What Does it Take to Call a Strike?
Three Biases in Umpire Decision Making

Etan Green, David P. Daniels
Stanford University,
Stanford, CA, USA, 94305
Email: eagreen@stanford.edu

Abstract

Do Major League Baseball umpires call balls and strikes solely in response to pitch location? We analyze all regular season calls from 2009 to 2011—over one million pitches—using non-parametric and structural estimation methods. We find that the strike zone contracts in 2-strike counts and expands in 3-ball counts, and that umpires are reluctant to call two strikes in a row. Effect sizes can be dramatic: in 2-strike counts the probability of a called strike drops by as much as 19 percentage points in the corners of the strike zone. We structurally estimate each umpire's aversions to miscalling balls and his aversions to miscalling strikes in different game states. If an umpire is unbiased, he would only need to be 50% sure that a pitch is a strike in order to call a strike half the time. In fact, the average umpire needs to be *64%* sure of a strike in order to call strike three half the time. Moreover, the least biased umpire still needs to be 55% sure of a strike in order to call strike three half the time. In other words, *every* umpire is biased. Contrary to their formal role as unbiased arbiters of balls and strikes, umpires are biased by the state of the at-bat when deciding whether a pitch intersects the strike zone.

1 Introduction

A growing literature in psychology and economics demonstrates that professional athletes change their behavior in response to extraneous factors. Basketball teams that are slightly behind at halftime are more likely to win [1]. Golf pros invest greater focus in par putts than in comparable birdie putts [2]. And baseball players exert additional effort to achieve batting averages just above, rather than just below, a salient round number, .300 [3]. While these behaviors violate standard economic assumptions of rational decision making [4], it is not clear that they are suboptimal for athletes seeking to maximize performance [5]. We study a context in which any deviations from rationality are clearly suboptimal: the calls of home plate umpires.

Major League Baseball (MLB) requires home plate umpires to call balls and strikes based solely on the position of the ball as it crosses the region above home plate. Do umpires follow this directive? If not, which extraneous factors bias their decisions? We estimate the probability of a strike call conditional on pitch location and examine whether these probabilities vary systematically with other observables. We find that the state of the at-bat distorts the probability of calling a strike. Umpires are influenced by what they have called previously and how their current decision bears on the outcome of the at-bat. The effects we document are strong enough to change the outcomes of games. Bias flips about 1% of calls. Almost once a game, an at-bat ends in something other than a strikeout after a third strike should have been called.

2 Evidence of Bias

Among contexts in which individuals make decisions based on objective evidence, baseball is unusual in that outside observers can evaluate whether decisions are consistent with the evidence. Stereoscopic cameras take a sequence of 60 images of each pitch from release until it crosses home plate. We use the locations of pitches when they cross the plane above the front of home plate to non-parametrically estimate the probability of a strike call from the location of the pitch on the plane.[1]
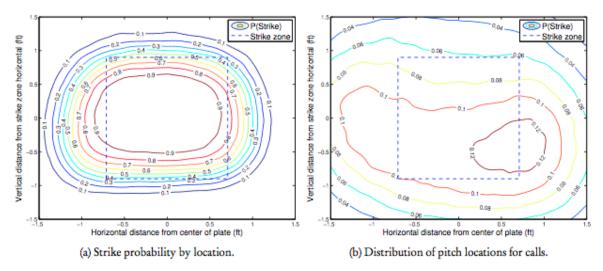
---

[1] For each location on the plane, our estimator takes a weighted average of all calls, weighting each observation inversely by its distance from the point. Our estimator is a kernel regression with a bivariate Gaussian kernel and Silverman's bandwidth for each axis.
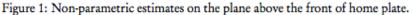
Figure 1a shows this estimate for calls made in the 2009-11 MLB regular seasons (over one million calls).[2] Pitches that cross the center of the official strike zone[3] are called strikes with probability greater than 99%. Those that cross the plane far from the official strike zone are called strikes with probability less than 1%. Probabilities change steeply across the edge of the strike zone: the ring in which the probability of a strike rises from 10% to 90% is a foot wide.[4] Figure 1b presents the distribution of calls across the plane. Since calls concentrate along the edges of the strike zone, this ring of uncertainty plays an outsized role in determining the outcomes of pitches, at-bats, and games.

