Impact Aversion: Agency Failure and Decision Bias at High Stakes

Etan Green    David P. Daniels
Stanford University

June 8, 2014

Abstract

Incentives are thought to solve principal-agent problems and to reduce decision biases. We test these propositions by analyzing over a million decisions made by Major League Baseball umpires. Even though MLB directs and incentivizes umpires to apply a consistent decision rule, we find that every umpire reveals an aversion to choices that would more strongly change the expected outcome of the game. We model the umpire as wanting to make the correct choice, but also wanting to avoid making a pivotal choice in error. When the correct choice is not obvious, the umpire will shade away from choices that represent greater departures from the current state. This impact aversion represents both an agency failure and a decision bias, and it results in distortions that increase with the stakes.

*Please direct all correspondence to eagreen@stanford.edu. The authors wish to thank Doug Bernheim, Nir Halevy, Dorothy Kronick, Jonathan Levav, Muriel Niederle, Roger Noll, Justin Rao, Peter Reiss, Al Roth, and Charlie Sprenger for helpful comments and suggestions on previous drafts. Green and Daniels also thank the Stanford University Graduate School of Business and a National Science Foundation Graduate Research Fellowship, respectively, for generous financial support. A previous version of this paper was presented at the 2014 MIT Sloan Sports Analytics Conference.
1 Introduction

Incentives are thought to align the actions of the agent with the goals of the principal (Grossman and Hart, 1983; Holmstrom and Milgrom, 1991; Laffont and Marimort, 2002) and to reduce decision biases (List, 2003; Hart, 2005; Levitt and List, 2008). In this paper, we test these propositions by analyzing over a million decisions made by home plate umpires in Major League Baseball. Even though MLB directs and incentivizes umpires to apply a consistent decision rule, we find that every umpire is impact averse—revealing an aversion to choices that represent greater departures from the current state. Specifically, when the correct choice is not obvious, every umpire shades away from choices that more strongly change the expected outcome of the game. In this setting, impact aversion represents both an agency failure and a decision bias, and it results in distortions that increase with the stakes.

Major League Baseball provides an ideal context for studying decision making. First, the umpire’s normative benchmark is clear: MLB directs umpires to make binary decisions according to a precise rule. Second, performance relative to this benchmark is observable; we use fine-grained data to determine whether an umpire’s decisions are consistent with this directive or whether they vary with normatively extraneous factors.

We show that the impacts of each the umpire’s choices, despite being normatively irrelevant, strongly affect his decisions. Consider a situation in which both of the umpire’s choices have equal impact on the likely outcome of the game, and the umpire is indifferent between them, selecting each choice 50% of the time. When the situation changes such that the impacts of those choices become asymmetric, the umpire will now distort his decisions by making the more pivotal choice as much as 20 percentage points less frequently, selecting the more pivotal choice only 30% of the time and the less pivotal choice 70% of the time. More generally, greater asymmetries in the impacts of the umpire’s choices induce more bias
towards the non-pivotal choice.

We model umpires as wanting to make a correct choice, but also wanting to avoid making a pivotal choice in error. Our main model makes two predictions. When the correct choice is obvious, the umpire will make it, regardless of how pivotal the choice would be. But when the correct choice is not obvious, the umpire will err towards the less pivotal choice. Our non-parametric and semi-parametric estimates are consistent with these predictions. We also structurally estimate this model’s coefficient of impact aversion separately for each umpire (allowing for possible impact neutrality or impact seeking), and we find that every umpire is impact averse.

In our setting, impact aversion is an agency failure (e.g. Prendergast, 1999; Ariely, Gneezy, Loewenstein and Mazar, 2009; Rebitzer and Taylor, 2011; Charness and Kuhn, 2011; Kamenica, 2012). Major League Baseball selects, trains, and rewards umpires for applying the rules consistently and disciplines umpires for applying the rules inconsistently. The explicit incentives faced by umpires disincentivize impact aversion. The league “strive[s] to make sure that umpires are consistent throughout,” according to the MLB VP of Baseball Operations overseeing umpires.\footnote{Baumbach (2014).} Because the impacts of the umpire’s choices typically change with each decision, an impact averse umpire will apply the rules inconsistently. We also consider the possibility that impact aversion is a response to incentives from other audiences—specifically, the crowd (Dohmen, 2008). However, we find that umpire behavior is inconsistent with this explanation. This suggests that impact aversion is a revealed preference, rather than a response to external incentives.\footnote{Revealed preferences might reflect normative preferences, decision-making errors, or both (Camerer, 2005; Beshears, Choi, Laibson and Madrian, 2008).}

Critics of behavioral economics have conjectured that expertise, competition, and high stakes will reduce biases (e.g. List, 2003; Hart, 2005; Levitt and List, 2008).\footnote{Of course, some of these authors have done behavioral work themselves (e.g. List, 2002; Kovash and Levitt, 2009).} In our setting,
impact aversion offers a counterexample to this claim, since it distorts high-stakes decisions by professionals in the field. Thus, this paper relates to a growing body of field studies that identifies systematic ways in which individuals violate standard economic assumptions (for a review, see DellaVigna, 2009), even in settings characterized by intense competition and high stakes (e.g. Northcraft and Neale, 1987; Berger and Pope, 2011; Pope and Simonsohn, 2011; Pope and Schweitzer, 2011). However, impact aversion not only distorts high-stakes decisions; it induces greater distortions as the stakes become more asymmetric. This suggests that decision biases may actually grow in importance as the stakes increase.4

Theoretically, impact aversion appears related to, and might help account for, several decision biases that have been previously studied in the lab, including status quo bias (Schweitzer, 1994), omission bias (Schweitzer, 1994), and choice deferral (Tversky and Shafir, 1992); see Anderson (2003) for a review. These biases have received relatively little attention in field studies, with the notable exception of the stickiness of default options in investment choices (Choi, Laibson, Madrian and Metrick, 2003) and organ donor registration (Johnson and Goldstein, 2003).

Impact aversion is distinct from arbitrator behaviors previously identified in the empirical literature. Studies in sports settings have documented evidence of player favoritism by arbitrators (Sutter and Kocher, 2004; Price and Wolfers, 2010; Parsons, Sulaeman, Yates and Hamermesh, 2011; Mills, 2013; Kim and King, Forthcoming; Zitzewitz, 2014). In contrast, an impact averse arbitrator will favor particular choices, not particular players. While labor arbitrators face external incentives to violate their directive (Bloom and Cavanagh, 1986; Klement and Neeman, 2013), it appears unlikely that external incentives generate the

---

4Experimental evidence has shown that in some circumstances, greater incentives produce more bias (for a review, see Camerer and Hogarth, 1999). Incentives can exacerbate bias when they motivate subjects to neglect a good heuristic (e.g. Arkes, Dawes and Christensen, 1986) or to experiment (Hogarth, Gibbs, McKenzie and Marquis, 1991), when the correct answer requires creative thinking (e.g. McGraw and McCullers, 1979), or when they cause the participant to “choke” (e.g. Ariely, Gneezy, Loewenstein and Mazar, 2009).
behavior we document. Much of the empirical literature on judicial decision making focuses on how the political ideology of the judge influences her rulings (e.g. Epstein, Landes and Posner, 2011). Recent evidence shows that judges display decision biases as well: experienced parole judges become discontinuously more likely to grant merciful rulings after food breaks (Danziger, Levav and Avnaim-Pesso, 2011). Though impact aversion is a decision bias, it depends on the choices presented to the arbitrator rather than on the arbitrator’s internal state.

The remainder of the paper is organized as follows. Section 2 describes the empirical context. Section 3 provides evidence that umpires are impact averse. Section 4 proposes and estimates a model of impact aversion and shows that every umpire is impact averse. Section 5 incorporates second-order risk aversion into the model, which predicts that impact aversion will increase when decisions are more difficult; we present evidence consistent with this prediction. Section 6 estimates the economic significance of impact aversion among umpires. Section 7 concludes, discussing how impact aversion may generalize beyond umpires to other arbitrators facing asymmetrically pivotal choices.

2 Background

2.1 Home plate umpires in baseball

Most plays in baseball begins with the pitcher throwing a pitch to the batter. When the batter chooses not to swing at a pitch, the home plate umpire makes a call—either a ball or a strike. The home plate umpire has a simple job: to decide whether the pitch intersects the strike zone. Major League Baseball defines the official strike zone as “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 kneecap.” Pitches that intersect the strike zone should be called strikes; pitches that do not intersect the strike zone should be called balls.

Umpires have not always adhered to this benchmark. As recently as the 1990s, the enforced strike zone varied greatly by umpire and often diverged significantly from the official strike zone. Pitches far beyond the side of home plate—that hitters would have to lunge for—were often called strikes. Meanwhile, high strikes—over the plate and above the hitter’s belt—were almost always called balls. Rather than assigning the best umpires to postseason games (the most visible and important games of the year), the umpires union mandated that all umpires split both postseason assignments and the extra pay—as much as half an umpire’s base salary over the entire postseason—equally among all umpires.

In 1999, MLB attempted to reduce the inconsistencies between the enforced strike zones and the official strike zone with three small measures: first, by reminding all umpires of the definition of the official strike zone; second, by instructing team officials to monitor each umpire’s enforced strike zone; and third, by suspending an umpire who physically confronted a player—the first suspension ever given to an umpire. A clumsy response by the umpires union paved the way for baseball to strengthen the incentive scheme faced by umpires. First, the union authorized a strike. Then, when it realized that its contract with MLB forbade a strike, the union persuaded 57 of its 66 members to resign so as to dissolve the union and negotiate with the league over a new contract. When a federal court negated this strategy, baseball used the already-tendered resignations to purge the umpires’ ranks of 22 of its members.

