IZA DP No. 3899

Strike Three: Umpires' Demand for Discrimination

Christopher A. Parsons
Johan Sulaeman
Michael C. Yates
Daniel S. Hamermesh

December 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Strike Three: Umpires' Demand for Discrimination

Christopher A. Parsons
University of North Carolina at Chapel Hill

Johan Sulaeman
Southern Methodist University

Michael C. Yates
Auburn University

Daniel S. Hamermesh
University of Texas at Austin,
NBER and IZA

Discussion Paper No. 3899
December 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3899
December 2008

ABSTRACT

Strike Three: Umpires' Demand for Discrimination[*]

We explore how umpires' racial/ethnic preferences are expressed in their evaluation of Major League Baseball pitchers. Controlling for umpire, pitcher, batter and catcher fixed effects and many other factors, strikes are more likely to be called if the umpire and pitcher match race/ethnicity. This effect only exists where there is little scrutiny of umpires' behavior – in ballparks without computerized systems monitoring umpires' calls, at poorly attended games, and when the called pitch cannot determine the outcome of the at-bat. If a pitcher shares the home-plate umpire's race/ethnicity, he gives up fewer hits, strikes out more batters, and improves his team's chance of winning. The general implication is that standard measures of salary discrimination that adjust for measured productivity may be flawed. We derive the magnitude of the bias generally and apply it to several examples.

JEL Classification: J44, J71

Keywords: wage equations, worker evaluation, strategic interactions, economics of sports

Corresponding author:

Johan Sulaeman
Southern Methodist University
P.O. Box 750333
Dallas, TX 75205-0333
USA
E-mail: sulaeman@smu.edu

---

[*] We thank Jason Abrevaya, Jeff Borland, J-C Bradbury, Jay Hartzell, Larry Kahn, William Mayew, Gerald Oettinger, Allen Sanderson, Alan Schwarz, John Siegfried, Paul Tetlock, Sheridan Titman, Deborah White and Justin Wolfers, and participants at seminars at several universities, for helpful discussions.

I. Introduction

Discrimination in the labor market can take many forms, including disparities in wages, promotion, hiring, or performance evaluation. Of these, the last is particularly troubling to economists because it distorts the benchmark: if workers are discriminated against when their performance is evaluated, then standard tests of discrimination will lack power. To see this, consider two co-workers A and B who perform a job where productivity is determined subjectively. The workers have the same true productivity, but due to discrimination by the evaluator, worker A's *measured* productivity is biased downward. An econometrician observing similar ratios of wages to measured productivity between the two workers would incorrectly dismiss discrimination when in fact it exists.

In this study we show how evaluators' preferences—in this case, for workers who share their race or ethnicity—can create a wedge between measured and actual productivity. To do this, we take advantage of a unique setting involving Major League Baseball (MLB) umpires evaluating the performance of pitchers. In certain situations, umpires have strong incentives to make *objectively* correct calls, while in others, such incentives are much weaker, allowing umpires' *subjective* biases to influence their evaluations. This distinction opens up two doors at once. It allows us to quantify directly how racial bias influences subjective performance evaluations, both in the games we study and in the labor market for baseball players. Additionally, it also provides a sharp test of incentive effects. We show that, although biased player evaluations can impact several aspects of the game, often in surprising ways, the bias can be mitigated by increasing the cost for umpires to express such bias.

We collect and analyze every pitch from three complete baseball seasons (2004-2006), paying particular attention to the race/ethnicity of the umpire, pitcher, batter and catcher. Our results are consistent with racial/ethnic bias influencing umpires' subjective performance evaluations. Specifically, pitchers who share the race/ethnicity of the home-plate umpire receive favorable treatment, as indicated by a higher probability that a pitch is called a strike rather than a ball. Stunningly, this effect shows up only when it is inexpensive for umpires to do so—that is, when the chance of or penalty for being caught is low. When umpires are reviewed by a computerized monitoring system, no race or ethnicity preference can be detected. Similarly, when the game is well attended, or when the pitch is particularly important, race/ethnicity plays no role in the umpire's evaluation. These effects are robust to a wide set of controls, including fixed effects for each pitcher, umpire, batter and catcher, suggesting that differences in umpire or player-specific characteristics are not driving the results.

In order to quantify the economic significance of these biases better, we shift our focus to the extent to which racial bias by umpires influences game events. Our analysis of at-bat outcomes confirms the previous evidence on individual pitches and generates new insights. For example, given that strikes are more likely to be called when the umpire and pitcher match race/ethnicity, it is tempting to predict that at-bats are more likely to end in strikeouts. Indeed, this is confirmed in the data, but interestingly, *called* third strikes do not drive the result, which is consistent with the earlier finding that heightened scrutiny of important pitches mutes the expression of umpire bias. Instead, it is *swinging* strikeouts that are more likely, which suggests that pitchers and/or batters are adjusting their strategies in anticipation of racial/ethnic bias by umpires. Likewise, when the batter puts the ball in play, he is less likely to get a hit. These intriguing results highlight how a small bias in one area of evaluation (i.e., called pitches) can, by causing agents to adjust their strategies, have a substantial *impact even when the evaluator's judgment and consequently biases are not directly involved.*

Baseball offers several advantages when studying discrimination. First, because every pitch is potentially subject to the home-plate umpire's discretion when it is thrown (several hundred times per game), there is sufficient scope for racial/ethnic discrimination to be expressed as well as for it to affect games' outcomes significantly. In addition, we have a very large number of independent pitch-level observations involving the interaction of four different races/ethnicities: White, Black, Hispanic, and Asian. The data thus allow us not only to explore an umpire's preference for players of his own race/ethnicity, but also to examine preferences between other races/ethnicities, e.g., whether a Black umpire penalizes Hispanic pitchers relative to White pitchers. An additional feature of baseball data is that, unlike other sports where a group dynamic among officials may alter the expression of individual biases, the home-plate umpire is exclusively responsible for calling every pitch in a typical baseball game.[1]

The most fortunate aspect of the data set is that it allows us to develop several independent proxies for the scrutiny of the umpire's decisions, and in so doing, to test for the existence of a price-sensitive demand for discrimination by umpires. To be sure, the time period 2004-2006 is special because only during this time were a portion of the ballparks outfitted with

---

[1] Umpires can be positioned behind home plate or at first, second or third base. The home-plate umpire (the umpire-

computers and cameras to monitor umpires' balls and strikes calls. Because umpires are randomly assigned to venues, observing differences in umpire behavior between parks with and without monitoring technology makes a convincing case that properly placed incentives can have the desired effect. These results allow us to not only describe how biases can influence subjective performance valuations, but also to offer prescriptive suggestions to minimize their impact.

Several studies (e.g., Garicano *et al*, 2005; Zitzewitz, 2006) have examined home-team preferences by referees/judges in sporting events, and another, Stoll *et al* (2004) examines racial match preferences in employment generally. Our study most closely resembles Price and Wolfers' (2007) work on NBA officiating crews' racial preferences. Although the first part of our empirical analysis corroborates their findings (but for a different sport), we are mainly interested in *when* racial or ethnic bias is most likely to be observed. Here, we offer two insights. First, we show that discrimination is price-sensitive, so that making it more costly to engage in will reduce its expression. Second, we show that, when quantifying how players are affected by biased performance evaluations, the direct effect is only part of the story. Because players will alter their strategies in response, even situations that are seemingly insulated from a biased evaluator (e.g., non-called pitches in baseball games) are affected. Each of these has policy implications relevant not only for baseball, but for the general labor market as well.

This research adds to a large literature on racial discrimination in sports, specifically in baseball, going back at least to Pascal and Rapping (1972) and Gwartney and Haworth (1974), with more recent examples being Nardinelli and Simon (1990), Findlay and Reid (1997) and Bradbury (2007). It includes studies of such outcomes as productivity, wages, customers' approbation of players, selection for honors, and others. There is some evidence of wage disparities among baseball players of different races, but the results are mixed, e.g., Kahn (1991). The conclusions of racial discrimination (or lack thereof) in this literature depend upon each player's productivity being accurately measured, as measured productivity is typically the crucial control variable. We suggest questioning this central assumption: If officials' judgments are themselves subject to racial/ethnic bias, adjusting for differences in the returns to *measured* productivity will not enable us to obtain proper measures of the extent of discrimination.

The results allow us to think about the deeper question of measuring discrimination generally. If, as we show here, evaluations of workers are affected by the match to the race/ethnicity of their evaluator, then the measured productivity of the worker will depend on the nature of that match. This difficulty has serious implications for measuring discrimination and is another manifestation of the problems in identifying discrimination pointed out by Donald and Hamermesh (2006).

In the following section we describe the pitch- and game-level data and explain our classification of umpires' and players' races/ethnicities. We then analyze individual pitches in Section III, presenting evidence suggesting that umpires evaluate pitchers who match their own race/ethnicity more favorably than pitchers who do not. In Section IV we show that umpires express these preferences strongly only in times of low-scrutiny—game- and pitch-level situations where monitoring of the umpire is less. We examine the impact of discrimination on the outcome of plate appearances, game outcomes and pitcher performance in Section V. In Section VI we provide robustness checks and consider some extensions to our primary analysis. Section VII derives the size of the effects of the bias in performance evaluation on the measurement of wage discrimination and applies the results to salaries of baseball pitchers.

II. Data and Institutions

There are 30 teams in Major League Baseball, with each team playing 162 games in each annual season. During a typical game each team's pitchers throw about 150 pitches, so that approximately 700,000 pitches are thrown each season. We collect pitch-by-pitch data from ESPN.com for every MLB game from 2004-2006.[2] Our final dataset consists of 2,120,166 total pitches.[3] For each pitch we identify the pitcher, pitcher's team, batter, batter's team, catcher, pitch count, score, inning, and pitch outcome. We classify each pitch into one of seven exhaustive and mutually exclusive categories: Called strike, called ball, swinging strike, foul, hit into play, intentional ball or hit by pitch. We supplement each pitch observation with other relevant information including the stadium name, home team, away team, team standings, and the identities and positions of all four umpires.

---

[2] The pitch-by-pitch information is from: http://sports.espn.go.com/mlb/playbyplay?gameId=NNNNNNNN&full=1, where NNNNNNNNN represents the nine-digit game ID. The first six digits correspond to the year, month and date of the game. The box score information is from http://sports.espn.go.com/mlb/boxscore?gameId=NNNNNNNNN .

[3] Due to their unusual nature, we exclude All-Star and post-season games from the sample.

In addition to pitch-level data, we collect the "outcome" of each plate appearance. Such outcomes include swinging strikeouts, called strikeouts, walks, hits, fly outs, etc.[4] For each pitcher's appearance in each game, we also note the exact number of innings pitched, the numbers of hits, runs and home runs allowed, walks given up, strikeouts, earned runs, and *GameScore*, a composite index designed to summarize the starting pitcher's performance.[5] We also obtain the final score of the game, so that we can identify the winning and losing team.

We next classify each position player, pitcher and umpire who appears in our dataset as White, Hispanic, Black or Asian. To begin this task, we collect country of birth for every player and umpire. Players or umpires are classified as Hispanic if they are born in one of the following countries: Colombia, Cuba, Curacao, Dominican Republic, Mexico, Nicaragua, Panama, Puerto Rico or Venezuela. Similarly, players from Japan, South Korea and Taiwan are classified as Asian. We classify an additional 69 players using an AOL Sports article which lists every African-American player on a MLB roster at the beginning of the 2007 season.[6] We also utilize a similar list of past and present Hispanic players in MLB from Answers.com.[7] All remaining unclassified players and umpires are classified by visual inspection of pictures found in internet searches.[8] Three of the four race/ethnic groups are represented among umpires (there are no Asian umpires in MLB), and all four are represented among pitchers.

Table 1 presents the pitch distribution across the seven possible pitch outcomes. The first row of the table summarizes all pitches, while subsequent rows sub-divide pitches based on the race/ethnicity of the pitcher, the batter and the home plate umpire, respectively. As Table 1

---

[4] The exhaustive list of possibilities is: strikeout swinging, strikeout looking, strikeout other (foul, etc.), fly, pop, or foul out, ground or line out, base hit, walk, hit by pitch, sacrifice, fielder's choice, bunt out, and safe on error. We exclude rare events such as when a batter's at-bat ends when a runner attempting to steal is thrown out.

[5] Developed by baseball statistician Bill James, GameScore is a composite metric designed to gauge the performance of a starting pitcher. Pitchers are rewarded for recording outs, innings (more points for later innings), and strikeouts, but are penalized for allowing hits, runs, and walks.