(a) Strike probability by location.   (b) Distribution of pitch locations for calls.

Figure 1: Non-parametric estimates on the plane above the front of home plate.
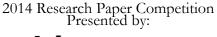
We call an umpire *inconsistent* when he makes different calls on pitches that cross the plane at the same location; we call him *biased* when these differences correlate with extraneous (non-location) factors. We look for biases induced by states of the at-bat, examining how the strike zone changes under four states: when the count has 3 balls, when the count has 2 strikes, when the last pitch in the at-bat was a ball, and when the last pitch in the at-bat was a called strike. These states represent either pivotal calls (the at-bat ends when the fourth ball or third strike is called) or calls that may be correlated across pitches.

To visualize the effect of a particular state, we non-parametrically estimate the probability of a strike twice: once for observations in which the state is true (e.g. 3 balls) and again for observations in which the state is false (e.g. < 3 balls). We plot the difference in Figure 2. For all states, the probability of a strike does not change either in the center of the strike zone or on the periphery, where pitches are obvious strikes or obvious balls. But the game state may systematically distort calls on borderline pitches. When the state raises the probability of a strike, we observe a ring of mountains. When the state lowers the strike probability, the mesh plot reveals a moat.

Baseball commentators have noted that the strike zone expands in 3-ball counts and contracts in 2-strike counts.[5] We find a small expansion in 3-ball counts and a large contraction in 2-strike counts. We also find a small expansion of the strike zone when the last pitch was a ball and a large contraction when the last pitch was a called strike. The effects of 2 strikes or a called strike on the last pitch can induce as much as a 20 percentage point drop, for the average umpire, in the probability of a strike. For some umpires, these states can shift the probability of a strike from a coin flip to almost zero. Robustness checks in Appendix A indicate that the apparent effect of a ball on the previous pitch is spurious but that sizable biases manifest in the other three states.

---

[2] We restrict the data to those umpires with at least 7500 calls over the observation window. This leaves 75 umpires and 1.03M calls.

[3] The official strike zone "is that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the [bottom] of the knees." (mlb.mlb.com/mlb/official_info/umpires/strike_zone.jsp) We normalize the vertical axis by the height of the batter's strike zone according to measurements taken by the camera operator at the beginning of each at-bat.

[4] The smoothing nature of the estimator may obscure a sharper boundary, though the bandwidth is small enough to minimize this concern. Figure 1a pools data across umpires; while individual strike zones are sharper, they are still imperfect. Analyses reported in Appendix A account for umpire heterogeneity.
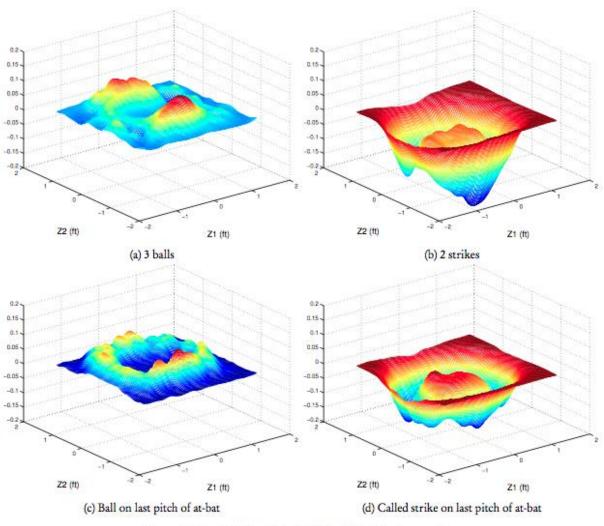
[5] For instance, see http://www.fangraphs.com/blogs/the-size-of-the-strike-zone-by-count/.

Figure 2: $P(\text{strike}|S) - P(\text{strike}|\neg S)$, S listed in figure titles.