Over subsequent seasons, MLB overhauled the umpires’ incentive regime. Home plate umpires in Major League Baseball now operate under a high degree of monitoring, incentives for good performance, possible punishment for poor performance, considerable training, and

---

5http://mlb.mlb.com/mlb/official_info/umpires/rules_interest.jsp.
6Callahan (1998) describes the state of umpiring in the late 1990’s.
7Callan (2012) details this sequence of events.
stringent screening.

The league now employs over a dozen officials to monitor and evaluate umpire performance. Most games are now overseen in person by a representative from the league, who files a report detailing blown calls in the game. The league uses pitch-tracking technology to evaluate the calls of home plate umpires. In the early 2000s, MLB installed the QuesTec system in half of its stadiums, which tracked the location of each pitch as it crossed the region above home plate. Prior to the 2009 season, MLB installed the more accurate PITCH F/X system in every park, which captures the location of each pitch 20 times along its trajectory. After each game, the home-plate umpire receives a breakdown of his performance, including a score that measures the consistency of his calls with the official strike zone.\footnote{Drellich (2012) describes the evaluation process currently in place for umpires.}

Recently, MLB has more closely tied rewards and discipline to performance. Umpires are subject to two evaluations each season that are based on the reports from umpire observers and the strike zone cameras. Those umpires who perform best are assigned to postseason games. “There have been situations where umpires have been disciplined” as a result of poor evaluations, according to Joe Torre, the Executive Vice President of Baseball Operations.\footnote{Callan (2012).}

After the 2009 season, baseball fired three of its umpire observers after a number of important missed calls during the postseason.\footnote{Nightengale (2010).} Since 2000, a handful of umpires have been suspended for inappropriate confrontations with players and managers. In 2013, baseball suspended a home plate umpire for forgetting a rule.\footnote{Hoffman (2013) details the recent history of umpire suspensions.}

Selection of Major League umpires is stringent and performance-based. A new major league umpires must attend one of three certified schools, graduate in the top fifth of his class, and then rise through four levels of the minor leagues before having a shot at filling in for a MLB umpire on vacation.\footnote{Caple (2011) details the process of becoming an umpire.} MLB employs 70 full-time umpires at any one time and 8
to 12 fill-ins from the minor leagues. Typically, only one fill-in is hired as a full-time MLB umpire after each season.\textsuperscript{13}

\section*{2.2 Data and descriptives}

Umpires are supposed to call balls and strikes based solely on the location of the pitch. We measure umpires’ adherence to this normative benchmark with fine-grained pitch location data from the PITCH F/X cameras—the same system used to monitor the calls of home plate umpires.\textsuperscript{14} We define the location of the pitch by its coordinates when it intersects the plane rising from the front of home plate, on which the official strike zone is implicitly rendered. The PITCH F/X system also provides estimates of the top and bottom borders of the official strike zone based on the batter’s stance prior to each pitch.\textsuperscript{15} We use these measurements to normalize the vertical location of the pitch. We merge pitch location data from MLB.com with game data from Retrosheet.org, including the identity of the home plate umpire, the number of balls and strikes in the count, the number of outs, and whether there is a runner on each base.

Our data comprise every pitch recorded by the cameras during the 2009-11 regular seasons, over 2 million pitches. Umpires make calls on 53\% of pitches in the sample. After eliminating the 47\% of pitches that are swung at, the 13,000 balls that were thrown intentionally, and the 50,000 calls made by the 21 umpires who each make fewer than 7,500 calls during the window, the remaining sample contains 1,036,355 calls made by 75 umpires. About two-thirds of calls are balls and the remaining third are called strikes. 6\% of calls occur in three-ball counts, 19\% of calls occur in two-strike counts, and 2\% of calls occur in

\textsuperscript{13}\cite{O'Connell2007}.

\textsuperscript{14}About 1\% of pitches are not captured by the cameras.

\textsuperscript{15}While the width of the official strike zone is fixed, the height of the official strike zone varies with the height and stance of the batter. According to Major League Baseball, “The strike zone shall be determined from the batter’s stance as the batter is prepared to swing at a pitched ball.” \url{http://mlb.mlb.com/mlb/official_info/umpires/rules_interest.jsp}
full counts (three balls and two strikes).\textsuperscript{16}

**Figure 1:** (a) $\hat{m}(X)$: the probability of a strike call when the batter does not swing, and (b) $\hat{f}(X)$: the distribution of calls. The plane at the base of each figure is the plane that rises from the front of home plate, which is shown from the umpire’s perspective. The red lines denote the boundaries of the official strike zone. As the left figure shows, pitches that cross the plane in the middle of the official strike zone are always called strikes; those that cross well outside the official strike zone are always called balls. Pitches that cross near the boundaries of the official strike zone are sometimes called strikes and sometimes called balls. As the right figure shows, these borderline pitches comprise a disproportionate share of calls.

From our sample of over a million calls, we non-parametrically estimate the probability of a called strike conditional on the location of the pitch. Figure 1a shows this estimate of the enforced strike zone. The plane at the base of the figure is the plane that rises from the front of home plate. The red lines on the axes indicate the boundaries of the official strike zone: the width of home plate on the horizontal axis, and the normalized distance

\textsuperscript{16}The count keeps track of the prior balls and strikes in the at-bat, or the sequence of consecutive pitches to the batter. Every at-bat begins with a count of zero balls and zero strikes. A ball is added when the umpire makes a ball call. A strike call is added when the umpire makes a strike call or when the batter swings—unless the count has two strikes and he makes contact with the pitch but does not put it in the field of play, in which case the count remains at two strikes. At-bats end most commonly when the batter swings and hits the pitch in the field of play, when the count reaches four balls, or when the count reaches three strikes.
from knees to chest on the vertical axis. The umpire stands behind home plate and looks
down through the plane, over the catcher’s head, and towards the pitcher. A right-handed
batter would stand to the umpire’s left, along the vertical axis.

Both the z-axis and the color of the plot denote \( \hat{m}(X) \): our estimate of the probability
of a called strike conditional on \( X = (x_1, x_2) \), the location of the pitch. This estimate is the
prediction from a kernel regression of an indicator for whether the call is a strike.\(^{17}\) Pitches
that intersect the middle of the official strike zone are obvious strikes, and umpires call them
strikes more than 99% of the time; pitches that cross far outside the official strike zone are
obvious balls, and umpires call them balls more than 99% of the time. In between, pitches
that intersect the plane at the same location are sometimes called strikes and sometimes
called balls. This band of uncertainty is wide: more than half a foot separates pitches that
are called strikes 90% of the time and those that are called strikes 10% of the time.\(^{18}\)

Figure 1b shows \( \hat{f}(X) \): our estimate of the distribution of calls by location.\(^{19}\) Since calls
disproportionately cluster near the borders of the official strike zone, the band of uncertainty
plays an outsized role in determining the outcomes of pitches, at-bats, and even games.

3 Evidence of Impact Aversion

3.1 Pivotal situations: non-parametric estimates

An umpire is \textit{inconsistent} if he makes different calls on pitches that cross the plane at the
same location; an umpire is \textit{biased} if these differences correlate with normatively extraneous
(non-location) factors. We first look for bias in two asymmetrically pivotal situations: when
the count has three balls or two strikes. The count tracks previous pitches in the at-bat, or

\(^{17}\)We use a bivariate Gaussian kernel and Silverman’s rule of thumb bandwidth for each axis.
\(^{18}\)The smoothing nature of the estimator may obscure a sharper boundary, though the bandwidth is small
enough to minimize this concern.
\(^{19}\)For the density estimate, we also use a bivariate Gaussian kernel and Silverman’s rule of thumb band-
width for each axis.
the sequence of pitches between pitcher and batter. A fourth ball would end the at-bat by walking the batter; a third strike would end the at-bat by striking him out. Unless there are three balls and two strikes (a full count), the umpire can extend the at-bat by calling a strike to avoid a walk or by calling a ball to avoid a strike-out. The count should not influence an umpire’s calls. According to Peter Woodfork, who oversees umpires as MLB Senior Vice President for Baseball Operations, Major League Baseball “strives[s] to make sure [umpires] are consistent throughout all at-bats, no matter the count.”

To visualize bias for a particular situation, we plot the difference between two non-parametric estimates of the enforced strike zone, \( \hat{m}(X|S) - \hat{m}(X|<3 \text{ balls } \& <2 \text{ strikes}) \): the first estimated on a subset of pitches for which the situation \( S \) is true (e.g. 3 balls \& < 2 strikes), and the second estimated on pitches in baseline counts with fewer than three balls and fewer than two strikes. Since the situations we consider are extraneous to the location of the pitch, the two enforced strike zones should be identical, and their difference should be zero across the plane.

Figures 2a and 2b show changes in the enforced strike zone when the count has three balls (2a), and when the count has two strikes (2b). In both graphs, pitches in full counts are excluded from the underlying strike zone estimates. If the enforced strike zones are the same, their difference will be a flat plane at zero. In each graph, the difference is near zero in both the center of the official strike zone and far outside of it. Even in three-ball or two-strike counts, obvious strikes are still called strikes, and obvious balls are still called balls. But where calls are not obvious, umpires enforce different strike zones. In three-ball counts, Figure 2a reveals a ring of mountains: the probability of a strike increases along the band of uncertainty—the strike zone expands. In two-strike counts, Figure 2b reveals a moat: the probability of a strike decreases along the band of uncertainty—the strike zone

\[^{20}\text{Baumbach (2014).}\]

\[^{21}\text{To attain a smoothed measure of the difference, we estimate each non-parametric strike zone using the maximum bandwidth along each axis across the two plots.}\]
Figure 2: $\hat{m}(X|S) - \hat{m}(X|<3 \text{ balls } & <2 \text{ strikes})$, for situation $S$ listed in figure titles. The change in the probability of a called strike when the count has (a) three balls, (b) two strikes, and (c) three balls and two strikes (full counts). The baseline case comprises calls in counts with fewer than three balls and fewer than two strikes. The enforced strike zone expands in three-ball counts and contracts in two-strike counts, particularly at the top and bottom. In full counts, the enforced strike zone contracts more moderately than with just two strikes.