[6] The complete list can be found at http://Blackvoices.aol.com/Black_sports/special/_a/african-american-players-in-mlb/20070413095009990001.

[7] The complete list can be found at http://www.answers.com/topic/list-of-hispanic-players-in-major-league-baseball.

[8] For a small number of umpires, no pictures were available on the internet. For each of these individuals, we watched past games in which the umpire worked to ascertain his race/ethnicity. Any such classification is necessarily ambiguous in a number of cases. To the extent that we have inadvertently classified pitchers umpires, or batters in ways different from how they might be treated on the field, all we have done is introduce measurement error into the matches and thus reduce the strength of any results that we generate.

demonstrates, approximately 47 percent of pitches elicit a swing from the batter, hit the batter, or are intentionally thrown out of the strike zone. Our pitch-level analysis focuses on the 53 percent of pitches (1.13 million) that result in called strikes and called balls, since these alone are subject to an evaluation by the home-plate umpire. Of these called pitches, about 32 percent are called strikes, and the rest are called balls.

Table 1 also reports the number of pitchers, batters and home-plate umpires in each of the four race/ethnicity categories. The percentages of White pitchers (71 percent) and batters (59 percent) are lower in our sample than the percentage of White umpires (91 percent). On the other hand, Hispanics, comprising 23 percent of pitchers and 27 percent of batters, are under-represented among umpires (only 3 percent). Black pitchers, batters and umpires comprise 3 percent, 11 percent, and 5 percent of the samples, respectively. Asian players comprise 3 percent each of pitchers and batters.[9]

Table 2 reports for each pitcher/umpire racial/ethnic combination the number of pitches thrown, the number of called pitches, the number of called strikes and the percentage of called pitches that are strikes. About two-thirds of the called pitches in our sample occur when the umpire and pitcher share the same race/ethnicity (mostly a White pitcher in a game called by a White home-plate umpire). While the percentage of pitches that are called is similar in situations where the umpire's and pitcher's race/ethnicity match and in situations where they do not (53.4 percent), a central difference is that the percentage of called pitches that are strikes is higher when they match (32.1 percent) than when they do not (31.5 percent).

III. Called Pitches and Umpire-Pitcher Matches

The summary statistics in Table 2 ignore possible differences inherent in the quality or "style" of pitchers by race/ethnicity. They also ignore the possible different outcomes generated by non-random assignment of pitchers to face different opponents, and of umpires to games

---

[9] While there are indeed a small number of minority umpires, this is not particularly important for either our main results or their interpretation. As we will see, the main finding is that umpires change how they treat pitchers of similar vs. different race/ethnicity based on whether they (the umpires) are themselves scrutinized. Thus, while it is interesting that *all* groups of umpires display racial/ethnic bias, the question can be posed for all umpires aggregated or for any one group separately. We discuss this point in more detail later.

played by particular teams.[10] To account for these and other potential difficulties, our central test for umpires' discrimination is the specification:

$$I(\text{Strike} \,\big|\, \text{Called Pitch})_i = \gamma_0 + \gamma_1 \text{UPM}_i + \gamma_2 \text{Controls}_i + \varepsilon_i, \tag{1}$$

where the dependent variable is an indicator of whether a called pitch is a strike, the γ are parameters, ε is a well behaved error term, and i indexes pitches. The main explanatory variable of interest is UPM, an indicator of whether the umpire (U) and pitcher (P) match (M) on race/ethnicity. In almost all of our tests, we include pitcher and umpire fixed effects, so that UPM picks up the *marginal* effect of a racial/ethnic match between the home-plate umpire and pitcher. That is, because any player or race-specific effects are swept out by the fixed effects, umpires' bias is identified purely via the interaction term, UPM.

In addition to these, we employ a number of important control variables. Pitch-count indicators, which record how many balls and strikes have accrued during a particular at-bat, are crucial because pitchers alter the location of their pitch based on the ball-strike count. Inning indicators are also included, because pitchers are usually less fatigued early in games, and because a pitcher who starts the game is often replaced by a "relief" pitcher in later innings, with a different (often reduced) accuracy.[11] Any home-field bias is captured by top-of-the-inning indicators**,** which account for whether the home (top=1) or visiting team (top=0) is pitching. Lastly, we include the pitcher's score advantage (defined as the number of runs, potentially negative, that the pitcher's team is ahead), because, if a pitcher is ahead in the game, he typically pitches more aggressively and is more likely to throw a pitch in the strike zone.[12]

Table 3 presents the results of estimating equations where the pitcher and umpire's race/ethnicity are allowed to influence the likelihood of a called strike. All the estimates are based on linear-probability models (but probit estimates present the same picture) with heteroskedasticity-robust standard errors. We account for any autocorrelation by clustering pitches by pitcher, although this adjustment makes almost no difference. The first three columns

---

[10] Examination of umpires' schedules indicates that while umpires typically travel as a four-person crew throughout much of the year, crews are randomly assigned across teams, ballparks, geography, and league (American or National). Furthermore, umpires rotate in a specific order, i.e., each serves as the home-plate umpire exactly every fourth game, resulting in random assignment of umpires to starting pitchers.

[11] In models with pitcher fixed effects, this second reason for inning indicators is obviously subsumed.

[12] The reason is that having a lead effectively reduces the pitcher's risk aversion. Relative to throwing a pitch likely to result in a walk, throwing a "hittable" pitch is risky—it increases the probabilities of both a very poor outcome for the pitcher (such as a home run) and a very good one (a fly out). We return to this issue in Section V.

show analysis separately for White, Black and Hispanic pitchers, respectively, controlling for umpire race/ethnicity and pitcher fixed effects. The next three columns show analysis separately for White, Black and Hispanic umpires, respectively, controlling for pitcher race/ethnicity and umpire fixed effects. The final three columns include all pitchers and umpires, with each column adding successive vectors of fixed effects.[13] There is some, albeit weak, evidence of favoritism by umpires for pitchers who match their race/ethnicity. Taking the results in Column (9) as the most indicative of the underlying behavior, it is quite clear, however, that there is no generally significant impact of the match (p=.17).[14]

Although the results with the broadest sets of fixed effects do not suggest a significant effect of the umpire-pitcher match, the point estimates imply that a given called pitch is approximately 0.27 percentage points more likely to be called a strike if the umpire and pitcher match race/ethnicity. Excluding (as we do) pitches where the batter swings, the likelihood that a given pitch is called a strike is 31.9 percent. Thus, when the umpire matches the pitcher's race/ethnicity, the rate of called strikes rises by slightly less than 1 percent above the rate when there is no match.[15]

IV. Called Pitches When Discrimination Is Costly to the Discriminator

One might examine the results in Table 3 and conclude that, while the point estimates are interesting, their statistical insignificance means that there is really little here. Given an economist's view that agents acting out their preferences will react to the price of an activity, however, it is worthwhile to examine the impacts of umpire-pitcher matches as the price of discrimination changes. Our data are particularly well suited to study this question, and it is our primary focus for much of the remaining analysis.

We begin by asking what factors affect the price of expressing racial or ethnic discrimination. Studies of cognitive behavior indicate that presenting the biased party with

---

[13] We include all pitchers in these regressions, although a case could be made that Asian pitchers should be excluded because they are never judged by an umpire of the same race. All the results are nearly identical if they are excluded.

[14] In unreported results, we estimated models with proxies for pitcher accuracy, e.g., earned run average (ERA) or walks/inning, with no qualitative change in the results.

[15] As a check on this issue we re-estimated the model including sequentially the race/ethnic match between the first-, second- and third-base umpire and the pitcher. None of these extensions materially changes our conclusions.

counter-examples of the stereotype of interest can reduce the severity and/or frequency of the biased behavior (Goodwin *et al*, 2000; Blair, 2002). In other words, simply making conscious a sub-conscious bias imposes a sufficient psychological cost to mitigate its expression. Another mechanism is to increase the visibility of the biased party's behavior, potentially exposing the offender to social or statutory penalties. In this section we proxy the price of discrimination by the extent to which an umpire's evaluations of pitchers will be scrutinized, and employ three different measures to examine whether a higher price of discrimination reduces the extent to which umpires engage in discriminatory behavior.

The first source of scrutiny is QuesTec, a computerized monitoring system intended to evaluate the accuracy and consistency of home-plate umpires' judgments. In 2003 MLB installed QuesTec in 11 of its 30 ballparks.[16] QuesTec's Umpire Information System (UIS) consists of four cameras that track and record the location of each pitch, providing information about the accuracy and precision of each umpire's ball and strike calls. Despite opposition from some umpires and players (perhaps most notably, pitcher Curt Schilling's assault on one of the cameras after a poor performance in 2003), the QuesTec system served as an important tool to evaluate umpires during our sample period. According to the umpires' union's agreement with MLB, QuesTec is the primary mechanism to gauge umpire performance. In particular, if more than 10 percent of an umpire's calls differ from QuesTec's records, his performance is considered substandard, which can influence his promotion to "crew chief," assignment to post-season games, or even retention in MLB.[17]

Because QuesTec is installed in roughly 35 percent of ballparks, and because umpiring crews are rotated randomly around the league's ballparks, virtually every umpire in our dataset calls a substantial number of pitches in parks both with and without QuesTec.[18] Additionally, both the umpires' and teams' schedules change every year, exposing each umpire to a wide

---

[16] These were the ballparks of the Anaheim Angels, Arizona Diamondbacks, Boston Red Sox, Cleveland Indians, Oakland Athletics, Milwaukee Brewers, Houston Astros, New York Mets, Tampa Bay Devil Rays, Chicago White Sox, and New York Yankees.

[17] An umpire's evaluation is not based solely on his performance as measured by QuesTec. If an umpire falls below the QuesTec standards, his performance is then reviewed by videotape and live observation by other umpires to determine his final evaluation score. No such measures are taken, however, if an umpire meets the QuesTec standards.

[18] The fraction of games in which QuesTec was installed was virtually identical for all umpires in our sample, differing for the few umpires calling only a handful of games.

cross-section of batters and pitchers in both QuesTec and non-QuesTec parks. Throughout the analysis we test whether greater scrutiny—the possibly higher cost of indulging in personal discretion in QuesTec parks—leads umpires to call strikes "by the book." Any role that racial/ethnic (or any other) preference plays in influencing pitch calls should be mitigated if the costs of being judged substandard by QuesTec are sufficiently high. Pitchers, however, may act strategically in response to the scrutiny of umpires, altering how they pitch depending on whether the game is in a QuesTec park or not.[19] For this reason, in all of the estimates in this part we include fixed effects not only for each pitcher, umpire and batter, but also for the presence or absence of QuesTec in each game, i.e., pitcher-QuesTec fixed effects, etc.

Figure 1 graphs the average percentage of called pitches that are strikes in ballparks with and without QuesTec for White and non-White pitchers respectively. The effect of monitoring on umpires' behavior is apparent, with both White and non-White pitchers being judged differently by umpires of the matched race/ethnicity depending on whether the pitch is thrown in a park with QuesTec installed. The difference in the called-strike percentage between QuesTec and non-QuesTec parks is significant for both White and non-White pitchers.

Table 4 contains the results of estimating (1) separately for QuesTec and non-QuesTec parks, with controls for inning, pitch count, pitcher score advantage, and top of the inning.[20] The results are remarkable: In ballparks with the umpire monitoring system, shown in Column (1), the coefficient on UPM is -0.35 percentage points and is not significantly different from zero. In parks without QuesTec, shown in Column (2), the same coefficient is 0.63 percentage points per pitch (p=.03). These differences make clear why UPM is not significant in the aggregate sample. The effects found in Table 3 average the statistically significant positive impact of an unscrutinized match (non-QuesTec) with the statistically insignificant negative impact of a scrutinized match (QuesTec). Thus, in the presence of price-sensitive discrimination, we should *expect* the point estimates in Table 3 to be low, given that the entire sample consists of a mix of high- and low-scrutiny games.

---

[19] For example, New York Mets pitcher Tom Glavine, known as a "finesse" pitcher who depends on pitches close to the strike zone border, complained publicly that QuesTec's influence on umpire calls forced him to change his style (Associated Press, July 9, 2003). Glavine reports that he was told, "[umpires do] not call pitches on the corners at Shea [his home ballpark] because they [the umpires] don't want the machine to give them poor grades."

[20] The direct effect of being in a QuesTec park is, of course, not directly observable, as it is subsumed in the pitcher-QuesTec fixed-effects terms.

