores. *Journal of Ethnic and Migration Studies 45*(11), 2026–2044.

Boyd-Swan, C. and C. M. Herbst (2019). Racial and ethnic discrimination in the labor market for child care teachers. *Educational Researcher 48*(7), 394–406.

Brown, E., R. Spiro, and D. Keenan (1991). Wage and nonwage discrimination in professional basketball: do fans affect it? *American Journal of Economics and Sociology 50*(3), 333–345.

Brown, R. W. and R. T. Jewell (1995). Race, revenues, and college basketball. *The Review of Black Political Economy 23*(3), 75–90.

Broyles, P. and B. Keen (2010). Consumer discrimination in the nba: An examination of the effect of race on the value of basketball trading cards. *The Social Science Journal 47*(1), 162–171.

Bryson, A. and A. Chevalier (2015). Is there a taste for racial discrimination amongst employers? *Labour Economics 34*, 51–63.

Burdekin, R. C., R. T. Hossfeld, and J. K. Smith (2005). Are nba fans becoming indifferent to race? evidence from the 1990s. *Journal of sports economics 6*(2), 144–159.

Burdekin, R. C. and T. L. Idson (1991). Customer preferences, attendance and the racial structure of professional basketball teams. *Applied Economics 23*(1), 179–186.

Bygren, M. (2010). The gender composition of workplaces and men's and women's turnover. *European Sociological Review 26*(2), 193–202.

Clochard, G.-J. (2024). A meta-analytic experiment on the contact hypothesis. Technical report, ISER Working Paper.

Combes, P.-P., B. Decreuse, M. Laouenan, and A. Trannoy (2016). Customer discrimination and employment outcomes: theory and evidence from the french labor market. *Journal of Labor Economics 34*(1), 107–160.

Depken II, C. A. and J. M. Ford (2006). Customer-based discrimination against major league baseball players: Additional evidence from all-star ballots. *The Journal of Socio-Economics 35*(6), 1061–1077.

Dey, M. S. (1997). Racial differences in national basketball association players' salaries: A new look. *The American Economist 41*(2), 84–90.

Doleac, J. L. and L. C. Stein (2013). The visible hand: Race and online market outcomes. *The Economic Journal 123*(572), F469–F492.

FitzGerald, C. and S. Hurst (2017). Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics 18*(1), 1–18.

Giuliano, L., D. I. Levine, and J. Leonard (2009). Manager race and the race of new hires. *Journal of Labor Economics 27*(4), 589–631.

Glover, D., A. Pallais, and W. Pariente (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics 132*(3), 1219–1260.

Goldin, C. and C. Rouse (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American economic review 90*(4), 715–741.

Gomez-Gonzalez, C., C. Nesseler, and H. M. Dietl (2021). Mapping discrimination in europe through a field experiment in amateur sport. *Humanities and Social Sciences Communications 8*(1), 1–8.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of participatory and Labor-managed Firms*, pp. 99–138. Emerald Group Publishing Limited.

Hauptman, D. (2009). The need for a worldwide draft to level the playing field and strike out the national origin discrimination in major league baseball. *Loy. LA Ent. L. Rev. 30*, 263.

Hedegaard, M. S. and J.-R. Tyran (2018). The price of prejudice. *American Economic Journal: Applied Economics 10*(1), 40–63.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hoang, H. and D. Rascher (1999). The nba, exit discrimination, and career earnings. *Industrial Relations: A Journal of Economy and Society 38*(1), 69–91.

Holzer, H. J. and K. R. Ihlanfeldt (1998). Customer discrimination and employment outcomes for minority workers. *The Quarterly Journal of Economics 113*(3), 835–867.

Jowell, R. and P. Prescott-Clarke (1970). Racial discrimination and white-collar workers in britain. *Race 11*(4), 397–417.

Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of economic perspectives 14*(3), 75–94.

Kahn, L. M. (2009). The economics of discrimination: Evidence from basketball. Technical report, IZA Discussion Paper.

Kahn, L. M. and M. Shah (2005). Race, compensation and contract length in the nba: 2001–2002. *Industrial Relations: A Journal of Economy and Society 44*(3), 444–462.

Kahn, L. M. and P. D. Sherer (1988). Racial differences in professional basketball players' compensation. *Journal of Labor Economics 6*(1), 40–61.

Kanazawa, M. T. and J. P. Funk (2001). Racial discrimination in professional basketball: Evidence from nielsen ratings. *Economic Inquiry 39*(4), 599–608.

Kelley, E. M., G. V. Lane, M. Pecenco, and E. A. Rubin (2024). Customer discrimination in the workplace: Evidence from online sales. Technical report, National Bureau of Economic Research.

Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics 137*(4), 1963–2036.

Kumar, S. and R. R. Pranesh (2021). Tweetblm: A hate speech dataset and analysis of black lives matter-related microblogs on twitter. Technical report, arXiv preprint arXiv:2108.12521.

Kuppuswamy, V. and P. Younkin (2020). Testing the theory of consumer discrimination as an explanation for the lack of minority hiring in hollywood films. *Management Science 66*(3), 1227–1247.

Landry, C. E., A. Lange, J. A. List, M. K. Price, and N. G. Rupp (2006). Toward an understanding of the economics of charity: Evidence from a field experiment. *The Quarterly journal of economics 121*(2), 747–782.

Lang, K. and A. K.-L. Spitzer (2020). Race discrimination: An economic perspective. *Journal of Economic Perspectives 34*(2), 68–89.

Laouenan, M. (2017). ‘hate at first sight’: Evidence of consumer discrimination against african-americans in the us. *Labour Economics 46*, 94–109.

Leonard, J. S., D. I. Levine, and L. Giuliano (2010). Customer discrimination. *The Review of Economics and Statistics 92*(3), 670–678.

Liebe, U. and H. Beyer (2021). Examining discrimination in everyday life: A stated choice experiment on racism in the sharing economy. *Journal of Ethnic and Migration Studies 47*(9), 2065–2088.

Lippens, L., S. Baert, and E. Derous (2021). Loss aversion in taste-based employee discrimination: Evidence from a choice experiment. *Economics Letters 208*, 110081.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics 119*(1), 49–89.

List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Technical report, National Bureau of Economic Research.

MacLeod, S. A. and P. W. Newall (2022). Investigating racial bias within australian rules football commentary. *PLoS one 17*(7), e0272005.

Maennig, W. and S. Q. Mueller (2021). Consumer and employer discrimination in professional sports markets–new evidence from major league baseball. Technical Report 69, Hamburg Contemporary Economic Discussions.

McCormick, R. E. and R. D. Tollison (2001). Why do black basketball players work more for less money? *Journal of Economic Behavior & Organization 44*(2), 201–219.

McLoughlin, D. (2020). Racial bias in football commentary (study).

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology 27*(1), 415–444.

Nardinelli, C. and C. Simon (1990). Customer racial discrimination in the market for memorabilia: The case of baseball. *The Quarterly Journal of Economics 105*(3), 575–595.

Paluck, E. L., S. A. Green, and D. P. Green (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy 3*(2), 129–158.

Parrett, M. (2011). Customer discrimination in restaurants: Dining frequency matters. *Journal of Labor Research 32*(2), 87–112.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review 62*(4), 659–661.

Potârcă, G. and M. Mills (2015). Racial preferences in online dating across european countries. *European Sociological Review 31*(3), 326–341.

Principe, F. and J. C. van Ours (2022). Racial bias in newspaper ratings of professional football players. *European Economic Review 141*, 103980.

Quillian, L., D. Pager, O. Hexel, and A. H. Midtbøen (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences 114*(41), 10870–10875.

Riach, P. A. and J. Rich (2002). Field experiments of discrimination in the market place. *The economic journal 112*(483), F480–F518.

Rosenman, E. T., S. Olivella, and K. Imai (2022). Race and ethnicity data for first, middle, and last names.

Ross, S. F. and M. James Jr (2015). A strategic legal challenge to the unforeseen anticompetitive and racially discriminatory effects of baseball's north american draft. *Colum. L. Rev. Sidebar 115*, 127.

Schmidt, A. and K. Coe (2014). Old and new forms of racial bias in mediated sports commentary: The case of the national football league draft. *Journal of Broadcasting & Electronic Media 58*(4), 655–670.

Silverman, A. (2020). Demographic data shows which major sports fan bases are most likely to support or reject social justice advocacy. Technical report, Morning Consult.

Steinberg, R. (2022). Mlb has lowest percentage of african american baseball players in 3 decades: Report.

Sur, P. K. and M. Sasaki (2020). Measuring customer discrimination: evidence from the professional cricket league in india. *Journal of Sports Economics 21*(4), 420–448.

Zussman, A. (2013). Ethnic discrimination: Lessons from the israeli online market for used cars. *The Economic Journal 123*(572), F433–F468.