(a) 3 balls & < 2 strikes

(b) 2 strikes & < 3 balls

(c) 3 balls & 2 strikes
contracts.\footnote{Baseball commentators and other researchers have previously shown that the enforced strike zone expands in three-ball counts and contracts in two-strike counts. For instance, see Moskowitz and Wertheim (2011) or \url{http://fangraphs.com/blogs/the-size-of-the-strike-zone-by-count/}. Our findings go beyond these in at least four ways. First, we measure the extent of the biases non-parametrically, semi-parametrically, and structurally. Second, we show that the bias is most pronounced in pivotal game states, not just pivotal counts. Third, we show that \textit{every} umpire exhibits impact aversion. Fourth, we show that decisions characterized by noisy signals induce even more impact aversion than comparable decisions characterized by non-noisy signals.}

With three balls and fewer than two strikes, a ball would be more pivotal than a strike. Similarly, with two strikes and fewer than three balls, a strike ends the at-bat while a ball prolongs it. The expansion of the enforced strike zone in three-ball counts and the contraction of the strike zone in two-strike counts suggest that umpires are averse to making the more pivotal call. By this logic, full counts (three balls and two strikes) should induce an intermediate effect—either a smaller expansion of the enforced strike zone than with three balls, or a smaller contraction than with two strikes. Because the umpire cannot avoid a pivotal call in a full count, he will distort the strike zone less than when he chooses between a pivotal call and an non-pivotal one. Figure 2c shows the difference in the enforced strike zone between calls in full counts and calls in counts with fewer than three balls and fewer than two strikes. Full counts induce a more moderate contraction of the strike zone than with just two strikes, which is consistent with an avoidance of asymmetrically pivotal calls. The fact that the enforced strike zone contracts in full counts relative to situations where neither choice is pivotal is consistent with our observation in Section 3.3 that third strikes tend to be more pivotal than fourth balls.

\subsection*{3.2 Semi-parametric estimates}

The non-parametric estimates in Figure 2 assume that all calls in a given situation (e.g. counts with three balls) are independent draws from an identical distribution. However,
the enforced strike zone varies across umpires\textsuperscript{23} and shifts left when a left-handed hitter is at bat.\textsuperscript{24} Accordingly, we model the umpire’s decision on call $i$, $y_i$, as the sum of a non-parametric prediction based on pitch location, $p_i$, and a linear term of situation-specific distortions, $[\omega(p_i) \cdot S_i] \beta$:

$$y_i = p_i + [\omega(p_i) \cdot S_i] \beta + \epsilon_i.$$ \hfill (1)

Here, $p_i$ is the baseline probability—an umpire- and batter handedness-specific measure of the probability of a strike call based on pitch location alone; $\omega(p_i)$ is a scalar weight; $S_i$ is a vector of situation indicators (e.g. three balls); $\beta$ is the vector of distortions induced by $S_i$; and $\epsilon_i$ is a mean-zero error.

We estimate $\beta$ in three steps. First, we estimate $p_i$ as $\hat{m}_{u^{(i)}, h^{(i)}, -S}(X_i)$, the predicted probability of a strike from a kernel regression of $y_i$. For call $i$, we estimate $m_{u^{(i)}, h^{(i)}, -S}$ only on pitches called by the same umpire $u^{(i)}$, and only for batters of the same handedness $h^{(i)}$. For $\hat{\beta}$ to be an unbiased measure of $\beta$, we need the baseline probability to measure the probability of a strike call in the absence of distortion. Hence, we measure $p_i$ only on pitches for which none of the states $S$ are true—i.e., for counts with fewer than three balls and fewer than two strikes. The baseline probability $p_i$ captures the likelihood of a strike call for pitches at location $X$ called by umpire $u^{(i)}$ to batters of handedness $h^{(i)}$. If only three-ball and two-strike counts induce bias, then the baseline probability identifies the likelihood of a strike call based solely on the location of the pitch.

Second, we calculate the weight of the linear component, $\omega(p_i)$, using our estimate of the

\textsuperscript{23}Anecdotal evidence suggests that small deviations from the official strike zone are tolerated so long as the umpire calls those deviations consistently. See, for instance: http://reds.enquirer.com/2001/02/25/red_high_time_for_new.html. Our data also show meaningful differences in the enforced strike zone across umpires.

\textsuperscript{24}Umpires position themselves differently behind the catcher based on the handedness of the hitter. The empirical strike zone is horizontally symmetric for right-handed hitters, but for left-handed hitters, umpires call strikes on outside pitches more frequently than on inside pitches.
baseline probability, \( p_i \). If \( \omega(\cdot) \) were equal to 1, our model would presume that the probability of a called strike is additive in \( p_i \) and \( S_i \beta \), and that the effect of a three ball or two strike count is uniform across the plane. Figure 2 shows that deviations are most pronounced at the borders of the official strike zone, so we define \( \omega \) in terms of the baseline probability of a strike: \( \omega(p_i) \equiv 1 - 2|p_i - 0.5| \). For certain balls or strikes (\( p_i \in \{0,1\} \)), \( \omega(p_i) = 0 \). For \textit{borderline pitches}, or locations in which balls and strikes are equally probable (\( p_i = 0.5 \)), \( \omega(p_i) = 1 \). When the situation indicators \( S \) are weighted by \( \omega(p_i) \), the interpretation of \( \beta \) is the percentage point change in the probability of a strike from the baseline probability when the baseline is 0.5—i.e., the bias on a borderline pitch.\(^{25}\) Finally, we estimate \( \beta \) by regressing \( y_i - p_i \), the component of the observed call not explained by pitch location, on \([\omega(p_i) \cdot S_i] \beta \).

Table 1 reports \( \hat{\beta} \), with standard errors clustered by \((u,h)\) tuple. The semi-parametric estimates echo the effects depicted non-parametrically in Figure 2. In three-ball counts, borderline pitches are called strikes more than 58\% of the time; in two-strike counts, borderline pitches are called strikes only 31\% of the time. In full counts (Model 2), the probability of a strike decreases by about 12 percentage points (0.09 - 0.19 - 0.02): 50/50 calls become 38/62 calls. The strike zone expands in three-ball counts, contracts in two-strike counts, and contracts to a lesser extent in full counts.

These estimates show that umpires violate their directive to call balls and strikes based solely on pitch location. However, the claim that asymmetrically pivotal counts cause changes in the enforced strike zone rests on an assumption of exogeneity with respect to omitted situational variables. While the structural context of umpire decision-making is relatively simple, we address some potential confounds by including additional situational variables interacted with the three-ball and two-strike indicator variables in Model 3.\(^{26}\) Specifically, we

\(^{25}\)Note that because the biases are greater at the top and bottom of the official strike zone than along the sides, \( \hat{\beta} \) will overstate the bias on the sides and understate the bias along the top and bottom.

\(^{26}\)We cannot include these situational variables directly because the linear component of the semi-
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-ball count</td>
<td>0.082***</td>
<td>0.088***</td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td>(17.21)</td>
<td>(15.67)</td>
<td>(4.47)</td>
</tr>
<tr>
<td>2-strike count</td>
<td>-0.19***</td>
<td>-0.19***</td>
<td>-0.20***</td>
</tr>
<tr>
<td></td>
<td>(-37.20)</td>
<td>(-35.40)</td>
<td>(-32.95)</td>
</tr>
<tr>
<td>Full count</td>
<td>-0.022</td>
<td></td>
<td>(-2.07)</td>
</tr>
<tr>
<td>Ball on last pitch in at-bat * 3-ball count</td>
<td></td>
<td></td>
<td>0.047**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.58)</td>
</tr>
<tr>
<td>Ball on last pitch in at-bat * 2-strike count</td>
<td></td>
<td></td>
<td>0.044***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(6.33)</td>
</tr>
<tr>
<td>Strike on last pitch in at-bat * 3-ball count</td>
<td></td>
<td>-0.056*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-3.29)</td>
</tr>
<tr>
<td>Strike on last pitch in at-bat * 2-strike count</td>
<td></td>
<td>-0.035***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-4.42)</td>
</tr>
<tr>
<td>Observations</td>
<td>1036335</td>
<td>1036335</td>
<td>1036335</td>
</tr>
</tbody>
</table>

$t$ statistics in parentheses

*p < 0.01, ** p < 0.001, *** p < 0.0001

**Table 1:** Semi-parametric regression on strike call. Coefficients of weighted linear component reported. Coefficient is percentage point change on the probability of a called strike for a borderline pitch under the given situation. Standard errors clustered by (umpire-batter handedness).
include indicator variables for whether the last pitch in the at-bat was a ball and whether
the last pitch in the at-bat was a called strike.\textsuperscript{27} The coefficient estimates on these terms
suggest negative autocorrelation in the umpire’s calls. A fourth ball is most likely following
a called strike and least likely after a third ball; strike three is most likely following a ball
and least likely after a called strike two. However, negative autocorrelation does not explain
impact aversion. The coefficients on the count indicators retain their signs and statistical
significance, though the umpire’s aversion to calling ball four can in part be explained by an
aversion to calling balls three and four in a row. Other alternative explanations are addressed
in the Appendix, which considers the possibility that impact aversion might be a response
to the umpire’s rational expectations of the forthcoming pitch.

3.3 A continuous measure of call impact

By expanding the strike zone in three-ball counts and shrinking it in two-strike counts,
umpires reveal an aversion to calls that end at-bats. But do they avoid these calls because
they are pivotal to the outcome of the at-bat, or because they are pivotal to the outcome
of the game? If umpires are averse to impacting the game, then the three-ball strike zone
should expand more when the bases are loaded (and a walk would score a run), and the
two-strike strike zone should contract more when there are two outs (and a strike-out would
end the inning).

To determine whether umpires avoid calls that affect the outcomes of games over and
above the outcomes of at-bats, we consider a continuous measure of how each call (ball or
strike) impacts the outcome of the half-inning.\textsuperscript{28} A baseball game comprises a series of half-

\textsuperscript{27} The omitted category contains the first pitch of the at-bat and pitches for which the previous pitch was
swung at.