Column (3) of Table 4 presents the results when the QuesTec indicator is interacted with UPM. When the pitcher and umpire match race/ethnicity, pitching in a QuesTec park reduces the likelihood that a called pitch is ruled a strike by almost 1 percentage point, more than offsetting the favoritism shown by umpires when they are not monitored by QuesTec. Each effect is statistically significant, implying that umpires implicitly indulge their apparent preference for matched pitchers when the pitches underlying their decisions are not recorded.[21]

QuesTec is an *explicit* monitoring technology that, as shown, can mitigate or eradicate the expression of bias by umpires. Here, we show that implicit monitoring can have a similar effect, suggesting that even subtle incentive mechanisms can have desirable effects. Our two measures for implicit scrutiny of umpires are crowd attendance (scaled by stadium attendance) and the "importance" of the pitch.[22]

The idea for the first is simple. The presence of many fans close to home plate presumably exposes the umpire to their scrutiny—a badly called pitch is unlikely to go unnoticed.[23] Figure 2 confirms that crowd attendance, like QuesTec, dramatically alters umpire behavior. A game is defined as "well-attended" if the crowd attendance is above the median percentage capacity for the sample, roughly 70 percent. Compared to well-attended games, umpires calling poorly-attended games appear to favor pitchers of matched race/ethnicity. In the case of White pitchers, both non-White and White umpires tend to call fewer strikes in poorly-attended games, but the reduction in strikes called by non-White umpires is over three times larger. The same effect is seen to an even greater degree among non-White pitchers. Umpires whose race/ethnicity matches non-White pitchers call nearly 1.5 percent more strikes in poorly-attended games, whereas unmatched umpires call fewer strikes.

In Columns (1) and (2) of Panel A in Table 5, we show the results of estimating Equation (1) separately for well- and poorly-attended games respectively. Each equation includes the same

---

[21] Even though the negative effect of a match in a QuesTec park is not statistically significant, what is intriguing is why umpires' decisions might favor unmatched pitchers when they are scrutinized.

[22] We scale by stadium capacity in an attempt to minimize the impact of differences between stadium sizes. Specifically, if we assume that stadiums populate relatively uniformly, then attendance/capacity is a good proxy for the *number* of fans close enough to judge pitch location. In any case, this scaling makes little difference in our results. If we use pure attendance instead, all coefficients of interest remain highly significant.

[23] Percentage attendance may also proxy the popularity of the participating teams or the importance of a particular game. Thus, not only might the umpire be exposed to more scrutiny from the additional fans present at well-attended games, but he may also face added scrutiny in the form of larger television audiences and increased air-time given to game highlights.

battery of control variables employed in Table 4, i.e., pitcher, umpire and batter fixed effects, pitch counts, inning indicators, clustering by pitcher, etc. As with the QuesTec results, the UPM variable is significant at the 5 percent level *only* in poorly-attended games, with an effect of 0.68 percentage points per pitch. During well-attended games there is no significant effect of an umpire-pitcher racial/ethnic match and, indeed, the point estimate is negative. Column (3) generalizes the results by aggregating all games and interacting UPM with an indicator for a game being well-attended. Compared to a pitch in a poorly-attended game when the umpire and pitcher do not match, a pitch called by an umpire of the same race/ethnicity as the pitcher is 0.52 percentage points more likely to be judged a strike. If the game is well-attended, a pitch is no more likely to be called a strike if the pitcher and umpire match race/ethnicity. The results for this completely different proxy of the price of discrimination are qualitatively identical to those obtained for the QuesTec vs. non-QuesTec distinction.

A third proxy for the scrutiny of umpires varies many times *within* each game. We separate pitches into two categories, “terminal” and “non-terminal.” A pitch is potentially terminal if the umpire’s next judgment can terminate the batter’s plate appearance. Specifically, a pitch that is thrown with two strikes and/or three balls is potentially terminal, as a third strike or fourth ball terminates the at-bat. In such situations, the umpire’s judgment is likely to be scrutinized more heavily by the pitcher, batter, catcher, managers and fans. An initial glimpse into the effects of this distinction is shown in Figure 3. Here we observe the same contrast as for the previous two proxies for scrutiny, as umpires appear to favor pitchers with whom they match only in non-terminal counts, when scrutiny is likely to be less.

Columns (4) and (5) of Panel B of Table 5 show estimates of (1) separately for terminal and non-terminal pitches, with pitcher, umpire and batter fixed effects and the now standard set of control variables. Each type of pitch is considered separately, with the result that the coefficients of UPM have opposite signs. For pitches of lesser importance, where the pitch cannot be terminal, the estimated coefficient of UPM is 0.43 percentage points (p=.16)—umpires favor pitchers who match their own race/ethnicity. For potentially terminal pitches, where scrutiny of the umpire is likely to be greater, umpires appear to judge pitchers of their own race/ethnicity insignificantly more harshly than unmatched pitchers. In Column (6) all pitches are aggregated, and UPM is interacted with an indicator for potentially terminal pitches. The results mimic those implicit in the estimates in Columns (4) and (5), as the coefficient on the interaction term is negative and significant at the 1 percent level.

In Columns (7) and (8) we consider more within-game variation in implicit scrutiny. We assume that because umpire evaluations are more likely to be pivotal late in games, scrutiny in the first few innings is likely to be comparatively lower. We thus designate the first third (three innings) of a game as "early," and the remainder "late." We expect that a terminal count will have a stronger effect on the outcome of a pitcher-umpire racial/ethnic match in early innings. Comparing the results across the two columns, we see that this is the case, with the magnitude of the interaction between terminal count and UPM being nearly twice as large in early as in late innings (-1.08 vs. -0.61 percentage points).[24]

Our proxies for scrutiny are not redundant. The correlation between QuesTec and attendance percentage is near zero, suggesting that these are indeed independent measures of scrutiny. Because the type of pitch (terminal or non-terminal) is a within-game measure, it is necessarily uncorrelated with either between-game measure. It is therefore not surprising that when all three interactions are included simultaneously, everything remains significant. We report the results of this specification, along with other robustness checks in Section VI.

Before proceeding to other issues, we briefly address an obvious question: Are umpires *intentionally* favoring pitchers of their own race or ethnicity? Although we cannot provide a definite answer, several factors argue for the bias being unconscious rather than conscious. Supposing for a moment that umpires are aware of their biases, it would be perhaps unsurprising to find that explicit monitoring by QuesTec results in umpires engaging in less discrimination. However, further assumptions are needed for the implicit measures of scrutiny to have similar effects. Here, umpires must also have in mind a mechanism whereby crowd attendance or pitch importance increases their probability of being discovered, and consciously adjust their behavior in response (pitch by pitch in the case of a terminal count). Moreover, one would expect that umpires consciously attempting to punish or reward pitchers would choose situations where their calls have the most impact on the game. Instead, we find the opposite, i.e., umpires show favoritism only in comparatively unimportant pitches, especially early in games.

---

[24] One could imagine still more indicators of the potential extent of the scrutiny of umpires. Indeed, we performed similar analyses for one of them, the closeness of a game, with results very much like those found for non-terminal and terminal pitches.

Of course, our previous analysis suggests that pitch importance proxies *simultaneously* both the umpire's price of expressing discrimination and the cost or benefit to the pitcher, making it impossible to refute the hypothesis of conscious bias. At the very least, this suggests that the impact of any conscious bias is just as easily eliminated as unconscious discrimination.[25] A more plausible hypothesis is that the unconscious bias of umpires is eradicated when they simply exercise more care. The three mechanisms that we study do exactly this, giving umpires strong incentives to make objectively correct calls, inadvertently swamping the effect of any latent bias. In any case, although this distinction is intriguing, the policy implications (either in baseball or employment relationships more generally) remain unchanged: To protect workers from the adverse effects of discrimination, increase the consequences for those that discriminate.

V. Impacts on Game Events

At first glance, the results in Tables 3-5 appear too small to affect the outcomes of at-bats, let alone entire games. What is perhaps more surprising, however, is that there is any detectable effect at all. MLB umpires are monitored much more intensively than most workers, even in the "low scrutiny" situations where perhaps only a few thousand people are watching. The fact that *additional* scrutiny can still affect behavior reflects the considerable difficulty in eradicating deeply ingrained, inherent biases.

Given these results, however, it is natural to attempt to quantify their economic significance. The dynamic between a pitcher and a batter implies two distinct channels through which an umpire's bias can have influence. The first is direct—if a pitch is more likely to be called a strike, the pitcher has an advantage. For example, all else equal, more called strikes will increase the probability of striking out and/or decrease the probability of walking. The pitch-level evidence, however, makes very clear that such direct effects are rather small. Of course, one can construct a scenario where the estimated direct effect of bias is fairly large—for *non-terminal* pitches thrown in the *early innings* of *poorly attended* games in parks *not equipped with QuesTec,* the marginal impact of an umpire-pitcher match on called pitches is nearly 2.5%, compared to an unconditional probability of 31.8%. But in most situations, the direct impact on

---

[25] In unreported results, we asked whether certain umpires consistently (i.e., across games) favor or penalize certain pitchers, a finding of which would suggest conscious motivation by the umpire. We find no evidence of such behavior.

called pitches is neither the largest nor the most interesting implication of racial/ethnic bias by umpires.

Instead, it is the *indirect* effects on player strategies that most influence outcomes of plate appearances and games. More specifically, the dynamic between a pitcher and batter is clearly affected by each party's beliefs about the umpire's evaluation in the event of a called pitch. If a pitcher expects favoritism, he will incorporate this advantage into his strategy, perhaps throwing pitches that would have little chance of being called a strike absent the umpire's bias. This in turn changes how the batter will optimally behave. If the batter expects such pitches to be called strikes, he is forced to swing at "worse" pitches, which reduces the likelihood of getting a hit.[26] This thought experiment illustrates how *even on pitches when the umpire's discretion is not directly involved*, the possibility of unfair evaluations can affect behavior and outcomes.

In the Appendix, we develop a stylized model of the interaction between the pitcher and batter. The model produces two important implications relevant for the next set of tests. First, when pitchers expect a racial/ethnic match with the umpire to result in more called strikes, his optimal response is to select pitch locations further from the center of the plate.[27] Intuitively, the umpire's bias reduces the cost of throwing pitches that are difficult for the batter to hit. The second implication follows directly: conditional on swinging, the batter is less likely to get a hit. Thus the pitcher holds an advantage, not only when balls and strikes are called, but also when the batter swings.

In Table 6 we present two panels where the dependent variables are various possible outcomes for a given plate appearance (the interaction between a pitcher and a batter). In Panel A, we consider how the match variable UPM influences the probability of the batter striking out. As with previous tests we include the full complement of fixed effects, inning indicators, clustering by pitcher, etc. Column (1) shows that the batter is more likely to strike out by 0.70 percentage points when the pitcher and umpire match race or ethnicity.

---

[26] The intuition for this can be seen by examining the coefficients on the count indicators in Table 3. When the pitcher has a substantial advantage in the count, he has little incentive to throw a "hittable" pitch, i.e., one near the middle of the plate. Instead, he usually throws pitches near the corners that are both less likely to be hit if the batter swings, and less likely to be called strikes if the batter does not. Such behavior translates into sizeable advantages for pitchers depending on the count. In 2004, batters got a hit 33 percent of the time when the count was 2-1 (two balls and one strike), but dropped to less than 18 percent when the count was 1-2.

[27] Chen (2007) presents evidence that justifies this assumption. Specifically, he examines how racial bias affected MLB umpires' balls and strikes calls in one season, as we did for 2004-06 in Section III, but also holding pitch location constant.

When we break out the type of strikeout, i.e., swinging vs. non-swinging, in Columns (2) and (3), the results are even more interesting. In a swinging strikeout, the batter swings and misses (or barely grazes the ball). In a non-swinging strikeout, the umpire judges a final pitch to be hittable, and the batter is called out. As seen, the impact of UPM on called strikeouts is virtually non-existent, while it remains significant for swinging strikeouts. This distinction both confirms the previous pitch-level results, and it supports the idea that batters are forced to swing at worse pitches.

Recalling that we found no effect of UPM on the probability of a called strike in terminal counts, it would be disturbing to find a significant effect on called third strikes. The fact that we do not is further confirmation that umpires do not display their biases in crucial situations, probably because they expend more care in such situations. Instead, we find that the strikeout effect is completely due to the batter's increased tendency to swing at difficult-to-hit pitches.

