Let The Boys Play: Omission Bias in MLB Umpires

Luke E. DeMartin

Advisor: Christopher Maxwell, Ph.D.

Boston College Department of Economics

Undergraduate Honors Thesis

May 2023

Acknowledgments

To the O's — As of May 2023, the Orioles have one of the best records in Major League Baseball. So, a special thanks to Adley Rutschman, Ryan Mountcastle, Cedric Mullins, Jorge Mateo, Gunnar Henderson, Felix Bautista, Grayson Rodriguez, Yennier Cano, Kyle Bradish, and the rest of the gang for making Baltimore baseball fun again after 22 years of misery.

Professor Maxwell — A massive thank you for your invaluable contribution to this project. It has been a joy to work with you over the past year.

To my family — Thank you for all of your love and support over the past four years.

Lastly, to my friends — I would like to explicitly snub all of you from this acknowledgment section. Of course I would rather chill on the roof and play beer die than write this paper. But the fact that you would even ask me that is extremely inconsiderate.

Abstract

This paper investigates the existence of omission bias in Major League Baseball's home plate umpires. Omission bias describes the human tendency to prefer harm caused by inaction, or acts of omission, over equal harm caused by action, or acts of commission. For umpires, I define an act of commission as a call made by the umpire that ends the at-bat and an act of omission as a call that does not end the bat. By analyzing over 1.5 million pitches thrown between the years 2018 and 2022, I find that MLB umpires display omission bias by systematically increasing the size of the enforced strike zone on three-ball counts and shrinking the size of the enforced strike zone on two-strike counts. Further, I find that omission bias exists separately from and is not impacted by other biases present in MLB umpiring, such as the biases favoring *home batters* and *star batters*.

Three Base Hit by James Daugherty, 1917

(via Whitney Museum of American Art)

Table of Contents

1 Introduction

In the world of professional baseball, it is often said that the best umpires are the ones whose names are not known. On its surface, this would seem to mean that an umpire should have zero impact on the outcome of a game outside of making the proper calls. But umpires are not robots (yet). They are imperfect human beings who are prone to making mistakes. Theoretically, if all umpires just made random errors, uncorrelated to in-game factors, there would be minimal impact on long-run outcomes. Existing literature suggests, though, that MLB umpires and officials across sports leagues display forms of "systematic bias" (Moskowitz and Wertheim, 2011), meaning that their errors are related to actual variables present in a given game, and thus impact the outcome of games in systematical ways. In fact, MLB umpires show bias in order to keep their names from being known.

In this paper, I investigate this systematic bias found in Major League Baseball called *omission bias*. Omission bias describes the human tendency to prefer harm caused by inaction, or acts of omission, over equal harm caused by action, or acts of commission (Baron and Ritov, 2004). When calling balls and strikes, MLB umpires display a form of omission bias by distorting the strike zone on counts with either three balls or two strikes to avoid ending the at-bat. On two-strike counts, a called third strike creates an action in the form of a strike-out. On three-ball counts, a called ball creates an action in the form of a walk. Thus, a biased umpire will distort the strike zone to avoid creating action, opting to let the at-bat continue. In other words, all else equal, umpires prefer acts of omission, letting the at-bat continue, to acts of commission, ending the at-bat with a called ball or strike.

Through my analysis, I confirm the existence of this bias and build on existing literature by investigating the factors that contribute to its magnitude. I aim to answer the following questions regarding omission bias:

1. Does omission bias exist in Major League Baseball umpiring and, if so, to what extent?
2. Is the bias related to direct external pressure caused by the presence of fans at the games? How does the bias differ for umpiring home and away teams?
3. How does the overall magnitude of the bias change as in-game factors and situational characteristics change? Do umpires' displays of omission bias increase or decrease during relatively crucial moments or when perceived superstars are up to bat?

The aim of this paper is to gain a better understanding of the sources of unconscious human biases generally, not just within the context of baseball. Baseball provides a context uniquely appropriate for examining behavioral biases. For one, home-plate umpires are put in a position to make hundreds of split-second judgments each game, providing a sample of hundreds of thousands of decisions over the course of just one season. Second, the in-game situation is constantly evolving, and each pitch provides a unique set of circumstances under which the umpire must make these decisions. Third, when calling balls and strikes, the umpire is either verifiably correct or incorrect due to the well-defined boundaries of the game's strike zone. Thanks to modern pitch-tracking technology, the precise location of each pitch throughout the season is known, and since the rules clearly define balls and strikes, there is little ambiguity as to whether the umpire made the correct or incorrect judgment. Due to these characteristics, Major League Baseball is a specifically advantageous setting to study the influences on human behavior.

Baseball For Dummies

In its rulebook, the MLB defines the strike zone as "the area over home plate from the midpoint between a batter's shoulders and the top of the uniform pants – when the batter is in his stance and prepared to swing at a pitched ball – and a point just below the kneecap" (Figure 1). If the pitched ball crosses over home plate anywhere within this zone, the pitch is, by definition, a strike. Likewise, if the ball crosses anywhere outside this zone, it is a ball. For the rest of the paper, I will refer to this defined area as the *defined strike zone*, or the strike zone as it is supposed to be enforced. In reality, though, the strike zone is not enforced strictly by these parameters. Umpires are less likely to call strikes that just graze the corner of the defined strike zone, for example. Thus, I will refer to the strike zone as it is actually enforced by umpires as the *enforced strike zone* for the remainder of the paper.

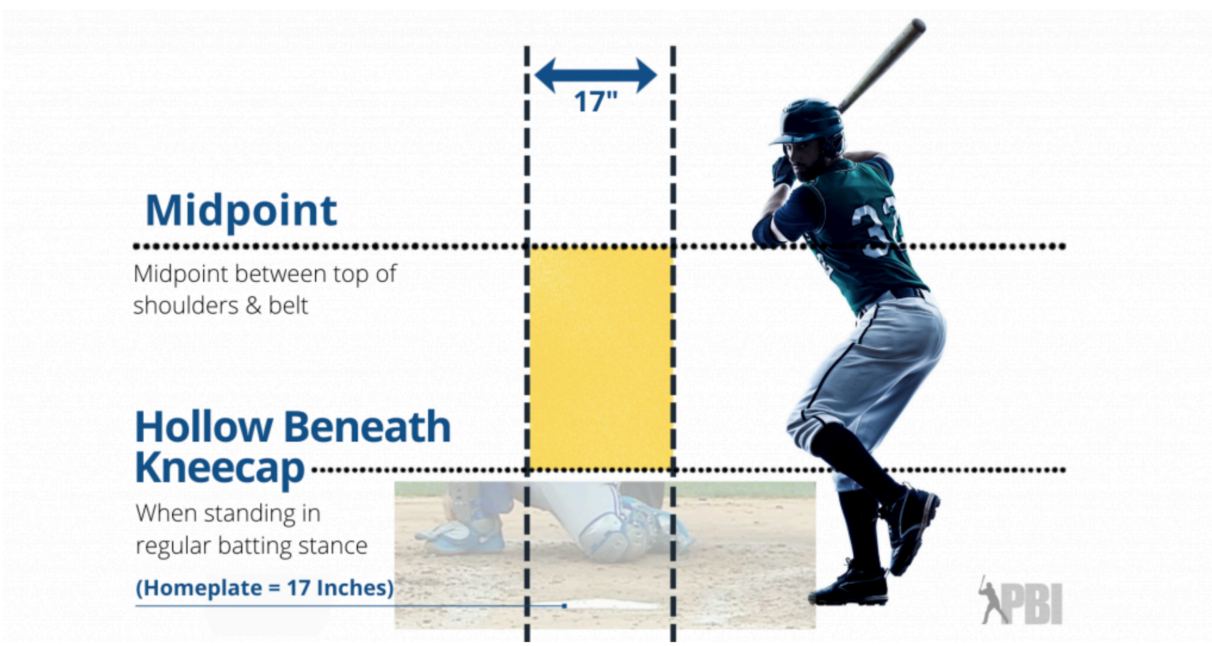

Figure 1: The Defined Strike Zone, According to the MLB Rule Glossary (Bernier, 2014)

When a batter swings at a pitch, there are dozens of potential outcomes, none of which involve a direct decision from the home-plate umpire, and are thus not relevant to umpire bias. When a batter does not swing at a pitch, there are generally two possible outcomes (ignoring if the batter is hit by the pitch, which is also not relevant). If the umpire deems the pitch to have crossed the plate outside of the defined strike zone, he will call the pitch a *ball*. If the umpire deems the pitch to have crossed the plate within the defined strike zone, he will call a *strike*. In other words, if the batter does not swing at the pitch, the home plate umpire must decide whether the pitch was a ball or a strike, and this decision will be either objectively correct or incorrect based on the defined strike zone. The batter is awarded first base if, within their at-bat, the pitcher throws four balls outside of the strike zone (resulting in a "walk"). The batter is called out if the pitcher throws three pitches in the strike zone (resulting in a "strikeout") within a single at-bat. I will refer to counts with fewer than three balls and fewer than two strikes as *neutral counts* since the umpire cannot end the at-bat with a subsequent ball or strike call.

Now that you understand the basics of how an at-bat works, close your eyes and imagine you are the home plate umpire for game seven of the World Series. Now, open your eyes so you can keep reading my paper. In this hypothetical dream world, it's the bottom of the ninth inning and the bases are loaded. The manager calls for a pinch hitter, and out comes none other than Babe Ruth who decided to rise from the dead to take this single at bat. As soon as the at-bat is over, he will return to his eternal rest. The first two pitches are right down the middle. You have no choice but to call strike one and strike two. The third pitch whizzes into the catcher's mitt at upwards of 100 miles per hour. Maybe the pitch barely clips the edge of the zone, but you are not sure. Unfortunately, you only have a split second to make the call. What then, dear reader, do you do? If you screw this up, it would be a pretty salient mistake, and you would have a direct link to the end of the at-bat. Do you (a) call strike three, effectively depriving the world of seeing The Great Bambino swing the bat one last time, or do you (b) call a ball and let the at-bat continue? If you picked the second option, you may suffer from omission bias.

My Approach

For each at-bat, the most impactful calls an umpire can make occur when the batter has either three balls or two strikes, since the call may result in the end of the at-bat via a walk or a strikeout. Thus, an umpire who displays omission bias will be more likely to call a ball when the batter has two strikes and more likely to call a strike when the batter has three balls. A consistent difference in the likelihood of an umpire calling a strike based on the current ball-strike count, controlling for the precise location of the pitch, would thus be evidence of omission bias. More specifically, holding pitch-location constant, umpires who display omission bias will be more likely to call strikes on three-ball counts than on neutral counts and less likely to call strikes on two-strikes counts than during neutral counts.

Next, if omission bias is, in some part, a response to direct external pressure, fan attendance should influence the occurrence of omission bias. I take advantage of within-season variance in fan attendance as well as the exogenous emptying of stadiums caused by the Covid-19 pandemic to evaluate the influence of fans on the occurrence of omission bias in umpires. Alongside many sports leagues worldwide, Major League Baseball held their truncated 2020 season entirely without fans. The season was delayed and ultimately shortened from 162 regular season games to just 60 for each team. Given that fan attendance is generally correlated with overall team success, previous literature has treated Covid-19's impact on sports leagues as a natural experiment to parse out the effect fans have on different sports outcomes, and I aim to implement a similar strategy in an attempt to identify the sources of umpire bias.

Lastly, if omission bias is influenced by the relative salience of a potential mistake, situational characteristics that influence the perceived stakes of a given in-game moment will influence when umpires display omission bias. Simply put, an umpire will be more biased when the stakes are higher. Factors such as runners on base, current inning, quality of teams, score, and notoriety of the batter and pitcher all impact the perceived gravitas of the in-game situation and thus would affect an umpire's reluctance towards making impactful calls.

I find conclusive evidence in support of the existence of omission bias in Major League Baseball, as umpires are less likely to call strikes on two-strike counts and more likely to call strikes on three-ball counts relative to neutral counts. Further, I find that the bias is primarily isolated to relatively uncertain pitches. I find little evidence that fan attendance or in-game situational characteristics influence the occurrence of omission bias in MLB umpires.

Baltimore Orioles' Manager Earl Weaver, following ejection for arguing Balls and Strikes, circa 1970

(via Sports Illustrated)

2 RELEVANT LITERATURE AND CONTEXT

Umpire Performance

Controversy surrounding umpire performance has led to the creation and growing popularity of the Twitter account and website, UmpScorecards, which is "an online platform dedicated to measuring the accuracy, consistency, and favor of MLB umpires." Created in 2019 by Ethan Singer, an undergraduate at Boston University, the Twitter account has amassed over 300,000 followers as of May 2023. Following each MLB game throughout the regular season and postseason, the account tweets a "scorecard" evaluating the performance of the game's home plate umpire. Utilizing pitch-level data from the MLB, the scorecard includes measures of accuracy, consistency, a list of "impactful missed calls," and an evaluation of "overall favor" that aims to measure which team benefitted from the Umpire's error and by how much they benefitted (Figure 2). The website also keeps track of umpire-specific statistics, such as strikes called, average run impact, and overall accuracy, and their database includes all games since 2015. To indicate whether a pitch was called correctly, they first calculate the likelihood that the pitch was a strike given the reported location of the pitch. They "consider a taken pitch to be incorrectly called if: the probability that the pitch was truly a strike was over 90%, and the umpire called it a ball; the probability that the pitch was a ball was over 90%, and the umpire called it a strike."

Thanks to the growing availability of and overall improvements to pitch-level data, Umpires are facing more scrutiny and pressure than ever before. Evidence shows, though, that this scrutiny may correlate with improved Umpire accuracy. Mills (2017) measures the change in umpire performance since the implementation of advanced pitch monitoring technologies in Major League Baseball. In 2001 the MLB began tracking the location of each pitch with a monitoring system called QuesTec, though it was not until 2004 that the umpires' union agreed to its use in evaluating umpire performance. In mid-2007, the MLB began using a more advanced system called PITCHf/x to track pitch locations, and, unlike with QuesTec, this data was made available to the public. In 2009, the umpires' union again agreed to the use of this data to train and evaluate the performance of MLB home-plate umpires. Figure 3 summarizes the timeline of the lead-up to the Pitch Tracking Era (2008 - Present). Using a difference-in-difference model with umpire, year, and stadium fixed effects, Mills estimates the impact of the installation of the monitoring systems as well as their usage as evaluation tools. Comparing performance in stadiums with and without the technology installed, Mills finds that "there was little to no continued development in performance" upon the implementation of QuesTec's tracking technology in 2001. On the other hand, following the implementation of the improved PITCHf/x system and corresponding training and evaluation in 2009, Mills finds that overall umpire accuracy improved "on average… nearly as much as a full percentage point per year." Further, in 2008, umpires called strikes correctly 78.7% of the time. But by 2014, umpires were more than 87% accurate on called strikes.

According to FanGraphs.com, this trend has only continued. Andrews (2023) finds that overall umpire accuracy has improved every year since the beginning of the Pitch Tracking Era from 81.3% overall accuracy in 2008 to 92.4% in 2022 (Figure 4). In fact, on October 29, 2022, in the second game of the World Series, umpire Pat Hoberg recorded the first "Perfect Game" according to UmpScorecards (Figure 5). Therefore, the data indicate that umpires are better than ever before. Still, they are quite far from perfect. Recently, the discussion of umpire performance in the MLB has become urgent, due to league-wide talks surrounding the implementation of the Automated Ball-Strike System (ABS) colloquially known as "robo umps." MLB commissioner Rob Manfred said in June of 2022 that technology that automatically calls balls and strikes could be implemented into Major League Ballparks as soon as 2024. In January 2023, it was announced that the ABS will be used in all AAA minor league ballparks, though in a somewhat limited capacity. Half of the games at this level “will be played with all of the calls determined by an electronic strike zone, and the other half will be played with an ABS challenge system” where teams have the option to force a review of the home plate umpire’s call (Olney, 2023).

