

SEEING STARS: MATTHEW EFFECTS AND STATUS BIAS IN MAJOR LEAGUE BASEBALL UMPIRING

Jerry W Kim

Columbia University

Columbia Business School

719 Uris Hall

3022 Broadway

New York, NY 10027

jwk2108@columbia.edu

Brayden G King

Northwestern University

Kellogg School of Management

2001 Sheridan Road

Evanston, IL 60208

b-king@kellogg.northwestern.edu

Abstract

This paper tests the assumption that evaluators are biased to positively evaluate high status individuals, irrespective of quality. Using unique data from Major League Baseball umpires' evaluation of pitch quality, which allow us to observe the difference in a pitch's objective quality and in its perceived quality as judged by the umpire, we show that umpires are more likely to over-recognize quality by expanding the strike zone, and less likely to under-recognize quality by missing pitches in the strike zone for high status pitchers. Ambiguity and the pitcher's reputation as a "control pitcher" moderate the effect of status on umpire judgment. Further, we show that umpire errors resulting from status bias lead to actual performance differences for the pitcher and team.

Keywords: Status, Bias, Performance

Forthcoming in *Management Science*

Social scientists have long understood status to be an indicator of hierarchical position and prestige that helps individuals and organizations procure resources and opportunities for advancement (e.g., Whyte 1943; Podolny 2001; Sauder, Lynn, and Podolny 2012). Status markers, like ranking systems in education (Espeland and Sauder 2007) or awards and prizes (Rossman et al. 2010), accentuate quality differences among actors and create greater socio-economic inequality. Past research has advanced our understanding of how status hierarchies emerge and persist (Berger, Rosenholtz, and Zelditch 1980; Webster and Hysom 1998; Ridgeway and Correll 2006), on the psychological rewards of status attainment (e.g., Willer 2009; Anderson et al. 2012), on the association of status with characteristics such as race and gender (Ridgeway 1991), or on status outcomes, such as the accumulation of power, price and wage differentials, and other economic and social benefits (Podolny 1993; Benjamin and Podolny 1999; Thye 2000; Correll et al. 2007; Stuart, Hoang, and Hybels 1999; Bothner, Kim, and Smith 2011; Pearce 2011; Waguespack and Sorenson, 2011).

In economic sociology, the classic statement of self-reproducing advantages is Robert Merton's (1968) theory of the Matthew Effect. Underlying the Matthew Effect is an assumption that individuals are biased to positively evaluate high status individuals irrespective of quality, and that this bias, rather than actual quality differences, perpetuates inequality between high status and low status actors. Scholars in status characteristics theory have made the case that this bias results from individuals' expectations and prior beliefs about competence and quality (Anderson et al. 2001; Ridgeway and Correll 2006; Berger et al. 1977; Ridgeway and Berger 1986), and experimental and audit studies have identified status characteristics as a mechanism that underlies discriminatory evaluations of low status individuals (e.g., Correll et al. 2007).

Despite the evidence emanating from the experimental research in status characteristics studies, scholars studying the Matthew Effect in real world settings have had a more difficult time demonstrating that status bias is a mechanism that leads to cumulative advantages for high status actors. Due, in part, to the empirical and conceptual challenges of establishing what the correct amount of recognition should be given to an actor's performance, past field research has been less concerned with status bias and instead

has focused on evidence supporting the presence of status advantages. Disentangling status and quality poses considerable empirical challenges for research based in field settings, as the very conditions that make status valuable are those that make it harder to observe actual quality.

Recent studies have exploited quasi or natural experiments involving exogenous shifts in status signals uncorrelated with quality to provide a clearer picture about status advantages, i.e., high status actors outperforming low status actors (Simcoe and Waguespack 2011; Azoulay, Stuart and Wang 2013). While these study designs have helped isolate the causal effect of status, the unsolved challenge of establishing true quality (i.e., the amount of recognition the actor “should have” received) has made it difficult to document the presence and direction of status bias outside of experimental settings. Imagine a low and high status scientist receiving 10 and 100 citations for the exact same discovery. We have clear evidence of a status advantage, but to identify bias, we must be able to answer the question: is the fundamental value of a scientific discovery worth 10 citations or 100 citations? Because of this empirical challenge, non-experimental research on status and inequality can only observe the actual outcomes of evaluative processes but cannot show that the evaluative processes are actually biased by status differences.

The purpose of this paper is to establish evaluative bias as a source of status-based advantages by comparing judgments made by evaluators to precise measurements of actual quality. Drawing on status characteristics theory, we argue that evaluators are biased to recognize quality in high status actors, which can at times lead to erroneous over-recognition of quality. High status structures the expectations of evaluators and leads them to “see” quality in high status actors, especially when a performance is ambiguous. Furthermore, because this status bias is driven by expectations, we argue that an actor’s reputation (i.e., history of past performance), which also establishes high expectations, should moderate the impact of status. High status actors that are also known for certain kinds of performance will receive more favorable evaluations than high status actors who do not carry those same expectations. Thus, an actor’s past performance helps to magnify future advantages gained from occupying a high status position

inasmuch as the expectations from an actor's reputation and status are aligned. When an actor's reputation is not well aligned with that actor's status, he or she may not benefit from the same expectations.

We utilize a unique dataset of judgments made under uncertainty, coupled with precise data on quality that allow us to examine the influence that status has on the rate and direction of status bias. Major League Baseball (MLB) operates four cameras placed around each ballpark to record the locations and trajectories of pitches throughout the game. Leveraging the considerable amount of detail provided by these cameras, we compare the umpiring decision on balls and strikes with the actual location at which the pitch crossed home plate. We find that the evaluator (i.e., umpire) is more likely to recognize quality when quality is absent (i.e., calling a strike when a pitch was actually a ball) if the actor (i.e., pitcher) has high status, measured by the number of All-Star game selections. In addition, we find evidence that umpires are less likely to withhold recognition of true quality (i.e., not calling a strike for a good pitch) when the pitcher is higher in status. We further find that ambiguity and players' reputations shape the effect of these status biases, consistent with the expectations mechanism. Finally, we demonstrate how this evaluative bias translates into real performance advantages for the pitcher.

Status Bias in the Evaluation of Quality

Our understanding of the role of status in the Matthew Effect is largely driven by egocentric considerations – e.g., high status actors get more resources, which makes it less costly to produce high quality performances – and underplays the altercentric aspects that stem from their evaluators' assessments of performance. Evidence from social psychology, however, clearly indicates that status characteristics shape the cognition of actors assessing performance. Members of groups perceive high status members as being more influential, more technically competent and defer more to them (Anderson et al. 2001; Ridgeway and Correll 2006; Berger et al. 1977; Ridgeway and Berger 1986). Even diffuse status characteristics, such as ethnicity or gender, can hamper fair evaluation (Berger et al. 1985).

Research in social psychology has demonstrated that cognitive biases profoundly shape judgment and decision-making (e.g., Gilovich and Griffin 2002; Shah and Oppenheimer 2008). For example, research on gender shows that gender stereotypes situationally influence individuals' perceptions of competence and cause them to modify their behavior (Deaux and Major 1987; Ridgeway 2001). This research suggests that status-based screening is driven by unconscious biases that are “non-functional” in that they do not select for better outcomes, and are discriminatory (Simcoe and Waguespack, 2011). By saying that status biases evaluations, we mean that differences in recognition or evaluative outcomes for two actors (or groups) with identical quality but different status result from implicit and subconscious sensitivities to status by the evaluator.

Status characteristics theory and expectation states theory have found evidence of this bias in experimental studies (e.g., Berger et al. 1972; Berger et al. 1977; Ridgeway 1991; Wagner and Berger 2002; Correll et al. 2007). Status bias occurs when certain status characteristics invoke expectations of high performance from evaluators, which in turn shapes the perception of the actions being judged so that evaluators “see” high quality. On the other hand, low status members of groups such as women or racial minorities evoke lower performance expectations, and thus, their actions are more likely to be perceived as low quality. More lenient evaluation standards for high status actors, and harsher evaluation standards for lower status actors together create conditions for bias and mistaken judgments by the evaluating actor.

Expectations are, of course, shaped by a multitude of factors beyond status positions and characteristics. In particular, an actor's reputation—“observed past behavior” (Raub and Weesie 1990: 629)—is both a reflection of the expectations held by audiences, but also a key factor in determining what evaluators expect regarding the performance of an actor. The relationship between status and reputation has been a heavily debated topic in the recent literature (Stern, Dukerich, and Zajac 2013), and we argue that the two concepts are interlinked by the expectations they create in the evaluator's mind. More specifically, we propose that reputations that are consistent with status expectations will accentuate the biases associated with status, while reputations that go against expectations will diminish bias, by virtue of creating greater focus on quality.

Empirical Setting and Hypotheses

Sports performance provides an intriguing window to assess the social dynamics of status bias as it allows you to observe objective differences in performance outcomes and, in a quasi-experimental fashion, control for other factors that should affect performance, such as skill or situational characteristics. Status is highly visible in most sports and has a considerable influence over the behavior of all actors involved. Baseball in particular is a sport that is steeped in tradition, and status distinctions such as the Hall of Fame and All-Stars have been an integral part of the sport's prominence in American culture (Allen and Parsons 2006). Our analysis focuses specifically on baseball umpires' calls of non-swinging pitches during at-bats in Major League baseball games. From a batter's point of view, the purpose of an at-bat is to hit the ball into play and avoid getting three strikes called against him. In order to get a strike against the batter, a pitcher must pitch the ball through the strike zone—an imagined box extending upward from the batter's knees to slightly below his chest and out from the center of the plate to the plate's edges¹—without the batter putting the ball into play. The batter hopes to either hit the ball or to advance to first by getting walked when a pitcher misses the strike zone four times during an at-bat. For all non-swinging pitches, the umpire is the arbitrator of pitch quality, determining whether a pitch is a strike or a ball. Our analysis focuses on the effect of pitchers' status on umpire calls because their attention is directed toward the pitcher as they try to determine whether a pitcher's pitch is high quality (i.e., a strike), not whether a batter performed well during the pitch.²

For a number of reasons, umpire decision-making is an ideal setting to observe status bias in quality evaluation. Baseball umpires' evaluations of non-swinging pitches are critical to a pitcher's performance. Unlike other sports in which the officials are mainly responsible for enforcing players' errors or fouls, baseball umpires are directly involved in determining the quality of a player's performance. Another reason for studying status effects on umpire decision-making is because their calls are inherently subjected to bias. Across a number of sports, officials exhibit perceptual and cognitive biases that vary their subjective assessments of player performance (e.g., Plessner 2005; Ford, Goodwin,

and Richardson 1996), and in fact, sabermetricians have demonstrated that baseball umpires vary in their calling of strike zones.³ Umpires are called upon to make instantaneous judgments of quality without the benefit of deliberation (e.g., Rainey et al. 1989), creating the possibility that subjective aspects of the game will bias their judgment. For example, recent studies have shown that professional sports officials were biased toward players of their own race (Price and Wolfers 2010; Parsons et al. 2011)

The final reason that umpiring is an ideal setting to assess status bias is because umpires have strong incentives to call the strike zone as accurately and as consistently as possible. MLB has made recent efforts to closely monitor and evaluate their accuracy. For most of MLB's history, the umpire had nearly complete discretion in determining the boundaries of the strike zone (Weber 2009), but concerns about subjectivity and fairness led the MLB to institute a uniform strike zone that would be monitored by computer technology. In the early 2000s, improvements in camera and computer technology made it possible for major league officials to set up a monitoring system (known as the Pitch f/x system) that would capture the exact location of a ball at the moment it crossed the plate, precisely determining if a pitch was objectively inside the boundaries of the strike zone. By 2007 a camera system was set up in every major league ballpark, giving MLB administrators and owners an unprecedented level of information about umpire accuracy. The new system was meant to encourage umpires to call a uniform strike zone and to avoid giving certain players more leniency. Thus, if we observe umpires making mistakes, it is likely the result of biased judgment and not purposeful calculation.

Using data produced by Pitch f/x, we can observe the accuracy of umpires directly. Assuming umpires judge a pitch's quality by its location in the strike zone, we can verify if an umpire has under-recognized quality (i.e., calling a ball that was really a strike) or over-recognized quality (i.e., calling a strike that was really a ball). The new monitoring technology revealed what many observers had always believed – umpires are not completely accurate. Our analysis of non-swinging, called pitches indicates that umpires make the wrong call – either type of error – 14.66% of the time. As we argue, a pitcher's status in the league, may bias the umpires' error rates.

Status-based Over-recognition of Quality

Even though umpires seek to create a fair strike zone, they are also not immune from social and cognitive biases. Umpires rely on keen perception to discern the borders of the strike zone, but players may influence this perception. Ozzie Guillen, a former player and manager in the league, summarized this bias:

Everybody has their own strike zone... This year they come to us and say, We're gonna call the strike zone from here to here – ‘ he sliced the air at the knees and the letters – ‘but they have to say something every year to make us think they're working. But if Roger Clemens or Pedro Martinez or Greg Maddux, Tom Glavine, is pitching, it's a strike. Jose Cruz pitching? It's a ball. Same way for hitting. You're fucking Wade Boggs? That's a ball. You're Frank Thomas? That's a ball. Ozzie Guillen hitting? Strike, get the fuck out (as quoted in Weber 2009: 179).

The quote expresses the belief that for an identical pitch, high-status pitchers are more likely to receive a strike call than low-status pitchers. But to what extent are high-status pitchers benefitting from unfairly called strikes? While Guillen's quote implies that high-status players receive undue favoritism, it is unclear whether calling more strikes for high-status pitchers is actually a mistaken judgment or not. High status pitchers, after all, should be expected to throw more strikes.

It is precisely this expectation that high status pitchers will throw strikes that leads to strike calls even when the pitch does not warrant this judgment. Greg Maddux, a renowned pitcher in the National League, is an example of how expectations can alter the judgment of the umpire. A four-time Cy Young winner and eight-time All-Star, Maddux often received favorable calls from umpires, usually by locating his pitch just outside the lower half of the zone away from the hitter. Bob Gibson, another multiple time All-Star, described the favoritism that umpires showed to Maddux:

Maddux can throw a pitch three or four inches outside, and the umpire will say, 'Striiiiike!' because he's always there. If the umpires call that one, Maddux will come right back to that spot,

or maybe stretch it out another half-inch. His control is just that good. The umpires know who's out there, and they have a tendency to give you a break if you have good control and you're always right there where you want to be (Gibson 2009: 163).

The high expectations that umpires have for high status pitchers should cause them to expand their strike zone, leading to greater over-recognition of quality. The likelihood that a low quality pitch (i.e., ball) will be seen as a high quality pitch (i.e., strike) should be greater for high-status actors. Therefore, we hypothesize that:

Hypothesis 1: The higher the status of the pitcher, the more likely the umpire will mistakenly call a real ball a strike

Status-based Under-Recognition of Quality

Merton describes the Matthew Effect as consisting of “the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark” (1968, p. 56). Much of the existing sociological work on status has focused on how high status individuals receive more positive outcomes, but relatively little work has been done to affirm the second component of the Matthew Effect, i.e., the withholding of recognition due to a lack of status. Making a highly important scientific breakthrough, only to have gatekeepers in the field fail to notice, or question the value of the discovery can be a frustrating experience for any scholar, regardless of prominence. The important question from an inequality perspective is whether status can be a contributing factor to this under-recognition of quality.