The final two columns of Panel A of Table 6 break down swinging strikeouts by the presence or absence of QuesTec. As expected, the point estimate in non-QuesTec parks (0.00821) is over twice as large as in QuesTec parks (0.00398). Although significant at only the 10 percent level, this distinction is noteworthy, as QuesTec cannot have any direct effect on a non-called pitch. This result thus emphasizes how a consideration of equilibrium behavior is needed to appreciate the full impact of biased subjective evaluations of outcomes of interest.

The remaining panel considers two additional outcomes of interest: walks and hits. In Column (6) of Panel B we see that walks are not significantly related to UPM, consistent with both the terminal-count pitch evidence as well as the called strikeout evidence in Column (2) of Panel A. Column (7) of Panel B considers what happens when a batter swings and hits the ball. He is either out (e.g., pops out, flies out, fouls out, grounds out, etc.) so that the dependent variable takes a value of zero, or he gets a hit (e.g., single, double, triple, or home run), and the dependent variable is one. Importantly, the sample here (balls put in play) is mutually exclusive to the positive outcomes (strikeouts) in Panel A, so that the same phenomenon is not simultaneously driving both sets of results. That is, it is not the increased likelihood of striking out that causes hits to be less likely.

Column (7) shows that, conditional on putting the ball in play, a hit is 0.33 percentage points less likely when the umpire and pitcher match, which is statistically significant at the 1% level. Like the swinging strike evidence, this finding suggests feedback between (in this case biased) umpire evaluations and player strategies. In Columns (8) and (9), we perform the same comparison between QuesTec and non-QuesTec parks, with by now familiar results. While the effect strengthens in magnitude and remains statistically significant in non-QuesTec parks, the effect weakens and becomes insignificant in parks equipped with umpire-monitoring technology. Quite clearly, the impact of incentives influences not only those situations it is directly designed to monitor, but also those that are indirectly affected.

As a final test of economic significance, we analyze a variety of *game-level* performance measures for each starting pitcher in our sample: Wins, hits, earned runs, home runs, strikeouts, walks, and *GameScore*.[28] Some of these, such as hits, walks, and strikeouts, have already been analyzed in Table 6, but others (e.g., *Gamescore*, earned runs) are new. Also, for reasons that will soon become clear, we aggregate these performance measures only for starting pitchers, whereas in Table 6 all pitchers in our sample are included. Figure 4 graphs each performance measure for the roughly 14,000 starting pitchers in the roughly 7,000 games in the three seasons in our sample. As in the previous figures, we display the results for White and non-White pitchers separately to highlight the magnified effect of racial/ethnic preference on non-White pitchers.

Perhaps unsurprisingly given the results in Table 6, we find that for virtually every measure of pitcher performance the impact of having a matched umpire benefits the starting pitcher. The composite measure, *GameScore*, is raised for both White and non-White pitchers when the home-plate umpire's race/ethnicity matches theirs. Similarly, both White and non-White pitchers allow fewer home runs (HR), hits, runs and walks, and have lower earned-run-averages (ERA), when a match occurs. Only strikeouts (K) among White pitchers do not accord with the observed racial/ethnic preferences by umpires, although the effect is minuscule.

While all of the pitching performance measures are informative, our main interest here is on the most important result in a game—who wins. Looking at the mean game outcomes in various instances of umpire-pitcher matches, the obvious benchmark is the case when both or neither starting pitcher matches the umpire's race/ethnicity. In that case, the home team wins 53.8 percent of the time, reflecting a slight home-field advantage. In 18.7 percent of the games

---

[28] Although most of our results are similar when we include all pitchers, starting pitchers are of particular interest because of their relative importance and because a team's starting pitcher generally interacts directly with the umpire far more than any other member of the team besides its catcher. In addition, *GameScore* is only calculated for starting pitchers.

only the home-team pitcher matches the umpire, while the opposite case, a match only between the visiting-team pitcher and the umpire, occurs 19.0 percent of the time.[29] In the former case, the home team wins 55.6 percent of its games. In the latter case the home team's winning percentage is unaffected—it remains 53.8 percent. These differences in the means suggest that there is an asymmetry in the impact of racial/ethnic matching: Matches are much more important between the umpire and the home-team's pitcher than between the umpire and the visiting team's pitcher.

The effect of racial/ethnic preferences on winning probabilities is even more striking when we disaggregate by umpire race/ethnicity. With White umpires the home team wins 54.4 percent of the time if its starting pitcher is White, but only 52.9 percent of the time if he is not. In the case of Black umpires, the home team's win percentage is 72.7 percent if the home team's pitcher is black and 55.1 percent if he is not, although there are only 11 games in which a Black starting pitcher is evaluated by a Black umpire. In the 36 games in which both the starting pitcher and the umpire are Hispanic the home team wins 61.1 percent of its games, compared to 52.0 percent if the pitcher is non-Hispanic.

We collected box score data for 7,124 games during the 2004-2006 seasons. For each of these games we compare the race/ethnicity of both starting pitchers to that of the umpire and analyze whether racial/ethnic relationships influence a particular outcome, adjusting for other characteristics. In Column (1) of Table 7 we present estimates with the dependent variable equaling one if the home team wins. We include the number of runs scored by the home pitcher's team and specify fixed effects for the pitcher, the umpire and the identity of the opposing team.[30] The coefficient on UPM is marginally significant (p = .07), with a magnitude of slightly over 4 percentage points.

Columns (2) and (3) examine the effect of an umpire-pitcher match on *GameScore* (with higher values of the dependent variable indicating a better performance) and the number of runs allowed by each starting pitcher (so that both variables are available for both starting pitchers in

---

[29] That these are nearly identical is further evidence of random matching between umpire and pitcher races/ethnicities.

[30] As the coefficient of interest, *UPM*, refers to the match between the home starting pitcher and umpire, it would clearly be inappropriate to include the number of runs scored by the away team in the regression. For this reason, we include only the runs scored by the *home pitcher's team*, so that the coefficient on *UPM* is the marginal probability of winning, conditional on the pitcher's own run support.

a game).[31] The results are qualitatively similar to those obtained on the probability of winning—there is a positive, albeit statistically insignificant benefit to a pitcher's *GameScore* if he matches the race/ethnicity of the umpire; and there is a marginally significant (p = .11) impact of the pitcher-umpire match on the number of runs allowed, even after adjusting for all the vectors of fixed effects.

In light of the evidence that the effects of an umpire-pitcher match are seen only when scrutiny is less, we can disaggregate the samples underlying the estimates in Table 7 and estimate the equations separately for games played in QuesTec and non-QuesTec parks. Given the small sub-samples and the tightness of the specification with the inclusion of all three vectors of fixed effects, even the estimated effects of an umpire-pitcher match become only slightly more significant when we use the non-QuesTec sub-samples. Nonetheless, for all three dependent variables the impacts of the match are larger in absolute value in these sub-samples than in the sub-samples for QuesTec parks, similar to the at-bat level evidence in Table 6.[32]

VI. Robustness Checks and Other Considerations

A. Accounting for Matches with Batters' and Catchers' Race/Ethnicity

It is natural to suppose that an umpire influenced by the race of the pitcher may also be influenced by that of the batter or the catcher, especially since in the latter case the umpire is in continuing close contact. We explore this possibility extensively, but find no evidence to support the argument. As shown in Column (1) of Table 8, estimating (1) substituting UBM, defined as a racial/ethnic match between umpire and batter, for UPM yields absolutely no effect. A catcher-umpire match, indicated by the analogously defined variable UCM, has a small insignificant positive effect on the probability of a called strike, as shown in Column (2). These results are unchanged when all three match variables, UPM, UBM and UCM, are included simultaneously (Column (3) of Table 8), and, indeed, the coefficients on all three match variables are essentially the same as when each is included separately.

---

[31] The increase in sample size is due to the fact that game-level outcomes are analyzed in Column (1), so that the number of observations is the number of games in our sample. In contrast, the remaining columns consider the performances of each starting pitcher as independent observations, roughly doubling the sample size.

[32] In most of the work in this section we have concentrated on the three outcome variables win probability, *GameScore* and runs allowed to avoid duplication. The results are qualitatively the same when we expand the analysis to the other measures depicted in Figure 4.

When interactions of UPM, UBM and UCM with QuesTec, high-attendance and terminal counts are included in Equation (1) sequentially, as shown in Columns (4)-(6) of Table 8, each indicator UPM is still statistically significant and positive, and each interaction with UPM is significantly negative. Moreover, none of the main effects of UBM or UCM approaches significance, nor do the interaction terms with those indicators. The results in Column (7), in which all the main-effect and interaction terms are included, give the same picture as the other results: Implicitly umpires engage in discrimination against unmatched pitchers, and each proxy for a higher price of discrimination reduces umpires' demand for discriminatory outcomes. Umpires appear focused on the pitchers they are judging—there is no evidence whatsoever that matches with other relevant players affect their judgment.

For at least two reasons the absence of any impact of UBM may not be as puzzling at it first appears. First, as suggested above, the per-pitch effect represents racial/ethnic discrimination only relatively infrequently and is concentrated in low-scrutiny situations. Both scrutiny and batters' race/ethnicity change frequently (many times within each game), so any effect may be swamped by the impact of scrutiny. We have no such concerns about statistical power with pitchers, who interact with each umpire over a hundred times within each game under varying degrees of scrutiny. The second possibility is more subtle, owing to the physical proximity of the umpire and batter relative to that of the umpire and pitcher. Psychological studies suggest that, although people may not recognize their own prejudice (Bargh, 1999, Devine and Monteith, 1999), the risk of being confronted reduces the frequency of biased behavior (Czopp *et al*, 2006). If physical proximity to the batter increases the probability of confrontation for an umpire, perhaps it acts as an additional check on the umpire's tendency to express discrimination.[33]

That there is only a small, insignificant positive effect of an umpire-catcher match, given the proximity of the umpire and catcher for half of each game, may be more surprising. It suggests, however, that umpires realize that they are judging the pitcher. A match with the catcher is less important, and additional interactions of UPM with UCM do not alter the conclusions about the effect of each first-order interaction individually. One can speculate about

---

[33] Batters' confrontations with the umpire are far more common than pitchers', lending support to this interpretation of the evidence.

the absence of a UCM effect, including the possibility that only the pitcher directly faces the umpire, but we cannot distinguish among possible interpretations.

B. Accounting for Umpire and City Characteristics

It may be that umpires' measurable characteristics (beyond their race/ethnicity) and those of the city where a game is played explain our results. We collect demographic information on each umpire from a variety of sources and include his age and experience, and in many cases both his state of birth and residence. For each ballpark we also obtain the racial/ethnic breakdown of the surrounding metropolitan statistical area.

While we find no evidence that the racial composition of an umpire's birthplace or residence predicts his propensity to penalize non-matching players, there is somewhat weak evidence that discrimination is more likely among younger and less experienced umpires. The coefficient on UPM in the re-estimation of Equation (1) among the upper half of umpires ranked by experience is less than half its magnitude in estimates for umpires in the lower half of the distribution of experience. In addition, the 18 "crew chiefs," veterans selected for their seniority and performance, do not appear influenced by the race/ethnicity of the pitcher. Indeed, if (1) is estimated separately for crew chiefs, the point estimate of the coefficient on UPM is nearly zero. This evidence is consistent with either a model of selection or learning. Perhaps discriminating umpires are not promoted and are dropped from the ranks. Alternatively, experience may teach umpires to restrain their own biases, so that highly experienced umpires are not likely to express racial/ethnic bias in their subjective calls.

We also re-estimated the basic equation for Blacks, and for Hispanics, separately, adding in each case main effects and interactions with UPM of the percentage of the minority group in the metropolitan area where the ballpark is located. Among Blacks the interaction was positive, but statistically insignificant; among Hispanics it was negative, but also statistically insignificant. Our conclusions are not affected by the racial/ethnic mix of the team's catchment area.

C. Other Issues

The overwhelming majority of minority pitchers are Hispanic. In our main tests, we aggregate them, but some are White Hispanics, while others are Black Hispanics. To allow for the possibility that the two different groups of minority umpires might treat Hispanic pitchers who match their own characteristics differently from other Hispanic pitchers, we visually inspected the Hispanic pitchers' pictures, divided the Hispanic aggregate into White and Black groups and re-defined UPM. This reclassification had almost no effect on the estimates produced in Tables 3-5. Implicitly, Hispanic and other umpires treat Hispanic pitchers the same regardless of the pitcher's racial identity.[34]

As the discussion has made clear, there is no objective measure of the quality of a pitch. We only have information on whether it is called and, conditional on that, if it is called as a ball or strike. It might, for example, be that pitchers, assuming that they will be treated worse if there is a racial/ethnic mismatch, are "rattled" and less likely to pitch strikes. We cannot refute this possibility with certainty, but one might argue that the absence of any mismatch effect on terminal pitches, when this effect would be most likely to prevail, suggests the argument is invalid. Even if it were valid, such a finding would still support the main result, although we would classify it as an "indirect" effect, similar to the effect if a pitcher intentionally altered his strategy in expectation of the umpire's bias.