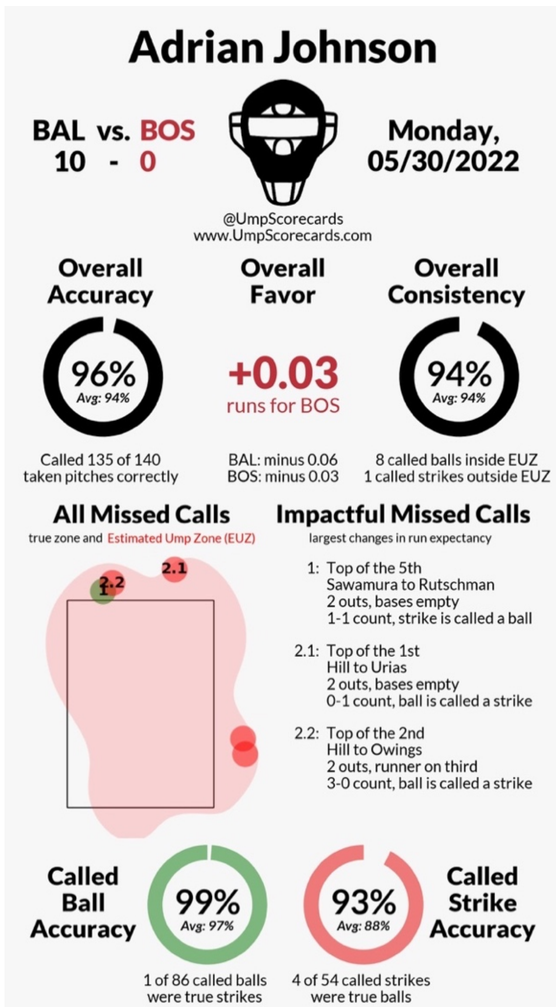

Figure 2: UmpScorecard from Baltimore Orioles vs. Boston Red Sox on May 30, 2022

Year	Event	Description
1996	New strike zone definition	MLB redefined the strike zone, moving lower end from top to bottom of knees
2001	Introduction of QuesTec	Introduction of technology to monitor umpire accuracy on ball-strike calls
2004	Ratification of QuesTec	Grievance by union settled; QuesTec used partially for evaluation; new CBA
2007	Introduction of PITCHf/x	Sportvision's PITCHf/x introduced for the Gameday website; data is publicly released
2009	Introduction of Zone Evaluation	Umpire union agrees to use of PITCHf/x data is for evaluation and training; new CBA

Figure 3: Timeline of Leadup to Pitch Tracking Era (Mills, 2017)

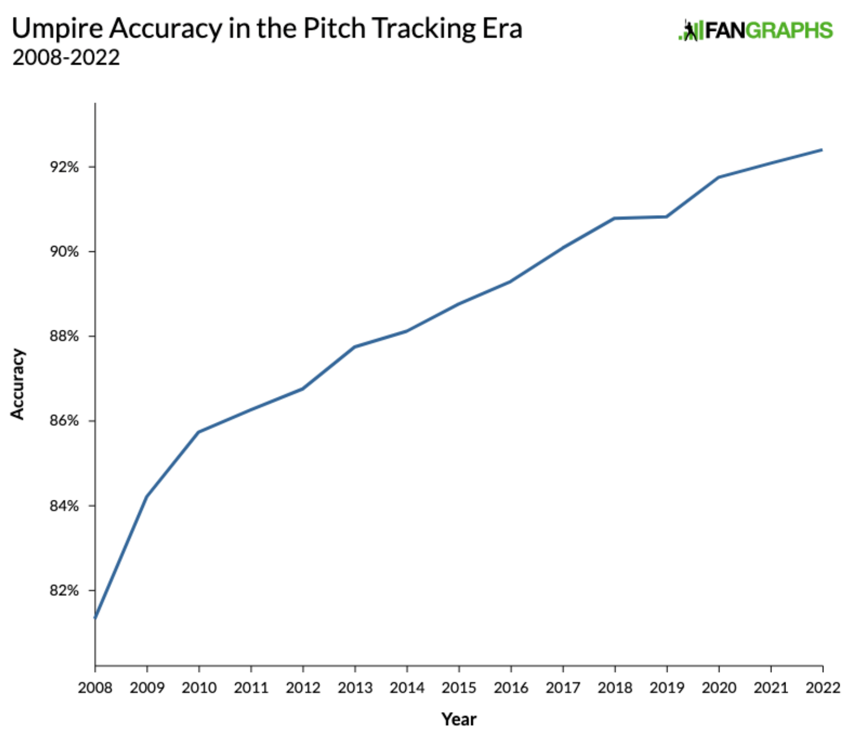

Figure 4: Improvement of Pitchers during the Pitch Tracking Era (Andrews, 2023)

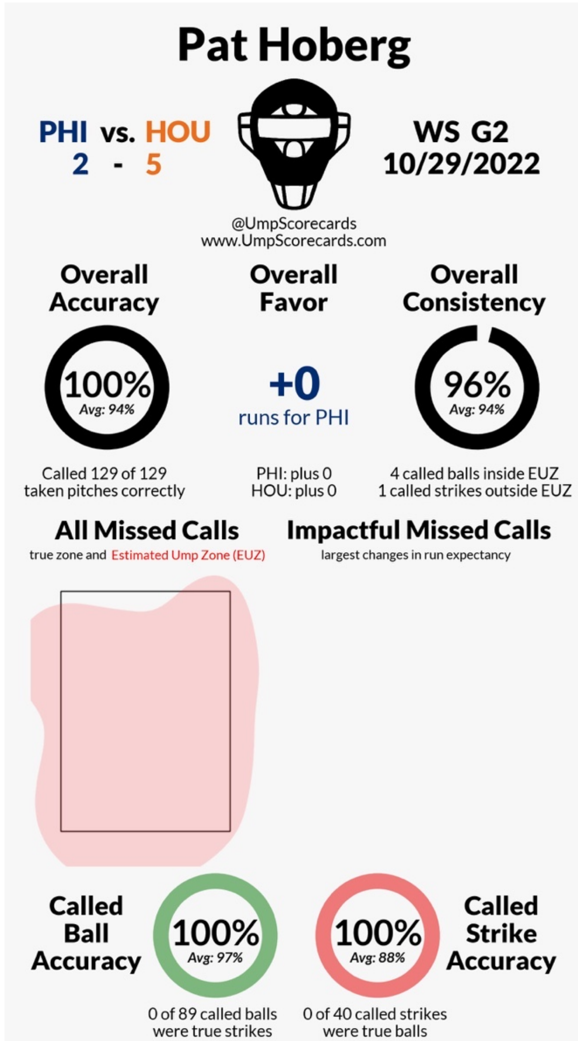

Figure 5: UmpScorecard from Pat Hoberg's "Perfect Game"

Home-Field Advantage and the Covid-19 Pandemic

Remember the pandemic? Well, just like every other facet of society, the MLB had to make adjustments and adapt to the new normal. After perpetual delays to the start of the 2020 season, the MLB opted to play a shortened 60-game season beginning on July 23, 2020. Several star players and 11 umpires opted out of participating in the season entirely. Due to the *Quarantine Act* in place in Canada, the Canadian government denied the Toronto Blue Jays the ability to play in their home parks, and they instead played their home games in Buffalo, New York. Like many other sports leagues around the world, the entirety of the season was played without any fans in attendance.

Citi Field in Queens, NY filled with Cardboard Cutout Fans, circa 2020

(via The New York Times)

The lack of fans in attendance removes one crucial factor in the jumbled equation that makes up home-field advantage. While home-field advantage is a proven phenomenon throughout practically all sports (the home team has won about 55% of all MLB games since 2000), there are countless factors theorized to contribute to its existence. Familiarity, lack of travel, and overall increased rest are posited factors. Perhaps even the cheering crowd affects the home team physiologically so that they perform better. Moskowitz and Wertheim (2011) argue that the most important factor, though, is that the fans psychologically influence the umpires into calling the game more favorably for the home team, citing disparities in the officials' treatment of home and away teams in the NFL, NHL, NBA, and MLB.

Bilalić, Gula, and Vaci (2021) investigate the home-field advantage effect across 12 different European soccer leagues and over 4,000 games pre- and post-Covid-19 and provide a model for the causes of home-field advantage called the Home Advantage Mediated (HAM) model (Figure 6). The HAM model exhibits the mechanisms through which the home-field advantage phenomenon could arise. They observe a significant spread between home and away teams' total fouls, yellow cards, and red cards prior to the pandemic that disappears completely when the games are played without fans during the pandemic, suggesting that Soccer referees are influenced by home fans.

Further, in the wake of the Covid-19 pandemic, several researchers have taken advantage of the exogenous variation in the number of fans in attendance to measure the impact fans have on sporting events. Nevertheless, a plurality of the research focuses on European soccer, and little is known about the impact on Major League Baseball. Cross and Uhrig (2022), for example, examine the effects of no fans on home-field advantage across European soccer leagues. The exogenous variation in attendance allows them to isolate the portion of home-field advantage attributable to the presence of fans. They find that "the change in home-field advantage decreases the probability a home team wins the game by 5.4 percentage points" (Figure 7), meaning that home-field advantage is 5.4 percentage points stronger when fans are in attendance. For their identification strategy, they control for Covid-19 case levels by geography, temperature to account for the delayed start to the season, and strength of schedule metrics. Further, they estimate the change in home-field advantage on realized and expected goals and find statistically significant decreases of 0.175 goals per game and 0.181 expected goals per game.

Still, they do not investigate why this decrease occurs when fans are not present. Thus, it is not clear whether this change is due to changes in referee or umpire behavior.

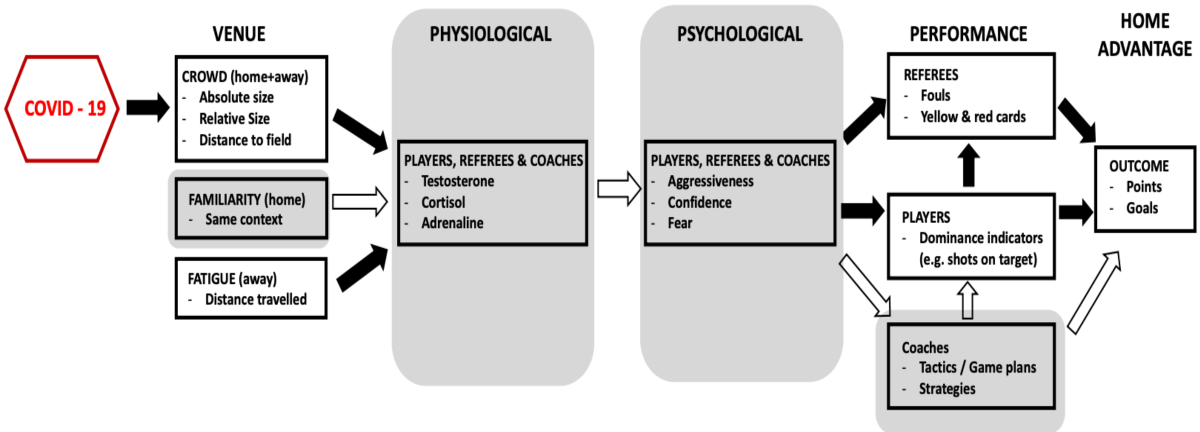

Figure 6: The Home Advantage Mediated (HAM) Model (Bilalić, Gula, and Vaci, 2021)

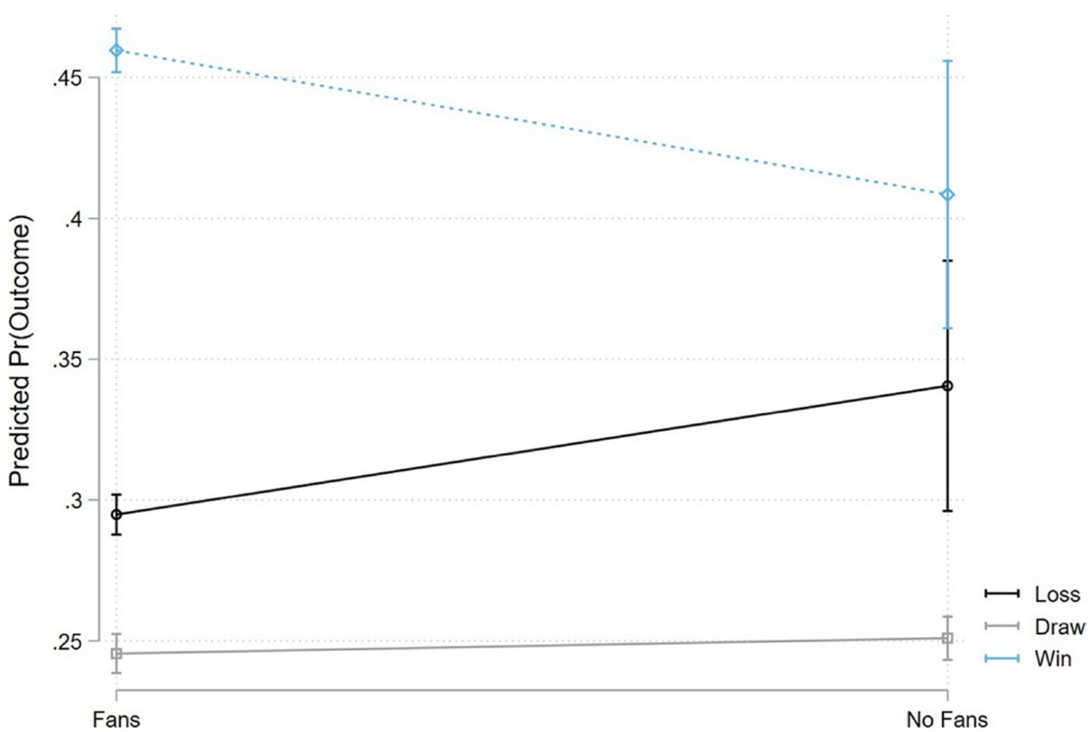

Figure 7: Effect of No Fans on Match Outcomes (Cross and Uhrig, 2022)

Gong (2022) examines NBA referee bias during the 2020-2021 season, when arenas were empty, to bias in the preceding seasons. He utilizes the NBA's Last Two Minute Reports which evaluate the performance of referees in the last two minutes of close games. Ostensibly, these games would contain primarily high-leverage situations given that only games that have a point differential of three or fewer at any point in the final two minutes of regulation. Further, the dataset indicates the correctness of both fouls called and non-calls (or moments where the ref swallowed their whistle). By looking at foul calls at the play level as opposed to at the game level, the estimation avoids suffering from endogeneity issues resulting from a change in player behavior caused by fans. The results suggest that "crowd support does not cause referees to treat home and away teams differently in crucial situations during the NBA regular season."