In the context of baseball, the under-recognition of quality manifests itself when the umpire misses a pitch that was in the strike zone, and calls it a ball. As with the over-recognition of quality, we argue that the status-based expectations held by umpires will have an impact on how likely this type of error is committed. Evaluators expect low quality performance (i.e., balls) from low-status actors, and this

leads to evaluators “seeing” poor quality. Pitches must be closer to the center of the zone for these low-status pitchers for umpires to see them as strikes, and this in turn means that umpires will be more likely to miss pitches that are actually strikes. In contrast, umpires expect high status pitchers to provide quality (i.e., throw strikes), and as a result, are more likely to accurately call a strike when a pitch is in the zone. Thus,

Hypothesis 2: The higher the status of the pitcher, the less likely the umpire will mistakenly call a real strike a ball.

Conditions that Heighten Status-Bias: Reputation and Uncertainty

If expectations drive status bias, then conditions that make expectations more salient should increase the magnitude of bias. In particular, a pitcher’s reputation and performance ambiguity are conditions that should moderate status bias. Sociologists define reputation as “a characteristic or attribute ascribed to him by his partners,” and the empirical basis of an actor’s reputation is “his observed past behavior” (Raub and Weesie 1990: 629). Players’ reputations are largely determined by their performance in certain statistical categories, which summarize their past performance and are often used to forecast future performance. Performance ambiguity refers to the extent to which the quality of an actor’s performance is difficult to measure. In the case of baseball pitches, performance ambiguity refers to the degree that pitches are difficult to evaluate because they are located near the edges of the strike zone.

The aforementioned All-Star pitcher Greg Maddux benefitted not just from his high status, but also because he had a reputation for being a control pitcher who could accurately put the pitch in a desired location. Because umpires believed Maddux could control pitch location, they were more likely to give him the benefit of the doubt. If reputation shapes the expectations umpires have about pitch quality, one should expect that reputation will moderate the effects of status. Having a reputation for consistently locating high-quality pitches (i.e., strikes) should bias the umpires to an even greater degree. Thus, high status pitchers with reputations as “control pitchers” (such as Maddux) are even more likely to benefit

from increased over-recognition and reduced under-recognition of quality than other high status pitchers. On the other hand, if a high status pitcher has a poor reputation for being able to control the location of the pitch—a track record of not being able to throw strikes despite being able to get batters out—the pitcher’s status advantage may be reduced by the umpire’s expectation that the pitcher will not be able to carefully locate the ball. Taken together, we hypothesize that:

Hypothesis 3: The effect that pitcher status has on the umpire’s over-recognition and under-recognition will increase with the pitcher’s prior performance (i.e., reputation) in throwing strikes.

We should further expect that umpires assessing high status pitchers will be more generous in their recognition of quality for high status pitchers when the pitch location has greater ambiguity.⁴ Podolny and Phillips (1996) argued that high status actors tend to benefit more when there is considerable ambiguity. One reason for this is that when there is greater ambiguity evaluators will resort to their prior expectations about performance. For umpires, pitches that are close to the border of the zone are extremely difficult to judge, leading to a greater reliance on expectations. Another reason to expect an interaction effect between ambiguity and status is that if status biases the perception of quality, then extreme and unambiguous demonstrations of quality are more likely to override these biases and lead to more accurate judgments.⁵ All in all, we expect:

Hypothesis 4: The effect that pitcher status has on the umpire’s over-recognition and under-recognition will increase when the pitch is located closer to the border of the strike zone.

Data and Methods

In order to assess the influence of status on umpire judgments, we evaluated the outcomes from pitches thrown during Major League Baseball games during the 2008 and 2009 seasons. The unit of analysis is a called pitch. Other types of pitches, including swinging strikes or hits, were not used in the analysis since they did not involve an umpire's judgment. We use pitch-by-pitch data from Major League Baseball games played in the 2008 and 2009 seasons to create our pitch-level and at-bat level variables. We obtained these data about each pitch and at-bat from MLB Gameday, a service that collects and stores data about each pitch from every game played during the regular season, including data about pitch location estimated by Pitch f/x. The unit of analysis is the pitch itself, yet we also merged data about the at-bat, player characteristics, and situational characteristics from other publicly available sources including Baseball Databank, Retrosheet, and Fangraphs in order to account for contextual determinants' of each pitch. Our data set included 756,848 observations (i.e., pitches), which took place over 313,774 at-bats and 4,914 games.

Dependent variable

Our dependent variable is the occurrence of an umpire mistake. This could take place in two different ways. An umpire over-recognizes quality by mistakenly calling an actual ball a strike.⁶ This is equivalent to a Type I error, false positive, or error of commission. The second type of mistake happens when an umpire under-recognizes quality when he mistakenly calls an actual strike a ball. This is also known as a Type II error, false negative, or error of omission. To operationalize these errors we used data about the location of each pitch and compared those locations against the strike zone. The Pitch f/x system calculates both the location of the pitch as it crosses the plate and the size of the strike zone. To ensure accuracy in their estimates, the Pitch f/x system takes twenty-five pictures of the ball after it leaves the pitcher's hand until it hits the catcher's glove. Thus, the system is not only able to determine the location of the pitch as it crosses the plate, but it also records the velocity, movement, and trajectory of the pitch. For our purposes, we are only interested in the location, although we also use information about movement and velocity to create control variables for the analysis.

A pitch varies in its location relative to the strike zone both vertically and horizontally. To count as an actual strike, a pitch must be located both vertically and horizontally inside the strike zone. To estimate vertical location relative to the strike zone, we compared the height of the pitch as it crosses the plate (labeled pz in the Pitch f/x data) with the upper and lower barriers of the strike zone, also measured in feet (labeled sz_top and sz_bottom , respectively). If the height of the pitch is either lower than the bottom of the strike zone or higher than the top of the strike zone, then the pitch is objectively a ball. To estimate horizontal location relative to the strike zone, we compared the distance of the pitch from the middle of the plate, measured in feet, from the outer boundaries of the strike zone (labeled px in the Pitch f/x data). Any pitch that is more than .83 feet from the middle of the plate is objectively a ball. For example, if a strike zone's upper bound is 3.6 and its lower bound is 1.4, and the height of the pitch is 2 and the horizontal location is .05, the pitch would objectively be a strike.

[Table 1 about here]

After we determined whether each pitch was an actual strike or an actual ball, we then compared the actual location to the umpire's call to assess whether he made an erroneous call. Table 1 shows the breakdown of all called pitches from the 2008 and 2009 seasons. The high frequency of pitches that were mistakenly called a ball reinforces the idea that under-recognition, or the "withholding of recognition" is an equally, if not more, important phenomenon as the over-recognition of quality most associated with Matthew Effects. Because the kind of mistake that an umpire makes is conditional on the type of pitch thrown by a pitcher, we do two different analyses. In the first analysis, we assess the likelihood of over-recognizing quality for all called pitches of actual strikes. In the second analysis we assess the likelihood of under-recognizing quality for all called pitches of actual balls.

Independent variable

The main independent variable in our analysis is the status of the pitcher. Although we include the batter's status in later analyses, we focus on the status of the pitcher since the umpire is evaluating the

performance of the pitcher, and not the batter, when assessing whether a pitch is a ball or a strike. We measure status using a pitcher's *number of All-Star appearances* in prior years. All-Star appearances are an adequate measure of status in professional baseball for a number of reasons. Status refers to a privileged position of esteem and distinction that elevates one actor above another (Weber 1978; Simmel 1950; Berger et al. 1972). Status distinctions are often measured through rankings (e.g., Sauder and Espeland 2009) or are made visible through deference patterns in relationships (Podolny 1993); however, the exact way in which status is manifest depends greatly on the context in which it is operating. In professional baseball, the main way in which achievement and esteem is ritually honored is by voting a player into the annual All-Star game – an exhibition game played every July in which each league (National and American) is represented by its sixty-eight “best” players. Players are selected as an All-Star in a combination of fan voting, player voting, and manager selection. Since 2002, fans cast votes for their favorite players and elect seventeen starters to the All-Star game. Managers and players vote to decide who will be the sixteen pitchers and sixteen reserves to fill each position in the field. Finally, the manager for each team fills out the remaining roster spots by choosing players he thinks deserve recognition. Given that All-Star appearances are determined by the accumulated acts of deference (i.e., votes) from fans, players, and managers, this measure is consistent with the sociological view of status as the ‘stock’ that corresponds to the ‘flow’ of deference (Podolny and Phillips, 1996).

Although in theory All-Stars are chosen based on the players’ performance on the field that year, other factors also account for who is actually chosen. Not all teams receive as much media coverage, have the same level of national visibility, or have the same amount of post-season success, leading to differential valuation of players’ abilities.⁷ But putting aside team differences, as is true with all measures of status, All-Star selection is also highly inertial and prone to self-reproduction. Once a player has established himself as an All-Star, he is more likely to be chosen in the future. “Perennial All-Stars,” which included high status pitchers like Roger Clemens and Roy Halladay, do not have to match the performance of their younger peers to make the All-Star team again. In contrast, players who have never

been All-Stars before may need consecutive years of high performance before they will be recognized with the honor. This “loose linkage” between status and quality (Podolny, 1993) further highlights the structural nature of status in the baseball context, and makes it an ideal measure of status as both an enabler and constraint on individual action.

Another way in which being an All-Star confers status is through labeling. Making an All-Star team just once forever labels that player as an All-Star, no matter how his performance varies in the future. The more All-Star appearances a player has, the more prestige associated with the player. Announcers routinely refer to elite players by the number of All-Star appearances they have made, using it as an honorific title. Being a “seven time All-Star” conveys high status in the same way that a Top Ten ranking would communicate the status of an elite university. In this sense, players differentiate themselves hierarchically by their number of All-Star appearances. We should expect status to increase incrementally with number of All-Star appearances. We obtained information about players’ All-Star appearances from the online Baseball Almanac, which contains complete histories of players’ statistics and accomplishments.

To create the interaction effects used to test Hypotheses 3 and 4 about the role of reputation and ambiguity, we included variables measuring the career performance of the pitcher and distance from the strike zone respectively. To measure the reputation of the pitcher as a control pitcher we include the pitcher’s *career average of walks per batter faced* by dividing the number of walks issued over the number of plate appearances from a pitcher’s debut to the season prior to the analysis. A high value of this variable indicates an inability of the pitcher to control the location of his pitch. Pitchers that are “wild,” or that have lots of walks, are known for lacking control or precision in their pitch location. This is not to say that these pitchers are poor pitchers. Many pitchers attain high status despite having superlative control. For example, Nolan Ryan, the former pitcher of California Angels, Houston Astros, and Texas Rangers, is considered one of the greatest pitchers to ever play and was an eight-time All-Star largely because of his ability to throw a hard and unpredictable fastball. Despite his dominant fastball,

Ryan was also notoriously prone to walking batters. Ryan ranks second all-time in the number of batters walked by a pitcher. His lack of control may have even given him an edge over batters because they were afraid of being hit by a Ryan fastball. Often pitchers who strike out many batters also walk a relatively high number of batters due to abnormal movement in their pitches. In short, a high career average of walks per game is not an indicator of poor quality, but it does indicate whether a pitcher lacks sharp control. A pitch's *total distance from strike zone* is measured as the absolute number of inches a pitch is from the boundaries of the strike zone when it crosses the plate. We also include a square term of total distance to account for the potential non-linearity in the relationship between distance and ambiguity (i.e., ambiguity being much greater the closer a pitch is to the border).

Control variables

Situational variables We account for a number of situational factors used in other statistical analyses of baseball that may influence umpires' decision-making. We included a dummy variable indicating if the pitcher's team was the *home team* to account for the ubiquitous home team advantage (Schwartz and Barsky 1977). The at-bat's *leverage* measures the importance of the at-bat to the outcome of the game. Leverage is defined as the expected change in probability of the pitcher's team winning the game relative to the average change in winning percentage for all situations taking into account the inning, score, outs, and number of runners on base. A leverage of 1 means that the at-bat has the same importance as a typical game event, while a leverage index of 2 means that the expected value of possible outcomes swing the winning percentage of the team by twice as much of a random state.⁸ Another situational measure that accounts for the urgency of the at-bat is *run expectancy* - the number of runs that are expected to score in that inning based on historical outcomes for identical runners-on-base and outs states. We included an estimate of the *fan attendance* to the game to account for the possibility that higher profile events (and the expected scrutiny associated with these events) will change umpires' accuracy. We also include dummy variables controlling for the *ball count* of the at-bat (e.g., no balls no strikes, three balls two strikes, etc.), and a dummy to indicate whether the batter bats from the left or right side of the plate (i.e., *Batter stands*),

and a control for the *inning* of pitch to account for potential umpire fatigue that influences accuracy. Finally, to control for the relative difficulty of assessing different speeds and spin rates of pitches, we include dummy variables indicating whether the pitch was classified by MLB as a breaking pitch (i.e., curve ball or slider), off-speed pitch (i.e., change-up, knuckleball), or unknown.

Player characteristics We also control for a number of player characteristics to account for differences in the quality and experience of the player. We obtained these data from the online Baseball Data Bank. *Pitcher's tenure* controls for the number of years that a pitcher has been in the major leagues. *Pitcher's hand* is a dummy that indicates if a pitcher throws with his left hand, which will affect the way the ball looks as it crosses the plate. The race/ethnicity of the players and umpires have been discussed as a factor in baseball, so we collected the profile pictures of the players and umpires in our sample from the MLB website, and classified each player into one of four categories (“White/Caucasian”, “Black/African American”, “Latino/Hispanic”, and “Asian”).

Statistical estimation

Because the dependent variable of the analysis is a binary outcome, we use logit regression to estimate the coefficients. To account for unmeasured heterogeneity across umpires, we included umpire fixed effects in all of the models. For simplicity of presentation, we do not show these coefficients in the tables below. To reduce collinearity bias, we mean centered all of the independent variables that we used to create interaction terms. After making this transformation, we found that the VIFs for all of the independent variables were within an acceptable range.

Results

Table 2 shows the descriptive statistics and correlations for our variables. The mean pitcher has 5.07 years of service, issues a base-on-balls to 8.9 percent of the batters he faces, and has been voted an All-Star 0.4

times. The correlations between situation and player variables are not very high, while player characteristics such as tenure and All-Star appearances show moderate correlation as one would expect.