The estimates in Table 7 would still be unbiased if managers were able to alter their starting pitchers' assignments to take advantage of the umpires' preferences that we have demonstrated exist. Nonetheless, it is interesting to inquire whether managers are implicitly both aware of these preferences and able to act upon them. The racial/ethnic endowments of umpires and starting pitchers in the 7,124 games in our sample would lead one to expect matches in 0.680 of the games. In fact, matches occur in only 0.677 of the games. The difference, aside from being in the unexpected direction, is statistically insignificant (t = -0.69). Quite clearly there is no evidence in our sample of non-random matching of umpires and starting pitchers.

VII. Biases in Measuring Wage Discrimination

In the previous sections we generate some evidence that presumably objective measures of a worker's (in this case, baseball pitcher's) activities can be subtly affected by his evaluator's racial/ethnic preferences, and that this effect in turn leads to reductions in his measured productivity (the game outcomes discussed in Section V). To the extent that pay is based on measured productivity, this finding carries important implications for measuring the extent of discrimination in baseball and in labor markets generally. In particular, it implies that estimates of the extent of discrimination will be understated.

---

[34] In addition, we investigated whether American-born Hispanic pitchers were treated differently from Hispanic pitchers born outside the U.S. We find no evidence that the Hispanic pitcher's birthplace affects the expression of any racial/ethnic bias by umpires.

Consider a simple earnings equation:

$$W_i = \alpha M_i + \beta P^*_i + \upsilon_i \,, \qquad (2)$$

where W is the logarithm of earnings, M an indicator of minority status, P* worker i's true productivity, and υ a random error in the determination of earnings. The parameter α is the true effect of minority status on earnings when productivity measurements are free of bias. Assume that the majority workers' productivity is measured without bias, but that minority workers are subject to a negative bias in their assessment by evaluators, which leads to a shortfall of their measured productivity P below their true productivity:

$$P_i = P^*_i - \varphi, \text{ if } M=1; \qquad (3)$$

$$P_i = P^*_i \,, \quad \text{if } M=0,$$

$\varphi>0$. Then we can rewrite (2) to obtain an estimating equation in observables:

$$W_i = [\alpha+\beta\varphi]M_i + \beta P_i + \upsilon_i \,, \text{ or} \qquad (2')$$

$$W_i = \alpha' M_i + \beta P_i + \upsilon_i.$$

The standard estimate of earnings discrimination adjusted for productivity differences, α', has a positive bias in the amount $\beta\varphi$.

To obtain some feel for the size of this bias in the particular case that we have examined, we can simulate the wage effects using the estimates of φ underlying Figure 4 and estimates of β from studies of salary determination in Major League Baseball. We are essentially estimating the reduction in minority pitchers' salaries as a result of the average amount of discrimination arising during the 2004-2006 seasons due to umpire-pitcher racial/ethnic matches. Kahn (1993, Table A2) estimates equations like (2') using a set of outcome measures that can be conformed to ours by including the percentage of games won and ERA. Making reasonable assumptions about the means of these outcomes for starting pitchers in 2006, applying the effects in Figure 4, and using his parameter estimates yields an estimated bias of $\beta\varphi = 0.039$. Gius and Hylan (1996, Table 6.2) use strikeouts/innings, walks/innings and winning percentage, all of which are also conformable with our outcome measures. The same method based on their parameter estimates produces an estimate of $\beta\varphi = 0.014$. Finally, using the estimates for starting pitchers by Krautmann *et al* (2003), the estimate of $\beta\varphi = 0.084$.[35]

---

[35] For the percentage of games won we use 0.5. The mean ERA is around 4.3, the mean strikeouts/inning by starting pitchers is around 0.7, and the mean walks/inning by them is around 0.3. We can take the estimates of the bias that we have produced as examples here to infer the dollar impacts of this subtle form of discrimination. In 2006 the average salaries of starting pitchers in MLB were $4.8 million. A bias to the estimated effect of minority status on

While we have demonstrated the extent of bias to estimated discrimination in earnings that arises because of biased evaluations of Major League Baseball pitchers, this effect is probably smaller than would be observed for workers generally. The scope for the expression of racial/ethnic preferences of umpires for/against pitchers is almost surely far less than in most workplaces. Evaluations of pitchers are made at discrete and very frequent times—when a pitch is thrown. These are not one-shot comments made at most monthly at the evaluator's leisure. Also, as our demonstrations of reduced bias when there is greater scrutiny suggest, there are quite stringent external limits on the expression of bias against unmatched pitchers. The relative lack of such limits in the general workplace suggests that the example here may provide a lower bound on the extent of bias to estimates of discrimination generally. The costs to the economy may be still higher: Analogous to our demonstration of pitchers' changed behavior in the face of umpires' discrimination, one might expect that workers who believe that they will be discriminated against in evaluations respond by altering their behavior in ways that reduce economic efficiency generally.

VIII. Conclusions

The analyses of individual pitches and game outcomes suggest that baseball umpires express racial/ethnic preferences in their decisions about players' performances. Pitches are more likely to be called strikes when the umpire shares the race/ethnicity of the starting pitcher, an effect that only is observable when umpires' behavior is not well monitored. The evidence also suggests that this bias is strong enough to affect pitchers' measured performance and games' outcomes. As in many other fields, racial/ethnic preferences work in all directions—most people give preference to members of their own group. The difference in MLB, as in so many other fields of endeavor, is that power belongs disproportionately to members of the majority—White—group.

The type of discrimination that we have demonstrated is disturbing because of its implications for the sports labor market. In particular, non-White pitchers are at a significant disadvantage relative to their White peers, even in the absence of explicit wage discrimination by teams. Although some evidence suggests such explicit discrimination exists, i.e., there is a wage

---

compensation of starting pitchers of between 1 and 8 percent suggests that those pitchers are underpaid relative to White pitchers by between $50,000 and $400,000 per year.

gap among baseball players of different races, the fact that over 90 percent of the umpires are White implies that the *measured* productivity of non-White pitchers may be downward biased. Implicitly, estimates of wage discrimination in baseball that hold measured productivity (at least of pitchers) constant will understate its true size.

More generally, our results suggest caution in interpreting any estimates of wage discrimination stemming from equations relating earnings to race/ethnicity, even with a large set of variables designed to control for differences in productivity. To the extent that supervisors' evaluations are among the control variables included in estimates of wage discrimination, or even if they only indirectly alter workers' objective performances, their inclusion or their mere existence contaminates attempts to infer discrimination from adjusted racial/ethnic differences in wages. If racial/ethnic preferences in evaluator-worker matches are important, standard econometric estimates will generally understate the magnitude of racial/ethnic discrimination in labor markets.

While the specific evidence of racial/ethnic match preferences is disturbing, our novel analysis of the demand for discrimination should be encouraging: When their decisions matter more, and when evaluators are themselves more likely to be evaluated by others, our results suggest that these preferences no longer manifest themselves. These findings imply that it should not be difficult for MLB to devise methods to eliminate the impacts of racial/ethnic match preferences.[36] Clearly, raising the price of discrimination in the labor market generally through analogous methods is more difficult; but these results may suggest measures that might have the desired effects.

---

[36] Whether the installation of a new strike-zone evaluation tool (ZE) in all baseball parks, projected during 2007, created the same incentives as QuesTec and vitiated apparent umpire discrimination is not clear.

References

Bargh, John, "The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects," in Shelley Chaiken and Yaacov Trope, eds., *Dual-Process Theories in Social Psychology*. New York: Guilford Press, 1999.

Blair, Irene, "The Malleability of Automatic Stereotypes and Prejudice," *Personality and Social Psychology Review*, 6 (2002): 242-261.

Bradbury, J.C., *The Baseball Economist: The Real Game Exposed*. New York: Dutton, 2007.

Chen, J., "What Does Baseball Teach Us About Reducing Racial Discrimination? Evidence From Two Natural Experiments," Stanford University working paper, 2007.

Czopp, Alexander, Monteith, Margo, and Mark, Aimee, "Standing Up for a Change: Reducing Bias Through Interpersonal Communication," *Journal of Personality and Social Psychology*, 90 (May 2006), 784-803.

Devine, Patricia., and Monteith, Margo, "Automaticity and Control in Stereotyping," in Shelley Chaiken and Yaacov Trope, eds., *Dual-Process Theories in Social Psychology*. New York: Guilford Press, 1999.

Donald, Stephen, and Hamermesh, Daniel, "What Is Discrimination? Gender in the American Economic Association," *American Economic Review*, 96 (September 2006): 1283-1292.

Findlay, David, and Reid, Clifford, "Voting Behavior, Discrimination and the National Baseball Hall of Fame," *Economic Inquiry*, 35 (July 1997): 562-578.

Garicano, Luis, Palacios-Huerta, Ignacio and Prendergast, Canice "Favoritism under Social Pressure," *Review of Economics and Statistics*, 87 (May 2005): 208-216.

Goodwin, Stephanie, Gubin, Alexandra, Fiske, Susan, and Yzerbyt, Vincent, "Power Can Bias Impression Processes: Stereotyping Subordinates by Default and By Design," *Group Processes and Intergroup Relations*, 3 (2000): 227-256.

Gwartney, James, and Haworth, Charles, "Employer Costs and Discrimination: The Case of Baseball," *Journal of Political Economy*, 82 (July-August 1974): 873-881.

Kahn, Lawrence, "Racial Discrimination in Professional Sports: A Survey of the Literature," *Industrial and Labor Relations Review*, 44 (April 1991): 395-418.

Krautmann, Anthony, Gustafson, Elizabeth, and Hadley, Lawrence, "A Note on the Structural Stability of Salary Equations: Major League Baseball Pitchers," *Journal of Sports Economics*, 4 (February 2003): 56-63.

Nardinelli, Clark, and Simon, Curtis, "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball," *Quarterly Journal of Economics*, 105 (August 1990): 575-595.

Pascal, Anthony, and Rapping, Leonard. "The Economics of Racial Discrimination in Organized Baseball," in Anthony Pascal, ed., *Racial Discrimination in Economic Life*. Lexington, MA: Heath, 1972.

Price, Joseph, and Wolfers, Justin, "Racial Discrimination Among NBA Referees," National Bureau of Economic Research, Working Paper No. 13206, June 2007.

Scully, Gerald, "Pay and Performance in Major League Baseball," *American Economic Review*, 64 (December 1974): 915-930.

Stoll, Michael, Raphael, Steven and Holzer, Harry, "Black Job Applicants and the Hiring Officer's Race," *Industrial and Labor Relations Review*, 57 (January 2004): 267-287.

Zitzewitz, Eric, "Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making," *Journal of Economics and Management Strategy*, 15 (Spring 2006): 67-99.

Appendix – A Model of Bias-Induced Changes in Player Strategies

Not for Publication

Consider the following simple representation of the interaction between the pitcher and hitter. Denote the horizontal distance from the center of the plate $\tilde{y}$. Assume for simplicity that the pitcher can control the width of pitches (i.e., the horizontal dimension), but not their height. Further suppose that the batter is left-handed, and that the pitcher never aims left of center, i.e., $\mu = 0$. This simplifying assumption is little more than a normalization, although a realistic one, as pitchers are usually cautious to avoid hitting the batter.

The game unravels as follows.

1. The pitcher moves first. He can select his aim, $\mu \geq 0$, but not the final pitch location, $\tilde{y}$, which is random. On average, the pitcher's aim is correct, i.e., $E(\tilde{y}) = \mu$.
2. The batter moves next. A batter must decide whether to swing or not soon after a pitch is thrown, but before it reaches its final location $\tilde{y}$. To capture this timing, the batter's swing decision is made immediately after observing μ.[37]
3. If the batter does not swing, two outcomes are possible. For a given value of μ, with probability $s(\mu)$, the pitch is called a strike, and confers the batter a payoff S. With probability $1 - s(\mu)$, the pitch is called a ball, with payoff $B > S$. We assume $s' < 0$, $s'' < 0$, i.e., that pitches aimed closer to the plate are more likely to be called strikes, and at an increasing rate.
4. If the batter swings, two additional outcomes are possible. With probability $h(\mu)$, the batter gets a hit, and enjoys a payoff H. With probability $1 - h(\mu)$, the batter does not get a hit, with payoff $N < H$.[38] Similar to the assumptions for s, we assume $h' < 0$, $h'' < 0$.