Whistle Swallowing

Moskowitz and Wertheim (2011) investigate a concept that they call "Whistle Swallowing" both in the MLB and throughout other professional sports. They argue that not only do referees tend to consistently display "omission bias" where officials "view acts of omission… as far less intrusive or harmful than acts of commission even if the acts are the same or worse", but also that fans and leagues encourage this behavior. Looking at pitch-level data from the MLB, the authors analyze the actual strike zone (that is, as it was actually called by umpires) against the rule-mandated strike zone. They find that when the batter has a 3-0 count, the actual strike zone grows significantly, and, likewise, when the batter faces an 0-2 count, the size of the actual strike zone shrinks significantly (Figure 8). The result is a strike zone on 3-0 counts that is "188 square inches larger than it is on 0-2 counts." Further, they find that on an 0-2 count, umpires have "more than twice" their overall error rate when calling balls and strikes "in favor of the batter." When the batter faces a 3-0 count, they find that umpires will "erroneously call strikes." 20 percent of the time, as opposed to their baseline error rate of 12.2 percent.

Similarly, Green and Daniels (2015) use a sample of over 1 million pitches from the years 2009 to 2011 to evaluate the effect of "Impact Aversion," or omission bias, on umpires' ball and strike calls. Specifically, they investigate the likelihood of a ball or strike call based on "the relative impacts of the umpire's two options" on "the expected number of runs that the batting team will score over the remainder of the half-inning." They find that every umpire in their sample displayed impact aversion at a significant level, meaning that each umpire tended to relatively favor the judgment that had a lesser impact on expected runs scored (Figure 9). They also conclude that, as the pitch got closer to the edges of the strike zone, umpires displayed more caution when calling balls and strikes. Further, they include "measures of scrutiny" into their estimation such as whether the game is nationally televised and game attendance. They hypothesize that the more people watching the game, the more impact averse an umpire will be. Their results suggest that "there is no meaningful difference in the enforced strike zones for the home and away teams, and umpires are equally impact averse when the home and away teams are at bat," but that "larger crowds are associated with a greater degree of impact aversion."

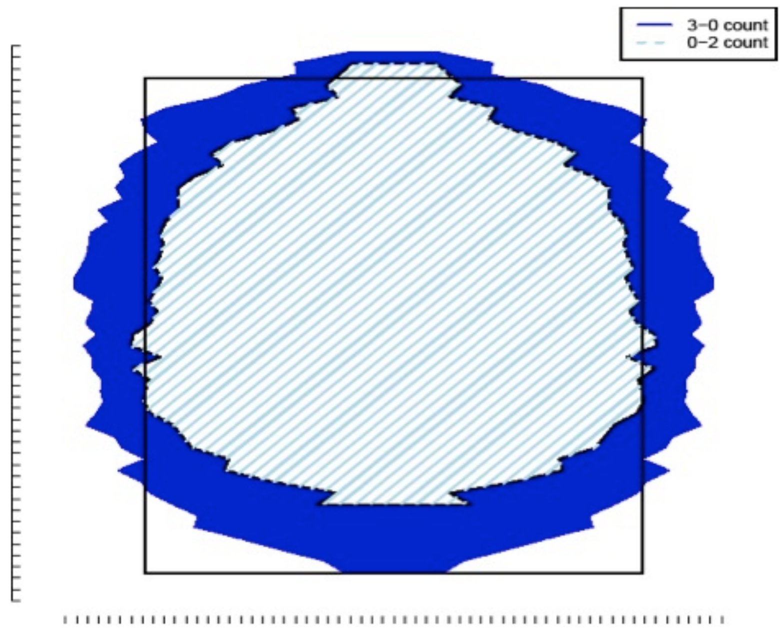

Figure 8: Plot of the empirical strike zone (defined as any pitch called a strike at least 50% of the time by MLB umpires) on 3-0 vs. 0-2 counts (Moskowitz and Wertheim, 2011)

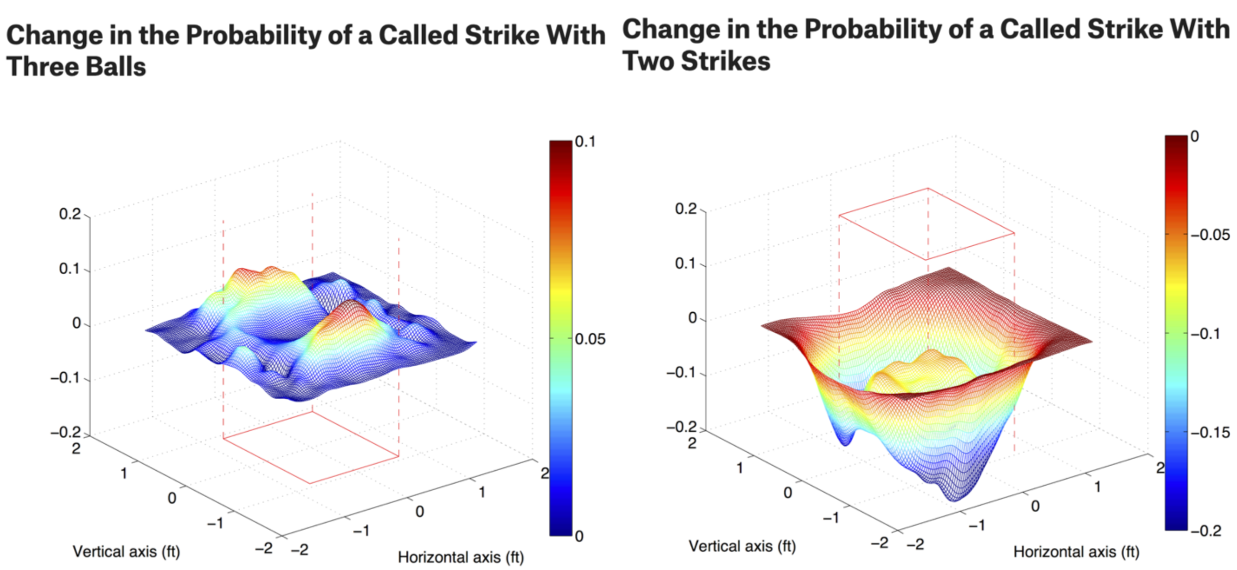

Figure 9: Figure 4: Change in Enforced Strike Zone with Three Balls and with Two Strikes (Green and Daniels, 2015)

Omission Bias in Behavioral Psychology

Both Moskowitz and Wertheim (2011) and Green and Daniels (2015) look to behavioral psychology to explain the behavior they find in MLB umpires. Kordes-de Vaal (1996) helps answer the following question related to umpires preference for inaction: Why are omissions and commission "evaluated differently, given that the consequences are the same"? Kordes-de Vaal argues that when omission is viewed as a "nondecision" as opposed to "a deliberate choice 'not to act'" and thus, a negative outcome viewed as a result of nothing is viewed less negatively than a negative outcome resulting from a deliberate action. Through a series of experiments, the paper shows that the omission bias stems primarily in the perceived causality of act of commission relative to acts of omission. In other words, since acts of commission are more salient than omission, the perceived causal link between the act and the outcome is higher for acts of commission.

The omission bias is closely tied to loss aversion, a concept first coined by Kahneman and Tversky (1979), who hypothesize that one's "response to losses is stronger than the response to corresponding gains." More relevant is the idea of the "action-effect," a specific form of loss aversion presented by Kahneman and Tversky (1982). They cite experimental evidence that suggests that "the regret associated with a loss that was incurred by an action tends to be more intense than the regret associated with inaction or a missed opportunity." They present the following hypothetical to support their hypothesis:

> Paul owns shares in Company A. During the past year he considered switching to stock in Company B, but he decided against it. He now finds that he would have been better off by $1,200 if he had switched to the stock of Company B. George owned shares in Company B. During the past year he switched to stock in Company A. He now finds that

he would have been better off by $1,200 if he had kept his stock in Company B. Who feels more regret?

They find that individuals like George, who took direct action that led to their misfortune, tend to feel far more regret than individuals like Paul who, through inaction, was worse off.

Another concrete example may help to clear up this concept. Brown et al. (2009) examines parents' hesitancy towards their children to receiving a swine flu vaccination. In a questionnaire, Brown et al. had parents of young children rate the probability of occurrence, symptoms, and duration of a hypothetical disease and of a "vaccine adverse event (VAE)." They find that parents view the vast majority of equivalent factors as less favorable when they arise from a VAE compared to a disease (Figure 10). In other words, they find that equal negative outcomes are perceived as worse when they arise from an act of commission (taking a vaccine) than an act of omission (denying a vaccine and contracting a disease).

Factor	Modal response (% of respondents choosing option)		Difference
	Disease	Vaccine reaction	
Likelihood ($n = 125$)			
Small	0.1% (40.8)	0.0001% (37.3)	$t(1,124) = 5.44^a$
Medium	1% (55.2)	1% (33.1)	$t(1,124) = 4.65^a$
Large	10% (78.4)	10% (53.5)	$t(1,124) = 2.98^b$
Symptoms/signs ($n = 99$)			
Loss of appetite	Mild (92.9)	Mild (70.1)	$t(1,98) = 5.16^a$
Irritability/crying	Mild (82.8)	Mild (72.4)	$t(1,98) = 2.28^c$
Fever <38.5 °C	Mild (93.9)	Mild (69.4)	$t(1,98) = 3.91^a$
Small pink rash on arms	Mild (62.6)	Moderate (56.7)	$t(1,98) = 3.37^b$
Fever >38.5 °C	Moderate (81.8)	Moderate (74.6)	$t(1,98) = 4.06^a$
Vomiting and diarrhoea	Moderate (74.7)	Moderate (71.6)	$t(1,98) = 4.78^a$
Joint pain	Moderate (69.7)	Moderate (66.4)	$t(1,98) = 2.24^c$
Blotchy red rash all over body	Moderate (66.7)	Moderate (67.2)	$t(1,98) = 4.64^a$
Brain damage	Severe (100.0)	Severe (97.8)	$t(1,98) = -1.68$
Death	Severe (99.0)	Severe (99.3)	$t(1,98) = 1.00$
Paralysis	Severe (97.0)	Severe (99.3)	$t(1,98) = 1.78$
Tissue swelling inhibiting breathing	Severe (81.8)	Severe (83.6)	$t(1,98) = 1.30$
Seizure/convulsion	Severe (65.7)	Severe (73.1)	$t(1,98) = 2.00^c$
Duration ($n = 97$)			
Short-term minimum	24 h (43.3)	< 5 min (42.6)	$t(1,96) = -7.68^a$
Short-term maximum	48 h (43.3)	24 h (39.5)	$t(1,96) = -6.27^a$
Medium-term minimum	48 h (33.0)	24 h (29.5)	$t(1,96) = -5.90^a$
Medium-term maximum	1 month (44.3)	1 week (36.4)	$t(1,96) = -3.13^b$
Long-term minimum	1 month (37.1)	1 year (22.5)	$t(1,96) = -3.93^a$
Long-term maximum	Lifelong (83.5)	Lifelong (90.7)	$t(1,96) = 0.74$

[a] $p < 0.001$.
[b] $p < 0.01$.
[c] $p < 0.05$.

Figure 10: Participant Ratings of Disease and Vaccine Reaction Characteristics (Brown et al., 2009)

Baron and Ritov (2004) argue that, rather than an innate psychological bias present in every individual, the omission bias is more like a heuristic. Heuristics can be defined as rough rules-of -thumb that are applied to decision making when there is a limited amount of available information. This is supported by the high variance in the of the omission bias between different individual. In other words, some people rely far more heavily on the heuristic than others. They also argue that the directness of causality is crucial in determining the magnitude of the omission bias. That is, holding outcomes from acts of commission and omission equal, if the act of commission is perceived the directly cause the negative outcome, the omission bias will, on average, be higher.

3 DATA DISCUSSION

MLB's Statcast Pitch Data

Beginning with the introduction of PITCHf/x in 2007, the MLB has kept track of every pitch in every game of every season. In 2017, PITCHf/x was replaced by TrackMan, as part of Major League Baseball's Statcast platform. The MLB's Statcast platform, first introduced in a limited capacity in 2014 using PITCHf/x tracking, measures practically every facet of the game as it is being played, beyond just pitch location including a pitch's maximum velocity, spin rate, and movement, as well as statistics of batted balls like launch angle, exit velocity, and hit distance.

In 2020, TrackMan was replaced with Hawk-Eye, a camera system best known for its use in the instant replay system in professional tennis. The TrackMan, and Hawk-Eye systems are functionally the same in how they track pitch information. Several cameras are mounted around the ballpark pointed at home plate and the pitcher's mound (Figure 11). Despite marginal improvements in accuracy mainly on batted balls, the two systems gather the same information.

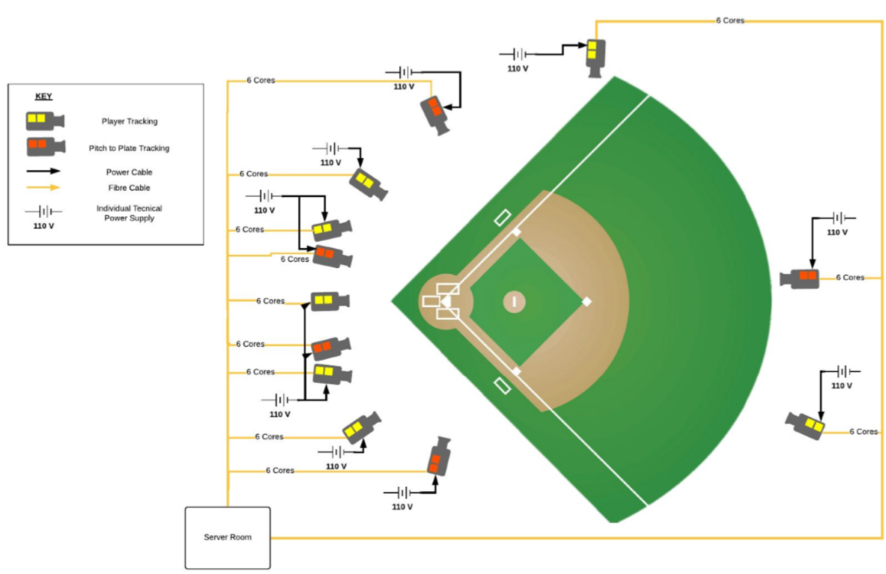

Figure 11: Typical Layout of Hawk-Eye Camera System (Jedlovec, 2020)

For my analysis, I will utilize Statcast's pitching data from 2018 to 2022. The full sample includes over 3.1 million pitches from 10,548 games over the five-season span. Since my analysis is concerned with umpire behavior, I eliminated any pitch for which the home plate umpire had no decision to make. Balls in play, foul balls, and swinging strikes, for example, do not require the home plate umpire to judge the pitch's location and determine whether it was a ball or a strike since the pitch was swung at. In total, batters swung at 47.13% of pitches in the full sample. Likewise, automatic/intentional balls, balls in the dirt, pitchouts, and hit-by-pitches were removed as well, since they also do not require a judgment call from the home plate umpire. These outcomes made up just 3% of the entire sample. The remaining pitches are called balls (33.38%) and called strikes (16.48%) which make up over 1.5 million pitches, slightly less than half the total pitches thrown (Table 1).

Pitch Outcome	*Freq.*	*Percent*	*Cum.*
Automatic Ball	10,690	0.34	0.34
Ball	1,050,198	33.38	33.72
Ball In Dirt	74,671	2.37	36.09
Called Strike	518,366	16.48	52.57
Foul	553,203	17.58	70.16
Foul Bunt	6,736	0.21	70.37
Foul Tip	28,900	0.92	71.29
Hit By Pitch	8,900	0.28	71.57
In play, no out	121,289	3.86	75.43
In play, out(s)	350,929	11.15	86.58
In play, run(s)	70,530	2.24	88.82
Missed Bunt	1,394	0.04	88.87
Pitchout	239	0.01	88.88
Swinging Strike	327,487	10.41	99.28
Swinging Strike (Blocked)	22,509	0.72	100.00
Total	3,146,041	100.00	

Table 1: Tabulation of Full Sample (2018-2022) by Pitch Outcome

Each pitch has a set of horizontal and vertical coordinates indicating where the center of the ball crosses home plate from the umpire's perspective, precise to the hundredth of an inch. The horizontal coordinates are measured relative to the center of the plate, which has a value of zero. For example, a pitch eight inches to the right of the center from the umpire's perspective would have a value of 8, while a pitch eight inches to the left of the center would have a value of -8. The vertical coordinates indicate the number of inches above the ground where the pitch crosses the plate. Figures 12 and 13 show the distribution of pitches by their horizontal and vertical coordinates, respectively. In addition to the two-dimensional coordinates, the MLB reports the release position of the pitch, the release speed, the spin rate and direction, and the ball's horizontal and vertical movement.