[Table 2 about here]

Over-Recognized Pitches (Actual Balls that were Mistakenly called Strikes)

Table 3 shows the results from logit models estimating the likelihood that an actual ball is mistakenly called a strike by the umpire. This captures the umpire's likelihood of offering greater recognition of a pitch despite the quality not being worthy of such recognition. The baseline regression (Model 1) shows that a wide range of situational characteristics and pitcher characteristics influence over-recognition, and act as important control variables for subsequent models focused on the status variables. Umpires exhibited a higher likelihood of mistakes when the pitcher was from the home team, with the odds of over-recognition increasing by 7.8 percent.⁹ Umpire over-recognition also increased as the game progressed, with the results implying a 4.7 percent increase in odds of over-recognition for the 9th inning relative to the 1st inning of the game. Also, umpires were significantly less likely to make over-recognition mistakes when the batter was right-handed, with a 41 percent decrease in the odds of mistake. The closer the pitch was to the border of the strike zone, the more likely the umpire was to make a mistake ($p < .01$), implying that uncertainty increases the difficulty of the judgment task. The relationship between distance and over-recognition exhibits a non-linear relationship with a rapid decline in the odds of mistake as distance increases. While the potential for runs to be scored (i.e., run expectancy) did not show any statistically significant effects, doubling the importance of the situation (i.e., an increase of leverage by 1) led to a 1.5 percent increase in the odds of a mistaken strike ($p < .05$), which, along with the positive effect for attendance ($p < .01$), contradicts the notion that potential scrutiny should reduce the mistakes by umpires (e.g., Parsons et al., 2011). The results also clearly show that the baseline rate of mistake varies significantly across pitch count—for example, the odds of an umpire mistakenly calling a strike is 62 percent lower when the count is 0-2 (i.e., no balls and two strikes, and thus, a strike would end

the at-bast), whereas the odds of over-recognition are 49 percent higher when the count is 3-0 (i.e., three balls and no strikes, a count that is known for generous strike calls)—validating the inclusion of count dummies.

For pitchers, each year of additional tenure in the league increased the odds of over-recognition by 2.1 percent ($p < .01$). Also, pitcher reputation had an effect on the likelihood of a mistaken strike, as the pitcher's tendency for control (i.e., lower BBs per batters faced) was associated with a higher rate of over-recognition—for every 10 percent increase in the percentage of batters walked in the pitcher's career leading up to the season, there is a 22 percent drop in the odds of over-recognition—consistent with a mechanism of expectations driving umpire judgment errors. African-American pitchers were less likely to benefit from over-recognized strike calls relative to Caucasian pitchers ($p < .01$). Race and tenure can both be considered as salient status characteristics, so we see evidence that is consistent with status bias in the baseline results.

[Table 3 about here]

Model 2 includes the pitcher's status in a regression of the likelihood of calling of over-recognizing a pitch. As hypothesized, umpires are more likely to mistakenly call a real strike a ball when assessing an All-Star pitcher. An additional appearance in an All-Star game leads to a 4.8 percent increase in the odds that an umpire will over-recognize a strike ($p < .01$). Based on the predicted probabilities, going from no All-Star appearances to one All-Star appearance increases the predicted probability of a mistaken strike from 12.8 percent to 13.2 percent, and with no overlap in the confidence intervals, this suggests statistically different probabilities. A player with 5 All-Star appearances has a 14.9 percent chance of a mistaken strike, a 16.4 percent increase over the probability that a pitcher with no All-Star appearances would receive the same favorable call. To facilitate comparisons across different variables, Figure 1 shows the increase in odds of an over-recognized pitch when there is a one standard deviation increase in

status (i.e., All-Star counts), reputation (i.e., Base-on-balls per Batters faced), tenure in the league, and other variables.

[Figure 1 about here]

In Hypothesis 3 we suggested that pitchers who have reputations as control pitchers (i.e., low percentages of walks given to batters) will benefit more from being high status. The interaction effect of status and a pitcher's career average of walks per game is negative and significant ($p < .05$), which indicates that players who have a reputation for being "wild" or lacking in control do not get as much status-based over-recognition as pitchers who are known for their precision pitching. In other words, a pitcher's reputation for being a control pitcher amplifies the effect of status on umpires' likelihood of over-recognition mistakes. Figure 2 shows the relationship between the percent of base-on-balls issued by the pitcher per batters faced, and the probability that a pitcher will receive a strike call when in fact the pitch is an objective ball for a pitcher with no All-Star appearances, one All-Star appearance, and five, with 95 percent confidence intervals for the estimated probabilities. Status's effect on the likelihood that an umpire mistakenly calls a ball a strike does not differ statistically when the pitcher issues a high proportion of base-on-balls (i.e., 15% of batters faced), but when pitchers have a low proportion of base-on-balls in their career (i.e., a reputation for better control), status amplifies the beneficial effect.¹⁰ Thus, the results show that when reputation and status are aligned by creating similar performance expectations, the effect of status bias is amplified.

[Figure 2 about here]

Contrary to Hypothesis 4, Model 4 shows that being closer to the border of the strike zone (i.e., increasing ambiguity) did not enhance the effect of status ($p > .10$). Similarly, situational factors, like pitching for the home team or attendance did not moderate the impact of status on the tendency for umpires to over-recognize quality. In models not shown, we included a number of other interaction effects to assess whether situational variables moderated the impact of status on the likelihood of umpires

under-recognizing quality. Notably, we did not find that situational effects such as game attendance, leverage, or run expectancy make status's influence any greater. The effect of status does not seem to be augmented by situational elements of the game. This further reinforces the notion that status bias is driven by instantaneous perceptions, as opposed to deliberate and strategic choices of evaluators.

Under-Recognized Pitches (Actual Strikes that were Mistakenly called Balls)

Table 4 contains the results of the logit model estimating the likelihood that a real strike is mistakenly called a ball. This captures the likelihood that the umpire will fail to recognize quality despite the pitch deserving recognition (i.e., a strike call). Consistent with the over-recognition of quality, the baseline results shown in Model 1 indicate that situational characteristics and individual attributes shape the mistakes made by umpires. Similar to over-recognition, umpires were more prone to make mistakes when the situation became more important (as reflected in leverage, run expectancy, and inning), although attendance did not have an effect on the likelihood of missing a good strike. Other situational variables such as home team, inning, pitch count and distance had similarly strong effects on the baseline effect. Pitcher characteristics such as tenure or prior reputation had effects as expected.

Model 2 shows the results when the status of the pitcher is included. Consistent with Hypothesis 2, pitchers with high-status (i.e., more All-Star appearances) are less likely to have their strikes missed. An additional All-Star appearance lowers the odds of under-recognition by 2.7 percent. For example, the model predicts that an average pitch within the strike zone thrown by a pitcher with no All-Star appearances will be mistakenly called a ball 18.9% of the time; in contrast, the same average pitch within the strike zone thrown by a pitcher with five All-Star appearances has only a 17.2% of being mistakenly called a ball. A summary of the effects of a one standard deviation change in key variables on under-recognition is shown in Figure 3.

[Figure 3 about here]

In Model 3 we include the interaction effect of pitcher's status and the pitcher's reputation. While the coefficient of the interaction term suggests that having a reputation of wildness (i.e., high percentage of base-on-balls) weakens the effect of status, the effect is not statistically different from zero. In Model 4, we show how the pitch's distance from the strike zone interacts with status. The effect is positive, indicating that when the pitch is located closer to the edges of the strike zone (i.e., a lower total distance), higher status players get a greater advantage in being evaluated accurately than do lower status players. For example, the probability that a pitch that is right down the middle of the zone, thus being a combined two feet from the border of the strike zone will have a 0.041 percent chance of being mistakenly called a ball for a pitcher with no All-Star game appearances, and 0.045 percent with 5 All-Star appearances, showing no statistical difference between the chances of mistake. In contrast, for pitches that are a combined three inches from the border of the strike zone, the predicted probability of a mistaken ball for the pitcher with no All-Star is 52.7 percent, and for a pitcher with five All-Star appearances 50.5 percent. Figure 3 depicts the relationship between predicted probability of a mistaken ball as a pitch approaches the border, and uncertainty increases. High status players, then, tend to benefit from ambiguity, inasmuch as umpires attend more to the quality of their pitches in the more ambiguous parts of the strike zone. This finding, then, indicates that high-status pitchers get more close strike calls to go their way, in part because the umpires are more accurate when there is uncertainty. On the other hand, low-status pitchers are not given the benefit of the doubt in these situations, leading to a higher number of under-recognized strikes relative to those with high-status.

[Figure 4 about here]

Batters and Catcher Effects

While our paper focuses on the effect that pitcher status has on umpire judgment, two additional actors are also involved in the pitch: batters and catchers. Players and managers have long held the suspicion that some batters receive more favorable judgments than others. In the aforementioned quote by

Ozzie Guillen, the former player and manager points out that it is the “same way for hitting” with Hall of Fame inductee Wade Boggs, and two-time league Most Valuable Player (MVP) Frank Thomas are more likely to receive “ball” calls for the same pitch. This suggests that the status of the batter should have the opposite effect of pitcher status in that high status batters are more likely to lead to less over-recognition of quality (i.e., balls called strikes) and more under-recognition of quality (i.e., strikes called balls). Furthermore, both Wade Boggs and Frank Thomas were well-known for their exceptional “batting eye”, or ability to tell apart a ball from a strike, so Guillen also implies that their reputations play a role in setting the expectations of umpires. To examine these mechanisms, we include both batter status (as measured by All-Star appearances) and the batter’s reputation for plate discipline (as measured by his career rate of base-on-balls per plate appearance) in Model 5 in Tables 3 and 4. The results show that batter status and reputation does indeed influence the judgment of umpires, mainly in a way that is unfavorable towards pitchers. For each additional All-Star appearance of the batter, the odds of over-recognition goes down by 1.3%, while the odds of under-recognition increases by 1.3%. The results of pitcher status remain robust to including these batter effects.¹¹

The catcher plays an even more direct role in the judgment of the pitch, through an act that is commonly known as “framing the pitch”. Catchers move around the home plate area to help guide a pitcher throw to specific locations, and this in turn, can influence how the pitch appears to the umpire. For example, if the pitcher and catcher agree upon throwing a pitch to the outside part of the strike zone, the catcher can shift his position to the outside so that the center of his glove would rest at the agreed upon location. Not only does this provide a visual cue for the pitcher (i.e., throw to the catcher’s glove), but it also helps “sell” the pitch to the umpire because the catcher does not have to reach to catch the ball, leading to an impression that the pitch was straight down the middle. This ability to “frame” the pitch has always been thought of as a valuable skillset in baseball, and if this is indeed a repeatable skill, could potentially explain the likelihood of mistaken judgment from the umpires.

We attempt to control for this heterogeneity in catching skill through two means: first, we include catcher fixed effects in our models, which does not alter the pattern of our results. Second, we use our data on mistakes to isolate the framing ability of catchers and include this as a control variable. In short, we looked at the rate at which catchers gained “extra strikes” for their pitchers relative to the pitchers’ own inherent ability to get these “extra strikes” and included the normalized variable in our regression models.¹² The logit estimates including this variable and a count of All-Star game appearances for catchers are shown in Model 6 in Tables 3 and 4. While catcher status (i.e., All-Star appearances) does not seem to influence the likelihood of over-recognition or under-recognition by umpires, the framing ability (or reputation of framing) has a strong and significant effect on calls ($p < .01$). Being a catcher that is one standard deviation higher in his ability to get extra calls leads to an increase in the odds of a over-recognized pitch by 7%, and a decrease of the odds of under-recognition by 10%. The effect of pitcher status remains robust, implying that the observed status advantages for pitchers were not driven solely by high-status pitches pairing with catchers who are more adept at framing.

In sum, we find that the identities of all actors involved (pitchers, batters, and catchers) influence the likelihood of a mistaken call (see Figures 1 and 3). Given the visual orientation of the umpire, pitcher status has a larger influence, with a one standard deviation increase in status leading to a 7 percent increase in odds of over-recognition, and a 4 percent decrease in odds of under-recognition, whereas batter status effects are roughly half the size at 3 percent decrease, and 3 percent increase respectively. Reputation (or past demonstrations of skill) has similar effects as status for both pitcher and batter, while catcher reputation/skill for framing had a considerable impact on umpire judgments, especially for avoiding missed strikes.

Disentangling Quality and Perception: Endogenous Locations

Our analyses above strongly support the hypothesis that pitchers with All-Star appearances benefit from greater over-recognition of their low-quality pitches and less under-recognition of their high-

quality pitches. We argue that this is the result of umpires being biased towards pitchers with high status. In other words, the status advantages derive from the perceptions of the evaluators, and not the quality differences of the actor. While pitch-level controls such as distance from the border or pitch type help us control for quality differences, it does not directly address the possibility that All-Star pitchers are different in their quality choices and execution from non-All-Star pitchers. If pitchers with more All-star appearances choose to locate their pitches in different areas than those with few All-star appearances, then the observed All-star effect may in fact be due to a difference in pitch quality as opposed to umpire perception. Furthermore, given the umpire-specific nature of the strike zone (i.e., “Everybody has their own strike zone”) it is also possible that All-star pitchers are more keenly aware of umpire-specific blind spots, and are better able to exploit these tendencies by locating pitches in those locations more frequently than low status pitchers.

To fully rule out the null hypothesis that status advantages are due to quality differences, we create a “matched” sample based on pitch location, umpire, and pitch characteristics, and examine the effect of being “treated” with status (i.e., All-star appearances).¹³ Due to the richness of the covariates upon which we intend to match, obtaining precise matches for pitches is impractical. Instead, we use the coarsened exact matching (CEM) approach (Iacus et al., 2011; 2012; Blackwell et al., 2009), a monotonic imbalance bounding method which has many desirable statistical properties for improving the estimation of causal effects in observational studies. Appendix 1 has details of the matching method, but the main intuition is 1) to create zones within and outside the strike zone (i.e., the coarsening of the location variable), 2) assign each pitch that was “treated” (i.e., All-stars) to a stratum based on the zone, umpire, and pitch characteristics, and 3) assign untreated pitches to these strata based on the covariates of the pitch. Strata that have no matching untreated pitches are pruned from the data, as well as untreated pitches that have no matching treated pitch. This pre-processing of the data also generates weights for the untreated pitches that allow us to obtain the average treatment effect on the treated.

The results from the logit analyses estimating the odds of over and under recognition with this matched sample show identical patterns to the results in Tables 3 and 4, confirming both the main effect of status, and the moderating effects of reputation and distance (i.e., uncertainty). (See Appendix Table 1 and 2 for the full results.) For pitches in a given location, umpire, pitch type and ball count, the effect of being in the “treatment condition” results in a 6.7 % increase in the odds of over-recognition, and a 5.7% decrease in the odds of under-recognition. We also find strong evidence that the effect of status is cumulative, with established All-Stars (i.e., those with five or more appearances) having a significantly larger effect on both biases (24% increase in odds of mistake) relative to those with just one All-Star appearance, which had effects that were in the expected direction, but not statistically different from zero. These results further solidify our conclusion that the higher rate of over-recognition and lower rate of under-recognition is attributable to biases in perception, and not a difference in skill.