---

[37] This strict timing assumption is not crucial. Instead, it is a simplified way of modeling that the batter makes his swing decision under imperfect information. For example, the batter could instead observe a noisy signal of $\tilde{y}$ without changing the results.

[38] Here, N captures the average payoff of swinging and missing, *(S)*, and hitting into an out.

The Batter's Problem:

To determine whether he swings at a pitch with expected location μ, the batter compares his expected payoff from swinging,

$$\pi(swing|\mu) = h(\mu)H + [1 - h(\mu)]N \tag{1}$$

with that from not swinging,

$$\pi(no\ swing|\mu) = s(\mu)S + [1 - s(\mu)]B. \tag{2}$$

Lemma 1. *Assume* $\pi(swing \mid \mu=0) > \pi(no\ swing) \mid \mu=0)$, *so that a batter always prefers to swing at a pitch aimed down the center of the plate. Then there exists a unique cutoff M whereby if: i)* $\mu < \hat{\mu}$, *the batter strictly prefers to swing, ii)* $\mu > \hat{\mu}$, *the batter strictly prefers to not swing, and iii)* $\mu = \hat{\mu}$, *the batter is indifferent between swinging and not.*

Proof. $\partial(\pi(swing|\mu)) / \partial\mu < 0$, *which follows from the assumptions that: i) called strikes are assumed to be more likely when thrown closer to the plate,* $s' < 0$, *ii) the batter's expected payoff from called balls is higher than that from called strikes,* $B > S$. *By similar logic,* $\partial(\pi(no\ swing|\mu)) / \partial\mu > 0$. *The convexity assumptions* s'', $h'' < 0$ *then guarantee a single crossing for* $(\pi\ (swing)\ |\mu)$ *and* $(\pi\ (no\ swing)\ |\mu)$, *which we denote* $\hat{\mu}$.

The intuition for Lemma 1 is straightforward. Batters will not attempt to hit pitches that have very little chance of being called a strike should they not swing, i.e., for sufficiently low values of μ .Moreover, the cutoff for swinging $\hat{\mu}$ is a function of the payoffs *S*, *B*, *H*, and *N* that correspond to the possible outcomes of the plate appearance. Generally, these payoffs will depend on game conditions, such as the score, the count, runners on base, or other factors that determine the payoffs to each outcome. For example, with runners on second and third base but no outs, the benefit of a hit (*H*) is substantial, where the cost of hitting into an out (*N*) is relatively small. In this situation, the batter will be less selective at the plate, which increases the swinging cutoff $\hat{\mu}$. We do not model differences in these payoffs across plate appearance, although the present set-up easily allows for this extension.

Our main interest is in how changes in the conditional strike function, $s(\mu)$, influence the batter's optimal behavior. Specifically, assume that the race/ethnicity match of the umpire and pitcher influences the probability that a pitch aimed at location μ will be called a strike. If the pitcher and umpire match (*M*), denote the conditional called strike probability $s_M(\mu)$. If they are different (*D*), the conditional strike probability becomes $s_D(\mu)$. To capture the idea that similar race or ethnicity helps the pitcher, we assume:

$$s_M(\mu) > s_D(\mu),\ \forall \mu \tag{3}$$

In other words, the same pitch has a different probability of being called a strike, conditional on whether the umpire and pitcher have the same or different races or ethnicities.

Lemma 2. *When the pitcher and umpire share the same race/ethnicity, the batter swings at pitches further from the center of the plate. That is, the cutoff location under a match is strictly greater than the cutoff location otherwise, i.e.,* $\hat{\mu}_M > \hat{\mu}_D$.

Proof. *Denote* $\hat{\mu}_M$ *as the cutoff swinging location when* $s(\mu) = s_M(\mu)$ *and* $\hat{\mu}_D$ *as that when* $s(\mu) = s_D(\mu)$. *Suppose* $s(\mu) = s_M(\mu)$ *and* $\mu = \hat{\mu}_M$. *From equation (2), when* $s(\mu)$ *changes to* $s_D(\mu)$ *the expected payoff of not swinging declines by* $[s_M(\hat{\mu}_M) - s_D(\hat{\mu}_M)](S - B) > 0$, *while the payoff from swinging is unchanged. We can now use the proof for Lemma 1. Because* $\partial(\pi(swing \mid \mu)) / \partial\mu < 0$ *and* $\partial(\pi(no\ swing \mid \mu)) / \partial\mu > 0$. *the new cutoff* $\hat{\mu}_D$ *is strictly less than* $\hat{\mu}_M$.

Lemma 2 indicates that when the batter anticipates judgments that favor the pitcher, his optimal strategy changes. Expecting the umpire's bias to reduce his payoff from not swinging, the batter takes matters into his own hands by swinging at pitches that he would otherwise let pass. Empirically, this implies a distinct advantage to the pitcher, not only for pitches that are called, but also for pitches that are hit. We complete this exercise by extending consideration to the pitcher's optimal strategy.

The Pitcher's Problem:

The pitcher's choice variable is μ, the expected location of the pitch. His expected payoff is the inverse of the batter's. If the batter swings, then the pitcher's expected payoff is $-h(\mu)H-[1-h(\mu)]N$. If the batter does not swing, then his expected payoff is $-s(\mu)S-[1-s(\mu)]B$.

Lemma 3. *The pitcher's optimal pitch location is* $\hat{\mu}$, *so that the batter is indifferent between swinging and not.*

Proof. *The batter will swing at any pitch aimed at* $\mu < \hat{\mu}$, *but because* $\partial(\pi(swing \mid \mu)) / \partial\mu < 0$, *the pitcher is always strictly better off increasing μ given that the batter will swing. The batter will not swing at any pitch aimed at* $\mu > \hat{\mu}$, *but because* $\partial(\pi(no\ swing \mid \mu)) / \partial\mu > 0$, *the pitcher* *will always decrease μ given that the batter will not swing. It follows then that the optimal pitch location must be* $\hat{\mu}$.

The model's main prediction is that the umpire's bias influences not only called strikes and balls, but also pitches where the umpire's judgment plays no direct role. Lemma 3 shows that the umpire's judgment influences the choice of pitch location, which in turn influences the batter's incentive to swing at the ball. It follows that *conditional on swinging*, the batter is less likely to hit the ball when the umpire and pitcher share race or ethnicity. As indicated by the model, this is because pitches are, on average, more difficult to hit in these situations.

Table 1: Pitch Outcome Summary

This table presents the pitch distribution across seven possible pitch outcomes for regular-season Major League Baseball games in the 2004-2006 seasons. The first row of the table summarizes all pitches, while subsequent rows sub-divide pitches based on the race/ethnicity of the pitcher, the batter and the home plate umpire, respectively. This table also reports the number of pitchers, batters and home-plate umpires in each of the four race/ethnicity categories.

	Called Strike	Called Ball	Swinging Strike	Foul	In Play	Intentional Ball	Hit by Pitch
All	360,809	771,314	188,989	362,381	417,211	13,956	5,506
Pitcher							
White (N=669)	260,601	552,545	132,574	259,752	301,718	10,018	3,883
Hispanic (N=219)	81,175	176,967	46,219	83,184	92,805	3,222	1,326
Black (N=27)	8,489	19,229	5,014	9,357	10,215	288	134
Asian (N=29)	10,544	22,573	5,182	10,088	12,473	428	163
Batter							
White (N=833)	189,239	401,755	98,314	185,183	208,976	6,601	3,156
Hispanic (N=385)	107,219	228,911	56,167	111,248	131,292	4,537	1,430
Black (N=154)	57,208	125,956	31,352	58,794	68,651	2,472	838
Asian (N=31)	7,143	14,692	3,156	7,156	8,292	346	82
Umpire							
White (N=85)	329,826	704,531	172,858	331,463	381,534	12,829	5,047
Hispanic (N=3)	10,681	22,884	5,471	10,488	12,198	402	174
Black (N=5)	20,302	43,899	10,660	20,430	23,479	725	285

Table 2: Summary of Umpires' Calls by Umpire-Pitcher Racial/Ethnic Match

This table reports the number of pitches thrown, the number of called pitches, the number of called strikes and the percentage of called pitches that are strikes for each pitcher/umpire racial/ethnic combination for regular-season Major League Baseball games in the 2004-2006 seasons. The final column reports the total percentage of called pitches that are strikes for each category of umpire race/ethnicity, while the final row reports the total percentage of called pitches that are strikes for each category of pitcher race/ethnicity. The unconditional percentage of called pitches that are strikes in the dataset is 31.87.

	Pitcher Race/Ethnicity				
	White	**Hispanic**	**Black**	**Asian**	**Total percent called strikes**
Umpire Race/Ethnicity					
White					
Pitches	1,388,318	445,107	47,797	56,866	
Called pitches	741,729	236,937	25,108	30,583	
Called strikes	237,798	74,564	7,686	9,777	
Percent called strikes	32.06	31.47	30.61	31.97	31.89
Hispanic					
Pitches	45,603	13,737	1,552	1,406	
Called pitches	24,592	7,323	845	805	
Called strikes	7,847	2,329	260	245	
Percent called strikes	31.91	31.80	30.77	30.43	31.81
Black					
Pitches	87,170	26,054	3,377	3,179	
Called pitches	46,825	13,882	1,765	1,729	
Called strikes	14,951	4,285	543	522	
Percent called strikes	31.93	30.87	30.76	30.19	31.62
Total percent called strikes	32.05	31.45	30.62	31.84	31.87

Table 3: Effects of the Relationship between Pitcher and Umpire Race/Ethnicity

This table presents the results of estimating equations where the pitcher and umpire's race/ethnicity are allowed to influence the likelihood of a called strike. The dependent variable is an indicator of whether a called pitch is a strike. All estimates are based on linear-probability models (but probit estimates present the same picture) with heteroskedasticity-robust standard errors. We account for any autocorrelation by clustering pitches by pitcher. The sample consists of all called pitches thrown during regular season MLB games in the 2004-2006 seasons. The first three columns show analysis separately for White, Black and Hispanic pitchers, respectively, controlling for umpire race/ethnicity and pitcher fixed effects. The next three columns show analysis separately for White, Black and Hispanic umpires, respectively, controlling for pitcher race/ethnicity and umpire fixed effects. The final three columns include all pitchers and umpires, with each column adding successive vectors of fixed effects. UPM is an indicator of whether the umpire and pitcher match on race/ethnicity. Pitcher Score Advantage is defined as the number of runs, potentially negative, that the pitcher's team is ahead at the time of the pitch. Top of Inning is an indicator which takes a value of 1 if the home team is pitching. P, U and B represent pitcher, umpire and batter fixed effects, respectively. Standard errors are in parenthesis.