\textsuperscript{28} It is difficult to develop a measure of how each call affects the outcome of the game, because the space
of game outcomes is too sparse to estimate reliably. Who wins a game is influenced not only by inning-level
indicators like the count, the number of outs, and the runners on base but also by the half-inning and the
score differential. Consider a game state defined by the combination of these variables. One can restrict the
innings in which one team pitches and the other team bats. When three outs are recorded, the half-inning ends and the teams switch roles in the next half-inning. Before a pitch, the state of the half-inning can be summarized by the expected number of runs the batting team will score over the remainder of the half-inning. We define a half-inning state as the tuple of the count, outs, and runners on base, of which there are \((4 \times 3) \times 3 \times 2^3 = 288\) combinations.

We estimate \(E[r_s]\), the expected number of runs to be scored over the remainder of each half-inning state \(s\), as \(\hat{R}_s = \frac{1}{||s||} \sum r_i\), the empirical average in corresponding states using 26 years and 16 million pitches of data.\(^{29}\)

<table>
<thead>
<tr>
<th>Half-inning state</th>
<th>Incidence (%)</th>
<th>(\hat{R}_s)</th>
<th>(\delta_{\text{ball}})</th>
<th>(\delta_{\text{strike}})</th>
<th>(\Delta = \delta_{\text{ball}} + \delta_{\text{strike}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 2-1, bases empty, 0 out</td>
<td>1.0</td>
<td>0.55</td>
<td>0.12</td>
<td>-0.07</td>
<td>0.047</td>
</tr>
<tr>
<td>b. 3-1, bases empty, 0 out</td>
<td>0.49</td>
<td>0.66</td>
<td>0.22</td>
<td>-0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>c. 2-2, bases empty, 0 out</td>
<td>1.2</td>
<td>0.48</td>
<td>0.10</td>
<td>-0.21</td>
<td>-0.11</td>
</tr>
<tr>
<td>d. 3-2, bases empty, 0 out</td>
<td>0.53</td>
<td>0.58</td>
<td>0.30</td>
<td>-0.31</td>
<td>-0.013</td>
</tr>
<tr>
<td>e. 2-1, bases loaded, 2 out</td>
<td>0.047</td>
<td>0.88</td>
<td>0.32</td>
<td>-0.21</td>
<td>0.11</td>
</tr>
<tr>
<td>f. 3-1, bases loaded, 2 out</td>
<td>0.027</td>
<td>1.2</td>
<td>0.53</td>
<td>-0.21</td>
<td>0.32</td>
</tr>
<tr>
<td>g. 2-2, bases loaded, 2 out</td>
<td>0.053</td>
<td>0.67</td>
<td>0.32</td>
<td>-0.67</td>
<td>-0.34</td>
</tr>
<tr>
<td>h. 3-2, bases loaded, 2 out</td>
<td>0.027</td>
<td>0.99</td>
<td>0.75</td>
<td>-0.99</td>
<td>-0.24</td>
</tr>
</tbody>
</table>

Table 2: The expected run measure \(\hat{R}_s\), the call impact measures \(\delta_{\text{ball}}\) & \(\delta_{\text{strike}}\), and the differential impact measure \(\Delta\) for selected half-inning states.

Table 2 lists properties for select half-inning states. If the bases are empty, there are no outs, and the count has two balls and one strike (row a), the batting team can expect to score 0.55 runs. This expectation rises to 0.66 runs if the umpire calls a ball (b), and it falls to 0.48 if the umpire calls a strike (c). A ball and a strike resets \(\hat{R}\) to 0.58 runs (d). With the bases loaded, two outs, and a count of two balls and two strikes (g), the batting team

---

\(^{29}\)These data comprise almost every pitch thrown during the 1988-2013 regular season. We observe the least common inning-state 688 times.
can expect to score 0.67 runs. This rises to 0.99 runs if the umpire calls a ball \((h)\). If the umpire calls a third strike in either state, the half-inning ends and \(\mathbb{E}[r_s] = 0\).

If umpires avoid calls that affect the outcome of the half-inning, then they will avoid making calls that greatly affect the number of runs the batting team can expect to score. We measure the impact of calling a ball or a strike as the change in the expected number of runs to be scored over the remainder of the half-inning as a result of the call. In all cases, calling a ball increases the number of runs the batting team can expect to score; in almost all cases, calling a strike decreases the number of runs the batting team can expect to score.\(^{30}\) We define the impact of a call, \(\delta\), as the difference in the expected runs measure between the half-inning state brought about by the call, \(s'\), and the current half-inning state, \(s\):

\[
\delta_{\text{ball}} = \hat{R}_{s'}^{\text{ball}} - \hat{R}_s \\
\delta_{\text{strike}} = \hat{R}_{s'}^{\text{strike}} - \hat{R}_s
\]

In Table 2, \(\delta_{\text{ball}}\) is positive and large in three-ball counts and even more positive with runners on base. Just a third ball with the bases loaded \((e \& g)\) increases the expected runs measure more than a walk (i.e. fourth ball) does when the bases are empty \((b \& d)\). Similarly, \(\delta_{\text{strike}}\) is negative and large in two-strike counts with zero outs and the bases empty \((c \& d)\), but it becomes even more negative with two outs and the bases loaded \((g \& h)\). In high-stakes states—two outs, bases loaded—a second strike \((e \& f)\) decreases the expected number of runs nearly as much as a third strike with the bases empty and zero outs \((c \& d)\).

Figure 3a shows the distribution of strike and ball impact in our sample of over a million calls. The graph contains one circle for each half-inning state, sized according to the relative incidence of that state. Most decisions are relatively non-pivotal regardless of whether a

\(^{30}\)In some three-ball and zero-strike counts with a runner on third and fewer than two outs, calling a strike increases the expected number of runs to be scored. We suspect that this is because hitters are instructed not to swing with three balls and zero strikes, but are allowed to swing with three balls and one strike. Since pitches in both counts are likely to be in the strike zone, swings with runners on third are likely to be beneficial for the batting team.
Figure 3: Distribution of half-inning states by strike and ball impact, $\delta_{\text{strike}}$ & $\delta_{\text{ball}}$, for calls made by umpires in our sample. The impact of a ball or a strike is the difference in the expected number of runs to be scored over the remainder of the half-inning from making that call. Sizes of circles (a) represent the relative incidence of states with associated impact. The differential impact of a call, $\Delta$, is $\delta_{\text{ball}} + \delta_{\text{strike}}$. For most calls, a strike and a ball are equally non-pivotal, creating a peak in the distribution of $\Delta$ at zero (b). But for some states, ball and strike impacts are asymmetric: one call is more pivotal than the other.

(a) Joint distribution of $\delta_{\text{ball}}$ & $\delta_{\text{strike}}$

(b) Distribution of $\delta_{\text{ball}} + \delta_{\text{strike}}$

ball or a strike is called; these calls have strike and ball impacts near zero. However, a considerable number of observed states are pivotal if strikes are called but non-pivotal if balls are called. Similarly, some observed states are pivotal for ball calls but not pivotal for strike calls. Strikes are generally more pivotal than balls; in fact, every state that involves a full count is asymmetrically strike-pivotal.

Umpires who are impact averse will tend to avoid the asymmetrically pivotal choice when the correct call is not obvious. We measure how asymmetrically pivotal a call is according to its differential impact, $\Delta$, or the sum of its ball and strike impacts: $\Delta = \delta_{\text{ball}} + \delta_{\text{strike}}$. For states that lie on the diagonal in Figure 3a (for which $\delta_{\text{ball}} = -\delta_{\text{strike}}$), the call impacts are symmetric and $\Delta = 0$. When impacts are asymmetric, $\Delta$ preserves the direction of the asymmetry. For asymmetrically ball-pivotal calls, $\Delta > 0$; for asymmetrically strike-pivotal calls, $\Delta < 0$. Figure 3b shows the distribution of differential impact in our sample.
The distribution peaks at zero—many calls are non-pivotal—but it has long tails in both directions. Most calls have near-equivalent impacts as balls or strikes, but some calls are asymmetrically pivotal as strikes by more than half a run, and some calls are asymmetrically pivotal as strikes by more than half a run.

With the bases loaded, two outs, and three balls and one strike (row $f$ of Table 2), a ball walks in a run, while a strike merely makes the count full. Indeed, this half-inning state is asymmetrically ball-pivotal: a ball increases the expected runs measure by three tenths of a run more than a strike decreases it. In this state, an impact averse umpire will avoid calling a ball. Were the count full ($h$), a ball would still walk in a run, but now a strike would end the half-inning. Here, a strike is asymmetrically pivotal: calling a strike decreases the expected runs measure by a quarter run more than a ball increases it. Now, an impact averse umpire will err away from calling a strike.

### 3.4 Umpires are averse to asymmetrically pivotal calls

We investigate whether umpires are averse to calls that affect the outcome of the half-inning by observing how the probability of a called strike changes with our differential impact measure $\Delta$. If umpires tend to avoid pivotal calls, we should observe that conditional on the location of the pitch, the probability of a called strike increases monotonically with $\Delta$. When $\Delta < 0$, a strike call is asymmetrically pivotal, and the probability of a strike call should decline; when $\Delta > 0$, a ball call is asymmetrically pivotal, and the probability of a strike call should increase.

We model the umpire’s decision on call $i$, $y_i$, as the sum of two non-parametric components and a mean-zero error term:

$$y_i = p_i' + g(\omega(p_i') \cdot \Delta_i) + \epsilon_i$$

(2)
Here, \( p_i' \) measures the baseline probability of a strike call based on pitch location alone, and \( g \) measures the effect of differential impact on the call.