Does Status Bias Influence Performance?

Our results suggest that the tendency to over-recognize quality in high status pitchers and under-recognize quality in low status pitchers causes umpires’ accuracy to fluctuate in such a way that high status pitchers receive performance advantages over their lower status peers. The strike zone becomes slightly larger for high status pitchers, and smaller for those with no status. But does this seemingly small advantage translate into more significant and enduring performance benefits for high status pitchers? To link status bias to systemic inequality, umpire inaccuracies must have a lasting impact on the trajectory of the performance of the pitchers affected. If status bias causes high status pitchers to get strike calls they do not deserve, then we should also expect these pitchers to have a lower likelihood of putting men on base and a lower threat of having runs scored against them. Using the same MLB data, we are able to empirically test this possibility.

[Table 5 about here]

Table 5 shows the regression results for two outcomes of an at-bat: total bases and Winning-Percentage-Added (WPA). The unit of analysis is the at-bat, and total bases refers to the number of bases the batter gains in that at-bat through a hit (i.e., 1 for a single, 2 for a double, 3 for a triple, and 4 for a home run) or a base-on-balls (i.e., 1 total base). Pitchers attempt to minimize the number of bases because the more bases batters get, the more likely runs will result in the current or future at-bats. WPA refers to the increase or decrease in likelihood of winning the game based on the outcome of the given at-bat. Since the impact of a hit or base-on-balls will differ depending on the context (i.e., a hit in the bottom of the ninth when the game is tied is a more damaging outcome to the pitcher than a hit in the bottom of the ninth when the pitching team is leading by ten runs), WPA provides a more accurate indicator of the value of a pitch in a given at-bat.

Model 1 shows the baseline model, which includes the pitcher's career Earned Run Average (ERA), and batter's career On-Base-Percentage plus Slugging Percentage (commonly referred to as OPS) as controls for the players' abilities, the leverage of the situation, whether the pitcher is pitching for the home team, and dummies for the number of outs. As expected, the higher the career ERA (i.e., the worse the pitcher's abilities), and the higher the batter's OPS (i.e., the better the batter's abilities), the lower the Winning-Percentage-Added by the pitcher, and the higher the average number of total bases that result from that at-bat. The leverage variable indicates that pitchers perform better when the stakes are high, and also when they are pitching in front of a home crowd. Also, as the number of outs increase, the more likely the at-bat will end in the pitcher's favor.

Including the umpire bias variables (Model 2) shows that each incremental increase in biased calls during an at-bat has a significant impact on the outcomes to the pitcher. An additional mistaken ball (i.e., under-recognition) decreases the probability that the pitcher's team will win by 0.3% ($p < .01$), while an additional mistaken strike (i.e., over-recognition) increases the win probability of the pitcher's team by 0.3% ($p < .01$). For total bases, an additional mistaken ball increases the resulting number of bases by 0.074 on average ($p < .01$), while a mistaken strike lowers the total bases by 0.067 ($p < .01$). In sum, umpire

mistakes are associated with worse outcomes for pitchers, suggesting that status bias can play a role in determining pitcher performance.

The impact that umpire errors have on pitcher performance further implies that the choice of a particular pitching strategy will be shaped by the status one enjoys. For a high status pitcher, pitching outside the strike zone, or if within the zone, as close to the border as possible is the optimal strategy, as umpires are more likely to call a strike under these circumstances. On the other hand, low status pitchers not only fail to get the outside pitch called a strike, but more problematic is the fact that their high-quality offerings (i.e., strikes) are also more likely to be missed, forcing the pitcher to throw pitches in less ambiguous locations (i.e., closer to the middle) to avoid an error of omission on the part of the umpire. As the likelihood of adverse events (e.g., extra base hits) increases as the pitch is thrown in the middle of the zone, in addition to having weaker skills, low status actors are further constrained by the fact that observers are biased and fail to recognize quality, leading them to engage in more risky actions to compensate for such bias. Thus, low status pitchers find it more difficult to obtain the same results for an at-bat as high status pitchers. The pitchers with the greatest advantage during any given at-bat are high status control pitchers, who receive more leniency from the umpires when throwing the ball on the edges of the strike zone.

Conclusion

The analyses in this paper point to status bias as a source of status-based advantages in evaluative decisions. The results suggest that paying attention to the biases of evaluators provides a more detailed picture of the accumulated advantages enjoyed by high status actors. Our analysis contributes to the literature on status in several different ways. First, our study provides a conclusive and systematic test of status bias in a real world situation by utilizing performance data that are extremely detailed and comprehensive. Second, by drawing on status research in social psychology, we integrate the economic

sociology literature on Matthew Effects with the status characteristics theory in social psychology, two literatures on status that have developed independently. Third, while past research has tended to show differences in outcomes for high status actors, our study is among the first to empirically examine differential evaluations among undeserving versus deserving performances by high status actors. Our study is able to show how evaluation bias leads high status players to be rewarded even when their performances are undeserving. And fourth, we show how reputation conditions status effects. The distinction between these two important concepts has been the subject of recent debates in the literature. Our analysis points to a mechanism – performance expectations – that links the two, and demonstrates how they combine to produce evaluation differences.

The result of these dynamics is that an actor's status leads to two distinct pathways to bias. First, under-recognition implies that high status actors are more likely to get noticed for what they already do well. The burden of proof is lesser for high status actors, inasmuch as evaluators already expect that they will do well. Second, over-recognition implies that high status actors are allowed more flexibility in their performance. They need not perform well every time in order to be rewarded. The high expectations set by their greater status allows them room for error that lower status actors do not have.

Of course, the value of understanding these socio-cognitive dynamics is that they are manifest in contexts other than baseball and by various types of status characteristics. In any context where status characteristics drive future interactions and performance evaluation, status bias contributes to persistence in inequality and the reproduction of status differences (see, for example, Ridgeway 2011 on gender as a status characteristic). Over-recognition of quality by evaluators may underlie the ability of high status actors to stand out among numerous candidates during job recruitment (Rivera 2012), the tendency for high status organizations to receive higher status affiliations (Benjamin and Podolny 1999), and the attractiveness of higher ranked law schools to law school applicants (Sauder and Lancaster 2006). Of course, status bias likely explains the differential rewards given to the Nobel Prize winners who Merton (1968) studied. Merton believed the Matthew Effect resulted from increased credit given to higher status

actors, but he did not delve into the psychological mechanisms that accounted for it. We believe that it is entirely consistent with Merton's findings to argue that Nobel Prize winners get more credit because their peers, grant providers, and university officials expect them to perform at higher levels than their colleagues, which makes their new research seem more interesting, more citable, and worthy of funding. Over-recognition means that even their low quality offerings will receive greater academic respect. Azoulay et al. (2013) find that once a scientist wins a prestigious grant, past papers subsequently generate new appreciation and suddenly spike in citations. At the same time, our results suggest that low status academics will be constrained from moving up the academic hierarchy because their high quality offerings will be consistently discounted in value, making it more difficult for them to get equal recognition for research findings of similar quality as their high status colleagues. Under-recognition, in this sense, impedes upward mobility in academic status hierarchies.

Over-recognition of quality, then, initiates the Matthew Effect (Merton 1968; Bothner et al. 2010; Correll et al. 2011). Once an individual is marked as high status, other valued qualities become more salient, and their attractiveness as potential partners increases. By shaping the recognition of quality when it is present, status helps differentiate otherwise evenly-distributed differences in quality. It also allows high status actors to be somewhat uneven in their performance without losing credibility or esteem in the eyes of evaluators. By expanding the zone of appreciation for high status actors, evaluators make it possible for them to receive more favors or to be given the benefit of the doubt when performance slacks. This cognitive bias may underlie the "halo effect" often accorded to high status actors (Sine, Shane and Di Gregorio 2003), the tendency of high status individuals to be seen as more competent and powerful (Ridgeway and Berger 1986), or demonstrations of interpersonal deference to high status actors (Sykes and Clark 1975). In contrast, our results also show that under-recognition of quality serves as a barrier to those who lack status. If an actor has not already established his or her position in the status hierarchy, they more likely to be discounted in the future, even when their performance equals that of their high

status counterparts. Thus, status bias “unjustifiably victimize[s]” low status actors (Merton 1968: 59), preventing them from being treated as if they were operating in an even playing field.

Our findings extend our understanding of the halo effect by showing that status is especially likely to lead to over-recognition of quality when status and reputation are aligned (Stern et al. 2013). The importance of aligning expectations in the evaluator’s mind may be one reason that other scholars have found that status has its greatest effect on economic performance advantages when there is consistency among multiple status signals (Zhao and Zhou 2011). Both status and reputation influence perceptions because they shape performance expectations. Thus, when they are aligned reputation enhances status bias but when they are not aligned the performance advantages of status bias disappear.

One implication of our analysis and status characteristics theory, more generally, is that evaluators need not be consciously aware of status differences for these two aspects of status bias to influence evaluations and performance outcomes. Umpires do not approach each at-bat thinking that they are going to evaluate an All Star pitcher differently than they would a rookie pitcher with no All Star appearances. If status bias was deliberate and calculated, one should expect situational factors such as the importance of the situation (i.e., leverage) or scrutiny (i.e., attendance) to moderate the effect of status, yet, our results suggest this is not the case. Umpires believe that they are being completely objective in their assessments. But umpires also evaluate pitches instantaneously and instinctively, which creates opportunities for status bias to creep in and influence their evaluations. Furthermore, umpires also know that the league, media, and fans have access to information regarding the accuracy of their judgments, so it is highly unlikely that umpires knowingly and deliberately make incorrect calls. Status bias seems to enter the equation, then, in a fairly automatic and unconscious manner. If status bias influences automatic responses even among highly trained and professionalized evaluators, then it is reasonable to expect that other types of evaluators in less high pressure contexts will be equally influenced by status bias. Performance expectations based on status differences are a pervasive feature of social interaction and quality evaluation that lead to the persistence of socio-economic differences associated with gender, race,

and other highly salient status differences. Despite changes in laws and organizational rules that are meant to enhance workplace justice and despite individuals' attempts to be "fair," status bias continues to shape evaluations in favor of higher status actors, thereby reproducing status hierarchies. As Ridgeway (2011: 185) argues, differences in beliefs and expectations formed by status differences may be at the very heart of inequality persistence.

In this paper we have focused mainly on the ways in which these differences in evaluator bias lead to performance advantages for high status actors, but it is reasonable to expect that in other contexts the same cognitive tendency could have negative consequences. For example, over-recognition may cause high status actors to become the targets of unwanted attention, as when social movement activists differentially target prominent corporations (King 2011; King and McDonnell forthcoming) or when the high status of an actor turns his or her transgressions into a public scandal (Adut 2005) or the tendency for audiences to give high status actors less leeway when they violate local norms of appropriateness (Phillips, Turco, and Zuckerman 2013). Our results suggest that audiences are more aware of the transgressions and norm violations of high status actors, in part, because they have strong priors about the kind of behaviors they expect from them. Although we have shown that evaluators tend to under-recognize the poor quality performances of high status actors, high expectations about performances may also lead evaluators to react more negatively when they are perceived to purposefully transgress local norms (e.g., pitchers throwing bean balls at hitters in retaliation). We might also expect that inasmuch as status increases performance expectations, actors may sometimes be put in situations where they cannot meet those expectations (Fine 2001). This is, to some degree, what happens to pitchers who are not known for their control. These pitchers face the pressure of not only being more intensely scrutinized but also of not being expected to precisely pitch around the edges of the strike zone. A bad day for a "wild" pitcher could lead to unfair assessments by umpires who are unwilling to give them close calls.

Despite this potential for backfire, having status leads to more positive outcomes for pitchers because it tends to influence evaluators' assessment of quality in a way that benefits those players.

Although we have known by past research that high status actors accumulate advantages of this type, a main contribution of this study is to explain why this is the case. By shifting the focus to the evaluator of status, and the various conditions under which status biases the evaluations of purportedly objective actors, we open a window into the micro-level processes that underpin the advantages that accrue to high status actors.

Acknowledgements

Authors in alphabetical order. We benefitted from helpful comments by Jeremy Freese, Bruce Kogut, Ko Kuwabara, Damon Phillips, Lauren Rivera, Michael Sauder, Olav Sorenson and Toby Stuart, as well as seminar participants at the University of California, Berkeley, Columbia Business School, and the Harvard Business School Strategy Conference.

Endnotes

¹ The official Major League Baseball rulebook describes the strike zone in this way: “The STRIKE ZONE is that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the hollow beneath the knee cap. The Strike Zone shall be determined from the batter’s stance as the batter is prepared to swing at a pitched ball” (Major League Baseball, 2011: 22-23).

² In fact, the umpire behind the plate does not make a decision about whether a player swung the bat or not so as not to distract him from assessing the location of the pitch. If there is any question of whether a batter swung the bat at a pitch, the first base umpire makes the call.

³ Mike Fast, in particular, has analyzed variation in umpires’ subjective assessment of the strike zone. See, for example, the following analyses at Baseball Prospectus:

<http://www.baseballprospectus.com/article.php?articleid=12965> and

<http://www.baseballprospectus.com/article.php?articleid=14098>.

⁴ Other factors such as pitch speed or break (i.e., spin of the ball) can also create uncertainty for umpires. However, the specific relationship can be ambiguous—for example, faster pitches may be harder to judge, but these pitches also tend to be the straightest, which reduces uncertainty.

⁵ We thank an anonymous reviewer for suggesting this mechanism.

⁶ We recognize that pitchers at times may intend to throw balls and, in those instances a “quality” pitch may indeed be a ball, but an umpire’s role is not to judge intent but to simply evaluate whether a pitch qualifies as a strike. Thus, the umpire is perceptually focused on the strike zone as the range of quality in which a pitch counts as a strike. Anything outside that zone, then, would be considered a ball – i.e., a low quality pitch.

⁷ See, for example, this analysis of teams that are under- or overvalued in All-Star voting:

<http://bleacherreport.com/articles/422086-subjectivity-objectified-measuring-mlb-fans-biases-with-all-star-votes>.

⁸ For a detailed explanation of the leverage index, refer to <http://www.hardballtimes.com/main/printarticle/crucial-situations/>

⁹ The odds of mistake are 1.0778 ($=\exp(0.075)$) times the baseline rate of mistakes

¹⁰ The increased effect for status for high-performing pitchers may be due to the non-linearity of status, or the “Winner-Take-All” effect where the highest status actors enjoy disproportionate returns (Frank and Cook, 1995). We include status as a quadratic term, and find that contrary to the idea of increasing returns to status, the effect of status exhibits a slight *decrease* in marginal returns to status. The peak of the effect occurs at 10 All-Star appearances according to this model.

¹¹ We also included interaction terms between pitcher status and batter status in models not shown here to see if pitcher status becomes more or less valuable when the batter is high status as well. The results were inconclusive, suggesting that the pitcher and batter effects operate independent of each other.