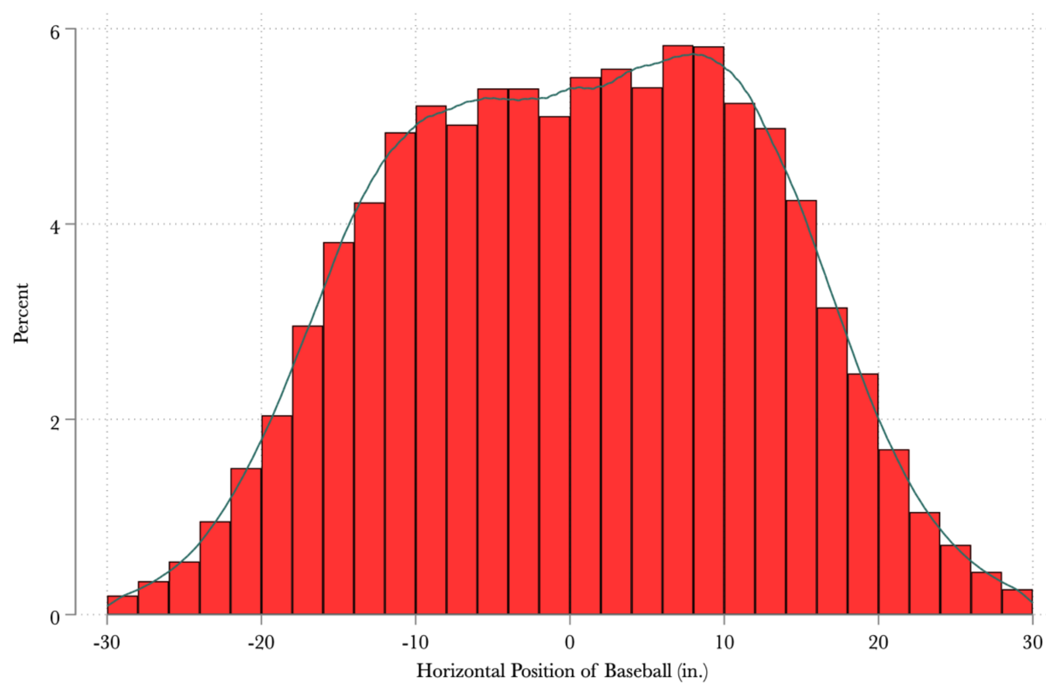

Figure 12: Histogram, Horizontal Position of Baseball (in.), 2-inch bins

The horizontal dimensions of the defined strike zone are fixed for every batter based on the 17-inch width of home plate. If any portion of the ball passes directly over home plate, it is, by rule, to be called a strike. Since the locational coordinates measure where the center of the ball passes home plate, I add the radius of a standard baseball, which is approximately 1.47 inches, to both the left and the right side of the strike zone to get the horizontal boundaries for the defined strike zone (Boyle, 2018). The total width of my defined strike zone is 19.94 inches, with a left boundary of -9.97 inches and a right boundary of 9.97 inches.

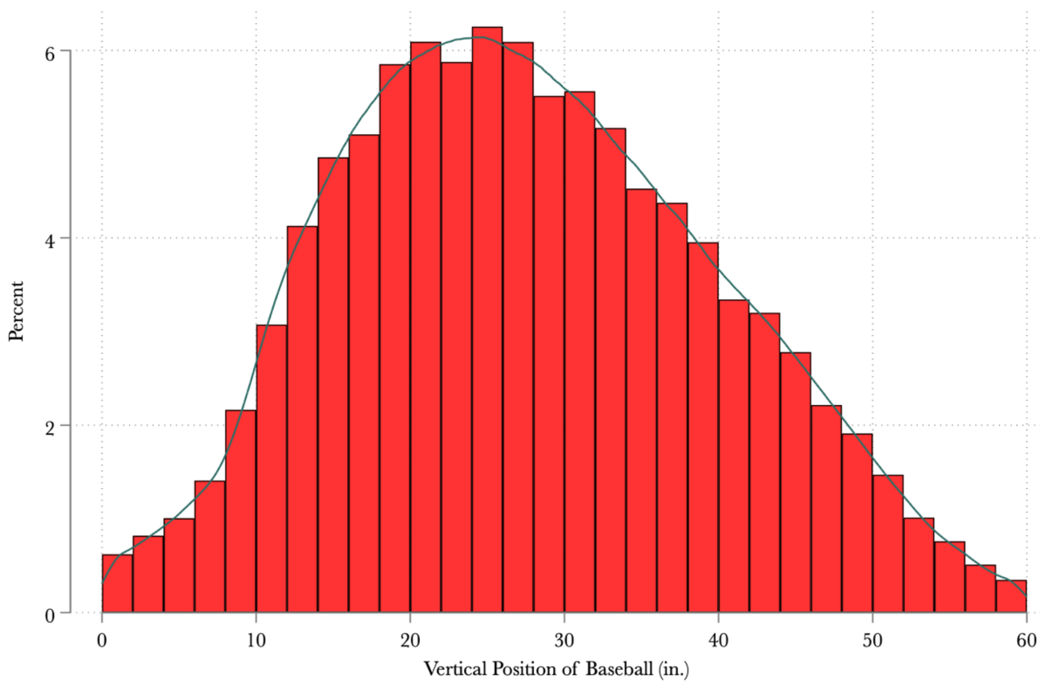

Figure 13: Histogram, Vertical Position of Baseball (in.), 2-inch bins

The vertical dimensions of the strike zone are slightly more complicated since they vary based on each batter's height and batting stance. Recall that, via the MLB's rulebook, the top boundary of the strike zone is "the midpoint between a batter's shoulders and the top of the uniform pants" and the bottom boundary is the "point just below the kneecap" (Figure 1). Both of these boundaries are set "when the batter is in his stance and prepared to swing at a pitched ball." Luckily, Statcast actually tracks these approximate boundaries for each pitch of each at-bat based on the batter's stance. The mean boundaries for the top and bottom of the strike zone are 40.6 inches and 19.2 inches respectively. In order to get a consistent strike zone for all batters, I standardized the vertical position of each pitch based on its distance from the middle portion of the defined strike zone. The top boundary of the standardized zone is 42 inches and the bottom boundary is 18 inches, making the standard center of the zone 30 inches high. To standardize the vertical position of each pitch, I first found the vertical midpoint of the pitch's captured strike zone. Depending on whether the pitch was thrown above or below this midpoint, I then calculated the distance from the midpoint relative to the nearest boundary. Using this calculation, I was then able to generate a vertical coordinate for every pitch relative to the standardized strike zone. Figure 14 plots the standardized locational data for all pitches from a single game relative to my defined strike zone, and Table 2 provides basic summary statistics for these coordinates.

Therefore, a pitch that has a standardized vertical point between 42 inches and 18 inches and a horizontal point between -9.97 inches and 9.97 inches is deemed a true strike. This designation is imperfect for a couple of reasons. First, the data itself are limited in that it measures in two dimensions, whereas, in reality, the strike zone has a third dimension of depth. Given that pitches typically have both horizontal and vertical movement and that a baseball has a diameter of about three inches, a ball could initially hit the plate outside of the zone but fully cross with a portion of the ball inside the zone. This scenario would technically be a strike that the data would not capture as such. Second, the tracking capabilities are quite good but not perfect. The current Hawk-Eye camera system boasts accuracy "within 0.25 inches on average"

(Jedlovec, 2020). With the system capturing both the point at which the ball crosses the plate as well as the upper and lower boundaries of each batter's strike zone, there is likely to be some level of measurement error.

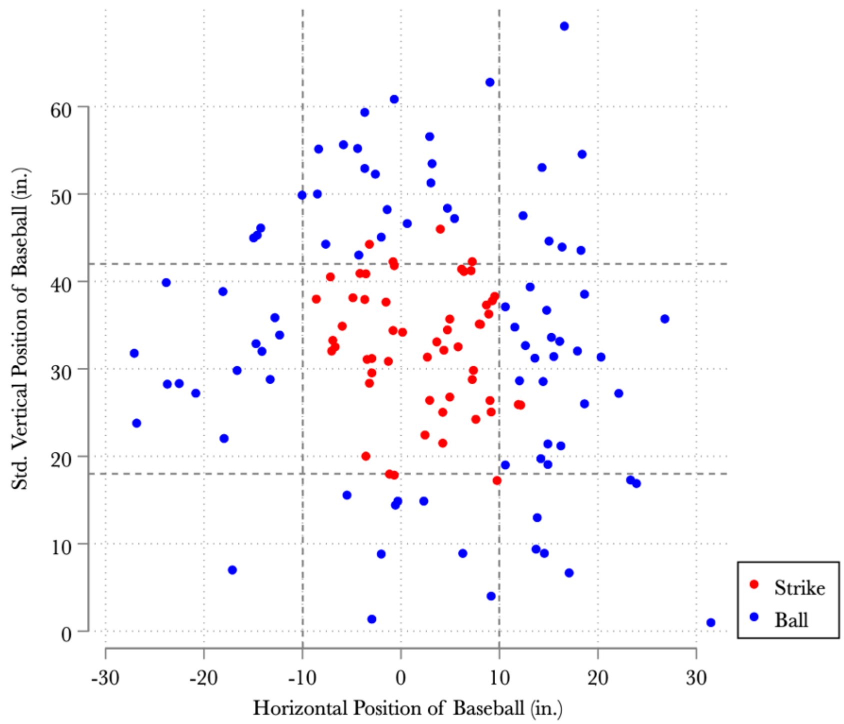

Figure 14: Called Pitch Locations for Baltimore Orioles vs. Boston Red Sox on May 30, 2022, Relative to Defined Strike Zone. Color Indicates Umpire's Call. The Defined Strike Zone Marked by the Dotted Lines. Does Not Include Pitches with Swings.

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Std. Vertical Position	1,511,978	28.244	13.919	0	158.06
Horizontal Position	1,511,978	.452	11.972	-116.04	109.32

Table 2: Summary Statistics for Horizontal and Standardized Vertical Coordinates

Table 3 provides summary statistics for three dummy variables indicating if the pitch was called a strike, if the pitch was inside the defined strike zone, and if the pitch was called correctly.

In total, 33.9% of pitches in the sample were called strikes, while just 28.9% were located in the defined strike zone. Umpires called just over 90% of pitches correctly over the five-year span. Further, Umpires called true strikes as strikes 78.9% of the time. On true balls, umpires called balls 96.7% of the time (Table 4). Umpires tend to be more accurate in calling balls since many pitches are well outside the zone and thus very clearly balls (Mills, 2017). Figure 15 shows the density of incorrect calls relative to the defined strike. The areas with the highest density of incorrect calls neighbor the border of the strike zone.

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
Strike Called = 1	1,518,110	.339	.473	0	1
Pitch In Strike Zone = 1	1,518,110	.289	.453	0	1
Pitch Correctly Called = 1	1,518,110	.907	.291	0	1

Table 3: Summary Statistics for if the Pitch was Called a Strike, if the Pitch was Inside the Defined Strike Zone, and if the Pitch was Called Correctly

	Umpire Called Pitch Correctly		
Umpire's Call	*Incorrect*	*Corrects*	*Total*
Ball	32,998	970,304	1,003,302
Strike	108,601	406,207	514,808
Total	141,599	1,376,511	1,518,110

Table 4: Cross Tabulation of Umpire's Call and Correctness of Call

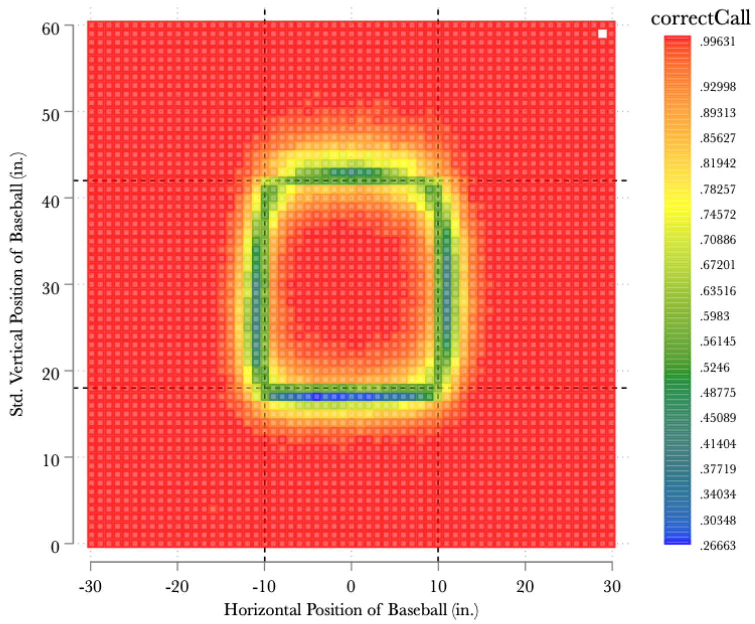

Figure 15: Percent of Pitches Called Correctly by Location

Table 5 shows the number of observations by the handedness of the batter, and Figure 16 displays the locational density of each pitch location based on horizontal and standardized vertical coordinates as well as the batters' handedness. A majority (58.19%) of pitches thrown in the sample are to right-handed batters. Further, the density plots for right and left-handed batters are near-mirror images of each other, as the lower, outside portion of the plate is the most populated for both groups of batters. This could imply that these are the most targeted zones by pitchers. Additionally, since the sample is limited to taken pitches, or pitches that the batter did not swing at, this could simply be a result of batters frequently opting not to swing at pitches in this portion of the zone.

Bat Side	*Freq.*	*Percent*	*Cum.*
Left	634,650	41.81	41.81
Right	883,460	58.19	100.00
Total	1,518,110	100.00	

Table 5: Tabulation of Sample by Batter Handedness

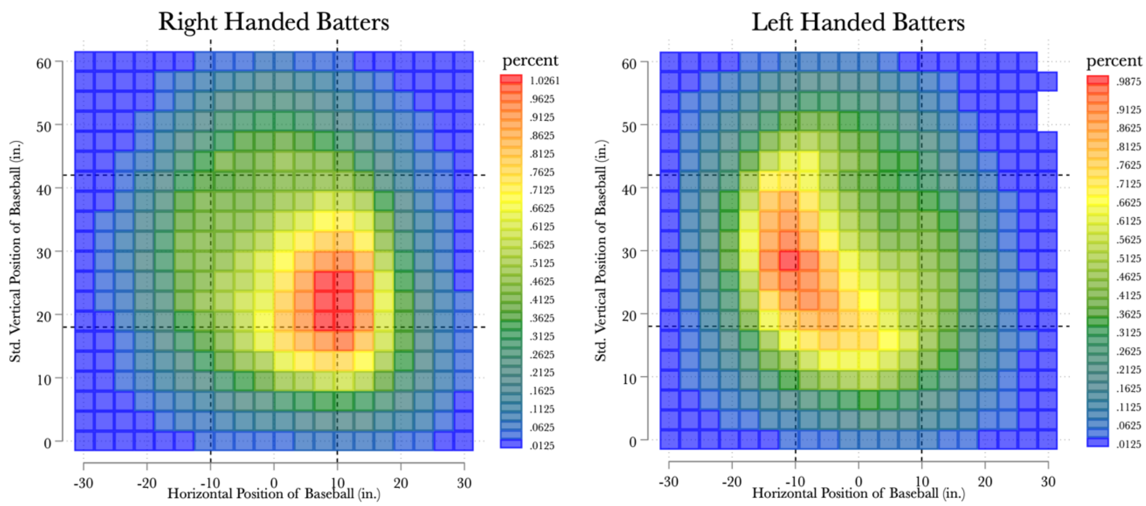

Figure 16: Density of Pitch Location by Batter Handedness. The Defined Strike Zone Marked by the Dotted Lines.