This semi-parametric specification differs from Equation 1 in two respects. First, we model the distortions to the location-specific component \( p_i' \) as a continuous function \( g \) of the weighted differential impact, rather than a sum of situation-specific distortions. We estimate \( g \) as the prediction from a univariate kernel regression of \( y_i - p_i' \) on \( \omega(p_i') \cdot \Delta_i. \)\(^{31}\) Our estimate, \( \hat{g} \), relates the effect of asymmetric call impact on the component of the observed outcome, \( y_i \), not explained by pitch location.

Second, we estimate the baseline probability \( p_i' \) as \( \hat{m}_{u(i),h(i)}(X_i, \Delta = 0) \), the prediction from a kernel regression of \( y_i \). As with \( p_i \) in Equation 1, \( p_i' \) is specific to the umpire \( u^{(i)} \) and the handedness of the batter \( h^{(i)} \). However, \( p_i' \) is the baseline probability when the umpire’s calls are symmetrically pivotal (\( \Delta_i = 0 \)), rather than when the count has fewer than three balls and fewer than two strikes as for \( p_i \). This change is motivated by the assumption that for symmetrically pivotal calls, the expected call depends on pitch location alone. Hence, \( g \) will be zero when \( \Delta_i = 0 \). In practice, call impact is never exactly symmetric, so we calculate \( \hat{m}_{u(i),h(i)}(X_i, \Delta = 0) \) using a three-dimensional Gaussian kernel: two dimensions for pitch location and one for \( \Delta \). Our estimate of \( p_i' \) is the slice of \( \hat{m} \) where \( \Delta = 0 \)—the strike zone that the umpire would enforce if the impacts of calling a ball and a strike were symmetric. If \( p_i' = 0.5 \), pitch \( i \) is borderline: the umpire calls a strike 50% of the time when his choices are symmetrically pivotal.

The form of the weighting function \( \omega \) is the same as in Section 3.2: \( \omega(p_i') \equiv 1 - 2|p_i' - 0.5|. \) Since the index of \( g \) is \( \omega(p_i') \cdot \Delta_i \), the interpretation of \( \hat{g}(z) \) is the change in the probability of a strike call for \( \Delta_i = z \) from the baseline probability \( p_i' \) when \( p_i' = 0.5 \)—i.e., the bias on \( 31 \)Since the distribution of \( \Delta \) is highly uneven (see Figure 3b), we use an adaptive bandwidth with a local bandwidth factor of the form \( \left( \hat{f}(x)/\exp\left(\frac{1}{N} \sum_{i=1}^{N} \log \hat{f}(X_i)\right)\right)^{-\alpha}. \) \( \hat{f}(x) \) is a density estimate using Silverman’s rule of thumb bandwidth. We use \( \alpha = 0.5 \) to balance smoothness and detail in the visual appearance of the function.
a borderline pitch. If umpires avoid making asymmetrically pivotal calls, then when \( \Delta > 0 \), a ball call is asymmetrically pivotal, and \( \hat{g} \) will be positive; when \( \Delta < 0 \), a strike call is asymmetrically pivotal, and \( \hat{g} \) will be negative.

**Figure 4:** \( \hat{g}(\omega(p_i) \cdot \Delta_i) \): the effect of differential impact on the probability of a strike call for a borderline pitch. States with slightly asymmetric call impacts produce sizable distortions in the enforced strike zone, beyond which the effect of differential impact is largely stable. Annotations refer to the states described in Table 2. Dotted lines denote 95% confidence intervals.

Figure 4 shows \( \hat{g}(\omega(p_i) \cdot \Delta_i) \), the effect of differential impact on the probability of a strike call for a borderline pitch. The function is generally increasing in \( \Delta \), but the steepest increases occur in a narrow band around zero. For highly asymmetric calls, \( \hat{g} \) is generally flat. When balls are asymmetrically pivotal, \( \hat{g} \) peaks initially at just \( \Delta = 0.05 \) for borderline pitches. This corresponds to half-inning states such as row a in Table 2, in which the bases are empty, there are zero outs, and the count has two balls and one strike. Here, a ball is pivotal because it creates a count favorable to the hitter, not because it walks the batter. Pitches that are called strikes 50% of the time when impacts are symmetric are instead called strikes about 55% of the time when \( \Delta = 0.05 \). Further increases in \( \Delta \) induce similar amounts of distortion.
When strikes are asymmetrically pivotal, \( \hat{g} \) falls quickly from \( \Delta = 0 \) until \( \Delta = -0.1 \). A differential impact of \(-0.1\) corresponds to states such as row \( c \) in Table 2, in which the bases are empty, there are zero outs, and the count has two balls and two strikes. Here, a strike is pivotal because it ends the at-bat, even though it decreases the expected runs measure by just a fifth of a run. When \( \Delta = -0.1 \), borderline pitches are called strikes only 35\% of the time. Further decreases in \( \Delta \) induce similar amounts of distortion.

The non-linearity of \( \hat{g} \)—a steep rise around zero with flat tails at extreme values of \( \Delta \)—may arise if umpires are as sensitive to moderate asymmetries in call impact as they are to large asymmetries. But this pattern may also arise if umpires greatly avoid making an impact on the at-bat but are less concerned about making an impact on the game. To determine whether impact aversion is restricted to at-bats, we estimate \( \hat{g} \) conditional on each of the twelve counts. As Table 2 shows, the same count can have different \( \Delta \)s depending on the number of outs and whether there are runners on base. For instance, a full count with the bases empty and zero outs has nearly symmetric ball and strike impact. But a full count with the bases loaded and two outs is asymmetrically strike-pivotal. We estimate:

\[
y_i = p_i' + \sum_{b=0}^{3} \sum_{s=0}^{2} g_{b(i),s(i)} (\omega(p_i) \cdot \Delta_i) + \epsilon_i, \quad (3)
\]

where \( g_{b(i),s(i)} \) is the effect of \( \Delta_i \) on the probability of a called strike for a borderline pitch when the count for call \( i \) has \( b \) balls and \( s \) strikes. If umpires define impact wholly by the count, then \( \partial \hat{g}_{b,s}(\Delta) / \partial \Delta \) will be zero for all \( b \) and \( s \), while \( \hat{g}_{b,s}(\Delta) \) will be uniformly positive in asymmetrically ball-pivotal counts and uniformly negative in asymmetrically strike-pivotal counts.\(^{32}\) For a given count, the predicted distortion \( \hat{g}_{b,s}(\Delta) \) will be independent of \( \Delta \). By contrast, if umpires are averse to making an impact on the half-inning over and above their impact on the at-bat, then \( \hat{g}_{b,s}(\Delta) \) will be increasing in every count. Alternatively, if umpires

\(^{32}\)For these comparative statics, we treat each pitch as a borderline pitch \( (\omega(p_i') = 1) \).
are averse to making an impact on the at-bat only by virtue of its impact on the half-inning, then \( \hat{g}_{b,s}(\Delta) \) will not vary across counts.

Figure 5 shows \( \hat{g}_{b,s} \). The slope of \( \hat{g}_{b,s}(z) \) is positive around \( z = 0 \) in eleven of twelve counts, while the level of \( \hat{g}_{b,s}(z) \) is similar across counts for moderately asymmetric states. When \( z = -0.1 \), for instance, the predicted change in the probability of a called strike on a borderline pitch is between \(-10\) and \(-20\) percentage points for six of the seven counts in which \( \Delta \leq -0.1 \) is observed. This suggests that umpires shade away from calls which impact the outcome of the half-inning, not just those that impact the outcome of the at-bat. As in Figure 4, umpires appear not to differentiate between calls that are moderately asymmetric in their impacts and those that are extremely asymmetric. In full counts (Figure 5l), for instance, \( \hat{g}_{3,2}(-0.1) \) and \( \hat{g}_{3,2}(-0.5) \) both imply about a 15 percentage point decrease in the probability of a strike call on a borderline pitch, even though these states portend considerably different outcomes for the half-inning.\(^{33}\)

### 3.5 Are umpires responding to external incentives?

Impact averse umpires defy the league’s directive and incentives to enforce a consistent strike zone. Here, we show that umpires do not respond to incentives from the crowd either. This suggests that impact aversion may be a revealed preference, rather than a response to external incentives.

The crowd can punish the umpire by drawing negative attention to him when it does not like a call. If the umpire prefers to avoid the crowd’s expressions of anger, he should

\(^{33}\)These figures reveal other interesting patterns. With three balls, the strike zone only expands when the count has zero strikes (Figure 5h), and then only at moderate levels of differential impact. When the count has three balls and one strike (Figure 5j), we estimate the distortions as a precise zero across the observed range of \( \Delta \). In addition, the most asymmetrically strike-pivotal calls, which occur in two-strike counts, induce dramatically different distortions depending on the number of balls. With zero balls (Figure 5e), the strike zone contracts by as much as 25 percentage points—50/50 calls become 25/75 calls. But with just one ball (Figure 5f), the bias is not statistically different from zero for the most asymmetrically strike-pivotal calls. For moderately strike asymmetric states in two-strike counts, the strike zone contracts by 10 to 20 percentage points regardless of the number of balls.
Figure 5: $\hat{g}_{b,s}$ for listed $b$ and $s$.

(a) 0 balls & 0 strikes

(b) 1 ball & 0 strikes

(c) 0 balls & 1 strike

(d) 1 ball & 1 strike

(e) 0 balls & 2 strikes

(f) 1 ball & 2 strikes
be especially impact averse when the pivotal call penalizes the home team (because crowds disproportionately support the home team). By this logic, umpires should be more impact averse when the asymmetrically pivotal call is a strike ($\Delta < 0$) and the home team is batting, or when the asymmetrically pivotal call is a ball ($\Delta > 0$) and the home team is pitching. To test these predictions, we estimate two non-parametric functions: $\hat{g}_{\text{Home}}$, a univariate kernel regression of $y_i - p_i'$ for calls $i$ when the home team bats; and $\hat{g}_{\text{Away}}$, a univariate kernel regression of $y_i - p_i'$ for calls $i$ when the away team bats. If the umpire tries to avoid the ire of the crowd, we should observe $\hat{g}_{\text{Away}} > \hat{g}_{\text{Home}}$ when $\Delta > 0$, and $\hat{g}_{\text{Home}} < \hat{g}_{\text{Away}}$ when $\Delta < 0$.