Appendices

A Descriptive statistics

Table A.1: Descriptive statistics

Variable	Mean	Std. Dev.	N
Panel A. Player race			
White	0,691	0,462	3248
Asian	0,005	0,070	3248
Black	0,167	0,373	3248
Hispanic	0,103	0,303	3248
Native American	0,005	0,072	3248
2 or more races	0,029	0,167	3248
Panel B. Metrics of player quality			
ML predicted quality	0,157	0,119	12886
Group evaluation	3,756	1,208	9940
Panel C. Signing bonus			
Signing bonus	567 691,04	913 114,63	2153
Panel D. Rounds of the draft			
Rounds 1-5	0,220	0,414	7762
Rounds 6-15	0,302	0,459	7762
Rounds 15+	0,478	0,500	7762

B Definitions of fan bias

Table B.1: Examples of comments with bias metric

Text	**Bias**
Love this team.	0
Over 600 likes better than I expected.	0
SO I guess the Blue Jays are a bunch of communists then? I sure hope lots of leftists watch baseball haha	1
I support the one player that did not join the BLM sheep.	1

Note: The "Bias" measure takes the value 1 if the algorithm estimates that the answer is biased. The estimation is done using Kumar and Pranesh (2021).

Table B.2: Restricting to the replies of followers

	Bias replies tweets (1)
Bias only followers	0.517***
	(0.088)
Constant	0.120***
	(0.021)
N	30
R^2	0.552

Note: The dependent variable is the average level of bias for replies to the tweets of teams involving Black Lives Matter. The independent variable is the same share, but restricted to comments posted by people following the team. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

Table B.3: Index of fan bias

Team	Index	Team	Index	Team	Index
Arizona Diamondbacks	-1,252	Houston Astros	2,090	Philadelphia Phillies	-0,041
Atlanta Braves	-0,038	Kansas City Royals	0,752	Pittsburgh Pirates	-1,220
Baltimore Orioles	1,449	Los Angeles Angels	-1,023	San Diego Padres	0,016
Boston Red Sox	1,367	Los Angeles Dodgers	-0,137	San Francisco Giants	1,417
Chicago Cubs	1,068	Miami Marlins	-0,243	Seattle Mariners	-1,086
Chicago White Sox	-0,237	Milwaukee Brewers	0,007	St. Louis Cardinals	
Cincinnati Reds	-1,134	Minnesota Twins	-1,024	Tampa Bay Rays	0,054
Cleveland Guardians	0,282	New York Mets	0,118	Texas Rangers	-1,258
Colorado Rockies	-0,978	New York Yankees	0,784	Toronto Blue Jays	-0,422
Detroit Tigers	1,859	Oakland Athletics	-0,021	Washington Nationals	-1,147

Note: The index of fan bias is defined using the reactions to the posts related to Black Lives Matter, as defined in Section 5.3. A higher coefficient means that fans reacted more negatively to the posts.

Table B.4: Fan bias in attendance

Team	Bias	Team	Bias	Team	Bias
Arizona Diamondbacks	-0.440	Houston Astros	1.809	Philadelphia Phillies	0.589
Atlanta Braves	-0.033	Kansas City Royals	1.242	Pittsburgh Pirates	0.292
Baltimore Orioles	-0.101	Los Angeles Angels	0.267	San Diego Padres	0.323
Boston Red Sox	-0.079	Los Angeles Dodgers	1.948	San Francisco Giants	-0.752
Chicago Cubs	-1.982	Miami Marlins	-1.379	Seattle Mariners	-0.776
Chicago White Sox	-1.229	Milwaukee Brewers	-2.794	St. Louis Cardinals	0.136
Cincinnati Reds	-0.258	Minnesota Twins	1.015	Tampa Bay Rays	0.514
Cleveland Guardians	0.705	New York Mets	0.276	Texas Rangers	-0.519
Colorado Rockies	0.592	New York Yankees	-0.099	Toronto Blue Jays	0.466
Detroit Tigers	0.277	Oakland Athletics	0.431	Washington Nationals	-0.436

Note: The attendance bias is defined as the normalized, inverted coefficient of the regression of the share of African American players on attendance. See description in Section 5.3. A higher coefficient means that when the share of African American players increase, fewer people attend games.