¹² Following work done by *Baseball Prospectus* writer Mike Fast, we first calculated a baseline rate of mistake for each pitcher for each pitcher during the 2007 season by counting the number of extra strikes (i.e., pitches called a strike despite being a ball), subtracting the number of extra balls (i.e., pitches called a ball despite being a strike), and dividing this net mistake figure by the total number of pitches called for that pitcher during the season. We then counted pitcher-catcher dyad specific net mistakes and subtracted the expected number of net mistakes (i.e., total called pitches for the pitcher-catcher dyad times the baseline rate of mistake for the pitcher), which gave us the total number of extra mistakes the catcher gained for the pitcher. We aggregated this figure for each catcher, and divided

it by the total number of pitches called, resulting in a rate statistic that captures the catcher's ability to get favorable mistakes, independent of the pitcher's own skills or status.

¹³ We are not aware of any methods to account for continuous treatment variables in matching methods, so we dichotomized our All-star variable into "pitcher appeared in at least one All-star game" (treated=1), and "pitcher did not appear in an All-star game" (treated=0). We believe that this is a conservative test of our arguments given that we believe status is accumulative. See Appendix for discussion of alternative methods of identifying the treatment effect.

References

- Adut, A. 2005. "A Theory of Scandal: Victorians, Homosexuality, and the Fall of Oscar Wilde." *American Journal of Sociology* 111:213-248.
- Allen, M. P., Parsons, N.L. 2006. "The institutionalization of fame: achievement, recognition, and cultural consecration in baseball." *American Sociological Review* 71(5):808-825.
- Anderson, C., O. P. John, D. Keltner, A. M. Kring. 2001. "Who attains social status? Effects of personality and physical attractiveness in social groups." *Journal of Personality and Social Psychology* 81:116-132.
- Anderson, C., R. Willer, G. Kilduff, C. Brown. 2012. "The Origins of Deference: When do People Prefer Lower Status?" *Journal of Personality and Social Psychology* 102 (5): 1077-1088.
- Azoulay, P., T. Stuart, Y. Wang. 2014. "Matthew: Effect or Fable?" *Management Science*. 60 (1): 92-109.
- Benjamin, B. A., J. M. Podolny. 1999. "Status, quality, and social order in the California wine industry." *Administrative Science Quarterly* 44:563-589.
- Berger, J., B. P. Cohen, M. Zelditch Jr. 1972. "Status characteristics and social interaction." *American Sociological Review* 37(3): 241-255.
- Berger, J., M. H. Fisek, R. Norman, D. Wagner. 1985. "The Formation of Reward Expectations in Status Situations." Pp. 215-261 in *Status, Rewards and Influence* Edited by J. Berger and M. Zelditch. San Francisco: Jossey-Bass.
- Berger, J., M. H. Fisek, R. Z. Norman, M. Zelditch, Jr. 1977. *Status Characteristics and Social Interaction: An Expectation States Approach*. New York: Elsevier.
- Berger, J., S. J. Rosenholtz, M. Zelditch. 1980. "Status organizing processes." *Annual review of sociology* 6:479-508.
- Blackwell, M., S. M. Iacus, G. King, G. Porro. 2009. cem: Coarsened exact matching in Stata. *Stata Journal*, 9(4), 524-546.
- Bothner, M. S., R. Haynes, W. Lee, E. B. Smith. 2010. "When Do Matthew Effects Occur?" *Journal of Mathematical Sociology* 34:80-114.
- Bothner, M. S., Y. Kim, and E. B. Smith. 2011. "How Does Status Affect Performance? Status as an Asset vs. Status as a Liability in the PGA and NASCAR." *Organization Science* 57(3):439-457.
- Correll, S. J., C. L. Ridgeway, E. W. Zuckerman, S. Bloch, S. Jank. 2011. "It's the conventional thought that counts: the origins of status advantage in third order inference." Working paper, Stanford University
- Correll, S. J., S. Benard, I. Paik. 2007. "Getting a job: Is there a motherhood penalty?" *American Journal of Sociology* 112:1297-1338.
- Deaux, K., B. Major. 1987. "Putting gender into context: An interactive model of gender-related behavior." *Psychological review* 94(3):369.

- Espeland, W. N., M. Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds1." *American Journal of Sociology* 113(1):1-40.
- Fine, G. A. 2001. *Difficult reputations : collective memories of the evil, inept, and controversial*. Chicago: University of Chicago Press.
- Ford, G. G., F. Goodwin, J. W. Richardson. 1996. "Perceptual factors affecting the accuracy of ball and strike judgments from the traditional American League and National League umpiring perspectives." *International Journal of Sport Psychology* 27:50-58.
- Gibson, B., R. Jackson, L. Wheeler. 2009. *Sixty feet, six inches : a Hall of Fame pitcher & a Hall of Fame hitter talk about how the game is played*. New York: Doubleday.
- Gilovich, T., D. Griffin. 2002. "Introduction-heuristics and biases: Then and now." In T. Gilovich, D. Griffin, and D. Kahneman (eds). *Heuristics and biases: The psychology of intuitive judgment* pp. 1-18. New York: Cambridge University Press.
- Iacus, S. M., G. King, G. Porro. 2011. Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106(493), 345-361.
- Iacus, S. M., G. King, G. Porro. 2012. Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1), 1-24.
- King, B. G. 2011. "The tactical disruptiveness of social movements: Sources of market and mediated disruption in corporate boycotts." *Social Problems* 58(4): 491-517.
- King, B., M. McDonnell. Forthcoming. "Good Firms, Good Targets: The Relationship between Corporate Social Responsibility, Reputation, and Activist Targeting." In Kiyoteru Tsutsui and Alwyn Lim (eds) *Corporate Social Responsibility in a Globalizing World: Toward Effective Global CSR Frameworks*. (Forthcoming).
- Merton, R. K. 1968. "Matthew Effect in Science." *Science* 159:56-63.
- Parsons, C. A, J. Sulaeman, M. C. Yates, D. S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." *American Economic Review* 101(4):1410-1435.
- Pearce, J. L. 2011. *Status in management and organizations*. Cambridge ; New York: Cambridge University Press.
- Phillips, D. J., C. J. Turco, E. W. Zuckerman. 2013. "Betrayal as market barrier: identity-based limits to diversification among high-status corporate law firms." *American Journal of Sociology* 118:1-32.
- Plessner, H. 2005. "Positive and negative effects of prior knowledge on referee decisions in sports." Pp. 327-342 in T. Betsch, S. Haberstroh (eds) *The routines of decision making*. Mahwah, N.J.: Erlbaum.
- Podolny, J. M. 1993. "A Status-Based Model of Market Competition." *The American Journal of Sociology* 98:829-872.
- Podolny, J. M. 2001. "Networks as the Pipes and Prisms of the Market." *American Journal of Sociology* 107(1):1-60.

- Podolny, J. M. D. J. Phillips. 1996. "The Dynamics of Organizational Status." *Industrial and Corporate Change* 5:453-471.
- Price, J., J. Wolfers. 2010. "Racial Discrimination among Nba Referees." *Quarterly Journal of Economics* 125:1859-1887.
- Rainey, D. W., J. D. Larsen, A. Stephenson, S. Coursey. 1989. "Accuracy and certainty judgments of umpires and nonumpires." *Journal of Sport Behavior* 12(1):12-22.
- Raub, W., J. Weesie. 1990. "Reputation and efficiency in social interactions: An example of network effects." *American Journal of Sociology* 96(3):626-654.
- Ridgeway, C. L. 1991. "The Social Construction of Status Value - Gender and Other Nominal Characteristics." *Social Forces* 70:367-386.
- Ridgeway, C. L. 2001. "Gender, status, and leadership." *Journal of Social Issues* 57(4):637-655.
- Ridgeway, C. L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World: How Gender Inequality Persists in the Modern World*. Oxford University Press, USA.
- Ridgeway, C. L., J. Berger. 1986. "Expectations, legitimation, and dominance behavior in task groups." *American Sociological Review* 51(5):603-617.
- Ridgeway, C. L., S. J. Correll. 2006. "Consensus and the creation of status beliefs." *Social Forces* 85:431-453.
- Rivera, L. 2012. "Hiring as cultural matching: The case of elite professional service firms." *American Sociological Review* 77(6): 999-1022.
- Rossmann, G., N. Esparza, P. Bonacich. 2010. "I'd Like to Thank the Academy, Team Spillovers, and Network Centrality." *American Sociological Review* 75(1):31-51.
- Sauder, M., W. N. Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74(1):63 -82.
- Sauder, M., R. Lancaster. 2006. "Do rankings matter? The effects of US News & World Report rankings on the admissions process of law schools." *Law & Society Review* 40:105-134.
- Sauder, M., F. Lynn, J. M. Podolny. 2012. "Status: insights from organizational sociology." *Annual Review of Sociology*, 38.
- Schwartz, B., S. F. Barsky. 1977. "The Home Advantage." *Social Forces* 55(3):641-661.
- Shah, A. K., D. M Oppenheimer. 2008. "Heuristics made easy: an effort-reduction framework." *Psychological bulletin* 134(2):207.
- Simcoe, T. S. D. M. Waguespack. 2011. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* 57:274-290.
- Simmel, G., K. H. Wolff. 1950. *The sociology of Georg Simmel*; Glencoe, Ill.: Free Press.

- Sine, W. D., S. Shane, D. Di Gregorio. 2003. "The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions." *Management Science* 49:478-496.
- Stern, I., J. M. Dukerich, E. Zajac. 2013. "Unmixed signals: How reputation and status affect alliance formation." *Strategic Management Journal*.
- Stuart, T. E., H. Hoang, R. C. Hybels. 1999. "Interorganizational endorsements and the performance of entrepreneurial ventures." *Administrative Science Quarterly* 44:315-349.
- Sykes, R. E., J. P. Clark. 1975. "A Theory of Deference Exchange in Police-Civilian Encounters." *American Journal of Sociology* 81:584-600.
- Thye, S. R. 2000. "A status value theory of power in exchange relations." *American Sociological Review* 65:407-432.
- Wagner, D. G., J. Berger. 2002. "Expectation states theory: An evolving research program." *New directions in contemporary sociological theory*:41-76.
- Waguespack, D. M., O. Sorenson. 2011. "The Ratings Game: Asymmetry in Classification." *Organization Science* 22(3):541-553.
- Weber, B. 2009. *As they see 'em: A fan's travels in the land of umpires*. New York: Scribner.
- Weber, M. 1978. *Economy and Society*. University of California Press.
- Webster, M., S. J. Hysom. 1998. "Creating status characteristics." *American Sociological Review* 63:351-378.
- Whyte, W. F. 1943. "Street corner society." Chicago: Chicago University Press.
- Willer, R. 2009. "Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem." *American Sociological Review* 74:23-43.
- Zhao, W. and X. G. Zhou. 2011. "Status Inconsistency and Product Valuation in the California Wine Market." *Organization Science* 22:1435-1448.

Table 1 Distribution of Biased Call in Called Pitches During MLB Games, 2008-09

	Called Ball	Called Strike	Total
Actual Strike	Under-Recognition (18.8%)	No error (81.2%)	250,156
Actual Ball	No error (87.1%)	Over-Recognition (12.9%)	533,823
Total	506,692	250,156	

Table 2 Descriptive Statistics and Correlations of Variables

	Mean	Std Dev	Obs	1	2	3	4	5	6	7	8				
1 Over-Recognized Pitch (Call: Strike; Actual: Ball)	0.091	0.28	756848	1.00											
2 Under-Recognized Pitch (Call: Ball; Actual: Strike)	0.055	0.22	756848	-0.08**	1.00										
3 Home Team	0.508	0.50	756848	0.01**	-0.00*	1.00									
4 Attendance (10K)	3.148	1.16	752210	0.01**	-0.00*	0.02**	1.00								
5 Total Distance from Border (ft)	0.730	0.51	741681	-0.33**	-0.06**	-0.00	0.00*	1.00							
6 Leverage	0.964	0.84	752379	0.00**	0.00*	-0.02**	0.00**	-0.01**	1.00						
7 Run Expectancy	0.490	0.35	752379	0.00	0.01**	-0.01**	-0.00	-0.01**	0.39**	1.00					
8 Inning	4.959	2.68	756848	-0.00*	-0.00	0.04**	0.00	0.02**	0.12**	0.00**	1.00				
9 Pitcher Tenure	5.065	4.40	756848	0.02**	-0.01**	-0.01**	0.08**	-0.01**	0.04**	-0.01**	0.02**	1.00			
10 Pitcher Career BBs per Batter Faced	0.089	0.02	756724	-0.02**	0.00**	0.01**	-0.00	0.02**	0.01**	0.02**	0.13**	0.08**	1.00		
11 Pitcher is Non-Caucasian	0.266	0.44	756848	-0.00	-0.00**	-0.00	0.02**	0.01**	0.02**	0.01**	0.08**	0.08**	0.08**	1.00	
12 Pitcher All-Star	0.446	1.20	756848	0.02**	-0.01**	-0.01**	0.08**	-0.01**	0.04**	-0.02**	-0.06**	-0.06**	-0.06**	-0.06**	1.00
13 Batter BBs per PA	0.084	0.03	755995	0.01**	0.00**	-0.00**	0.03**	-0.02**	0.02**	0.02**	0.02**	-0.00	-0.00	-0.00	1.00
14 Batter is Non-Caucasian	0.457	0.49	756848	-0.00**	-0.00	-0.00	0.05**	0.00**	0.00	0.00	-0.01**	-0.01**	-0.01**	-0.01**	1.00
15 Batter All-Star	0.924	1.98	756848	-0.01**	0.00	0.00	0.11**	0.00**	0.01**	0.00	0.00	-0.03**	-0.03**	-0.03**	1.00
16 Catcher All-Star	0.792	2.22	743825	0.00	-0.00	-0.01**	0.07**	0.00	0.01**	0.01**	0.01**	0.00	0.00	0.00	1.00
17 Catcher Framing Ability	0.000	1.00	648380	0.01**	-0.02**	-0.01**	0.07**	0.00	0.01**	-0.01**	-0.01**	0.01**	0.01**	0.01**	1.00

Table 2 Descriptive Statistics and Correlations of Variables (Continued)

	9	10	11	12	13	14	15	16
9 Pitcher Tenure	1.00							
10 Pitcher Career BBs per Batter Faced	-0.13**	1.00						
11 Pitcher is Non-Caucasian	-0.05**	0.09**	1.00					
12 Pitcher All-Star	0.52**	-0.19**	-0.00	1.00				
13 Batter BBs per PA	-0.00	-0.00**	0.00**	0.00	1.00			
14 Batter is Non-Caucasian	-0.01**	-0.00**	-0.01**	-0.01**	-0.14**	1.00		
15 Batter All-Star	-0.01**	-0.01**	-0.01**	-0.00*	0.22**	0.16**	1.00	
16 Catcher All-Star	0.07**	-0.02**	0.03**	0.06**	0.00	0.00	-0.01**	1.00
17 Catcher Framing Ability	0.02**	-0.01**	-0.02**	0.05**	-0.01**	0.00	0.00	0.05**