Figure 6: $\hat{g}(\omega(p_i) \cdot \Delta_i)$: the effect of differential impact on the probability of a strike call for a borderline pitch. Estimates of the distortion when the home or away team is batting in (a). The difference, with 95% confidence intervals, in (b). Distortions induced by impact aversion are not meaningfully different when the crowd favors the batting team compared to when the crowd favors the pitching team.

(a) Distortion when the home or away team is batting

(b) Difference in distortion between home and away teams

Figure 6 shows $\hat{g}_{\text{Home}}$ and $\hat{g}_{\text{Away}}$ separately (6a) as well as the difference $\hat{g}_{\text{Home}} - \hat{g}_{\text{Away}}$ with 95% confidence intervals (6b). There is no meaningful difference between the two

---

34 The variance of difference between two random variables is the sum of the variances of each random variable minus twice the covariance. Rather than compute the covariance between two nonparametric esti-
estimates. Umpires are equally impact averse when the crowd supports the home team as when it supports the away team. The observed pattern of impact aversion is inconsistent with a response to incentives from Major League Baseball or from the crowd. This makes it more likely that impact aversion is a revealed preference, rather than a response to external incentives.

4 A Model of Impact Aversion

We propose and estimate a single parameter, state-based utility model of umpire decision making. We use this model to characterize the heterogeneity in impact aversion among umpires. The model also provides a template for estimating coefficients of impact aversion in other settings. In our model, umpires derive utility from making calls that are consistent with their interpretations of the strike zone. Umpires gain utility when they make self-consistent (correct) calls, and they lose utility when they make self-inconsistent (incorrect) calls.

Our model presumes that umpires prefer to make the correct call. But it also allows for umpires to have preferences about making the more pivotal call in error. If an umpire calls a self-consistent ball or strike, he receives a fixed amount of utility regardless of the impact of that call. But if he calls a self-inconsistent ball or strike, the amount of disutility he receives depends on how pivotal that call is. Consider the hypothetical utilities in Figure 7. If the umpire’s call is correct according to his idiosyncratic strike zone, he receives a utility of 1. The impact of his call does not affect the utility he gains from making the self-consistent call.

mates, we make the assumption that the covariance is zero. Since the estimates follow each other closely, the true covariance is almost certainly positive. Assuming it to be zero means that the confidence interval shown is wider than the true confidence interval.

35There is also no meaningful difference between the enforced strike zone when the home team bats and the enforced strike zone when the away team bats, even late in close games. Umpires appear relatively immune to the expressed sentiments of the crowd.

36In allowing for asymmetric valuation of outcomes, we follow Kahneman and Tversky (1979) and others.
Figure 7: Hypothetical utilities. Calling a self-consistent ball or strike generates a fixed amount of utility regardless of the impact of that call. However, the disutility generated by calling a self-inconsistent ball or strike depends on how pivotal that call is. In this example, a self-inconsistent ball or strike generates more disutility when the impact of that call is high. We estimate the slope of $U_{self-inconsistent}$ for each umpire.

Call. If he calls an incorrect ball or strike, and the impact of that call is zero, he receives a symmetric disutility of $-1$. But if the absolute impact of his call, $|\delta_{call}|$, is greater than zero, his disutility rises in proportion to the absolute impact he makes. Our model measures the slope of this disutility, which can be interpreted as the disappointment the umpire anticipates when he makes a wrong call that affects the outcome of the game. If an umpire is impact neutral, this slope will be zero. But if he is impact averse, this slope will be negative. (If he is impact seeking, the slope will be positive.)

Prior to each pitch, the umpire forms beliefs about likely game outcomes if he calls a strike or if he calls a ball. We measure these beliefs by the strike and ball impact of the call, $\delta_{i,strike}$ and $\delta_{i,ball}$. The umpire then observes the location of the pitch, which signals the probability that the pitch is a strike according to his personal strike zone. We measure this signal $p''_i$ as $\hat{m}_{u(i),h(i)}(X_i, \delta_{ball} = 0, \delta_{strike} = 0)$: the probability that umpire $u^{(i)}$ calls a strike on a batter with handedness $h^{(i)}$ when neither call is pivotal. Unlike the baseline
probability $p'_i$ in Section 3.4—which measures the probability of a called strike when the impacts are symmetric—$p''_i$ measures the probability of a called strike when the impacts are not only symmetric but also both equal to zero. Accordingly, $\hat{m}$ is a kernel regression in four dimensions: two for the location of the pitch, one for $\delta_{\text{ball}}$, and one for $\delta_{\text{strike}}$. The signal $p''_i$ is the likelihood that the umpire would call a strike were he not influenced by the impact of either call. For notational simplicity, we stop indexing the call by $i$, and we denote $p''$ as $p$.

We model umpires as maximizing the signal-weighted utilities of making the self-consistent and self-inconsistent calls:

$$U_{\text{strike}} = p \cdot U_{\text{self-consistent}} + (1 - p) \cdot U_{\text{self-inconsistent}}$$

With probability $p$, a strike call is self-consistent; with probability $1 - p$, a strike call is self-inconsistent. The reverse is true for calling a ball:

$$U_{\text{ball}} = (1 - p) \cdot U_{\text{self-consistent}} + p \cdot U_{\text{self-inconsistent}}$$

Given these utilities, the umpire calls a strike in expectation if $U_{\text{strike}} > U_{\text{ball}}$, and he calls a ball in expectation if $U_{\text{strike}} < U_{\text{ball}}$.

We normalize $U_{\text{self-consistent}} = 1$. Symmetrically, we fix $U_{\text{self-inconsistent}} = -1$ when the impact of the associated call is zero. When the associated call is pivotal, we allow $U_{\text{self-inconsistent}}$ to vary linearly according to its impact. We also allow the slope of this relationship to vary by umpire.

$$U_{\text{strike}}(p) = p - (1 - p)(1 - \lambda_u \delta_{\text{strike}})$$
$$U_{\text{ball}}(p) = (1 - p) - p(1 + \lambda_u \delta_{\text{ball}})$$

If the umpire observes an obvious ball or strike ($p \in \{0, 1\}$), he receives a utility of 1 for
making the obviously right call and a utility of 0 for making the obviously wrong call. He makes the right call in expectation. But if the signal is indeterminate ($p \in (0, 1)$), his call depends on the amount of disappointment he expects to feel when making the wrong call. If $\lambda_u = 0$, the umpire is not influenced by the impact of the call, and he receives a utility of $2p - 1$ for calling a strike and $1 - 2p$ for calling a ball. Again, he makes the self-consistent call in expectation. But if $\lambda_u > 0$, he may choose the self-inconsistent call in expectation if it is the less pivotal choice. Consider a call for which $p = 0.6$, $\delta_{\text{strike}} = -0.1$ and $\delta_{\text{ball}} = 0$. Here, a strike is the self-consistent call in expectation, but the umpire calls a ball in expectation if $\lambda_u > 10$.

**Figure 8:** Distribution of $\hat{\lambda}_u$ across umpires. For an unbiased umpire, $\hat{\lambda}_u = 0$. The lowest estimate of $\lambda_u$ is our sample of 75 umpires is 10.

We estimate $\lambda_u$ separately for each umpire. We add an IID type I extreme value error term to each of the utilities; we then estimate the $\hat{\lambda}_u$ that maximizes the resulting logistic likelihood function. Calls with asymmetric impact identify $\hat{\lambda}_u$. Figure 8 shows the distribution of $\hat{\lambda}_u$ across the 75 umpires in our sample. Each $\hat{\lambda}_u$ is considerably greater than zero, in both the statistical and economic senses. The least biased umpire has a $\hat{\lambda}_u = 10$ with a standard error of 0.55, and the largest standard error for any umpire’s $\hat{\lambda}$ is 1.1. Every umpire in our
sample shades away from the more pivotal call when the self-consistent call is not obvious.\footnote{Over the observation window, 65 of the 75 umpires receive postseason assignments, while 13 do so in the World Series. We cannot reject the null of no difference in the estimates of \( \lambda_u \) between those who umpire playoff games and those who do not, nor can we reject the null of no difference between those who umpire World Series games and those who do not.}

To see the distortion of the strike zone implied by a particular \( \hat{\lambda}_u \), consider a counterfactual prediction: the signal an umpire would need to receive in order to be indifferent between calling a ball and a strike in expectation. An unbiased umpire is indifferent when he receives a signal of \( p = 0.5 \), but a biased umpire \((\lambda_u > 0)\) may require a different signal when choosing between calls with asymmetric impact. Let \( \tilde{p}_u \) be a \textit{strike threshold}: the signal \( p \) at which umpire \( u \) with parameter \( \lambda_u \) is indifferent between calling a ball and calling a strike:

\[
\tilde{p}_u = \{ p : U_{\text{strike}} = U_{\text{ball}}; \lambda_u \}
\]

Substituting from equations 4 & 5 and solving for \( \tilde{p}_u \):

\[
\tilde{p}_u = \frac{2 - \lambda_u \delta_{\text{strike}}}{4 + \lambda_u (\delta_{\text{ball}} - \delta_{\text{strike}})} \tag{6}
\]

For an umpire averse to making pivotal calls \((\hat{\lambda}_u > 0)\), \( \tilde{p}_u > 0.5 \) when \( \delta_{\text{ball}} < -\delta_{\text{strike}} \), and \( \tilde{p}_u < 0.5 \) when \( \delta_{\text{ball}} > -\delta_{\text{strike}} \). When a strike is more pivotal than a ball, the biased umpire needs a signal greater than 50\% in order to call a strike in expectation; he is ball-biased. But when a ball is more pivotal than a strike, the biased umpire calls strikes in expectation when he is less than 50\% sure that the pitch is actually a strike; he is strike-biased. By construction, \( \tilde{p}_u = 0.5 \) when \( \delta_{\text{ball}} = -\delta_{\text{strike}} \) (i.e. \( \Delta = 0 \)): the umpire is unbiased when the call impacts are symmetric.