Situational Characteristics

Beyond tracking the paths of the ball and the player, Statcast also provides situational variables at the pitch, at-bat, inning, and game level. At the pitch level, most crucial to my analysis is the pre-pitch count, or the number of balls and strikes prior to the pitch being thrown. Table 6 provides the frequency at which each count occurs in the sample. In total, there are 12 different combinations of balls and strikes possible for any given pitch. Unsurprisingly, 0-0 counts (0 balls and 0 strikes) are the most common, making up 35.18% of the sample, since every at-bat must begin at this count. 3-0 counts are the least common and only make up 1.90% of the sample. 0-2 counts make up 5.48% of the sample. In total, 20.29% of pitches came with two strikes while just 6.43% of pitches came with three balls. Pitches on counts with 3 balls and fewer than 2 strikes make up 3.84% of the sample, while pitches on counts with 2 strikes and fewer than 3 balls make up 17.7% of the sample.

At-bat Count	*Freq.*	*Percent*	*Cum.*
0-0	534,143	35.18	35.18
0-1	188,661	12.43	47.61
0-2	83,170	5.48	53.09
1-0	170,019	11.20	64.29
1-1	135,303	8.91	73.20
1-2	106,689	7.03	80.23
2-0	58,134	3.83	84.06
2-1	65,516	4.32	88.38
2-2	78,852	5.19	93.57
3-0	28,806	1.90	95.47
3-1	29,433	1.94	97.41
3-2	39,384	2.59	100.00
Total	1,518,110	100.00	

Table 6: Tabulation of Pre-Pitch Ball and Strike Count (*# of balls - # of strikes)*

Like handedness, the pitch count tends to influence the locations of pitches in the sample (Figure 17). 0-2 pitches, for example, are most densely populated in the areas surrounding the strike zone. 3-0 pitches, on the other hand, are highly condensed within the strike zone. In theory, a pitch will aim for the strike zone on 3-0 counts to avoid walking the batter. In an 0-2 count, the pitcher will avoid the zone in an attempt to get the batter to swing at a pitch that is harder to hit.

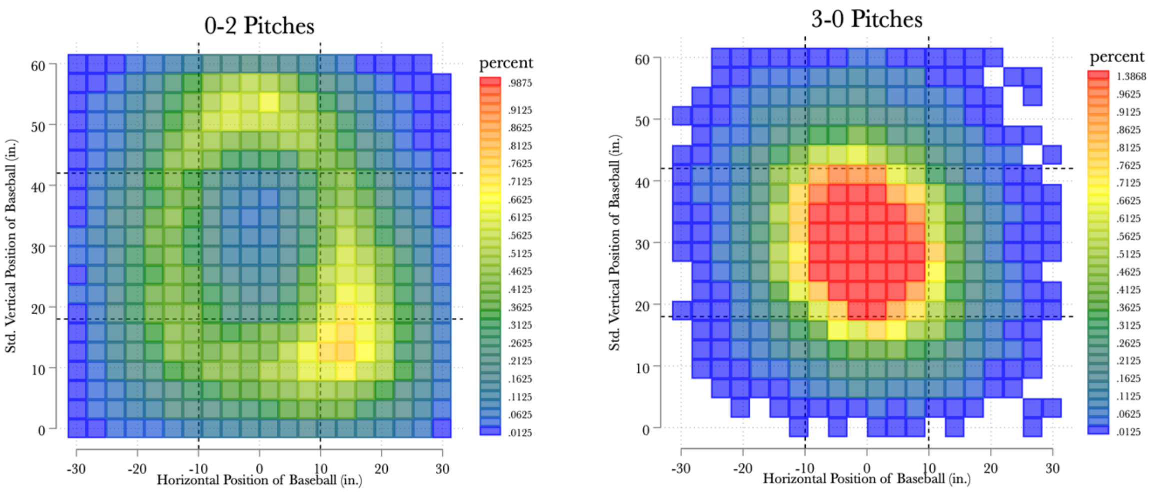

Figure 17: Density of Pitch Location by Pre-Pitch Count. The Defined Strike Zone Marked by the Dotted Lines.

At the at-bat level, the current inning, the number of outs, the runners on base, and the score contribute to the stakes of the in-game situation. All games in the sample have at least nine innings. In 2020 and 2021, when teams had to play two games in a single day (called a doubleheader), each game would only last seven innings. These games were dropped from the sample. If a game is tied after nine innings, the teams continue to play until one team leads after the completion of an inning. 98.46% of pitches in the sample come within the standard nine innings. Within each inning are half-innings. During the top half of the inning, the visiting team bats while the home team pitches. The reverse occurs in the bottom half of the inning. If the home team is leading after the top of the ninth inning, they win the game and do not bat in the bottom of the ninth. Since each batting team gets three outs per inning, the minimum number of pre-at-bat outs is zero and the maximum is two, and pitches per number of outs in the entire sample are roughly evenly distributed.

A common measure of each at-bat's importance towards the outcome of a game that incorporates the four aforementioned in-game discrete variables is the Leverage Index (LI) created by Tom Tango. The LI for an at-bat depends on four factors: the inning, the run differential, the number of outs, and the number and location of runners on base. Tango (2006) arrived at the following for the index's calculation:

> "You take the current base-out state, inning, and score and you find the possible changes in Win Expectancy that could occur during this particular plate appearance. Then you multiply those potential changes by the odds of that potential change occurring, add them up, and divide by the average potential swing in [win expectancy] to get the Leverage Index."

Simply put, the LI is a useful proxy for an at-bat's stakes. The higher the LI, the more important the moment. An LI of 1 is considered the average, neutral moment. The highest LI in the sample is 10.9 which occurred for 51 at-bats. All of these at-bats featured a one-run deficit for the home team in the bottom of the ninth (or later) with the bases loaded. Tango (2006) provides LI values for at-bats up to the ninth inning and with a run differential of four runs or fewer. Thus, at-bats beyond the ninth inning were treated as though they occurred in the ninth inning when assigning them an LI. Further, at-bats that occurred when the run differential was greater than four were assigned an LI of 0. The average LI in the five-year sample was 0.906, just below the neutral value of 1. 75% of at-bats had an LI below 1.2, and less than 10% of at-bats had an LI above 2 (Table 7).

	N	*Mean*	*SD*	*Min*	*p25*	*Median*	*p75*	*Max*
Leverage Index	1,516,770	.906	0.822	0	.4	.8	1.2	10.9
Number of Outs	1,518,110	.97	0.820	0	0	1	2	2
Inning	1,518,110	4.94	2.638	1	3	5	7	19
On First	1,518,110	.299	0.458	0	0	0	1	1
On Second	1,518,110	.162	0.368	0	0	0	0	1
On Third	1,518,110	.079	0.270	0	0	0	0	1

Table 7: Summary Statistics for At-bat level Situational Characteristics

The Statcast data includes the identity of the batter for every Pitch. For each batter, I gathered attributes illustrating their overall ability and perceived ability or their *star power*. The three measures of star power I use are (1) the batter's salary, (2) if the batter was selected to the all-star game, and (3) the batter's Batting Wins above Replacement (bWAR). These values hold constant for each batter during a single season, meaning, for example, that a single batter will have the same bWAR value in their first at-bat of a single season as in their final at-bat. Further, they are meant to be interpreted as proxies for star power. Table 8 and Table 9 provide summary statistics and a pairwise correlation matrix for these three variables, respectively. The three variables are all positively correlated with each other. Note that salary data is somewhat incomplete due to a lack of availability.

	N	*Mean*	*SD*	*Min*	*p25*	*Median*	*p75*	*Max*
All-Star	1,517,720	.138	0.345	0	0	0	0	1
Batting WAR	1,517,720	1.784	2.055	-3.848	.195	1.33	2.98	11
Salary (in millions)	1,038,819	7.634	7.937	.1	1.275	4.667	11.7	40

Table 8: Summary Statistics for All-Star Status, Batting WAR, and Yearly Salary (in millions)

Variables	*(1)*	*(2)*	*(3)*	*(4)*
(1) Strike Called	1.000			
(2) All-Star	-0.019	1.000		
(3) Batting WAR	-0.021	0.553	1.000	
(4) Salary (in millions)	-0.012	0.262	0.250	1.000

Table 9: Pairwise Correlation Matrix for Strike Called, All-Star Status, Batting WAR, and Yearly Salary (in millions)

All-Star team data was pulled from MLB.com for each of the five seasons. For each season, the nine starting hitters are chosen through fan voting, where the top vote-getter at each position is selected. Voting takes place between early June and early July of the season. The rest of the all-star team, the reserves, are selected using a combination of voting from fans, players, and the Commissioner's Office. There are 20 roster spots for position players (meaning players that bat) on each of the American League and National League teams, so 40 total players are all-stars each year. In 2020, despite the cancellation of the All-Star Game, the MLB released an unofficial starting lineup for the American and National League teams. Thus, only 18 batters are considered all-stars from that season. Overall, 13.8% of all pitches in the sample were thrown to all-star hitters. This number is positively skewed relative to the percent of all-stars in the total player population since all-stars tend to see more at-bats than non-all-stars.

Batting Wins Above Replacement, or bWAR aims to measure how many more wins a batter's hitting is worth than a "replacement-level" player. In short, it combines all elements of a batter's performance hitting the ball during a single season and estimates the number of wins that he is worth. For example, Aaron Judge has the highest bWAR in the sample for his 2022 season at 11. Thus, his hitting was estimated to be worth 11 games won. The statistic is not precise but aims to capture and summarize a hitter's total contribution to the season (Slowinski, 2010). Additionally, the statistic is standardized to a full 162-game season. A list of the top and bottom ten batters' seasons in the sample is included in the appendix (Table A1).

For each game, the MLB reports the total number of fans in attendance. As mentioned prior, for every game in the year 2020, no fans were in attendance due to the Covid-19 pandemic. Additionally, in 2021, several stadiums had capacity restrictions to start the season. By the end of July 2021, though, all stadiums were operating at full capacity. Still, 2021 had the lowest average attendance of any season other than 2020 in the sample due to the ongoing pandemic. Average attendance was at its highest in 2018 (Table 10).

year	*N*	*Mean*	*SD*	*Min*	*Max*
2018	353,370	28,910.867	10,838.495	2,429	56,310
2019	354,256	28,389.787	11,220.081	2,503	59,659
2020	130,450	0	0.000	0	0
2021	340,836	18,890.912	10,898.923	1,107	53,114
2022	339,198	26,905.319	11,293.918	2,467	53,432

Table 10: Summary Statistics for Attendance by Year

Lastly, Statcast includes the identity of the home plate umpire for each game. The sample includes games called by 114 different umpires, and the umpires have varying degrees of total pitches called and call accuracy. There is a 4.3 percentage point difference between the accuracy of the "best" umpire (Alex MacKay, 93.10% accurate) and the "worst" umpire (Ted Barrett, 88.80% accurate) (Table 11).

Rank	*Home Plate Umpire*	*N*	*% Correct*
1	Alex MacKay	857	93.10%
2	Brock Ballou	1,016	92.50%
3	John Libka	16,945	92.40%
4	Adam Beck	8,853	92.20%
5	Lew Williams	578	92.20%
6	Pat Hoberg	18,622	92.00%
7	Edwin Moscoso	12,354	91.90%
8	Alex Tosi	6,238	91.80%
9	Jansen Visconti	19,272	91.80%
10	Alan Porter	18,473	91.70%
…	…	…	…
105	Paul Nauert	9,443	89.60%
106	Rob Drake	15,385	89.60%
107	Kerwin Danley	10,119	89.50%
108	Brian Gorman	8,066	89.40%
109	Laz Diaz	17,967	89.30%
110	Ed Hickox	15,265	89.20%
111	Joe West	14,662	89.20%
112	Doug Eddings	18,638	89.10%
113	Mike Winters	6,480	89.10%
114	Ted Barrett	18,936	88.80%

Table 11: Most Accurate and Least Accurate Umpires

4 METHODOLOGY

Baseline Probability Estimation

Before trying to investigate the existence of the omission bias, I calculate the baseline probability of a pitch being called a strike based on location and batter handedness alone. To do so, I select a 40-inch by 40-inch vertical plane directly above the front of home plate spanning vertically from 10 inches above the ground to 50 inches above the ground and horizontally from 20 inches to the left of the center of home plate to 20 inches to the right of home plate. I then divide this plane into 1,600 1-inch by 1-inch squares. I repeat this process for pitches to left-handed batters and pitches to right-handed batters to create 3,200 location- and batter-handedness-specific zones. There are no called strikes in any location outside of these recognized zones in the sample. From there, I calculate the percentage of pitches that were called a strike within each of these zones. In this calculation, I include only pitches from neutral counts and thus exclude pitches that occurred during counts with two strikes and fewer than three balls as well as counts with three balls and fewer than two strikes. If three-ball and two-strike counts do, in fact, create umpire bias, then the baseline probability would also have contained this bias had I included these counts. On average, fewer than 400 pitches occupy each locational zone. Thus, to create a more robust estimate for the baseline probability that a pitch is called a strike, I use a simple weighted average including the zones directly surrounding the exact location of the pitch (Equation 1):

$$(1)\quad Prob(Strike_i) = \frac{1}{16}[4 * \bar{X}_i + 2 * (\bar{X}_L + \bar{X}_R + \bar{X}_U + \bar{X}_D) + 1 * (\bar{X}_{UR} + \bar{X}_{DR} + \bar{X}_{UL} + \bar{X}_{DL})]$$

where, for pitches receive by batters of the same handedness, $\overline{X}_i$ is the percentage of strikes called in the exact locational zone of the pitch. Further, $\overline{X}_L$, $\overline{X}_R$, $\overline{X}_U$, and $\overline{X}_D$ represent the percentage of strikes called in the locational zone directly to the left, to the right, underneath, and above the exact zone, respectively. Likewise, $\overline{X}_{UR}$, $\overline{X}_{DR}$, $\overline{X}_{UL}$, and $\overline{X}_{DL}$ represent the percentage of strikes called in the locational zone up and to the right, down and to the right, up and to the left, and down and to the left of the exact zone, respectively. Figure 18 provides a visual representation of the weights used in calculating the baseline probability of a strike being called based on location and handedness.