* $p < .05$ ** $p < .01$

Table 3 Determinants of Over-Recognized Pitch (*Actual Ball that was mistakenly called a Strike*)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Situational</i>						
Home Team	0.075** (0.010)	0.074** (0.010)	0.074** (0.010)	0.074** (0.010)	0.074** (0.010)	0.068** (0.011)
Attendance (10K)	0.026** (0.004)	0.024** (0.004)	0.023** (0.004)	0.024** (0.004)	0.026** (0.004)	0.022** (0.005)
Total Distance from Border (ft)	-8.291** (0.120)	-8.292** (0.120)	-8.291** (0.120)	-8.290** (0.120)	-8.298** (0.120)	-8.325** (0.130)
Total DistanceXTotal Distance	-2.226** (0.134)	-2.224** (0.134)	-2.224** (0.134)	-2.220** (0.134)	-2.233** (0.134)	-2.222** (0.145)
Leverage	0.015* (0.006)	0.012 (0.006)	0.013* (0.006)	0.012 (0.006)	0.013* (0.006)	0.009 (0.007)
Run Expectancy	-0.025 (0.015)	-0.020 (0.015)	-0.020 (0.015)	-0.020 (0.015)	-0.018 (0.015)	-0.013 (0.017)
Inning	0.006** (0.002)	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)	0.007** (0.002)
Batter stands Right-handed	-0.532** (0.010)	-0.532** (0.010)	-0.532** (0.010)	-0.532** (0.010)	-0.549** (0.011)	-0.533** (0.011)
Year is 2009	-0.033** (0.010)	-0.030** (0.010)	-0.030** (0.010)	-0.030** (0.010)	-0.032** (0.010)	-0.039** (0.011)
<i>Ballcount (Balls-Strikes)</i>						
0-1	-0.593** (0.017)	-0.594** (0.017)	-0.594** (0.017)	-0.594** (0.017)	-0.594** (0.017)	-0.595** (0.018)
0-2	-0.966** (0.034)	-0.969** (0.034)	-0.969** (0.034)	-0.969** (0.034)	-0.968** (0.034)	-0.975** (0.037)
1-0	0.110** (0.015)	0.110** (0.015)	0.110** (0.015)	0.110** (0.015)	0.113** (0.015)	0.117** (0.016)
1-1	-0.400** (0.018)	-0.400** (0.018)	-0.400** (0.018)	-0.400** (0.018)	-0.397** (0.018)	-0.396** (0.020)
1-2	-0.726** (0.026)	-0.727** (0.026)	-0.728** (0.026)	-0.727** (0.026)	-0.727** (0.026)	-0.725** (0.028)
2-0	0.260** (0.023)	0.261** (0.023)	0.261** (0.023)	0.261** (0.023)	0.265** (0.023)	0.247** (0.025)
2-1	-0.242** (0.024)	-0.242** (0.024)	-0.242** (0.024)	-0.242** (0.024)	-0.235** (0.024)	-0.247** (0.026)
2-2	-0.566** (0.027)	-0.567** (0.027)	-0.567** (0.027)	-0.567** (0.027)	-0.561** (0.027)	-0.562** (0.029)
3-0	0.401** (0.033)	0.402** (0.033)	0.402** (0.033)	0.402** (0.033)	0.408** (0.033)	0.398** (0.036)
3-1	-0.171** (0.033)	-0.170** (0.033)	-0.170** (0.033)	-0.170** (0.033)	-0.165** (0.033)	-0.201** (0.036)
3-2	-0.748** (0.037)	-0.747** (0.037)	-0.747** (0.037)	-0.747** (0.037)	-0.738** (0.037)	-0.775** (0.040)
<i>Pitch Type</i>						
Breaking pitch	-0.139** (0.012)	-0.138** (0.012)	-0.138** (0.012)	-0.138** (0.012)	-0.136** (0.012)	-0.137** (0.013)
Offspeed Pitch	-0.228** (0.017)	-0.226** (0.017)	-0.226** (0.017)	-0.226** (0.017)	-0.226** (0.017)	-0.236** (0.018)
Unknown Pitch	2.893**	2.901**	2.904**	2.901**	2.927**	3.072**

	(0.614)	(0.614)	(0.614)	(0.615)	(0.617)	(0.748)
<i>Pitcher</i>						
Pitcher Tenure in League	0.021** (0.001)	0.014** (0.001)	0.014** (0.001)	0.014** (0.001)	0.014** (0.001)	0.014** (0.001)
Pitcher BBs per Batters Faced	-2.783** (0.210)	-2.471** (0.211)	-2.590** (0.218)	-2.474** (0.212)	-2.488** (0.212)	-2.516** (0.226)
Pitcher is African American	-0.157** (0.034)	-0.164** (0.034)	-0.162** (0.034)	-0.164** (0.034)	-0.164** (0.034)	-0.121** (0.037)
Pitcher is Hispanic	-0.007 (0.012)	-0.014 (0.012)	-0.013 (0.012)	-0.014 (0.012)	-0.013 (0.012)	0.003 (0.013)
Pitcher is Asian	0.025 (0.033)	0.024 (0.033)	0.026 (0.033)	0.024 (0.033)	0.022 (0.033)	0.045 (0.036)
Pitcher is Right-handed	0.041** (0.012)	0.048** (0.012)	0.046** (0.012)	0.048** (0.012)	0.044** (0.012)	0.046** (0.012)
<i>All-Star</i>						
Pitcher All-Star Appearances		0.047** (0.005)	0.038** (0.006)	0.045* (0.019)	0.047** (0.005)	0.047** (0.005)
Pitcher All-Star XBBs per Batter			-0.507* (0.231)			
Pitcher All-Star X Distance				-0.060 (0.087)		
Pitcher All-Star X Distance ^2				-0.103 (0.100)		
<i>Batter</i>						
Batter All-Star Appearances					-0.014** (0.003)	
Batter BBs per Plate Appearance					-0.991** (0.171)	
Batter is African American					-0.035* (0.014)	
Batter is Hispanic					0.032** (0.012)	
Batter is Asian					-0.006 (0.030)	
<i>Catcher</i>						
Catcher All-Star Appearances						-0.003 (0.002)
Catcher Framing Ability						0.066** (0.005)
Constant	-4.379** (0.053)	-4.374** (0.053)	-4.375** (0.053)	-4.374** (0.053)	-4.292** (0.056)	-4.361** (0.057)
N	489264	489264	489264	489264	488814	421530

* $p < .05$ ** $p < .01$

Table 4 Determinants of Under-Recognized Pitch (*Actual Strike that was mistakenly called a Ball*)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Situational</i>						
Home Team	-0.036** (0.012)	-0.036** (0.012)	-0.036** (0.012)	-0.036** (0.012)	-0.035** (0.012)	-0.041** (0.013)
Attendance (10K)	-0.008 (0.005)	-0.006 (0.005)	-0.006 (0.005)	-0.006 (0.005)	-0.009 (0.005)	-0.002 (0.006)
Total Distance from Border (ft)	-3.320** (0.022)	-3.320** (0.022)	-3.320** (0.022)	-3.319** (0.022)	-3.320** (0.022)	-3.312** (0.023)
Total DistanceXTotal Distance	-0.825** (0.057)	-0.824** (0.057)	-0.824** (0.057)	-0.822** (0.057)	-0.823** (0.057)	-0.784** (0.061)
Leverage	0.017* (0.008)	0.019* (0.008)	0.018* (0.008)	0.019* (0.008)	0.018* (0.008)	0.024** (0.009)
Run Expectancy	0.153** (0.019)	0.150** (0.019)	0.150** (0.019)	0.150** (0.019)	0.146** (0.019)	0.143** (0.021)
Inning	0.005* (0.002)	0.005* (0.002)	0.005* (0.002)	0.005* (0.002)	0.005* (0.002)	0.006* (0.002)
Batter stands Right-handed	-0.048** (0.012)	-0.047** (0.012)	-0.047** (0.012)	-0.047** (0.012)	-0.023 (0.013)	-0.038** (0.013)
Year is 2009	-0.192** (0.012)	-0.192** (0.012)	-0.193** (0.012)	-0.193** (0.012)	-0.192** (0.012)	-0.193** (0.013)
<i>Ballcount (Balls-Strikes)</i>						
0-1	0.673** (0.020)	0.673** (0.020)	0.674** (0.020)	0.673** (0.020)	0.672** (0.020)	0.668** (0.022)
0-2	1.095** (0.041)	1.097** (0.041)	1.097** (0.041)	1.097** (0.041)	1.104** (0.041)	1.094** (0.045)
1-0	-0.097** (0.019)	-0.097** (0.019)	-0.097** (0.019)	-0.097** (0.019)	-0.101** (0.019)	-0.097** (0.021)
1-1	0.482** (0.023)	0.482** (0.023)	0.482** (0.023)	0.483** (0.023)	0.476** (0.023)	0.482** (0.024)
1-2	0.827** (0.034)	0.828** (0.034)	0.828** (0.034)	0.828** (0.034)	0.825** (0.034)	0.838** (0.036)
2-0	-0.352** (0.031)	-0.352** (0.031)	-0.352** (0.031)	-0.352** (0.031)	-0.355** (0.031)	-0.336** (0.033)
2-1	0.275** (0.030)	0.275** (0.030)	0.275** (0.030)	0.275** (0.030)	0.265** (0.030)	0.259** (0.033)
2-2	0.769** (0.035)	0.769** (0.035)	0.769** (0.035)	0.770** (0.035)	0.763** (0.035)	0.782** (0.038)
3-0	-0.749** (0.044)	-0.749** (0.044)	-0.750** (0.044)	-0.750** (0.044)	-0.759** (0.044)	-0.768** (0.048)
3-1	0.005 (0.041)	0.005 (0.041)	0.005 (0.041)	0.005 (0.041)	-0.003 (0.041)	0.009 (0.044)
3-2	0.693** (0.046)	0.693** (0.046)	0.693** (0.046)	0.693** (0.046)	0.681** (0.046)	0.693** (0.050)
<i>Pitch Type</i>						
Breaking pitch	-0.109** (0.015)	-0.109** (0.015)	-0.109** (0.015)	-0.110** (0.015)	-0.114** (0.015)	-0.100** (0.016)
Offspeed Pitch	0.209** (0.021)	0.208** (0.021)	0.208** (0.021)	0.208** (0.021)	0.205** (0.021)	0.192** (0.023)
Unknown Pitch	2.245**	2.244**	2.243**	2.230**	2.251**	2.032*

	(0.798)	(0.798)	(0.798)	(0.793)	(0.801)	(0.870)
<i>Pitcher</i>						
Pitcher Tenure in League	-0.013** (0.001)	-0.010** (0.002)	-0.010** (0.002)	-0.010** (0.002)	-0.010** (0.002)	-0.011** (0.002)
Pitcher BBs per Batters Faced	1.194** (0.249)	1.049** (0.251)	1.124** (0.266)	1.046** (0.251)	1.060** (0.251)	1.162** (0.269)
Pitcher is African American	0.002 (0.040)	0.008 (0.040)	0.007 (0.040)	0.008 (0.040)	0.007 (0.040)	-0.035 (0.043)
Pitcher is Hispanic	-0.031* (0.015)	-0.028 (0.015)	-0.029 (0.015)	-0.028 (0.015)	-0.028 (0.015)	-0.034* (0.016)
Pitcher is Asian	-0.090* (0.042)	-0.089* (0.042)	-0.090* (0.042)	-0.089* (0.042)	-0.085* (0.042)	-0.121** (0.045)
Pitcher is Right Handed	-0.055** (0.014)	-0.059** (0.014)	-0.058** (0.014)	-0.059** (0.014)	-0.054** (0.014)	-0.073** (0.015)
<i>All-Star</i>						
Pitcher All-Star Appearances		-0.028** (0.006)	-0.024** (0.008)	-0.040** (0.008)	-0.028** (0.007)	-0.027** (0.007)
Pitcher All-Star X BBs per batters faced			0.272 (0.318)			
Pitcher All-Star X Distance				0.032 (0.018)		
Pitcher All-Star X Distance ^2				0.150** (0.046)		
<i>Batter</i>						
Batter All-Star Appearances					0.013** (0.003)	
Batter BBs per Plate Appearance					1.309** (0.210)	
Batter is African American					0.019 (0.017)	
Batter is Hispanic					-0.042** (0.015)	
Batter is Asian					0.086* (0.038)	
<i>Catcher</i>						
Catcher All-Star Appearances						0.007* (0.003)
Catcher Framing Ability						-0.102** (0.007)
Constant	-1.109** (0.055)	-1.112** (0.055)	-1.111** (0.055)	-1.112** (0.055)	-1.221** (0.059)	-1.162** (0.059)
N	221291	221291	221291	221291	220925	190583

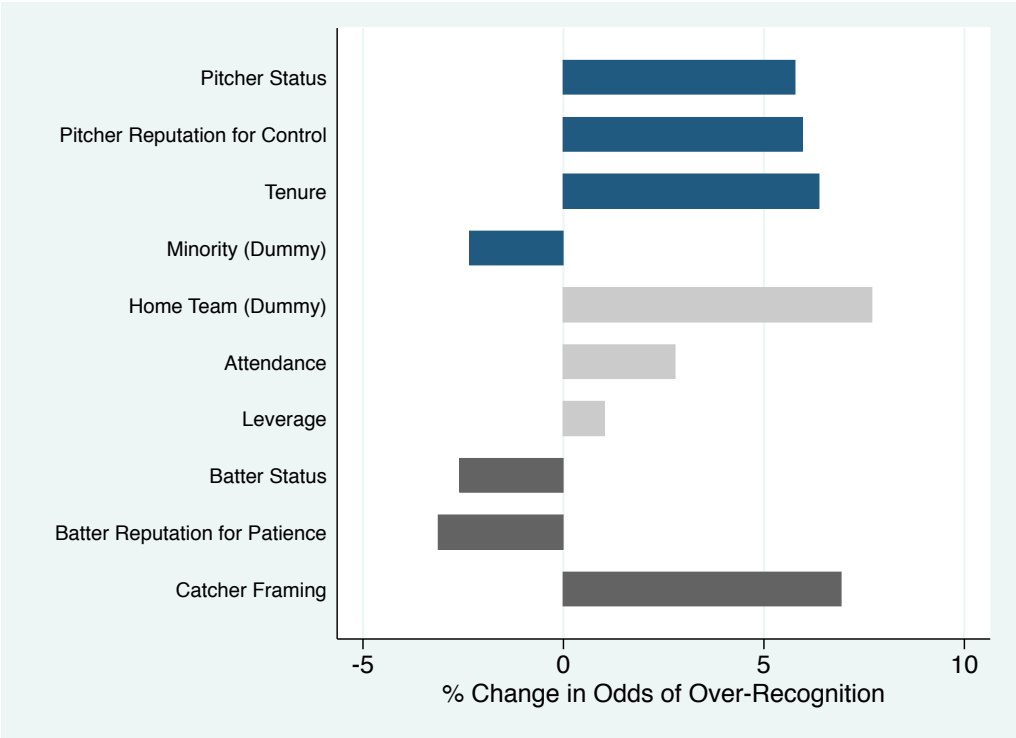
* $p < .05$ ** $p < .01$

Table 5 Determinants of Outcomes of an At-Bat: Winning Percentage Added (WPA) and Total Bases