Figure 9 shows strike thresholds for the lowest observed \( \hat{\lambda}_u \) (9a) and the highest observed \( \hat{\lambda}_u \) (9b). For both the least and most biased umpires, the strike threshold deviates greatly from 0.5 with moderate amounts of impact asymmetry. Half-inning state \( a \) from Table 2 has
Figure 9: Strike thresholds for the minimum (a) and maximum (b) $\lambda$ as computed using Equation 6. By construction the strike threshold is 0.5 for calls with symmetric impact. For calls with asymmetric impact, the strike zone expands or contracts as a function of the magnitude of the asymmetry. When a strike is more pivotal than a ball, the strike threshold is greater than 0.5: the strike zone contracts, and both the least biased and the most biased umpires need a signal of greater than 50% to call a strike 50% of the time. Annotated letters correspond to half-inning states from Table 2.

(a) $\tilde{p}(\delta_{\text{ball}}, \delta_{\text{strike}}; \hat{\lambda}_{\text{min}} = 10.4)$

(b) $\tilde{p}(\delta_{\text{ball}}, \delta_{\text{strike}}; \hat{\lambda}_{\text{max}} = 20.9)$

nearly symmetric call impacts with $\Delta < 0.05$. Still, the strike threshold ranges from 37% to 42% in the population—no umpire needs to be more than 42% confident that a pitch is a strike in order to call a strike in expectation. Heterogeneity in impact aversion is small relative to the magnitude of impact aversion for the least biased umpire. For each of the five half-inning states plotted on the figures, the difference in the strike thresholds between the most and least biased umpires is smaller than the difference between the strike threshold of the least biased umpire and the unbiased threshold of 0.5. Not only are all umpires impact averse, but they are all impact averse to similar degrees.
5 Noisy Signals

In this section, we extend the model from Section 4 to address situations in which umpires observe a noisy signal rather than a point probability. We also assume umpires are second-order risk averse (e.g. Nau, 2006; Abdellaoui, Klibanoff and Placido, Forthcoming). In Section 5.1, we show that second-order risk aversion implies that impact aversion is increasing in the noisiness of the signal. In section 5.2, we examine three empirical situations characterized by noisy signals. In all three cases, consistent with the extended model’s predictions, impact aversion is greater for decisions with noisy signals than for comparable decisions with non-noisy signals.

5.1 Extended model with second-order risk aversion

A non-noisy signal $p$ is the point probability that a strike is consistent with the umpire’s idiosyncratic strike zone. Let a noisy signal $F_p$ be a symmetric distribution around $p$—i.e. a mean-preserving probability spread. In the model from Section 4, a noisy signal $F_p$ and a non-noisy signal $p$ produce the same behavior. Because the expected utilities in Equations 4 and 5 are linear in $p$, $\int_{p-\epsilon}^{p+\epsilon} U(q)dF(q) = U(p)$ for both $U_{\text{strike}}$ and $U_{\text{ball}}$.

This is no longer the case when the umpire is second-order risk averse. We incorporate second-order risk aversion by introducing the concave and strictly increasing function $v(\cdot)$:

\[
U_{\text{strike}}(p) = v\left(p - \gamma_{\text{strike}} \cdot (1 - p)\right) \tag{7}
\]
\[
U_{\text{ball}}(p) = v\left((1 - p) - \gamma_{\text{ball}} \cdot p\right) \tag{8}
\]

where $\gamma_{\text{strike}}$ and $\gamma_{\text{ball}}$ are the choice specific coefficients of impact aversion $1 - \lambda \delta_{\text{strike}}$ and $1 + \lambda \delta_{\text{ball}}$, respectively.

Abdellaoui, Klibanoff and Placido (Forthcoming) provides experimental evidence that individuals are second-order risk averse. Nau (2006) presents a theoretical analysis of second-order risk aversion.
The concavity of the utilities affects the expected choice when the signal becomes noisy. Consider a simple noisy signal $F_p$ that realizes $p - \epsilon$ with probability $\frac{1}{2}$ and realizes $p + \epsilon$ with probability $\frac{1}{2}$, for $\epsilon > 0$. Under this Bernoulli noisy signal,

$$
\int_{p-\epsilon}^{p+\epsilon} U(q) dF(q) = \frac{1}{2} U(p - \epsilon) + \frac{1}{2} U(p + \epsilon)
$$

$$
= \frac{1}{2} v(a - b\epsilon) + \frac{1}{2} v(a + b\epsilon)
$$

$$
< v(a) = E[U(p)],
$$

where $a$ is a choice-specific function of $p$ and $\gamma$, and $b = 1 + \gamma$. Because $v$ is concave, the utility of a choice decreases as signal noise $\epsilon$ increases. Moreover, this decrease is sharper for more pivotal choices, or those with higher $\gamma$. A noisy signal introduces the symmetric second-order risks that a pivotal choice is more likely to be right and that it is more likely to be wrong. With concave second-order utility, an impact averse umpire will overweight the second-order risk that a pivotal choice is more likely to be wrong relative to the second-order risk that it is more likely to be right—and he will overweight the downside risk more for more pivotal choices. Hence, second-order risk aversion makes an impact averse umpire err even more towards the less pivotal choice when the signal is noisy than when the signal is not noisy.

### 5.2 Impact aversion is increasing in the noisiness of the signal

We assume that the signal is more noisy when the location of the pitch with respect to the official strike zone is more difficult to observe. We examine three situations in the data characterized by noisy signals: pitches near the top and bottom borders of the official strike zone, which move up and down based on the hitter’s height and stance; off-speed pitches, which follow a curved trajectory rather than a straight line; and pitches in which
the umpire must make his call instantaneously, rather than be allowed to take his time. In all three cases, impact aversion is greater under noisy signals than under comparable non-noisy signals. These findings are consistent with the predictions of the extended model in Section 5.1.

5.2.1 The top and bottom of the official strike zone

The location of the pitch with respect to the official strike zone is more uncertain at the top and bottom of the official strike than along its sides for two reasons. First, the width of the official strike zone is fixed, but the height varies both with the height of the batter and with the stance he takes for each pitch. Second, the vertical location of the pitch is more difficult to observe than its horizontal location. Standing behind home plate, the umpire can more easily tell whether a pitch passes over the white of the plate than whether it crosses between the bottom of the batter’s knees and the midline of his chest.

Difficulty in determining the location of the pitch with respect to the top and bottom of the official strike zone creates uncertainty about the probability that the pitch is a strike. If a pitch passes over the edge of home plate at the hitter’s belt, it is clearly a borderline pitch, or a strike 50% of the time. But if it passes over the center of home plate at the level of the batter’s knees, it might be a borderline pitch, but depending on the batter’s stance and the umpire’s perception, it might be a certain strike or a certain ball instead. Pitches near the top and bottom of the official strike zone carry noisier signals than pitches along the sides.

If umpires become more impact averse as the signal becomes noisier, then we should observe greater bias at the top and bottom of the official strike zone than along the sides. This is what we see in Figures 2a and 2b: the expansion of the strike zone in three-ball counts and the contraction of the strike zone in two-strike counts are both greater at the top and bottom of the official strike zone than along its sides. In both figures, the distortions along the top and bottom are twice as large as along the sides. Where the location of the
pitch is more uncertain, umpires display greater impact aversion.

5.2.2 Off-speed pitches

The ease of identifying the location of the pitch also varies by the type of pitch. The locations of off-speed pitches, which tend to move vertically or laterally from the umpire’s perspective, are more difficult to observe than the locations of fastballs, which trace a more linear path from the pitcher’s hand to the catcher’s mitt. MLB classifies each pitch into one of more than a dozen types. We reduce this taxonomy to two types: fastballs, which comprise 64% of calls, and off-speed pitches, which comprise the remaining 36%. About two-thirds of off-speed pitches are either curveballs or sliders, two pitch types that pitchers spin upon release in order to induce vertical or lateral movement. On average, fastballs drop 5.0 vertical inches from release until crossing home plate, while off-speed pitches fall 9.5 inches.

We estimate two non-parametric functions: \( \hat{g}_{\text{Offspeed}} \), a univariate kernel regression of \( y_i - p_i' \) for calls \( i \) made on off-speed pitches; and \( \hat{g}_{\text{Fastball}} \), a univariate kernel regression of \( y_i - p_i' \) for calls \( i \) made on fastballs. If impact aversion increases for noisier signals, we will observe \( \hat{g}_{\text{Offspeed}} > \hat{g}_{\text{Fastball}} \) when \( \Delta > 0 \), and \( \hat{g}_{\text{Fastball}} > \hat{g}_{\text{Offspeed}} \) when \( \Delta < 0 \).

Figure 10 shows \( \hat{g}_{\text{Offspeed}} \) and \( \hat{g}_{\text{Fastball}} \) separately (10a) as well as the difference \( \hat{g}_{\text{Fastball}} - \hat{g}_{\text{Offspeed}} \) with 95% confidence intervals (10b). Impact aversion is stronger for off-speed pitches than for fastballs: the bias is more negative when the call is asymmetrically strike-pivotal and generally more positive when the call is asymmetrically ball-pivotal. Noisier signals induce greater impact aversion.

5.2.3 Time pressure

For most calls, play stops, and the umpire renders his verdict about a second after the catcher catches the pitch. But for about 1.5% of calls, the umpire must announce his

\[ \text{\textsuperscript{39}The t-statistic for this difference is of the order } 10^{3}. \]
Figure 10: \( \hat{g}(\omega(p_i) \cdot \Delta_i) \): the effect of differential impact on the probability of a strike call for a borderline pitch. Estimates of the distortion for fastballs and for off-speed pitches in (a). The difference, with 95% confidence intervals, in (b). Distortions induced by impact aversion are generally greater (i.e. farther from zero) for off-speed pitches.