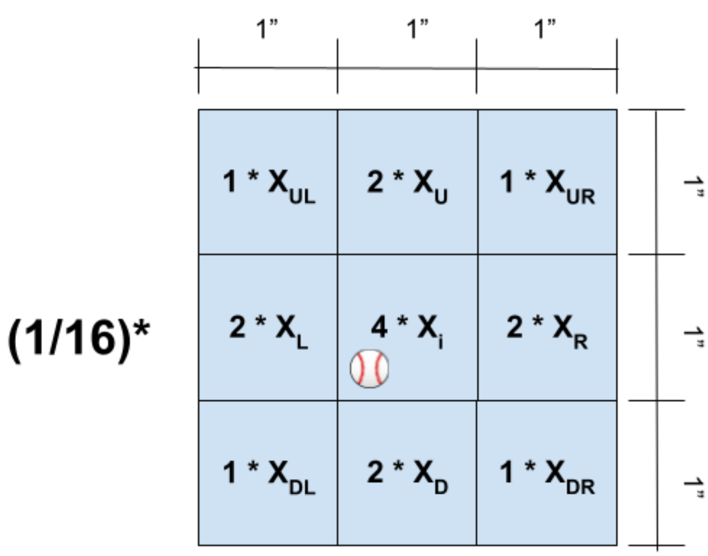

Figure 18: Visual Representation of Weights Used for Baseline Probability Estimation

Figure 19 illustrates the value of the baseline probability by location and batter handedness. Notably, the enforced strike zone for left- and right-handed batters differs significantly along the edges of the defined strike zone. For left-handed batters, the edge of the plate surrounding -9.97 inches is more densely populated with called strikes than it is for right-handed batters. This region represents the outside edge of the zone for lefties and the inside edge for righties. Further, the reverse is true on the opposite edge of the zone surrounding 9.97 inches from the center of the plate. Thus, the enforced strike zone is, in general, skewed towards the outside part of the zone, or the edge furthest from the batter, and fewer strikes are called on the inside part of the zone, or the edge closest to the batter.

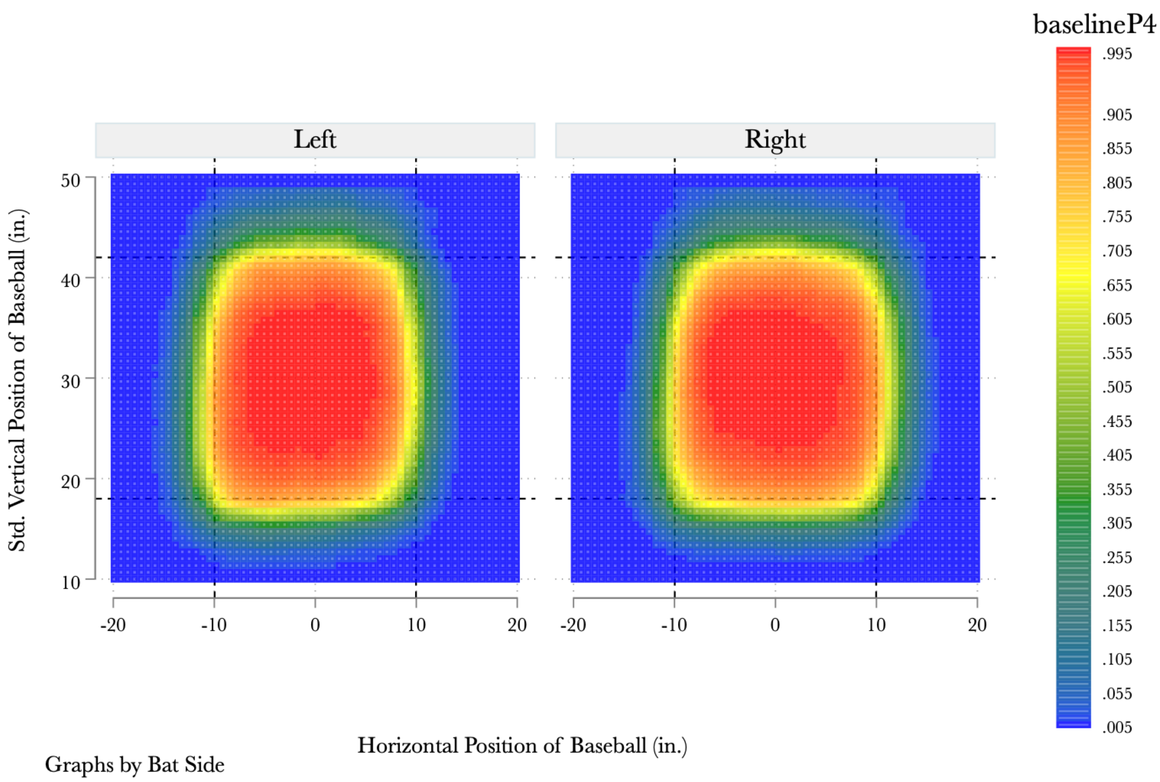

Figure 19: The Baseline Probability of a Called Strike based on Pitch Location and Batter Handedness

Over 50% of all pitches in the sample have a baseline probability of less than 10% that they are called a strike. Additionally, over 20% of all pitches have a probability of greater than

90% that they are called a strike (Figure 20). Hence, a majority of pitches in the sample are either near-certain balls or near-certain strikes based on location alone.

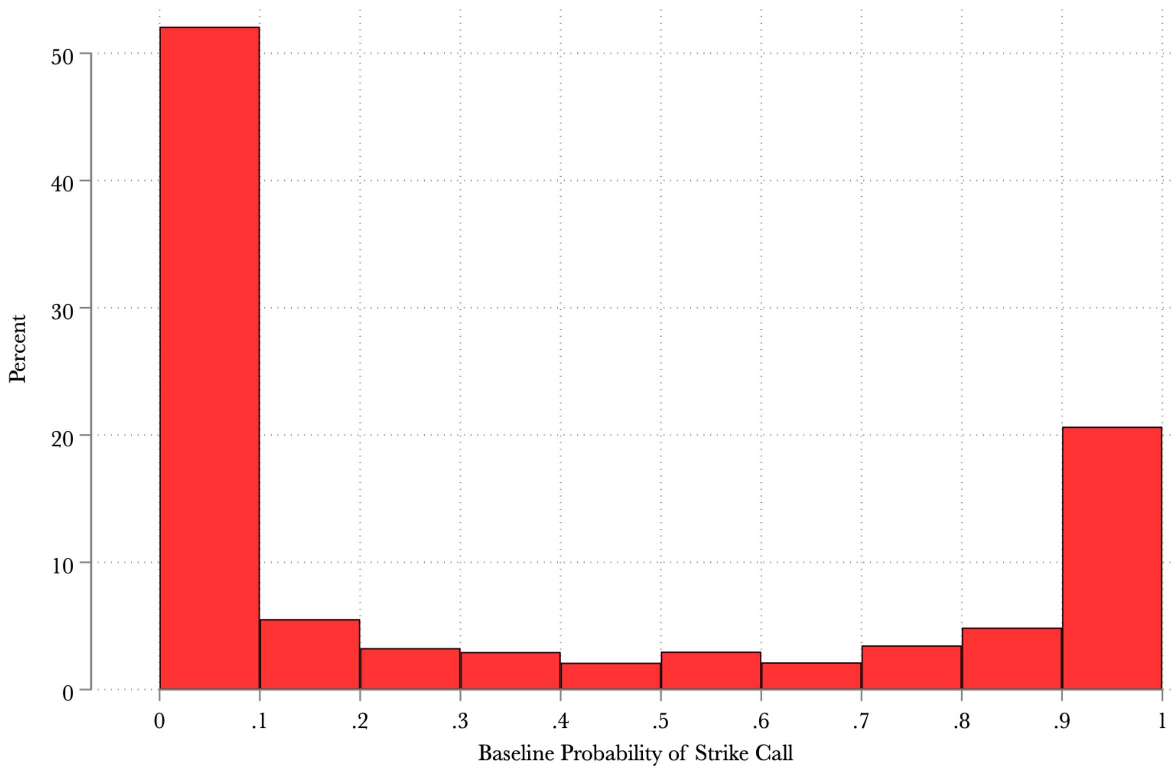

Figure 20: Histogram of Baseline Probability of Strike Call

Empirical Strategy

In order to investigate the impact of the omission bias on MLB umpires, I use a linear probability model to measure the distortion of the strike zone during two-strike and three-ball counts:

Equation 2: $$p(y_i = 1) = p_i + X_i\beta_1 + w(p_i) * X_i\beta_2 + Z_i\beta_3 + w(p_i) * Z_i\beta_4 + \epsilon_i$$

where y_i is an indicator for if the pitch is called a strike, p_i is the baseline probability estimated using Equation 1, X_i is an indicator for counts with 2 strikes. (excluding full counts), and Z_i is an indicator for counts with 3 balls (excluding full counts). I use a linear probability model as opposed to logit or probit models due to my large sample size. To account for changes in the omission bias relative to the uncertainty of the call, I weight the call indicators by w, a function of the baseline probability (Green and Daniels, 2015), where:

Equation 3: $$w(p_i) = 1 - 2|p_i - 0.5|$$

and whose value I will refer to as the Borderline Index for a given pitch. For the most borderline pitches, or pitches with a baseline strike probability of 50%, the Borderline Index equals 1. Pitches with baseline strike probabilities of 0% or 100%, on the other hand, will have a Borderline Index of 0. The more borderline a pitch is, meaning the closer its baseline strike probability is to 50%, the closer its Borderline Index value is to 1, and all pitches have a Borderline Index value between 0 and 1. Thus, $\beta_1 + \beta_2$ can be interpreted as the change in probability of a called strike on a two-strike count for a pitch that is otherwise called a strike half of the time. Likewise, $\beta_3 + \beta_4$ can be interpreted as the change in probability of a called strike on a three-ball count for a pitch that is otherwise called a strike half of the time. To control for the asymmetry in umpire skill and behavior (Table 11), I also utilize umpire fixed effects. Further, I incorporate a vector of pitch-level controls to account for the potentiality of increased umpire uncertainty caused by different types of pitches. The control variables included are the pitch's spin rate, horizontal movement, vertical movement, horizontal (x-dimension) velocity, vertical (z-dimension) velocity, and velocity towards the batter (y-dimension).

Next, I estimate the changes to the magnitude of the omission bias caused by immediate external pressure using the following equation:

$$\text{Equation 4:} \quad p(y_i = 1) = p_i + X_i\beta_1 + H_i\beta_2 + \mathrm{A_i}\beta_3 + (X_i * H_i)\beta_4$$
$$+[X_i * H_i * \mathrm{A_i}]\beta_5 + \epsilon_i$$

where p_i is the baseline probability estimate, X_i is a count-specific indicator, H_i is an indicator for if the batter is on the home team, and A_i is the total attendance. Here, the models are specific to the effects on either two-strike or three-ball counts. Additionally, the sample is limited to pitches with a baseline strike probability between 25% and 75% (Figure 21). The model is run separately for pitches from the year 2020. In these iterations, β_3 and β_5 are dropped, since attendance always equals zero during that season. In theory, the effects of the crowd on the umpire will only be realized in favor of the home batter or against the away batter (Bilalić, Gula, and Vaci, 2021). Thus, I use interaction terms to separate the home and away effects. Similar to before, I incorporate umpire fixed effects and pitch-characteristic controls into my estimation.

Lastly, to estimate the change in the omission bias caused by in-game characteristics, I use the following:

$$\text{Equation 5:} \quad p(y_i = 1) = p_i + X_i\beta_1 + S_i\beta_2 + (\mathrm{X_i} * \mathrm{S_i})\beta_3 + \epsilon_i$$

where p_i is the baseline probability estimate, X_i is a count-specific indicator, and S_i is a situational characteristic. I repeat this model for three-ball counts and two-strike counts, and for the following situational characteristics: Batter is an All-Star, Batter's Salary, Batter's Wins Above Replacement, and at-bat's Leverage Index. The first three characteristics aim to capture the importance of the batter's perceived star power on the umpire's behavior, whereas the final characteristic aims to capture the influence of the relative stakes of an at-bat on the umpire's behavior. Again, the sample is limited to pitches with a baseline strike probability between 25% and 75% (Figure 21), and I incorporate umpire fixed effects and pitch-characteristic controls into my estimation.

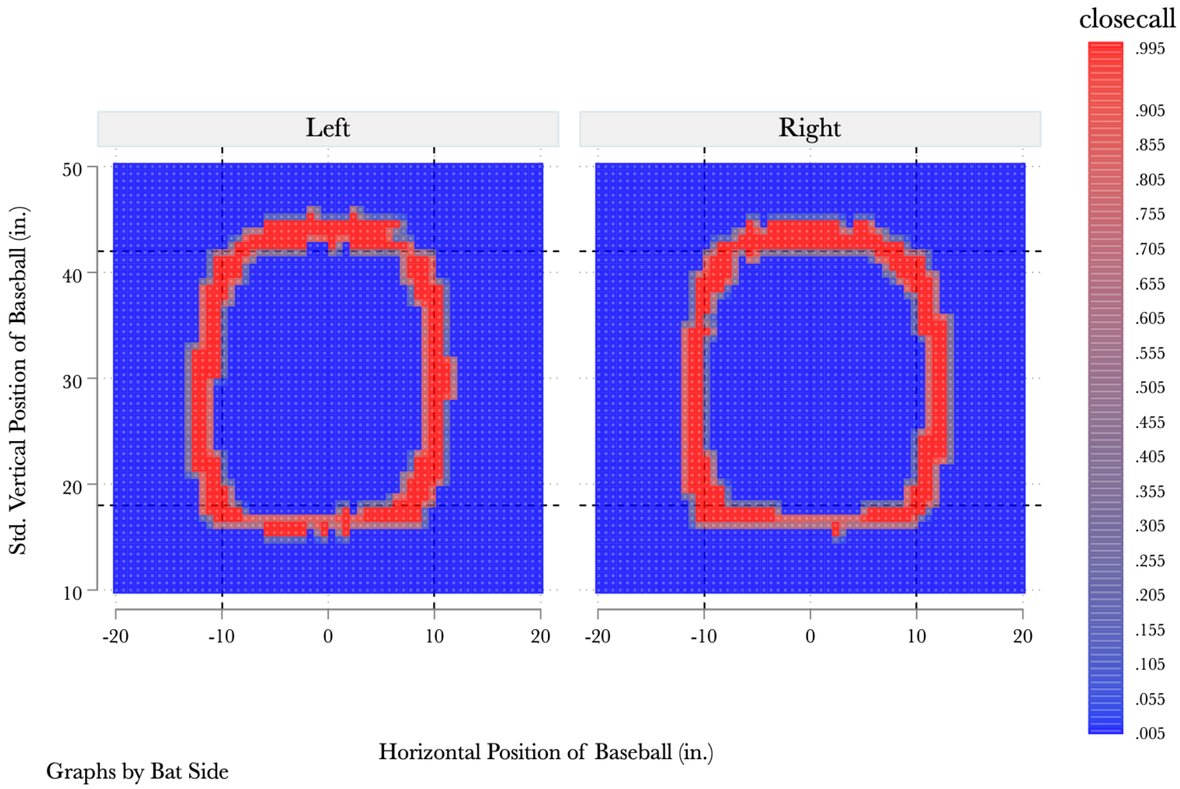

Figure 21: Zones With a Baseline Strike Probability Between 25% and 75% (in Red)

5 RESULTS

Existence of the Omission Bias

Table 12 provides the results for the first set of regressions aimed at investigating the existence of omission bias in MLB umpires. Models (1.1-6) build towards Model (1.7), the full model presented in Equations Three. Overall, umpires are, on average, 2.7% less likely to call a strike on a two-strike count than on a neutral count and 2.8% more likely to call a strike on a three-ball ball count. Both of these estimates are statistically significant from zero at a 99% confidence level.