	WPA		Total Bases	
	(1)	(2)	(1)	(2)
Pitcher Career Era	-0.00172** (0.00012)	-0.00166** (0.00012)	0.02841** (0.00180)	0.02709** (0.00180)
Batter Career OBP+SLG	-0.00966** (0.00069)	-0.00972** (0.00069)	0.24903** (0.00974)	0.25031** (0.00972)
Leverage	0.00732** (0.00012)	0.00731** (0.00012)	-0.02074** (0.00169)	-0.02065** (0.00168)
Home Team	0.00095** (0.00021)	0.00091** (0.00021)	-0.01688** (0.00296)	-0.01602** (0.00295)
Outcount (1 out)	0.00198** (0.00024)	0.00201** (0.00024)	0.00144** (0.00344)	0.00079** (0.00343)
Outcount (2 out)	0.02623** (0.00027)	0.02611** (0.00027)	-0.44295** (0.00377)	-0.44030** (0.00377)
Under-Recognized Pitch (Call: Ball; Location: Strike)		-0.00324** (0.00023)		0.07440** (0.00408)
Over-Recognized Pitch (Call: Strike; Location: Ball)		0.00326** (0.00029)		-0.06741** (0.00321)
Cons	0.00732** (0.00081)	0.00683** (0.00081)	0.16969** (0.01141)	0.17905** (0.01142)
F	2555.8	1962.77	3094.02	2428.8
R-Square	0.0655	0.067	0.0783	0.0816
N	218621	218621	218621	218621

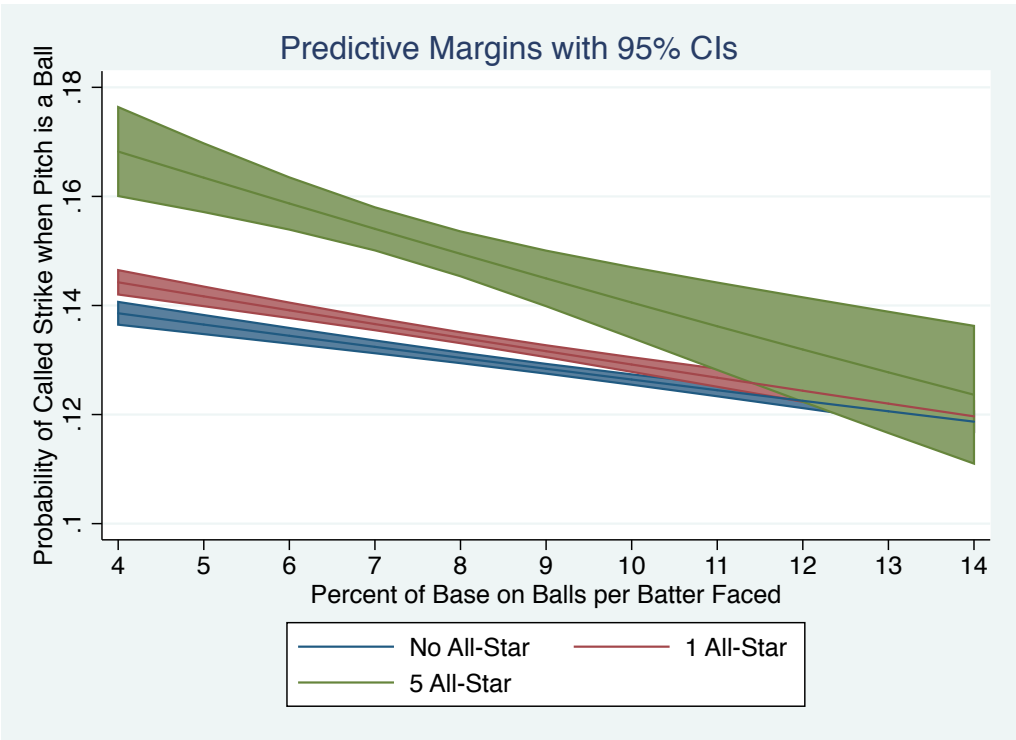
* $p < .05$ ** $p < .01$

Figure 1 Effects of One Standard Deviation Change in Selected Variables on Odds of Over-Recognition



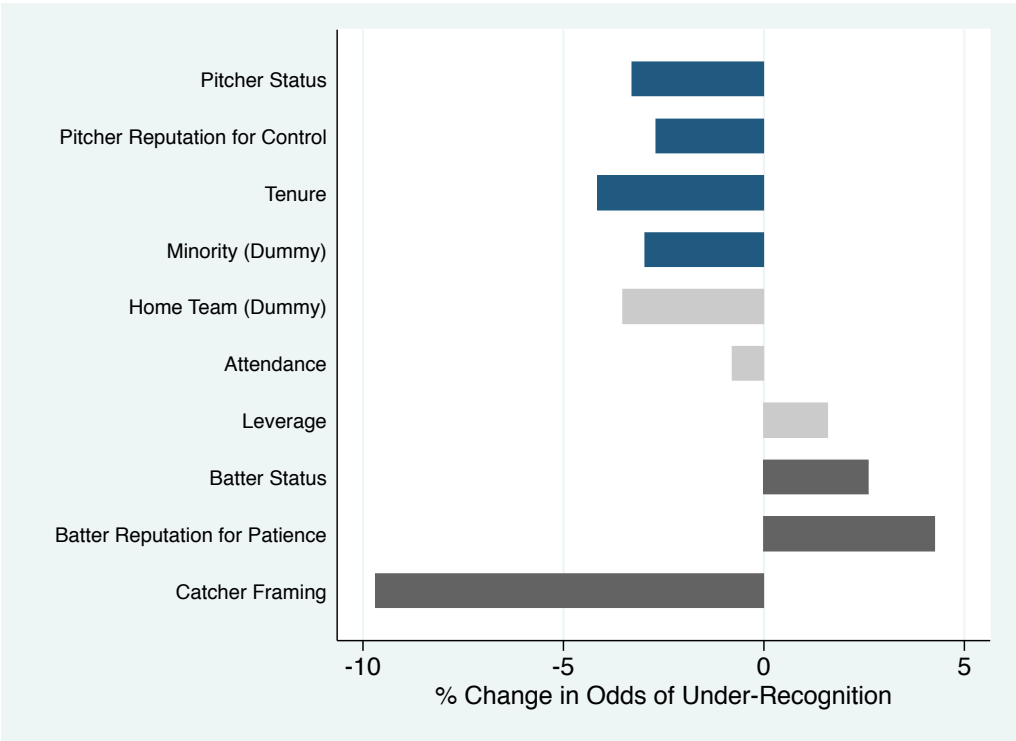
Notes: Changes in the odds of over-recognition are shown for a one standard deviation change in selected variables. The dark blue bars correspond to pitcher characteristics, light gray to situational context variables, and the dark gray to batter and catcher effects. Changes in odds are calculated based on a model with all control variables and the status variable (i.e., equivalent to Model 2 in Table 3, but with key variables standardized to allow comparison) except for batter and catcher effects, which were based on Model 6.

Figure 2 Interaction between Reputation and All-Star Appearances on Probability of Over-Recognized Pitch



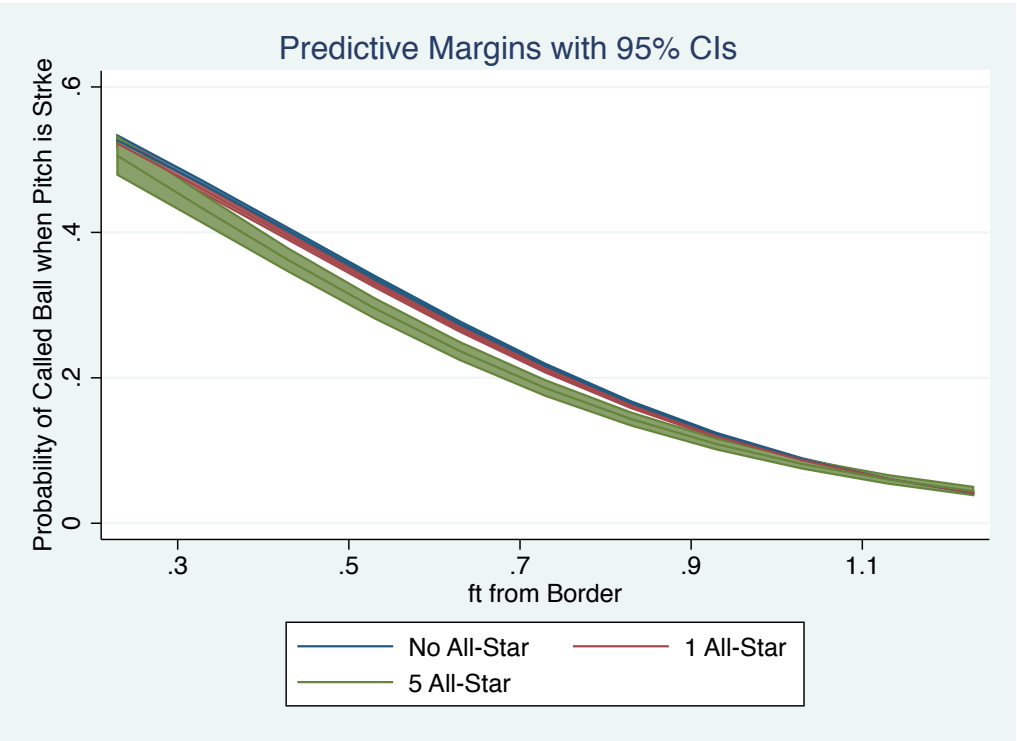
Notes: The effect of reputation for control (i.e., career percent of base on balls per batter faced) at different levels of status is shown. When pitcher has a track record of precise control (i.e., not walking batters), the effect of status is magnified, such that there is a considerable gap in the predicted probabilities of an over-recognized strike. In contrast, when a pitcher has a track record of poor control (i.e., walking many batters), status' effect on the probability of over-recognition diminishes and becomes indistinguishable.

Figure 3 Effects of One Standard Deviation Change in Selected Variables on Odds of Under-Recognition



Notes: Changes in the odds of under-recognition are shown for a one standard deviation change in selected variables. The dark blue bars correspond to pitcher characteristics, light gray to situational context variables, and the dark gray to batter and catcher effects. Changes in odds are calculated based on a model with all control variables and the status variable (i.e., equivalent to Model 2 in Table 4, but with key variables standardized to allow comparison), except for batter and catcher effects, which were based on Model 6.

Figure 4 Interaction between Distance from Border and All-Star Appearances on Probability of Under-Recognized Pitch



Notes: The effect of distance from the strike zone border (i.e., uncertainty) at different levels of status is shown. When a pitch is closer to the strike zone, the effect of status on under-recognition is magnified, but as the pitch reaches 1 foot or more from the border, and closer to the middle of the strike zone, there is no difference between high and low status.

Appendix: Coarsened Exact Matching Based on Location

In this appendix, we provide details for the coarsened exact matching (CEM) procedure we conduct to rule out heterogeneity in pitch location as an alternative to our observed status bias. CEM is a monotonic imbalance-reducing matching method in which data are pre-processed by temporarily coarsening variables into groups, performing exact matches on these coarsened data, and then applying statistical estimators to the data after matching. See Iacus, King, & Porro (2012, 2011) for an overview of the method, and Blackwell, Iacus, King, and Porro (2009) for implementation in Stata.

The matching process starts by creating a stratum for each unique combination of (coarsened) covariates, and placing observations into the corresponding strata. Strata that do not have at least one treated and one control unit are dropped. The strata serve as the basis for calculating the treatment effect via a weighting scheme that takes into consideration the number of controls and treated in the sample and stratum.

The treatment variable for CEM is typically binary, so we consider pitches that were thrown by pitchers with All Star appearances as “treated”, while pitches that were thrown by pitchers with no All Star appearances as part of the “control” group. (See below for different configurations of the treatment variable.)

We construct the matching sample based on the following covariates:

1. Pitch Zone
2. Umpire
3. Pitch Classification
4. Ballcount

Pitch Zone: We create 49 pitch zones based on the coordinates of the pitch as a method to “coarsen” the location variable to allow matching. In essence, we divide the horizontal plane into 7 slices (cutoffs at -1.38, -.83, -.28, .28, .83, 1.38 feet from the center of home plate), and the vertical plane into 7 slices (cutoff at 1.18, 1.74, 2.3, 2.86, 3.42, 3.98, 4.54 feet from the ground). The end result is that the strike zone is divided into 9 zones (following baseball convention), with two layers of zones surrounding the strike zone (total of 40 zones).

Umpire: We also match on umpire, given the possibility that pitchers may adjust their location strategy based on umpire specific “blind spots” or preferences.

Pitch Classification: We find matches on pitch type, which accounts for the speed and break of the pitch. All-star pitchers may choose different mixes of pitches depending on the location or umpire, so we consider this covariate in matching pitches.

Ballcount: All-star pitchers also might find themselves in disproportionately more favorable counts, so we balance out the distribution of counts by matching on this dimension.

We create one stratum for each pitch zone, umpire, pitch classification, and ball count combination, leading to a total of 51,138 matched strata. Of the 161,345 pitches in the treatment condition (i.e., thrown by a pitcher with at least one All Star appearance), we were able to find at least one matching control observation for 153,701 pitches, resulting in 7,644 unmatched treatment, and 109,546 control pitches.

We estimate logit models similar to our analyses in the main text using this matched sample and the generated CEM weights, with the results shown in Table 1 and 2 below. The main difference from the main models is that the status variable is a binary variable indicating whether the pitcher had at least one All-star appearance or not.

[Appendix Tables 1 and 2 here]

As seen in the tables, the effect of status is robust, with the treatment effect of having been an All Star resulting in a 6.7 percent increase in the odds of over-recognition, and 5.7 percent decrease in the odds of under-recognition. (The coefficients for models with interaction effects appear to be less consistent due to collinearity.). The persistence of

the status effect after matching suggests that the observed All Star effects are not the result of All Stars having a different location profile of pitches (i.e., throwing pitches in tougher to judge locations) relative to non All Stars. Furthermore, because we match on umpire as well as location, this also eliminates the possibility that All Star pitchers are better at exploiting individual umpire tendencies to see pitches.

We have also tried other combinations of covariates for matching purposes, including 1) more finer grained slices of the coordinates (20 slices each for horizontal and vertical coordinates), 2) matching on pitcher characteristics including tenure, pitcher BBs per batter faced, handedness in addition to the location and pitch characteristics used above, and find identical results.