(a) Distortion for fastballs and off-speed pitches

(b) Difference in distortion between off-speed pitches and fastballs

40 When the count has three balls and a walk would advance the runner(s) but a called strike would not end the inning, the call tells the catcher how he should address a potential steal. If the call is a strike, the catcher should make a play on the runner. But if the call is a ball, the runner advances and the catcher can only err by trying to make a play. Since the home plate umpire’s focus is on the pitch rather than the runners, he must make his call immediately in case a play needs to be made, even if no runners are trying to advance.

41 We compare calls under time pressure to calls in three-ball counts without time pressure because time...
Figure 11: $\hat{g}(\omega(p_i) \cdot \Delta_i)$: the effect of differential impact on the probability of a strike call for a borderline pitch. Estimates of the distortion with time pressure and without time pressure in (a). The difference, with 95% confidence intervals, in (b). Distortions induced by impact aversion are generally greater (i.e. farther from zero) under time pressure.

(a) Distortion under time pressure and in 3-ball counts without time pressure

(b) Difference in distortions between time pressure conditions

Figure 11 shows $\hat{g}_{TP}$ and $\hat{g}_{\neg TP (3balls)}$ separately (11a) as well as the difference $\hat{g}_{TP} - \hat{g}_{\neg TP (3balls)}$ with 95% confidence intervals (11b). As with off-speed pitches, calls under time pressure generally exhibit greater impact aversion than calls not under time pressure. In both cases, noisy signals induce greater impact aversion than non-noisy signals.

6 Economic Significance of Umpires’ Impact Aversion

On the free agent labor market, MLB teams spend $6.5M on average for each win that the acquired player is expected to contribute. With nearly 2500 games in each season, impact aversion need only affect the outcomes of a small number of games in order to significantly affect the economic fortunes of teams. We use Equation 3 to quantify the effect of impact pressure implies three balls.

\footnote{Silver (2014).}
aversion on games. In Section 6.1, we measure the number of calls that reverse in expectation as a result of the bias. In Section 6.2, we measure the mean amount of distortion, and its corresponding dollar value, induced by each call.

### 6.1 Call reversals

A call *reverses* in expectation if $p' < 0.5$ and $p' + \hat{g}_{b,s}(\omega(p') \cdot \Delta) > 0.5$ or $p' > 0.5$ and $p' + \hat{g}_{b,s}(\omega(p') \cdot \Delta) < 0.5$. In the first case, the pitch is a ball in expectation according to its location, but the umpire calls it a strike more than half the time; in the second case, the pitch is a strike in expectation according to its location, but the umpire calls it a ball more than half the time.

In an average game, impact aversion reverses four calls in expectation, or one call in every forty. Calls in counts with zero balls and one strike flip most frequently, at 5.4%, followed by calls in counts with three balls and zero strikes, which reverse in 4.4% of calls. In two-strike counts, calls flip between 3.4% and 4.1% of the time. An average game comprises eighty at-bats. About half of these reach a two-strike count, and about half of those include a call with two strikes. Among at-bats in which a call is made in a two-strike count, 5.8% include at least one call that flips in expectation from a strike to a ball. Absent impact aversion, these at-bats likely would have ended in strikeouts; but 42% of these at-bats end in something other than a strikeout. Once a game, on average, an expected third strike is called a ball because of impact aversion. And once every other game, an at-bat ends in something other than a strikeout after a third strike should have been called.

### 6.2 Mean distortion

Our estimate of the distortion induced by impact aversion is $\hat{g}$. We define the *mean distortion* as $\frac{1}{N} \sum_{i=1}^{N} |\hat{g}_{b(i),s(i)}(\omega(p'_i) \cdot \Delta_i)|$, or the average absolute deviation in the observed calls from
their baseline probabilities. The mean distortion is 2.9 percentage points for all calls, which implies that for every 100 calls, 2.9 are called differently because of impact aversion. This figure is higher in more asymmetrically pivotal counts. When the count has zero balls and one strike or three balls and zero strikes, the mean distortion is 5.7 percentage points. In two-strike counts, the mean distortion varies from 4.9 to 5.4 percentage points; on average, the looming impact of a strikeout makes umpires about five percentage points less likely to call strike three.

We use the mean distortion measure to quantify the financial consequences of impact aversion. If teams are willing to pay $6.5M to turn a loss into a win, then a risk-neutral team is willing to pay up to $(dp \cdot 6.5)M for a call that increases its probability of winning by $dp$ over the opposite call. Assume that the probability of winning is a linear function of the number of runs a team scores. We regress an indicator for whether a team wins on how many runs it scores using 26 years of game data. According to this model, an extra run increases the probability of winning by 8.6 percentage points. Hence, $dp = 0.086 \cdot \left| \hat{R}_{s'} - \hat{R}_{s^{\prime \prime}} \right|$, where $\hat{R}_s$ is the expected runs measure in half-inning state s, and $s'$ is the state that follows the associated call. For the average call, the absolute difference in the win probability resulting from the umpire’s choices, or $\frac{1}{N} \sum_{i=1}^{N} dp_i$, is 1.2 percentage points. This implies that on average, $75,000 hangs in the balance for each call.

We are interested in the fraction of this amount that is attributable to impact aversion. We calculate this quantity as:

$$\frac{6.5M}{N} \sum_{i=1}^{N} \left| dp_i \cdot \hat{g}_{b(i),s(i)}(\omega(p'_i) \cdot \Delta_i) \right| \approx 3,000$$

Here, we weight the change in the win probability by the amount of distortion associated with the current count $(b, s)$ and the differential impact $\Delta$. If $dp$ and $\hat{g}$ were independent, this figure would be the product of $75,000 and the mean distortion estimate of 2.9%, or
about $2,000. The true figure is higher because the calls that greatly affect which team is likely to win are subject to higher levels of distortion. On average, impact aversion distorts about $3,000 of team value every call.

7 Conclusion

Major League Baseball umpires are impact averse. Despite a directive and incentives from MLB to call balls and strikes based solely on pitch location, every umpire is averse to making the choice that more greatly changes the expected outcome of the game. We suspect that other arbitrators may also be averse to choices that represent greater departures from the current state. Consider judges. Supreme Court Chief Justice John Roberts stated in his confirmation hearing that “Judges are like umpires…it’s my job to call balls and strikes.”43 The American Bar Association states on its website that “Judges are like umpires in baseball…Like the ump, they call ’em as they see ’em.”44 These beliefs reflect a common perspective that judges make decisions by objectively applying legal principles (Sunstein, 2013).

In contrast, an emerging literature on the psychology of judges argues that judicial decisions are context-dependent (Guthrie, Rachlinski and Wistrich, 2007; Bordalo, Gennaioli and Shleifer, Forthcoming) and can be influenced by normatively irrelevant factors such as the amount of time since the last food break (Danziger, Levav and Avnaim-Pesso, 2011). If judges are like umpires, then perhaps judges show impact aversion too. One way this could happen is through decisions on the proceedings of a trial, such as decisions about whether evidence is admissible. These procedural rulings are supposed to be decided without regard to their impacts on the expected outcome of the case. But each time a judge makes a pro-

cedural ruling, she does so knowing which party is more likely to win the case based on
the evidence already presented. If judges are impact averse, they will distort their rulings
by avoiding choices that more greatly shift the expected outcome of the case. In this way,
impact aversion may reverse case outcomes.
Appendix: Rational Expectations

We consider the possibility that evidence of impact aversion can be explained by umpires’ rational expectations of the forthcoming pitch. Umpires might form expectations from the long-run distribution of pitches thrown in particular counts. If pitchers tend to throw strikes in three-ball counts, umpires might expect a strike in those counts; if pitchers tend to throw balls in two-strike counts, umpires might expect balls in those counts.

**Figure 12:** $\hat{f}(X|S) - \hat{f}(X|< 3 \text{ balls} \& < 2 \text{ strikes})$, for situation $S$ listed in figure titles. The change in pitch density when the count has (a) three balls and fewer than two strikes, and (b) two strikes and fewer than three balls. The base case comprises pitches in counts with fewer than three balls and fewer than two strikes.

Indeed, pitchers do throw more strikes in three-ball counts, and fewer strikes in two-strike counts. But these deviations are limited to the center of the official strike zone, where the call is obvious. As Figure 12 shows, pitches on the edge of the official strike zone—where the biases are strongest in Figure 2—are thrown just as frequently in pivotal counts as in non-pivotal counts. Umpires may expect more strikes in three-ball counts and fewer strikes in two-strike counts, but they can rationally expect those deviations only where strikes are obvious. Where the correct call is uncertain—i.e. where umpires display the greatest bias—
pitcher tendencies do not inform the umpire’s rational expectations about the forthcoming pitch.

Rational expectations may also be informed by whether the batter swings. Specifically, a batter’s decision not to swing may signal to the umpire that the pitch is a ball. Our results cannot be explained by swing signaling directly because umpires only make calls when the batter does not swing—the enforced strike zone varies, but the signal does not. Still, the rate at which batters swing in certain states may inform the umpire of the likelihood of a strike in those states. If in asymmetrically strike-pivotal states, batters swing more often, then the decision not to swing may signal that the pitch is a ball. However, the argument is uni-directional: choosing not to swing can only signal that the pitch is a ball, but in asymmetrically ball-pivotal states, we find that umpires are more likely to call strikes. Swing rates cannot explain the expansion of the strike zone when a ball would be pivotal. As with pitch location, swing rates cannot fully account for impact aversion.
References


Kim, Jerry W and Brayden G King (Forthcoming) “Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring,” *Management Science*.


Moskowitz, Tobias and L Jon Wertheim (2011) *Scorecasting: The hidden influences behind how sports are played and games are won*: Random House LLC.