Umpires	***All***	***All***	***All***	***White***	***Black***	***Hispanic***	***All***	***All***	***All***
Pitchers	***White***	***Black***	***Hispanic***	***All***	***All***	***All***	***All***	***All***	***All***
	(1)	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***	***(9)***
Black Umpire	-0.0025	0.0019	-0.0041						
	(-0.002)	(-0.009)	(-0.004)						
Hispanic Umpire	-0.0040	0.0034	0.0076						
	(-0.003)	(-0.013)	(-0.006)						
Black Pitcher				-0.0187	-0.0196	-0.0124			
				(-0.007)	(-0.011)	(-0.017)			
Hispanic Pitcher				-0.0069	-0.0110	0.0040			
				(-0.003)	(-0.006)	(-0.007)			
Asian Pitcher				-0.0056	-0.0151	-0.0307			
				(-0.006)	(-0.007)	(-0.020)			
UPM							0.0034	0.0028	0.0027
							(-0.002)	(-0.002)	(-0.002)
Pitch Count									
0&1	-0.2270	-0.2140	-0.2140	-0.2230	-0.2180	-0.1990	-0.2240	-0.2240	-0.2240
	(-0.003)	(-0.010)	(-0.005)	(-0.002)	(-0.006)	(-0.008)	(-0.002)	(-0.002)	(-0.002)
0&2	-0.3540	-0.3450	-0.3440	-0.3490	-0.3340	-0.3500	-0.3510	-0.3510	-0.3530
	(-0.003)	(-0.014)	(-0.006)	(-0.003)	(-0.007)	(-0.008)	(-0.003)	(-0.003)	(-0.003)
1&0	-0.0282	-0.0324	-0.0177	-0.0274	-0.0256	-0.0385	-0.0258	-0.0255	-0.0245
	(-0.002)	(-0.010)	(-0.004)	(-0.002)	(-0.006)	(-0.009)	(-0.002)	(-0.002)	(-0.002)
1&1	-0.1920	-0.1990	-0.1860	-0.1910	-0.1860	-0.1810	-0.1900	-0.1900	-0.1890
	(-0.003)	(-0.009)	(-0.005)	(-0.002)	(-0.006)	(-0.010)	(-0.002)	(-0.002)	(-0.002)
1&2	-0.3290	-0.3140	-0.3150	-0.3250	-0.3080	-0.3220	-0.3250	-0.3250	-0.3240
	(-0.003)	(-0.016)	(-0.005)	(-0.003)	(-0.007)	(-0.008)	(-0.003)	(-0.003)	(-0.003)
2&0	0.0430	0.0122	0.0507	0.0407	0.0498	0.0303	0.0447	0.0452	0.0461
	(-0.003)	(-0.011)	(-0.007)	(-0.003)	(-0.010)	(-0.013)	(-0.003)	(-0.003)	(-0.003)
2&1	-0.1570	-0.1900	-0.1440	-0.1580	-0.1380	-0.1540	-0.1540	-0.1540	-0.1500
	(-0.003)	(-0.012)	(-0.007)	(-0.003)	(-0.010)	(-0.013)	(-0.003)	(-0.003)	(-0.003)
2&2	-0.2940	-0.2860	-0.2900	-0.2940	-0.2730	-0.2950	-0.2920	-0.2930	-0.2890
	(-0.003)	(-0.014)	(-0.005)	(-0.003)	(-0.007)	(-0.010)	(-0.003)	(-0.003)	(-0.003)
3&0	0.2060	0.1520	0.2120	0.1980	0.2410	0.1860	0.2060	0.2070	0.2110
	(-0.005)	(-0.025)	(-0.009)	(-0.004)	(-0.013)	(-0.020)	(-0.004)	(-0.004)	(-0.004)
3&1	-0.0644	-0.0376	-0.0574	-0.0669	-0.0379	-0.0726	-0.0611	-0.0605	-0.0586
	(-0.005)	(-0.031)	(-0.007)	(-0.004)	(-0.014)	(-0.018)	(-0.004)	(-0.004)	(-0.004)
3&2	-0.2600	-0.2560	-0.2510	-0.2610	-0.2350	-0.2680	-0.2580	-0.2570	-0.2520
	(-0.004)	(-0.014)	(-0.007)	(-0.003)	(-0.011)	(-0.014)	(-0.003)	(-0.003)	(-0.003)
Inning									
2	-0.0058	-0.0150	-0.0060	-0.0048	-0.0129	-0.0156	-0.0057	-0.0057	-0.0114
	(-0.002)	(-0.013)	(-0.004)	(-0.002)	(-0.008)	(-0.009)	(-0.002)	(-0.002)	(-0.002)
3	-0.0163	-0.0136	-0.0152	-0.0154	-0.0155	-0.0193	-0.0156	-0.0155	-0.0262
	(-0.002)	(-0.009)	(-0.004)	(-0.002)	(-0.007)	(-0.010)	(-0.002)	(-0.002)	(-0.002)
4	-0.0341	-0.0375	-0.0269	-0.0308	-0.0353	-0.0525	-0.0317	-0.0317	-0.0339
	(-0.002)	(-0.008)	(-0.004)	(-0.002)	(-0.008)	(-0.010)	(-0.002)	(-0.002)	(-0.002)
5	-0.0262	-0.0329	-0.0254	-0.0262	-0.0172	-0.0344	-0.0258	-0.0259	-0.0349
	(-0.002)	(-0.008)	(-0.004)	(-0.002)	(-0.007)	(-0.011)	(-0.002)	(-0.002)	(-0.002)
6	-0.0332	-0.0351	-0.0308	-0.0318	-0.0305	-0.0548	-0.0329	-0.0330	-0.0361
	(-0.002)	(-0.015)	(-0.004)	(-0.002)	(-0.007)	(-0.011)	(-0.002)	(-0.002)	(-0.002)
7	-0.0256	-0.0189	-0.0232	-0.0237	-0.0184	-0.0391	-0.0249	-0.0251	-0.0294
	(-0.002)	(-0.014)	(-0.004)	(-0.002)	(-0.007)	(-0.011)	(-0.002)	(-0.002)	(-0.002)
8	-0.0254	-0.0387	-0.0202	-0.0216	-0.0153	-0.0382	-0.0245	-0.0249	-0.0284
	(-0.003)	(-0.014)	(-0.005)	(-0.003)	(-0.007)	(-0.011)	(-0.002)	(-0.002)	(-0.002)
9+	-0.0152	-0.0151	-0.0172	-0.0094	0.0042	-0.0305	-0.0157	-0.0163	-0.0203
	(-0.003)	(-0.016)	(-0.006)	(-0.003)	(-0.008)	(-0.011)	(-0.003)	(-0.003)	(-0.003)
Pitcher Score Advantage	0.0018	0.0028	0.0017	0.0024	0.0017	0.0009	0.0018	0.0018	0.0018
	0.0000	(-0.001)	0.0000	0.0000	(-0.001)	(-0.001)	0.0000	0.0000	0.0000
Top of Inning	0.0077	0.0175	0.0047	0.0066	0.0070	0.0073	0.0071	0.0071	0.0065
	(-0.001)	(-0.003)	(-0.002)	(-0.001)	(-0.004)	(-0.006)	(-0.001)	(-0.001)	(-0.001)
Observations	812,745	27,721	258,562	1,034,379	64,201	33,565	1,132,145	1,132,145	1,132,145
R^2	0.09	0.08	0.09	0.09	0.08	0.08	0.09	0.09	0.09
Fixed Effects	P	P	P	U	U	U	P	PU	PUB

Table 4: Explicit Monitoring of Umpires and Racial Discrimination

This table presents the results of estimating equation (1) separately for ballparks installed with QuesTec, a system of cameras intended to rate the umpire's performance, and those ballparks without it. The dependent variable is an indicator of whether a called pitch is a strike. All estimates are based on linear-probability models with heteroskedasticity-robust standard errors. The sample consists of all called pitches thrown during regular season MLB games in the 2004-2006 seasons. The first column shows analysis for pitches thrown in ballparks with QuesTec installed, while the second column analyzes only pitches in non-QuesTec ballparks. The final column aggregates all called pitches. UPM is an indicator of whether the umpire and pitcher match on race/ethnicity. All columns include the same set of control variables shown in Table 3—the indicators for inning, count, pitcher score advantage and the top of the inning. All columns also include fixed effects: 1) For each pitcher interacted with whether he pitched in a QuesTec ballpark, i.e., two fixed effects for each pitcher who pitched in both a ballpark where QuesTec was and was not installed; 2) For each umpire, i.e., umpire-QuesTec fixed effects, and 3) For each batter. Standard errors are in parenthesis.

	QuesTec ***(1)***	***Non-QuesTec*** ***(2)***	***All*** ***(3)***
Umpire-Pitcher Match (UPM)	-0.0035	0.0063	0.0064
	(0.0038)	(0.0029)	(0.0029)
QuesTec*UPM			-0.0098
			(0.0049)
Observations	420,125	712,020	1,132,145
R^2	0.09	0.09	0.09

Table 5: Implicit Monitoring of Umpires and Discrimination

This table presents the results of estimating equation (1), where the dependent variable is an indicator of whether a called pitch is a strike. All estimates are based on linear-probability models with heteroskedasticity-robust standard errors. The sample consists of all called pitches thrown during regular season MLB games in the 2004-2006 seasons. The first column of Panel A shows analysis for pitches thrown in games with above-median percentage attendance, while the second column analyzes only pitches in games with below-median percentage attendance. The final column of Panel A aggregates all called pitches. Column (4) of Panel B shows analysis for pitches thrown in terminal counts, i.e., counts with three balls or two strikes, such that the umpire's judgment can potentially end the at-bat. Column (5) analyzes only pitches in non-terminal counts, and Column (6) aggregates all called pitches. Column (7) analyzes only pitches throw in the first three innings of games, while Column (8) analyzes pitches thrown after the third inning. UPM is an indicator of whether the umpire and pitcher match on race/ethnicity. All columns include the same set of control variables shown in Table 3—the indicators for inning, count, pitcher score advantage and the top of the inning. All columns also include pitcher, batter and umpire fixed effects. Standard errors are in parenthesis.

Panel A. Distinguishing by Game Attendance

	High Attendance ***(1)***	***Low Attendance*** ***(2)***	***All Games*** ***(3)***
Umpire-Pitcher-Match (UPM)	-0.0021	0.0068	0.0052
	(0.0034)	(0.0031)	(0.0024)
Well Attended (>69% capacity)			0.0062
			(0.0015)
Well Attended*UPM			-0.0051
			(0.0019)
Observations	546,855	585,290	1,132,145
R^2	0.09	0.09	0.09

Panel B. Distinguishing by Terminal Count and Inning

	Terminal ***(4)***	***Non-Terminal*** ***(5)***	***All Pitches*** ***(6)***	***Early Inning*** ***(7)***	***Late Inning*** ***(8)***
UPM	-0.00184	0.00426	0.00453	0.00612	0.00385
	(-0.004)	(-0.003)	(-0.002)	(-0.004)	(-0.003)
Terminal Count			-0.00790	-0.0108	-0.00606
***UPM**			(-0.003)	(-0.004)	(-0.003)
Observations	261,670	870,475	1,132,145	396,438	735,707
R^2	0.17	0.04	0.09	0.09	0.08

Table 6: Examination of Outcomes of Individual Plate Appearances

This table presents regressions of outcomes of plate appearances by batters. Panel A considers the specific outcome of striking out. In Column (1), the dependent variable in an indicator that takes a value of one for all strikeouts, whereas in Columns (2) and (3), the dependent variable takes a value of one only for "called" and "swinging" strikeouts, respectively. In swinging strikeouts, the batter either swings and misses on the final third strike or fouls into the catcher's mitt. In called strikeouts, the batter does not swing, but the umpire calls a final third strike. Columns (4) and (5) consider the effect of swinging strikeouts in parks equipped (not equipped) with QuesTec, a system of cameras intended to rate the umpire's performance. Panel B considers a larger family of outcomes. In Column (6), the dependent variable is whether the batter walked. In Column (7), the dependent variable is whether the batter gets a hit or not, conditional on having put the ball in play. In Columns (8) and (9), the same specification in Column (7) is considered separately for QuesTec and non-QuesTec parks. UPM is an indicator of whether the umpire and pitcher match on race/ethnicity. All specifications include the same set of control variables shown in Table 3—the indicators for inning, count, pitcher score advantage and the top of the inning. All columns also include pitcher, batter and umpire fixed effects. Standard errors are clustered by pitcher and are reported in parenthesis.

Panel A. Striking Out

Dependent Variable	***Strikeout All***	***Strikeout Called***	***Strikeout Swinging***	***Strikeout Swinging***	***Strikeout Swinging***
	All (1)	***All (2)***	***All (3)***	***QuesTec (4)***	***Non-QuesTec (5)***
UPM	0.00709	0.00040	0.00669	0.00398	0.00821
	(0.0036)	(0.0019)	(0.0034)	(0.0056)	(0.0048)
Observations	552,770	552,770	552,770	204,873	347,897
R^2	0.001	0.001	0.001	0.001	0.001

Panel B. Walks and Hit

Dependent Variable	***Walk***	***Hit***	***Hit***	***Hit***
	All (6)	***In Play (7)***	***In Play QuesTec (8)***	***In Play Non-QuesTec (9)***
UPM	-0.00093	-0.00330	-0.00258	-0.00360
	(0.0025)	(0.0012)	(0.0020)	(0.0016)
Observations	552,770	292,777	107,953	184,824
R^2	0.001	0.001	0.001	0.001

Table 7: Effect of Umpire and Starting Pitcher Race/Ethnicity on Performance

This table examines the effect of a racial/ethnic match between the pitcher and umpire on three measures of performance for starting pitchers in 7,124 Major League Baseball games during the 2004-2006 seasons. The dependent variable in Column (1) is an indicator of a win by the home team. In Column (2), the dependent variable is *GameScore*™, a composite metric calculated only for starting pitchers. The dependent variable in Column (3) is the number of runs allowed by each starting pitcher. UPM is an indicator of whether the umpire and pitcher match on race/ethnicity. Each specification contains controls for the home team's runs scored, as well as fixed effects for both the home and away starting pitchers, the opposing team, and the home plate umpire. Standard errors are reported in parenthesis.