Models (1.2) and (1.5) limit the sample to pitches with a baseline strike probability between 25% and 75%. In this subsample, the estimated size of the omission bias grows for both two-strike and three-ball counts. On average, umpires are 13.2% less likely to call pitches within these zones as strikes on two-strike counts than they are on neutral counts and 8.4% more likely to call pitches within these zones as balls on three-strike counts. Again, both of these estimates are statistically significant from zero at a 99% confidence level.

Model (1.7) is simply just a combination of models (1.3) and (1.6), and the coefficients do not change substantially between these models. Using the full sample and controlling for pitch-characteristics and umpire asymmetry, I find that a pitch that is typically called a strike 50% of the time during a neutral count is called a strike 31.3% of the time when the pitch occurs during a two-strike count. Further, I find that a pitch that is typically called a strike 50% of the time during a neutral is called a strike 60.2% of the time when the pitch occurs during a three-ball count. Figure 22 illustrates the average enforced strike zone for three-ball and two-strike counts. Figure 23 illustrates the difference between the strike during three-ball counts and two-strike counts versus neutral counts.

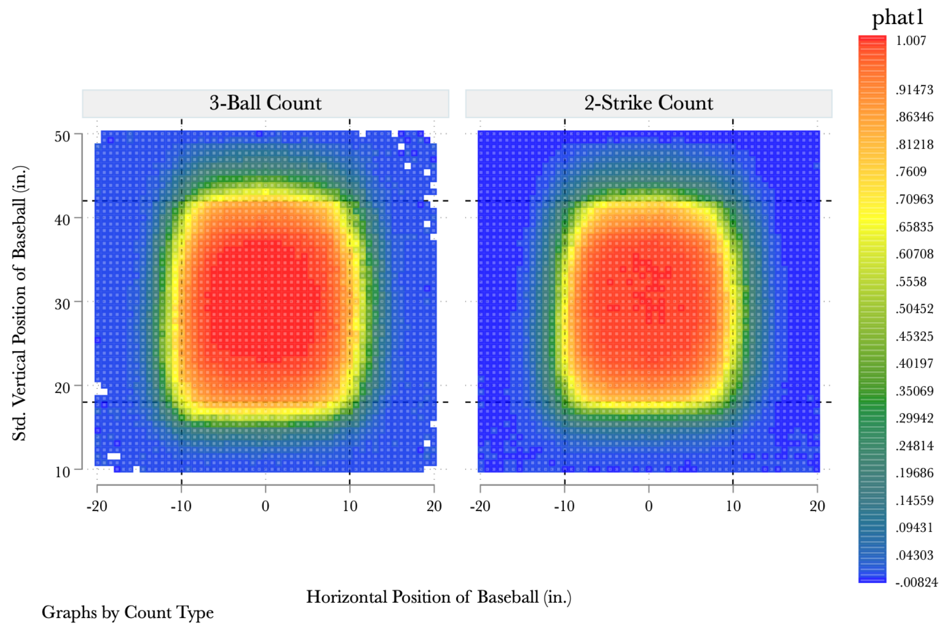

Figure 22: Predicted Strike Zone For 3-Ball Counts and 2-Strike Count, Model (1.7). Results are Averaged by Handedness.

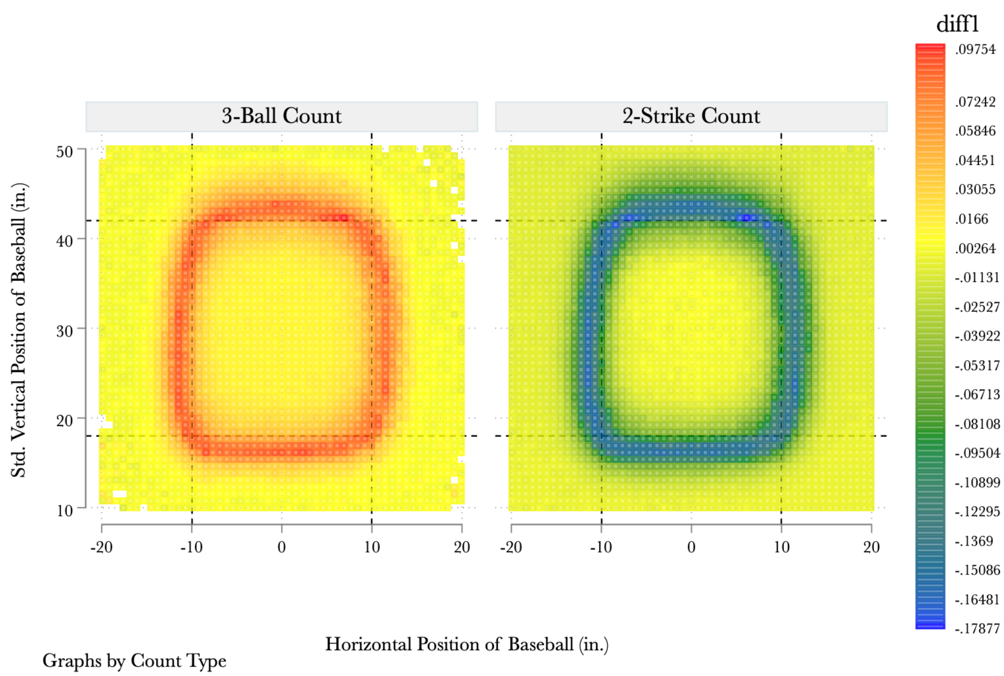

Figure 23: Difference in Predicted Strike Zone For 3-Ball Counts vs. Neutral Counts and 2-Strike Counts vs. Neutral Counts, Model (1.7). Results are Averaged by Handedness.

Influence of Immediate External Pressure

Table 13 contains the results for the next set of models aimed at measuring the impact of immediate external pressure from fans in the stadium on strike calls, irrespective of the pitch count, but still controlling for umpire asymmetry and pitch characteristics. Model (2.1) finds that, for pitches with a baseline strike probability between 25% and 75% in all years of the sample except for 2020, umpires are 0.9% less likely to call a strike if the batter is on the home team. This effect disappears for pitches thrown during the shortened, fan-less 2020 season, as shown in Model (2.2).

Model (2.3) incorporates total fan attendance, finding that, as the home crowd grows, so does the disparity in how umpires call strikes (Figure 24). When there are fewer than 20 thousand fans in attendance, I estimate no significant difference between the likelihood of a called strike for a home batter and a visiting batter. As the number of fans in attendance exceeds 25 thousand, though, I find that the likelihood of a called strike differs for home batters and away batters significantly at a 95% confidence level. Still, when there are 50 thousand fans in attendance, I estimate that umpires are only 3% less likely to call a strike for home batters than they are for away batters, holding all else equal.

Table 14 contains the results for the models related to the effect of immediate external pressure on the omission bias. All three models are limited to pitches with a baseline strike probability between 25% and 75% and incorporate umpire fixed effects and pitch-characteristic controls. Model (2.4) shows no statistically significant impact of home-field advantage on the omission bias for either two-strike counts or three-ball counts for all seasons except 2020. Unsurprisingly, Model (2.5) likewise shows no statistically significant impact of home-field advantage on the omission bias.

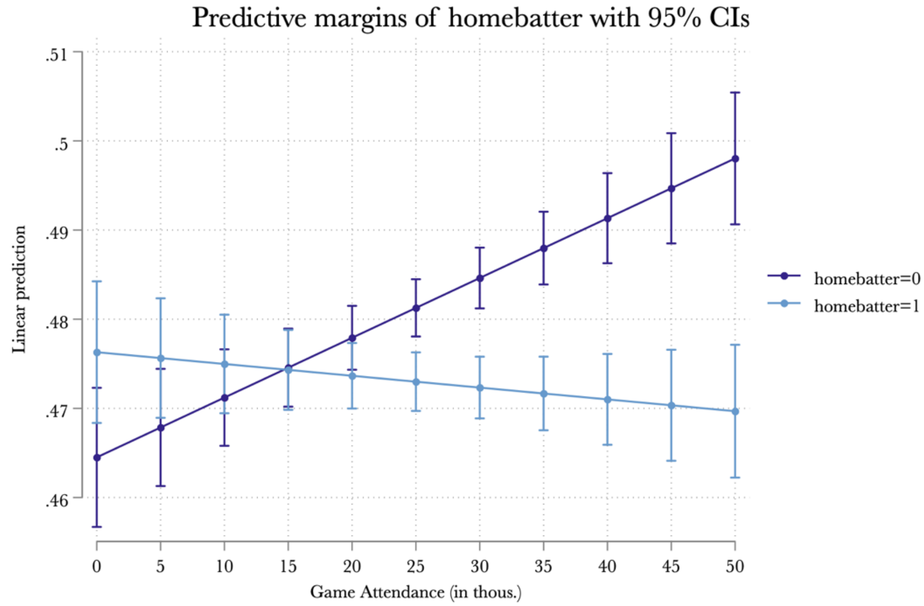

Figure 24: Effects of Game Attendance on Strike-Call Probability, Home vs. Away

Model (2.6), represented by Equation 4, incorporates total attendance and finds no evidence that home-field advantage influences the magnitude of umpires' omission bias during three-ball counts, even as the number of fans in attendance increases. I estimate that home attendance increases the magnitude of the omission bias on two-ball counts, reducing the probability of a called strike for the home batter by 0.1% for every thousand fans, on average. This effect is significant at a 90% confidence level, but lacks any practical significance in magnitude.

Influence of Situational Characteristics

Table 15 contains the results for the set of models aimed at measuring the impact of situational characteristics on strike calls, irrespective of the pitch count, but still controlling for umpire asymmetry and pitch characteristics. For pitches with baseline strike probabilities between 25% and 75%, all three proxies for a batter's perceived star power have negative coefficients and are statistically significant at the 95% level or higher. Called strikes are, on average, 1.3% less likely for all-star batters, 0.1% likely for every million-dollar increase in a batter's salary, and 0.1% less like for every full-point increase in Batter Win Above Replacement. Likewise, I estimate that for every full-point increase in an at-bat's Leverage Index, the probability of a strike falls by 0.2%.

Table 16 contains estimates for the influence of selected situational characteristics on umpires' omission bias during two-strike counts. I find that the batter's all-star status, the batter's salary, and the at-bat's leverage index are not associated with a significant change in the magnitude of the umpire's omission bias for two-strike counts. Every full-point increase in the batter's bWAR, on the other hand, is associated with a -0.3% decrease in the likelihood of a called strike, all else equal.

Table 17 contains estimates for the influence of selected situation characteristics on umpires' omission bias during three-ball counts. I find that the batter's all-star status, the batter's salary, the batter's bWAR, and the at-bat's leverage index are not associated with a significant change in the magnitude of the umpire's omission bias for two-strike counts.

	(1.1)	(1.2)	(1.3)	(1.4)	(1.5)	(1.6)	(1.7)
	strike	*strike*	*strike*	*strike*	*strike*	*strike*	*strike*
Two-Strike Count	-.027***	-.132***	-.003***				-.003***
	(.001)	(.003)	(.001)				(.001)
2 Strikes * Borderline Index			-.185***				-.184***
			(.002)				(.002)
Three-Ball Count				.028***	.084***	.012***	.007***
				(.001)	(.005)	(.001)	(.001)
3 Balls * Borderline Index						.095***	.095***
						(.004)	(.004)
Observations	1518110	206493	1499257	1518110	206493	1499257	1499257
R-squared	.746	.129	.753	.746	.123	.751	.753
Controls & Ump FE			yes			yes	yes

Standard errors are in parentheses
**** $p<.01$, ** $p<.05$, * $p<.1$*

Table 12: Results, Part 1. Models (1.1) and (1.4) show overall estimate of omission bias for all two-strike and all three-ball pitches respectively. Models (1.2) and (1.5) show overall estimate of omission bias for all two-strike and all three-ball pitches with a baseline strike probability ≥0.25 and ≤0.75, respectively. Models (1.3) and (1.6) introduce Borderline Index weights, Umpire Fixed Effect, and Pitch-Characteristic Controls to full sample estimation. Model (1.7) is full model found in Equation 2.

	(2.1)	(2.2)	(2,3)
	strike	*strike*	*strike*
Home Batter	-.009***	.001	.012**
	(.002)	(.008)	(.006)
Game Attendance (in thous.)			.001***
			(0)
Home Batter * Attend.			-.001***
			(0)
Observations	163735	15471	163735
R-squared	.105	.111	.106
Covid Year (2020)	no	yes	no

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 13: Results, Part 2. Models (2.1) and (2.2) exclude pitches from 2020. Model (2.2) is limited to pitches from 2020. Models (2.1), (2.2), and (2.3) include only pitches with a baseline strike probability ≥0.25 and ≤0.75 and incorporate Umpire Fixed Effect and Pitch-Characteristic Controls.

	(2.4) *strike*	(2.5) *strike*	(2.6) *strike*
Home Batter	-.009***	.001	-.009***
	(.003)	(.008)	(.003)
Two-Strike Count	-.122***	-.137***	-.122***
	(.005)	(.017)	(.005)
2-Strike Count * Home Batter	-.005	-.008	.014
	(.007)	(.024)	(.013)
Three-Ball Count	.07***	.077***	.07***
	(.008)	(.025)	(.008)
3-Ball Count * Home Batter	.009	.016	.008
	(.011)	(.036)	(.02)
Game Attendance (in thous.)			.0003***
			(0)
3-Ball Count * Home Batter * Attend.			0
			(.001)
2-Strike Count * Home Batter * Attend.			-.001*
			(0)
Observations	163735	15471	163735
R-squared	.113	.121	.113
Covid Year (2020)	no	yes	no

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 14: Results, Part 2 (cont.). All Models are limited to pitches with a baseline strike probability ≥0.25 and ≤0.75. Models (2.4) and (2.6) exclude pitches from 2020. Model (2.5) is limited to pitches from 2020. All Models include only pitches with a baseline strike probability ≥0.25 and ≤0.75 and incorporate Umpire Fixed Effect, and Pitch-Characteristic Controls

	(1) *strike*	(2) *strike*	(3) *strike*	(4) *strike*
All Star	-.013***			
	(.003)			
Salary (in mil.)		-.001***		
		(0)		
Batting WAR (per 162 games)			-.001**	
			(.001)	
Leverage Index				-.002*
				(.001)
Observations	204903	139810	204903	204751
R-squared	.129	.13	.129	.129

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 15: Results, Part 3. Effect of Situational Characteristics on Called Strike Probability. Model includes only pitches with a baseline strike probability ≥0.25 and ≤0.75 and incorporate Umpire Fixed Effect and Pitch-Characteristic Controls.

	(1)	(2)	(3)	(4)
	strike	*strike*	*strike*	*strike*
All Star	-.013***			
	(.003)			
Two-Strike Count	-.128***	-.135***	-.124***	-.131***
	(.003)	(.005)	(.004)	(.005)
2 Strikes * All Star	-.012			
	(.01)			
Salary (in mil.)		-.001***		
		(0)		
2 Strikes * salary		0		
		(0)		
Batting WAR (per 162 games)			-.001**	
			(.001)	
2 Strikes * bWAR			-.003**	
			(.002)	
Leverage Index				-.003**
				(.001)
2 Strikes * LI				.002
				(.004)
Observations	204903	139810	204903	204751
R-squared	.136	.137	.136	.136

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 16: Results, Part 3 (cont.). Effects of Situational Characteristics on Omission Bias for Two-Strike Counts. All Models include only pitches with baseline strike probability ≥0.25 and ≤0.75 and incorporate Umpire Fixed Effect and Pitch-Characteristic Controls.