C Internal Bias: Additional Results

Table C.1: Same race as the scouting director and probability of being drafted, by years

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-15 (3)	Rounds 16+ (4)
Panel A. Baseline: All years combined				
Same race as scouting director	0.139**	0.052	−0.001	0.330***
	(0.070)	(0.129)	(0.127)	(0.123)
Panel B. Early seasons (2008-2014)				
Same race as scouting director	0.309**	0.089	0.126	0.617**
	(0.134)	(0.229)	(0.271)	(0.247)
Panel C. Late seasons (2015-2020)				
Same race as scouting director	0.062	0.163	−0.081	0.147
	(0.085)	(0.168)	(0.146)	(0.146)

Note: The outcome variable is the probability of being drafted by each team. Controls include player race fixed effects, distance to the team and the ML prediction of player quality. Panel A represents the baseline estimation, with all years included. Panels B and C represent the results from similar estimations, with players divided with a median split with respect to the season they are drafted. In Panel B, all players drafted before the season 2015 are used. For Panel C, all players drafted after the season 2015 are used. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table C.2: Race of scouts and player quality

	ML quality predictor
	(1)
Black scout	-0.027^{***}
	(0.004)
Hispanic scout	-0.011^{***}
	(0.004)
Constant	0.173^{***}
	(0.001)
N	10,883
R^2	0.005

Note: The dependent variable is the predicted quality of the prospect, coming from the ML algorithm taking as output high school or college data and trained on the players who made it to the MLB. A higher value of the outcome is associated with a higher quality of prospect. The omitted category is White scouts. * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table C.3: Same race as scouts, by draft round

	All rounds	Rounds 1-5	Rounds 6-15	Rounds 16+
	(1)	(2)	(3)	(4)
Same race as scout	-0.254^{***}	-0.188^{**}	-0.232^{**}	-0.241^{**}
	(0.054)	(0.087)	(0.096)	(0.100)
2 or more races	-0.280^{**}	-0.191	-0.222	-0.219
	(0.137)	(0.169)	(0.395)	(0.397)
Asian	-0.313	-0.450	-0.149	-0.143
	(0.307)	(0.377)	(0.773)	(0.777)
Black	-0.243^{***}	-0.173	-0.471^{***}	-0.432^{***}
	(0.071)	(0.118)	(0.117)	(0.123)
Hispanic	-0.530^{***}	-0.362^{***}	-0.470^{***}	-0.447^{***}
	(0.081)	(0.135)	(0.141)	(0.145)
Native American	-0.402	-0.014	-0.213	0.461
	(0.286)	(0.311)	(0.774)	(1.095)
ML quality prediction	-4.592^{***}	-2.957^{***}	-4.141^{***}	-4.110^{***}
	(0.143)	(0.196)	(0.333)	(0.341)
Constant	4.803^{***}	3.716^{***}	5.104^{***}	5.094^{***}
	(0.059)	(0.102)	(0.105)	(0.110)
N	2,493	780	787	720
R^2	0.301	0.236	0.178	0.180

Note: The dependent variable is the group scoring from the internal reports for a partner MLB team. A lower value means a better evaluation. The omitted category is White players. Each column corresponds to different subsamples with respect to the drafting round. * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table C.4: Relationship between race and internal scouting evaluation, evolution with time

	Baseline (All seasons) (1)	Early seasons (2008-2014) (2)	Late seasons (2015-2020) (3)
Same race as scout	-0.254^{***}	-0.154^{**}	-0.253^{***}
	(0.054)	(0.075)	(0.071)
2 or more races	-0.280^{**}	-0.188	-0.232
	(0.137)	(0.159)	(0.202)
Asian	-0.313	0.349	-0.496
	(0.307)	(0.499)	(0.373)
Black	-0.243^{***}	-0.021	-0.255^{***}
	(0.071)	(0.090)	(0.098)
Hispanic	-0.530^{***}	-0.453^{***}	-0.543^{***}
	(0.081)	(0.104)	(0.110)
Native American	-0.402	-0.592	-0.309
	(0.286)	(0.434)	(0.356)
ML quality prediction	-4.592^{***}	-3.058^{***}	-5.292^{***}
	(0.143)	(0.177)	(0.203)
Constant	4.803^{***}	4.038^{***}	5.095^{***}
	(0.059)	(0.085)	(0.076)
N	2,493	913	1,580
R^2	0.301	0.264	0.309

Note: The dependent variable is the group scoring from the internal reports for a partner MLB team. A lower value means a better evaluation. The omitted category is White players. In column 1, all the sample of players is included. In column 2, players drafted before 2015 (the median split of our sample) are used. In column 3, players drafted after 2015 are used. * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

D External Bias: Robustness check with the inclusion of controls

To test for the robustness of our findings, we controlled for two types of factors: (1) indicators of racial and political composition of fanbases. (2) indicators of the local situation of teams with the Share of African-Americans in the county of the team and the vote share for Donald Trump in the 2020 presidential election; and (3) a dummy variable on whether any of the teams' scouting directors over the period 2000-2020 was African-American; (4) the number of years the team has been under an African-American scouting director over the same period. Results are presented in Table D.1. Including controls of fans' demographic or political variables does not change the influence of fan bias on teams' recruitment of African-American players in the early stages of the draft (column 1). Neither does the inclusion of the share of African-Americans and Republicans at the county level (column 2). We also find that adding whether the team has an African-American scouting directors does not change the results (columns 3 and 4).