One cost to dichotomizing the All Star variable is that it prevents us from capturing the cumulative effect of status as seen in the main analysis. To examine whether the status effect is cumulative (i.e., having more status is better), we compare the treatment effect across different criteria for consideration of treatment. More specifically, in one scheme, we defined treatment as a pitch being thrown by a pitcher with five or more All-star appearances, and matched these treated pitches with control pitches that were thrown by pitchers with no All-star appearances, and calculated the treatment effect. In another, pitches that were thrown by pitchers with 2-4 All Star appearances were compared with control pitches. And finally, we calculate the treatment effect when one-time All Stars were compared with no All Stars.

Figure 1 shows the comparison between these different levels of status. Being an established All Star (i.e., five or more appearances) has a considerable effect in both increasing over-recognition and reducing under-recognition. For the same location, umpire, pitch type and ballcount, a pitch that was thrown by an established All Star showed a 24.7 percent increase ($p < .01$) in the likelihood of over-recognition, and a 24.4 percent decrease ($p < .01$) in the likelihood of under-recognition relative to a pitcher with no All Star appearances. In contrast, just having one appearance increased the odds of over-recognition by 2.8 percent, and decreased the odds of under-recognition by 1.7 percent, but both effects were not statistically different from zero. The “middle status” zone of All Star appearances between 2 and 4 showed effects in between the established and first-time All Stars, with a 16.1 percent increase of over-recognition, and 12 percent decrease in odds of under-recognition, both statistically significant effects ($p < .01$).

In sum, the pattern of results suggest that status is in fact cumulative, even after accounting for location endogeneity by CEM matching, consistent with the main results.

Appendix Table 1 Determinants of Over-Recognized Pitch (*Actual Ball that was mistakenly called a Strike*)

	(1)	(2)	(3)	(4)	(5)
<i>Situational</i>					
Home Team	0.075** (0.011)	0.075** (0.011)	0.075** (0.011)	0.075** (0.011)	0.071** (0.011)
Attendance (10K)	0.023** (0.004)	0.022** (0.004)	0.023** (0.004)	0.025** (0.004)	0.020** (0.005)
Total Distance from Border (ft)	-8.967** (0.140)	-8.965** (0.140)	-9.124** (0.167)	-8.971** (0.140)	-8.823** (0.149)
Total Distance X Total Distance	-2.911** (0.154)	-2.907** (0.154)	-3.040** (0.182)	-2.916** (0.154)	-2.700** (0.164)
Leverage	0.014* (0.007)	0.016* (0.007)	0.014* (0.007)	0.015* (0.007)	0.015* (0.007)
Run Expectancy	-0.019 (0.016)	-0.020 (0.016)	-0.019 (0.016)	-0.017 (0.016)	-0.015 (0.018)
Inning	0.008** (0.002)	0.009** (0.002)	0.008** (0.002)	0.008** (0.002)	0.007** (0.002)
Batter stands Right-handed	-0.511** (0.011)	-0.510** (0.011)	-0.511** (0.011)	-0.525** (0.011)	-0.509** (0.012)
Year is 2009	-0.040** (0.011)	-0.037** (0.011)	-0.040** (0.011)	-0.040** (0.011)	-0.053** (0.012)
<i>Ballcount (Balls-Strikes)</i>					
0-1	-0.596** (0.018)	-0.596** (0.018)	-0.596** (0.018)	-0.596** (0.018)	-0.597** (0.019)
0-2	-0.962** (0.036)	-0.962** (0.036)	-0.962** (0.036)	-0.962** (0.036)	-0.978** (0.039)
1-0	0.110** (0.016)	0.111** (0.016)	0.110** (0.016)	0.113** (0.016)	0.114** (0.017)
1-1	-0.401** (0.019)	-0.401** (0.020)	-0.401** (0.019)	-0.398** (0.020)	-0.387** (0.021)
1-2	-0.738** (0.027)	-0.738** (0.027)	-0.738** (0.027)	-0.739** (0.027)	-0.730** (0.029)
2-0	0.256** (0.026)	0.257** (0.026)	0.256** (0.026)	0.259** (0.026)	0.241** (0.028)
2-1	-0.191** (0.026)	-0.191** (0.026)	-0.191** (0.026)	-0.185** (0.026)	-0.213** (0.028)
2-2	-0.576** (0.029)	-0.576** (0.029)	-0.576** (0.029)	-0.569** (0.029)	-0.586** (0.031)
3-0	0.477** (0.040)	0.479** (0.040)	0.477** (0.040)	0.484** (0.040)	0.475** (0.044)
3-1	-0.271** (0.040)	-0.270** (0.040)	-0.271** (0.040)	-0.268** (0.040)	-0.319** (0.044)
3-2	-0.751** (0.043)	-0.749** (0.043)	-0.751** (0.043)	-0.743** (0.043)	-0.745** (0.046)
<i>Pitch Type</i>					
Breaking pitch	-0.130** (0.013)	-0.130** (0.013)	-0.130** (0.013)	-0.128** (0.013)	-0.126** (0.014)
Offspeed Pitch	-0.306** (0.018)	-0.306** (0.018)	-0.306** (0.018)	-0.307** (0.018)	-0.318** (0.020)
<i>Pitcher</i>					

Pitcher Tenure in League	0.017** (0.001)	0.017** (0.001)	0.017** (0.001)	0.017** (0.001)	0.017** (0.001)
Pitcher BBs per Batters Faced	-2.507** (0.229)	-1.967** (0.244)	-2.509** (0.229)	-2.531** (0.229)	-2.512** (0.244)
Pitcher is African American	-0.147** (0.036)	-0.140** (0.036)	-0.147** (0.036)	-0.146** (0.036)	-0.093* (0.039)
Pitcher is Hispanic	-0.018 (0.013)	-0.017 (0.013)	-0.018 (0.013)	-0.018 (0.013)	-0.011 (0.014)
Pitcher is Asian	0.057 (0.035)	0.066 (0.035)	0.057 (0.035)	0.057 (0.035)	0.083* (0.037)
Pitcher is Right-handed	0.031* (0.012)	0.032** (0.012)	0.031* (0.012)	0.027* (0.012)	0.028* (0.013)
<i>All-Star</i>					
Pitcher All-Star>1 Appearance	0.065** (0.015)	0.031* (0.016)	0.216** (0.069)	0.065** (0.015)	0.058** (0.015)
Pitcher All-Star X BBs per Batter		-4.173** (0.666)			
Pitcher All-Star X Distance			0.539 (0.308)		
Pitcher All-Star X Distance ^2			0.436 (0.342)		
<i>Batter</i>					
Batter All-Star Appearances				-0.014** (0.003)	
Batter BBs per Plate Appearance				-0.985** (0.183)	
Batter is African American				-0.055** (0.015)	
Batter is Hispanic				0.015 (0.013)	
Batter is Asian				-0.002 (0.033)	
<i>Catcher</i>					
Catcher All-Star Appearances					-0.003 (0.002)
Catcher Framing Ability (norm)					0.068** (0.006)
Constant	-4.468** (0.061)	-4.469** (0.061)	-4.512** (0.064)	-4.381** (0.064)	-4.427** (0.065)
N	383193	383193	383193	382828	331396

Notes: Logit models estimating the likelihood of an actual ball being called a strike are shown. The sample is pre-processed using a coarsened exact matching (CEM) scheme based on pitch location, pitch type, umpire and ball count.

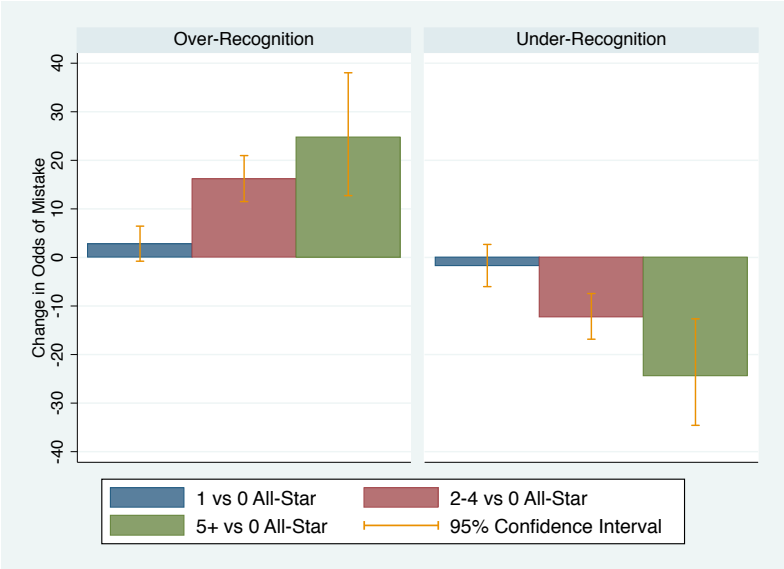
Appendix Table 2 Determinants of Under-Recognized Pitch (*Actual Strike that was mistakenly called a Ball*)

	(1)	(2)	(3)	(4)	(5)
Situational					
Home Team	-0.050** (0.013)	-0.050** (0.013)	-0.051** (0.013)	-0.050** (0.013)	-0.052** (0.014)
Attendance (10K)	-0.007 (0.006)	-0.007 (0.006)	-0.007 (0.006)	-0.010 (0.006)	-0.004 (0.007)
Total Distance from Border (ft)	-3.348** (0.024)	-3.362** (0.027)	-3.348** (0.024)	-3.347** (0.024)	-3.337** (0.026)
Total Distance X Total Distance	-0.775** (0.062)	-0.876** (0.072)	-0.775** (0.062)	-0.773** (0.063)	-0.732** (0.067)
Leverage	0.022* (0.009)	0.022* (0.009)	0.020* (0.009)	0.021* (0.009)	0.022* (0.009)
Run Expectancy	0.143** (0.021)	0.143** (0.021)	0.144** (0.021)	0.140** (0.021)	0.147** (0.023)
Inning	0.007** (0.003)	0.007** (0.003)	0.006* (0.003)	0.007** (0.003)	0.008** (0.003)
Batter stands Right-handed	-0.064** (0.014)	-0.064** (0.014)	-0.065** (0.014)	-0.047** (0.014)	-0.058** (0.015)
Year is 2009	-0.172** (0.014)	-0.172** (0.014)	-0.174** (0.014)	-0.171** (0.014)	-0.170** (0.015)
Ballcount (Balls-Strikes)					
0-1	0.691** (0.022)	0.691** (0.022)	0.692** (0.022)	0.691** (0.022)	0.685** (0.023)
0-2	1.152** (0.048)	1.152** (0.048)	1.152** (0.048)	1.159** (0.048)	1.189** (0.051)
1-0	-0.105** (0.021)	-0.105** (0.021)	-0.105** (0.021)	-0.108** (0.021)	-0.100** (0.023)
1-1	0.465** (0.025)	0.466** (0.025)	0.465** (0.025)	0.461** (0.025)	0.463** (0.027)
1-2	0.810** (0.037)	0.809** (0.037)	0.810** (0.037)	0.805** (0.037)	0.840** (0.040)
2-0	-0.343** (0.035)	-0.343** (0.035)	-0.344** (0.035)	-0.342** (0.035)	-0.327** (0.038)
2-1	0.302** (0.034)	0.302** (0.034)	0.302** (0.034)	0.294** (0.034)	0.302** (0.037)
2-2	0.784** (0.040)	0.784** (0.040)	0.785** (0.040)	0.779** (0.040)	0.785** (0.043)
3-0	-0.684** (0.051)	-0.684** (0.051)	-0.685** (0.051)	-0.693** (0.052)	-0.693** (0.056)
3-1	-0.012 (0.049)	-0.013 (0.049)	-0.014 (0.049)	-0.017 (0.049)	-0.012 (0.053)
3-2	0.716** (0.057)	0.716** (0.057)	0.715** (0.057)	0.704** (0.057)	0.714** (0.062)
Pitch Type					
Breaking pitch	-0.088** (0.016)	-0.088** (0.016)	-0.088** (0.016)	-0.091** (0.016)	-0.075** (0.017)
Offspeed Pitch	0.233** (0.022)	0.233** (0.022)	0.233** (0.022)	0.232** (0.022)	0.226** (0.024)
Pitcher					

Pitcher Tenure in League	-0.010** (0.002)	-0.010** (0.002)	-0.010** (0.002)	-0.010** (0.002)	-0.012** (0.002)
Pitcher BBs per Batters Faced	0.943** (0.277)	0.941** (0.277)	0.632* (0.294)	0.950** (0.277)	0.910** (0.296)
Pitcher is African American	-0.029 (0.044)	-0.029 (0.044)	-0.033 (0.044)	-0.030 (0.044)	-0.078 (0.048)
Pitcher is Hispanic	-0.018 (0.016)	-0.018 (0.016)	-0.019 (0.016)	-0.018 (0.016)	-0.028 (0.018)
Pitcher is Asian	-0.048 (0.046)	-0.048 (0.046)	-0.053 (0.046)	-0.046 (0.046)	-0.088 (0.049)
Pitcher is Right-handed	-0.055** (0.015)	-0.055** (0.015)	-0.056** (0.015)	-0.052** (0.015)	-0.067** (0.016)
All-Star					
Pitcher All-Star Appearances	-0.059** (0.019)	-0.092** (0.023)	-0.038+ (0.020)	-0.059** (0.019)	-0.054** (0.020)
Pitcher All-Star X BBs per Batters Faced			2.780** (0.853)		
Pitcher All-Star X Distance		0.062 (0.055)			
Pitcher All-Star X Distance ^2		0.431** (0.146)			
Batter					
Batter All-Star Appearances				0.016** (0.004)	
Batter BBs per Plate Appearance				0.984** (0.229)	
Batter is African American				0.016 (0.019)	
Batter is Hispanic				-0.037* (0.016)	
Batter is Asian				0.044 (0.042)	
Catcher					
Catcher All-Star Appearances					0.005 (0.003)
Catcher Framing Ability (normalized)					-0.106** (0.007)
Constant	-1.149** (0.064)	-1.141** (0.064)	-1.149** (0.064)	-1.225** (0.068)	-1.216** (0.069)
N	197664	197664	197664	197334	170654

Notes: Logit models estimating the likelihood of an actual strike being called a ball are shown. The sample is pre-processed using a coarsened exact matching (CEM) scheme based on pitch location, pitch type, umpire and ball count.

Appendix Figure 1 Status Effects and All-Star Appearances for Matched Samples



Notes: Figure shows the status effect when comparing pitches thrown by pitchers with one All-Star appearance with pitches thrown by pitchers with no All-Star appearances (blue), pitches thrown by pitchers with 2-4 All-Star appearances with no All-Star appearances (red), and pitches thrown by pitchers with 5-plus appearances with no All-Star appearances. All effects are based on samples undergoing a coarsened exact matching process to reduce imbalance and calculate the average treatment effect.