Dependent Variable	***Win*** ***(1)***	***GameScore*™** ***(2)***	***Runs Allowed*** ***(3)***
UPM	0.0424	1.0238	-0.1635
	(0.0236)	(0.8219)	(0.1026)
Pitcher's Run Support	0.1345	-0.0728	0.0029
	(0.0033)	(0.0454)	(0.0056)
Observations	6979	14,229	14,229

Table 8: Batter-Umpire and Catcher-Umpire Race/Ethnicity Matches

This table presents the results of estimating equations where matches between the race/ethnicity of umpires with pitchers, batters or catchers are allowed to influence the likelihood of a called strike. The dependent variable is an indicator of whether a called pitch is a strike. All estimates are based on linear-probability models with heteroskedasticity-robust standard errors. The sample consists of all called pitches thrown during regular season MLB games in the 2004-2006 seasons. UPM, UBM and UCM are indicators of whether the umpire matches the race/ethnicity of the pitcher, batter and catcher, respectively. The QuesTec variable is an indicator of whether the pitch is thrown in a ballpark installed with QuesTec, a system of cameras intended to rate the umpire's performance. High Attendance is an indicator of whether a pitch is thrown in a game with above-median percentage attendance. Terminal Count is an indicator of whether a pitch is thrown in a terminal count, i.e., a count with three balls or two strikes, such that the umpire’s judgment can potentially end the at-bat. All columns include the same set of control variables shown in Table 3—the indicators for inning, count, pitcher score advantage and the top of the inning. All columns also include fixed effects for each pitcher-QuesTec, umpire-QuesTec, and batter-QuesTec combination. Standard errors are in parenthesis.

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Umpire Pitcher Match (UPM)			0.00265	0.00635	0.00540	0.00444	0.0106
			(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Umpire Batter Match (UBM)	-0.00005		-0.00006	0.00048	-0.00009	-0.00074	
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)	
Umpire Catcher Match (UCM)		0.00081	0.00077	0.00018	-0.00082	0.00130	
		(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	
Questec*UPM				-0.00994			-0.00966
				(0.005)			(0.005)
Questec*UBM				-0.00135			-0.00132
				(0.004)			(0.003)
Questec*UCM				0.00184			0.00093
				(0.002)			(0.002)
High attendance					0.00463		0.00497
					(0.002)		(0.002)
High attendance*UPM					-0.00562		-0.00546
					(0.002)		(0.002)
High attendance*UBM					0.00013		0.00022
					(0.002)		(0.002)
High attendance*UCM					0.00323		0.00234
					(0.002)		(0.002)
Terminal count*UPM						-0.00776	-0.00768
						(0.003)	(0.003)
Terminal count*UBM						0.00306	0.00310
						(0.002)	(0.002)
Terminal count*UCM						-0.00228	-0.00277
						(0.002)	(0.002)
UPM*UBM							-0.00077
							(0.002)
UPM*UCM							0.00081
							(0.002)
Observations	1,132,145	1,132,145	1,132,145	1,132,145	1,132,145	1,132,145	1,132,145
R^2	0.09	0.09	0.09	0.09	0.09	0.09	0.09

Figure 1: Race and Called Strike Percentage in QuesTec and Non-QuesTec Ballparks
This figure graphs the average percentage of called pitches that are strikes in ballparks with and without QuesTec, a system of cameras intended to rate the umpire's performance, for White and non-White pitchers, respectively. For each pair of bars, the bar on the left represents pitches for which the umpire matches the race/ethnicity of the pitcher, while the bar on the right represents pitches for which there is not a match.

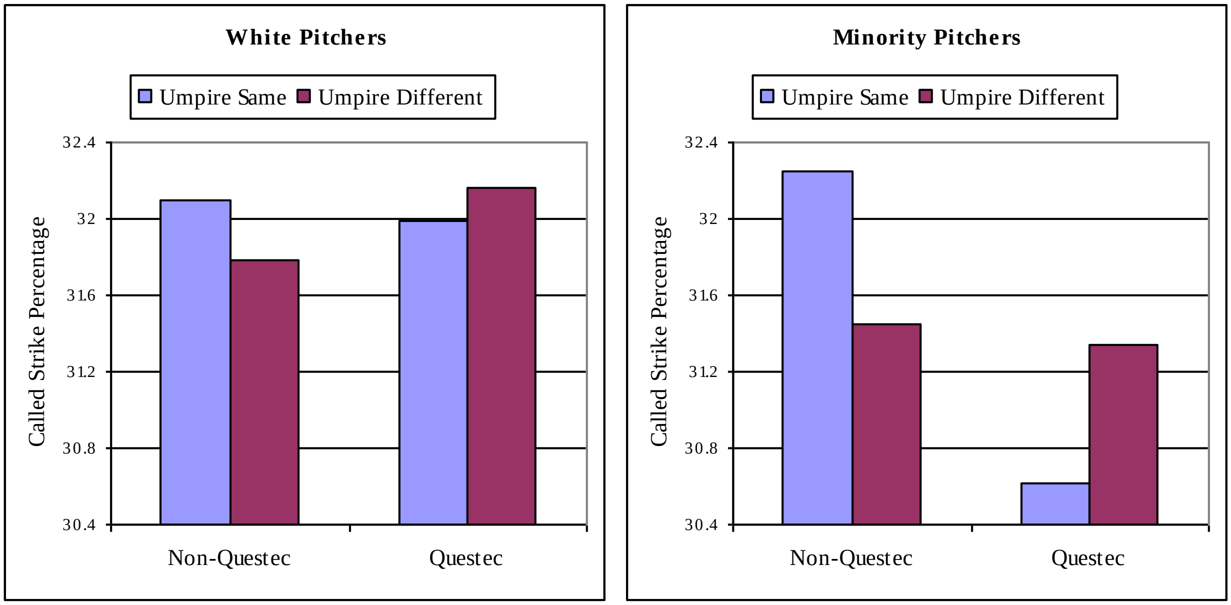

Figure 2: Race and Called Strike Percentage by Game Attendance
This figure graphs the average percentage of called pitches that are strikes when percentage attendance is below the median and above the median for White and non-White pitchers, respectively. For each pair of bars, the bar on the left represents pitches for which the umpire matches the race/ethnicity of the pitcher, while the bar on the right represents pitches for which there is not a match.

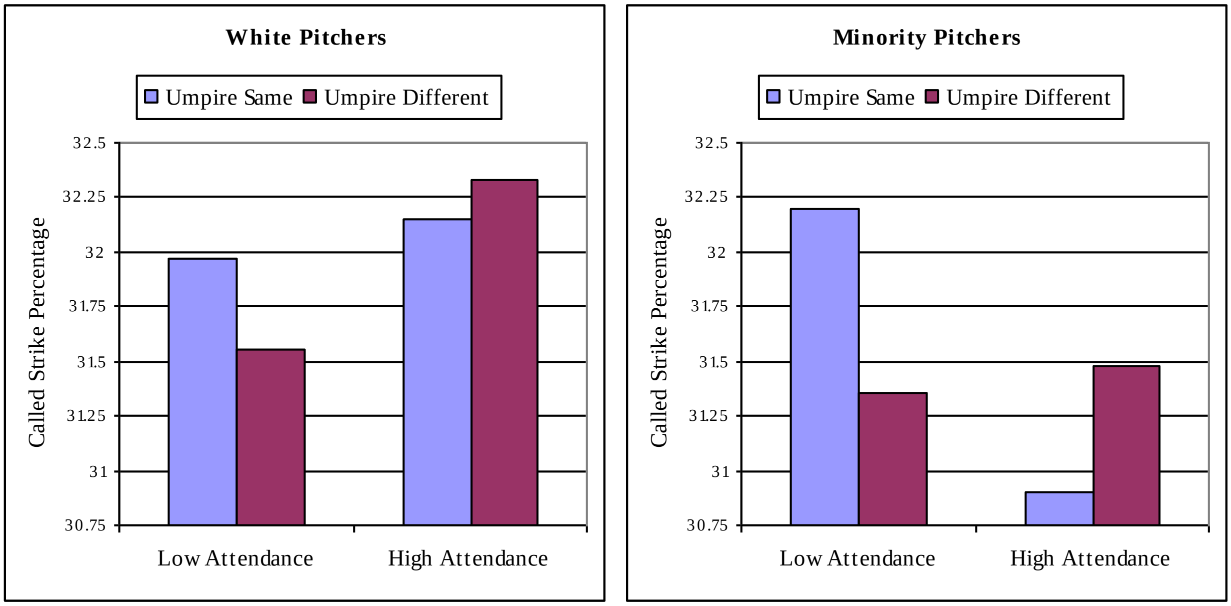

Figure 3: Race and Called Strike Percentage in Terminal and Non-Terminal Counts

This figure graphs the average percentage of called pitches that are strikes for pitches thrown in non-terminal and terminal counts for White and non-White pitchers, respectively. A terminal count is defined as a count with three balls or two strikes, such that the umpire's judgment can potentially end the at-bat. For each pair of bars, the bar on the left represents pitches for which the umpire matches the race/ethnicity of the pitcher, while the bar on the right represents pitches for which there is not a match.

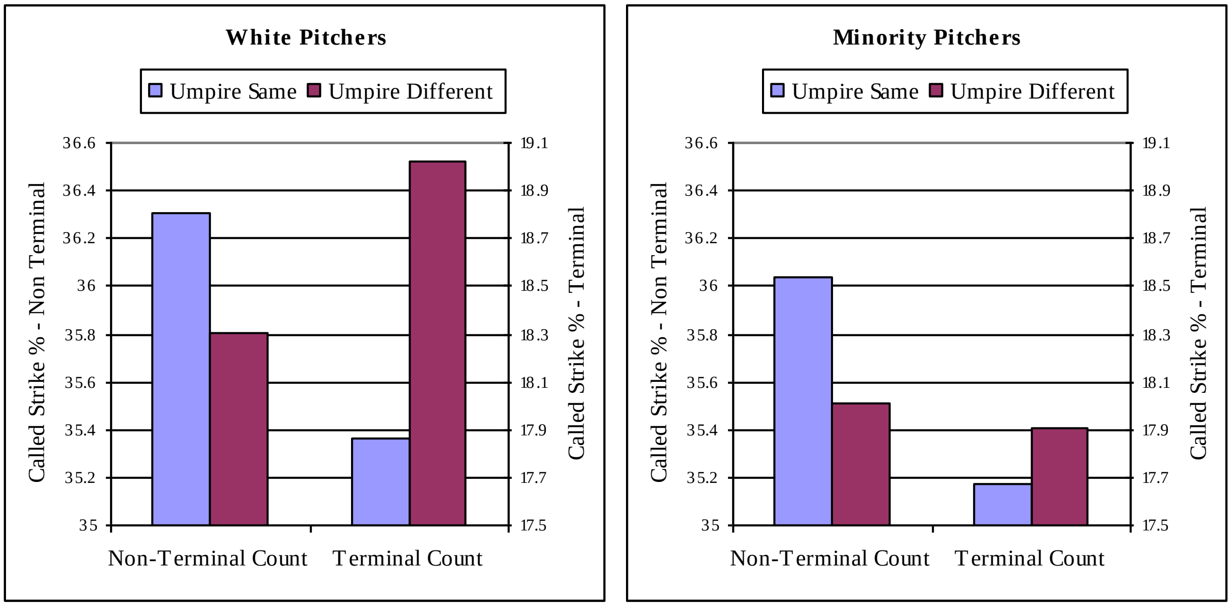

Figure 4: Change in Pitcher Performance When Umpire Matches Race/Ethnicity

This figure graphs the percentage change in starting pitcher performance when the pitcher and umpire match race/ethnicity versus the baseline of a race/ethnicity mismatch. We examine eight performance measures for the 14,248 starting pitchers in 7,124 regular-season games in the 2004-2006 Major League Baseball seasons. GameScore is a composite index designed by Bill James to summarize the starting pitcher's performance. K represents the number of strikeouts by the starting pitcher, HR is the number of homeruns allowed by the pitcher, and ERA is the pitcher's earned run average.