Table D.1: Robustness of the relationship between fan bias and draft decisions

	Fans indicators (1)	Local indicators (2)	Any Black SD (3)	Years of Black SD (4)
Index	-0.049^{**}	-0.048^{*}	-0.048^{*}	-0.054^{**}
	(0.023)	(0.023)	(0.024)	(0.024)
Fans Black	1.214			
	(1.067)			
Fans Republicans	0.527			
	(0.827)			
Local Blacks		0.083		
		(0.183)		
Local Republicans		0.338		
		(0.213)		
Any Black Scouting Director			0.009	
			(0.059)	
Years of Black Scouting Directors				-0.005
				(0.009)
Constant	-0.223	0.029	0.138^{***}	0.144^{***}
	(0.344)	(0.085)	(0.025)	(0.024)
N	29	28	29	29
R^2	0.201	0.241	0.154	0.163

Note: The dependent variable is the share of African American players in Rounds 1-5 of the draft. The dependent variable Index is the metric of fans' bias. The other dependent variables are: in Column 1, the shares of African Americans and Republicans among fans; in Column 2, the share of African Americans and Republicans at the county level; in Column 3, a dummy variable on whether the Scouting Director for the team was African American at any point between 2000 and 2020; in Column 4, the number of years the Scouting Director has been African American over the period 2000-2020. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

E External Bias: Robustness check with Attendance Bias

Table E.1: Correlation measures of fan bias

	Bias in attendance
Index	0.096
	(0.195)
Constant	−0.005
	(0.192)
N	29
R^2	0.009

Note: The dependent variable is the bias of fans in attendance, measured as the correlation between attendance and the share of African American players in the team. A higher value of the index means that increasing the share of African American players is associated with lower attendance. The independent variable is the index of bias, measured using a textual analysis of replies to posts related to the Black Lives Matter. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

Table E.2: Attendance bias and draft strategy

	Share of African Americans drafted (1)	Including local controls (2)	Two bias indices (3)
Attendance bias	-0.036^{**}	-0.039^{***}	-0.034^{**}
	(0.014)	(0.014)	(0.013)
Local Blacks		-0.115	
		(0.106)	
Local Republicans		0.156	
		(0.131)	
Index			-0.025^{*}
			(0.014)
Constant	0.172^{***}	0.154^{***}	0.170^{***}
	(0.014)	(0.052)	(0.013)
N	30	29	29
R^2	0.193	0.311	0.291

The dependent variable is the share of African American players drafted in any round of the draft. The dependent variable attendance bias is measured as the correlation between attendance and the share of African American players in the team. The additional independent variable in column 3 is the index of fan bias from reactions to Black Lives Matter posts. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

Table E.3: Correlation fan bias and draft of Black players by round of draft

	All rounds (1)	Rounds ≤ 5 (2)	Rounds 6+ (3)
Attendance bias	-0.036^{**}	-0.026	-0.017
	(0.014)	(0.023)	(0.019)
Constant	0.172^{***}	0.135^{***}	0.180^{***}
	(0.014)	(0.023)	(0.018)
N	30	30	30
R^2	0.193	0.042	0.029

The dependent variable in Column 1 is the share of African American players drafted in any round of the draft. In Column 2, it is the share of African American players drafted in the first five rounds. In Column 3, it is the share of African American players drafted in later rounds (6+). The independent variable is the index of fan bias in attendance. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

F Cost of racial bias

Table F.1: Cost of racial bias with only players who made the majors

	All rounds (1)	Rounds 1-5 (2)	Rounds 6-15 (3)	Rounds 16+ (4)
Fan bias	0.164*	0.202	0.194	0.164
	(0.096)	(0.134)	(0.194)	(0.515)
White players × Fan bias	−0.182*	−0.232	−0.304	−0.066
	(0.107)	(0.155)	(0.204)	(0.524)
Same race as scouting director	0.077	0.119	0.262	−0.272
	(0.161)	(0.249)	(0.268)	(0.277)
Constant	0.165	0.235	0.003	0.247
	(0.166)	(0.262)	(0.283)	(0.277)
Race-fixed effects	Yes	Yes	Yes	Yes
Prospect quality metric	Yes	Yes	Yes	Yes
N	464	219	102	95
R^2	0.045	0.076	0.072	0.134

Note: The dependent variable is the WAR (Win Above Replacement) value for the first year the players has played in the MLB. The index of fan bias is defined in Section 3. Player's race-fixed effects are included (White is the omitted category). * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.