	(1) *strike*	(2) *strike*	(3) *strike*	(4) *strike*
All Star	-.014***			
	(.003)			
Three-Ball Count	.09***	.087***	.092***	.087***
	(.005)	(.008)	(.007)	(.008)
3 Balls * All Star	.01			
	(.014)			
Salary (in mil.)		-.001***		
		(0)		
3 Balls * salary		0		
		(0)		
Batting WAR (per 162 games)			-.001**	
			(.001)	
3 Balls * bWAR			0	
			(.002)	
Leverage Index				-.003**
				(.001)
3 Balls * LI				.005
				(.006)
Observations	204903	139810	204903	204751
R-squared	.131	.131	.131	.131

Standard errors are in parentheses
**** $p<.01$, ** $p<.05$, * $p<.1$*

Table 17: Results, Part 3 (cont.). Effects of Situational Characteristics on Omission Bias for Three-Ball Counts. All Models include only pitches with baseline strike probability ≥0.25 and ≤0.75 and incorporate Umpire Fixed Effect and Pitch-Characteristic Controls.

6 Discussion

My results indicate that umpires significantly alter their behavior based on the ball-strike count. All else equal, on three-ball counts, umpires are more likely to call a strike, and, on two-strike counts, umpires are more likely to call a ball. In other words, umpires prefer inaction to action. Additionally, instances of omission bias are generally isolated to the borderline of the strike zone. Umpires remain fairly consistent on pitches well within or outside of the strike zone regardless of the count. When the pitch is near the borderline, though, omission bias has a sizable impact on umpires' calls. Intuitively, this result makes sense. On the most obvious pitches, umpires have little to no discretion. Errors on clear balls or strikes will be salient enough on their own, regardless of the ball-strike count. Further, these pitches are easier to call correctly, so there is little need to rely on an additional rule of thumb. On the other hand, pitches located around the border of the strike zone allow for more umpire discretion, given the high variability in umpires' enforced strike zone in these areas. For neutral counts, errors on these pitches are much less salient than errors on pitches well within or outside the zone. It is only when the error leads to an action, either a strike-out or a walk, that errors of borderline pitches become salient. Thus, when umpires have some magnitude of discretion, they generally prefer inaction to action. Likewise, if the umpire is truly unsure of the correct call, they may rely on their aversion to impact as a rule of thumb when deciding what to call. The closer the pitch is to the border of the strike zone, the more likely that the umpire is uncertain of the correct call.

Ignoring their impact on umpires' omission bias, my results indicate a slight bias in favor of home batters, or a home-field advantage, on borderline pitches overall. Umpires were slightly less likely to call borderline pitches as strikes for home batters than they were for away batters. This advantage disappeared during the 2020 season, when fans were not in attendance. Further, home-field advantage via a reduction in called strikes increases when more fans were in attendance. Nevertheless, the actual magnitude of the advantage given to home batters by umpires is rather small, as there was less than a 1% difference in the likelihood of a strike call on borderline pitches for home and away batters for games with an average attendance. Beyond this initial advantage, there is no further advantage granted to home batters on two-strike or three-ball counts. On three-ball and two-strike counts, umpires distort the strike zone similarly for both home and away teams. My results indicate that, perhaps, the strike zone shrinks slightly more for home batters on two-strike counts as attendance increases. Still, the practical significance of this additional distortion is little to none.

Similarly, my results indicate that star players get slightly favorable treatment from home plate umpires overall, but that there is little difference in umpires' omission bias towards superstar batters. Additionally, in higher leverage moments, umpires are slightly less likely to call strikes, but this bias is unchanged by the count. On any given count, all-star batters are, on average, 1.3% less likely to receive a called strike relative to non-all-stars on borderline pitches. Similarly, every additional million dollars in salary and each additional bWAR point are associated with a 0.1% decrease in the likelihood of a called strike on borderline pitches. On three-ball counts, there is no statistically significant difference in the distortion of the enforced strike zone for all-star batters, batters with higher salaries, or batters with a higher bWAR. On two-strike counts, neither all-star status nor salary are associated with a significant difference in the distortion of the enforced strike zone. Each additional point of a batter's bWAR, on the other hand, is associated with a 0.3% reduction in the likelihood of a called strike on borderline pitches coming on two-strike counts.

7 Conclusion

Omission bias is the most statistically and practically significant bias I observe. The pre-pitch count has a larger impact on the enforced strike zone than the game's stakes, the stardom of the batter, the game's attendance, and if the home team is batting. Further, there is little evidence to suggest that omission bias is impacted by external pressure or in-game characteristics. Instead, I find that omission bias exists separately from and is not impacted by other biases present in MLB umpiring, such as the biases favoring home batters and star batters. Lastly, instances of omission bias are largely isolated to the border of the strike zone and increase in frequency at locations with relatively higher levels of uncertainty, suggesting that omission bias is used as a rule of thumb when the umpire is otherwise uncertain of the correct call (Baron and Ritov, 2004).

On March 30, 2023, I had the opportunity to go to the Red Sox opening day at Fenway Park against the Baltimore Orioles. If you have not figured it out by now, I am a pretty big Orioles fan. It was freezing cold that day and the wind was absolutely ripping. My buddy Ben and I were sitting pretty far up in the grandstand past foul territory in right field. That game came over six months after I began this project. Lance Barksdale was the home plate umpire, and he is typically pretty good, with a correct call rate of 91.2%.

In the 7th inning, Red Sox relief pitcher Kaleb Ort was facing off against Orioles infielder Adam Frazier. The first three pitches of the at-bat were easy balls. On the fourth pitch, Ort tosses a four-seam fastball chest-high. The pitch looks close, but, from my angle, it was a ball. Barksdale, of course, calls a strike, screwing the Orioles out of a baserunner. Of course, I was furious, given that I had just run some preliminary regressions confirming the existence of this bias. To be fair though, I was likely four or five Sam Adams deep at that point, and we did not have the best sightline to the strike zone (our view was blocked almost completely by one of those giant green poles. Side note, but why do people think Fenway is so great? I get it has history or whatever, but Camden Yards is far superior in terms of actual experience!). Long story short, it did not matter, as Adam Frazier smacked a double into right-center field a couple of pitches later, and the Orioles went on to win the game 10 to 9. In other words, Barksdale let the boys play it out, and something exciting happened. No one remembers that call from opening day (except for me), and, in that moment, Barksdale escaped criticism from everyone in the ballpark (except from me).

3-0 Pitch from Ort to Frazier, Called A Strike. You be the judge!

(via MLB Film Room)

References

Andrews, Davy. 2023. "We May Never Find Out How Good Umpires Can Be." *FanGraphs Baseball* (blog). February 17, 2023. https://blogs.fangraphs.com/we-may-never-find-out-how-good-umpires-can-be/.

Associated Press. 2022. "Robot Umpires to Be Used in Triple-A in 2022." ESPN.Com. January 20, 2022. https://www.espn.com/mlb/story/_/id/33110161/robot-umpires-home-plate-moving-triple-2022-one-step-away-major-leagues.

Baron, Jonathan, and Ilana Ritov. 2004. "Omission Bias, Individual Differences, and Normality." *Organizational Behavior and Human Decision Processes* 94 (2): 74–85. https://doi.org/10.1016/j.obhdp.2004.03.003.

Bernier, Doug. 2014. "What Is the Strike Zone in Baseball? - Pro Baseball Insider." Https://Probaseballinsider.Com/. 2014. https://probaseballinsider.com/what-is-the-strike-zone-in-baseball/.

Bilalić, Merim, Bartosz Gula, and Nemanja Vaci. 2021. "Home Advantage Mediated (HAM) by Referee Bias and Team Performance during Covid." *Scientific Reports* 11 (1): 21558. https://doi.org/10.1038/s41598-021-00784-8.

Brown, Katrina F., J. Simon Kroll, Michael J. Hudson, Mary Ramsay, John Green, Charles A. Vincent, Graham Fraser, and Nick Sevdalis. 2010. "Omission Bias and Vaccine Rejection by Parents of Healthy Children: Implications for the Influenza A/H1N1 Vaccination Programme." *Vaccine* 28 (25): 4181–85. https://doi.org/10.1016/j.vaccine.2010.04.012.

Cross, Jeffrey, and Richard Uhrig. 2022. "Do Fans Impact Sports Outcomes? A COVID-19 Natural Experiment." *Journal of Sports Economics*, May, 15270025221100204. https://doi.org/10.1177/15270025221100204.

Foolish Baseball, dir. 2022. What Bad Umpiring Looks Like | Baseball Bits. https://www.youtube.com/watch?v=_nX_ZphDNQM.

Gong, Hua. 2022. "The Effect of the Crowd on Home Bias: Evidence from NBA Games During the COVID-19 Pandemic." *Journal of Sports Economics* 23 (7): 950–75. https://doi.org/10.1177/15270025211073337.

Green, Etan. 2014. "Four Strikes And You're Out." *FiveThirtyEight* (blog). April 3, 2014. https://fivethirtyeight.com/features/four-strikes-and-youre-out/.

Green, Etan, and David Daniels. 2015. "Impact Aversion in Arbitrator Decisions." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.2391558.

———. 2021. "Bayesian Instinct." SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.2916929.

Hauser, Marc. 2006. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong. New York, NY, US: Ecco/HarperCollins Publishers.

Jedlovec, Ben. 2020. "Introducing Statcast 2020: Hawk-Eye and Google Cloud." Medium. July 20, 2020. https://technology.mlblogs.com/introducing-statcast-2020-hawk-eye-and-google-cloud-a5f5c20321b8.

Kahneman, Daniel, and Dale T. Miller. 1986. "Norm Theory: Comparing Reality to Its Alternatives." *Psychological Review* 93 (2): 136–53. https://doi.org/10.1037/0033-295X.93.2.136.

Kahneman, Daniel, and Amos Tversky. 1982. "The Psychology of Preferences." *Scientific American* 246 (1): 160–73. https://doi.org/10.1038/scientificamerican0182-160.

KordesdeVaal, J. H. 1996. "Intention and the Omission Bias: Omissions Perceived as Nondecisions." *Acta Psychologica* 93 (1): 161–72. https://doi.org/10.1016/0001-6918(96)00027-3.

Mills, Brian M. 2017. "Technological Innovations in Monitoring and Evaluation: Evidence of Performance Impacts among Major League Baseball Umpires." *Labour Economics* 46 (June): 189–99. https://doi.org/10.1016/j.labeco.2016.10.004.

Moskowitz, Tobias J. 2011. *Scorecasting: The Hidden Influences behind How Sports Are Played and Games Are Won*. 1st ed.. New York: Crown Archetype.

Nathan, Alan. n.d. "The PITCHf/x System." Accessed December 8, 2022. http://baseball.physics.illinois.edu/pitchtracker.html.

Olney, Buster. 2023. "Sources: All AAA Parks to Have Robot Umps in '23." ESPN.Com. January 13, 2023. https://www.espn.com/mlb/story/_/id/35434317/sources-all-aaa-parks-use-electronic-strike-zone-23.

Paine, Neil. n.d. "GitHub - NeilPaine538/MLB-WAR-Data-Historical: JEFFBAGWELL Wins above Replacement (WAR) and Various Other Metrics for MLB since 1901." Accessed April 13, 2023. https://github.com/NeilPaine538/MLB-WAR-data-historical.

Perry, Dayn. 2020. "Timeline of How the COVID-19 Pandemic Has Impacted the 2020 Major League Baseball Season." CBSSports.Com. July 29, 2020. https://www.cbssports.com/mlb/news/timeline-of-how-the-covid-19-pandemic-has-impacted-the-2020-major-league-baseball-season/.

Samuelson, William and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk & Uncertainty* 1 (1): 7–59. https://doi.org/10.1007/BF00055564.

Shallow, Christopher, Rumen Iliev, and Douglas Medin. 2011. "Trolley Problems in Context." *Judgment and Decision Making* 6 (7): 593–601. https://doi.org/10.1017/S1930297500002631.

Spranca, Mark, Elisa Minsk, and Jonathan Baron. 1991. "Omission and Commission in Judgment and Choice." *Journal of Experimental Social Psychology* 27 (1): 76–105. https://doi.org/10.1016/0022-1031(91)90011-T.

"Statcast Search CSV Documentation." n.d. Baseballsavant.Com. Accessed April 7, 2023. https://baseballsavant.mlb.com/csv-docs.

"Strike Zone | Glossary | MLB.Com." n.d. Accessed November 28, 2022. https://www.mlb.com/glossary/rules/strike-zone.

Tango, Tom. 2006. "Crucial Situations." *The Hardball Times* (blog). May 1, 2006. https://tht.fangraphs.com/crucial-situations/.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk & Uncertainty* 5 (4): 297–323. https://doi.org/10.1007/BF00122574.

"Umpire Scorecards (@UmpScorecards) / Twitter." 2023. Twitter. 2023. https://twitter.com/UmpScorecards.

Verducci, Tom. 2019. "The Technology Boom Is Fundamentally Altering Baseball." Sports Illustrated. March 28, 2019. https://www.si.com/mlb/2019/03/29/technology-revolution-baseball-trackman-edgertronic-rapsodo.

Appendix:

Figure A1: Photo I took of the Baltimore Orioles and Boston Red Sox Game on May 30, 2022.
Final Score: Orioles 10, Red Sox 0

Rank	*Batter/Team/Year*	*Plate Appearances*	*bWAR*
1	Aaron Judge/Yankees/2022	696	11.00
2	Mookie Betts/Red Sox/2018	614	10.59
3	Mike Trout/Angels/2018	608	9.71
4	Alex Bregman/Astros/2019	690	8.495
5	Cody Bellinger/Dodgers/2019	661	8.165
6	Mike Trout/Angels/2019	600	8.145
7	Alex Bregman/Astros/2018	705	7.95
8	José Ramírez/Guardians/2018	698	7.785
9	Nolan Arenado/Cardinals/2022	620	7.59
10	Marcus Semien/Athletics/2019	747	7.495
…	…	…	…
761	Maikel Franco/Orioles/2021	403	-1.13
762	Spencer Torkelson/Tigers/2022	404	-1.135
763	Hunter Dozier/Royals/2022	500	-1.18
764	Miguel Cabrera/Tigers/2022	433	-1.245
765	Lewis Brinson/Marlins/2018	406	-1.25
766	Kole Calhoun/Rangers/2022	424	-1.31
767	Victor Martinez/Tigers/2018	508	-1.49
768	Alcides Escobar/Royals/2018	531	-1.53
769	Hunter Dozier/Royals/2021	543	-1.77
770	Chris Davis/Orioles/2018	522	-2.93

Table A1: Best and Worst bWAR Seasons, 2018-2022 (min. 400 Plate Appearances)

Rank	*Home Plate Umpire*	*N*	*% Strike Called*
1	John Libka	255	24.7%
2	Gerry Davis	119	25.2%
3	Carlos Torres	256	25.4%
4	David Rackley	234	26.1%
5	Alan Porter	234	27.4%
6	Jordan Baker	272	27.9%
7	Shane Livensparger	147	27.9%
8	Mark Ripperger	255	28.6%
9	Alfonso Marquez	296	29.1%
10	Scott Barry	165	29.1%

Table A2: Most Biased Umps on Two-Strike Borderline Pitches (Min. 100 Pitches)

Rank	*Home Plate Umpire*	*N*	*% Strike Called*
1	Joe West	91	70.3%
2	Nick Mahrley	88	69.3%
3	Lance Barrett	81	69.1%
4	Chad Whitson	79	67.1%
5	Sean Barber	78	66.7%
6	Ted Barrett	99	66.7%
7	Vic Carapazza	88	65.9%
8	Bill Welke	81	65.4%
9	Mark Ripperger	75	65.3%
10	Ryan Additon	80	65.0%

Table A3: Most Biased Umps on Three-Ball Borderline Pitches (Min. 75 Pitches)