3 Time Pressure

Evidence of bias raises the question of whether the bias is deliberative or intuitive. We arbitrate by observing how the magnitudes of the biases change when an umpire is under time pressure. For about 2% of calls, the umpire must make his call immediately, because the call tells the catcher whether to attempt a play on a potential baserunner. These calls occur in 3-ball counts with a runner on first, except for calls with 2 strikes and 2 outs.[6] If under time pressure the bias weakens, then the bias is likely deliberative—bias is created when the umpire thinks through his decision. If under time pressure the bias strengthens, then the bias is likely intuitive—bias is a consequence of snap judgment.

---

[6] When the count has 3 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.

Since time pressure implies a 3-ball count, we compare calls under time pressure with those in a 3-ball count but not under time pressure. We examine two biases under time pressure.[7] Figure 3 shows that time pressure accentuates the 2-strike bias. With time pressure (3b), the graph depicts a deeper moat than without time pressure (3a). In the case of 3 balls and a called strike on the last pitch (not shown), there is no apparent bias with or without time pressure. These results provide evidence in favor of the intuitive interpretation.
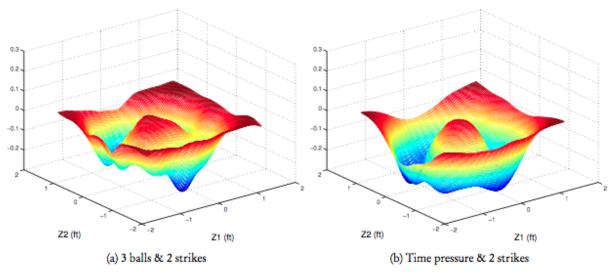

(a) 3 balls & 2 strikes   (b) Time pressure & 2 strikes

Figure 3: $P(\text{strike}|S_1, S_2) - P(\text{strike}|S_1, \neg S_2)$, S_1 & S_2 listed in figure titles.

4   Strike Thresholds

An unbiased umpire calls a strike when the probability of a strike based solely on the location of the pitch is greater than 50%. An umpire biased towards calling balls requires a higher threshold; one biased towards strikes requires a lower threshold. Bias flips close calls. A pitch that an umpire calls a strike 50.1% of the time in an unbiased state will likely be called a ball in a ball-biased state. We structurally estimate each umpire's state-specific aversions to miscalling balls and his aversions to miscalling strikes. Then we use these estimates to simulate the strike probability that would produce an umpire's observed calls in a biased state were he actually in the unbiased state.[8] We define this probability as the *strike threshold*. If a state does not induce bias, the strike threshold is 50%. Appendix B details the estimation procedure.

Strike thresholds document the extent and universality of the biases. Each curve in Figure 4 is the distribution of umpires' strike thresholds for a particular state. When the last pitch was a ball, for instance, the modal umpire has an unbiased strike threshold of 50%. But when the last pitch was a strike, the modal umpire's strike threshold rises to 60%. In 3-ball counts, the modal umpire's strike threshold falls to 35%. For 2-strike counts, the modal umpire's strike threshold rises above 60%; that is, pitches the modal umpire calls strikes half the time in 2-strike counts would be called strikes more than 60% of the time with fewer than 2 strikes. For two of the three states in which the modal umpire is biased, every umpire is biased. Despite their professional directive and expertise, umpires err in their decision making; their mistakes are systematic, sizable, and pervasive.

---

[7] Since time pressure implies a 3-ball count, we cannot observe the difference in the strike zone under time pressure between counts with 3 balls and counts with fewer than 3 balls. We do not consider the apparent last-ball bias depicted in figure 2, as this effect is shown to be spurious in Appendix A.

[8] Here, we use the terms unbiased state and baseline state interchangeably; a baseline state is one in which the count has fewer than 3 balls and fewer than 2 strikes, and the last pitch in the at-bat was not a call.
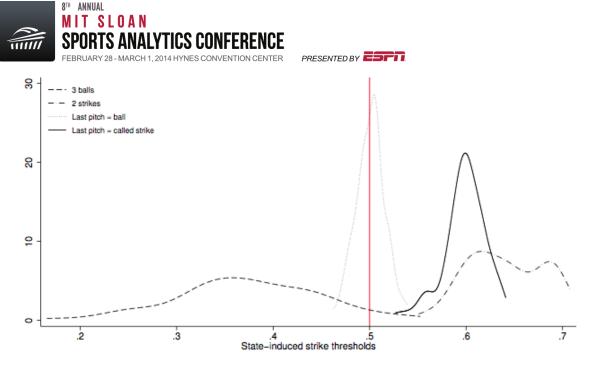
Figure 4: What does it take to call a strike? Distributions of umpires' strike thresholds show that some states induce pervasive bias. In principle, an unbiased umpire has a strike threshold of 0.5 for all states. Most observed umpires have lower strike thresholds in 3 ball counts, and all observed umpires have higher strike thresholds in 2 strike counts or when the last pitch was a called strike.

5 Acknowledgements

6 References

[1] J. Berger and D. Pope, "Can Losing Lead to Winning?," *Management Science*, vol. 57, pp. 817–827, Apr. 2011.

[2] D. G. Pope and M. E. Schweitzer, "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, vol. 101, no. 1, pp. 129–157, 2011.

[3] D. G. Pope and U. Simonsohn, "Round numbers as goals: evidence from baseball, SAT takers, and the lab.," *Psychological science*, vol. 22, pp. 71–9, Jan. 2011.

[4] C. F. Camerer, G. F. Loewenstein, and M. Rabin, eds., *Advances in Behavioral Economics*. Princeton: Princeton University Press, 2004.

[5] C. F. Camerer, "Three cheers–psychological, theoretical, empirical–for loss aversion," *Journal of Marketing Research*, vol. 42, no. 2, pp. 129–133, 2005.

[6] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, vol. 47, p. 263, Mar. 1979.

A Semi-parametric estimation

While Figure 2 provides illustrative evidence of state-induced deviations, the ascription of bias to the state relies on a number of strong assumptions. These assumptions include invariant strike zones across umpires, invariance in each umpire's strike zone between left and right-handed hitters,[9] and zero correlation between states.[10] In addition, it is analytically and computationally burdensome to estimate standard errors for a two-dimensional kernel regression, and a design challenge to visualize them. For these reasons we estimate the following semi-parametric model:

---

[9] The strike zone shifts sideways based on the handedness of the hitter because the umpire positions himself differently behind the catcher. This is why the strike zone in Figure 1 is left-shifted (from the umpire's perspective).

[10] This is obviously false for 3-ball counts which can only be first achieved when the last pitch was a ball.

$$y_{iub} = p_{ub}(z_{1i}, z_{2i}) + \omega_i \mathbf{x}_i^T \beta + \epsilon_i,$$

where y_{iub} is and indicator for whether call i by umpire u for batter handedness b is a strike. The call is modeled as an additive function of a non-parametric component, $p_{ub}(z_{1i}, z_{2i})$; a series of weighted indicators for game states, $\omega_i \mathbf{x}_i^T \beta$; and an IID mean-zero error term, ϵ_i.

The non-parametric component, $p_{ub}(z_{1i}, z_{2i})$, captures the baseline probability that umpire u calls a strike on a pitch that intersects the plane at coordinates (z_{1i}, z_{2i}). To estimate this probability, we run a bivariate kernel regression for each (u,h) tuple. In order to establish a baseline probability, we consider only calls that are neither pivotal nor are preceded by a call on the previous pitch—that is, those made in counts with fewer than 3 balls and fewer than 2 strikes, and when the last pitch was either swung at or ended the previous at-bat. If only the states we identify induce bias, this baseline probability represents the likelihood of a strike call based solely on the location of the pitch.

In the linear component $\omega_i \mathbf{x}_i^T \beta$, $\mathbf{x}_i$ is a $k \times 1$ vector of state indicators, β is a $k \times 1$ vector to be estimated, and ω_i is a scalar weight. If $\omega = 1$, state-induced deviations are estimated as uniform inflations or deflations across the plane. Figure 2 shows that deviations are most pronounced at the border of the strike zone, so we define ω in terms of the baseline probability of a strike: $\omega_i = 1 - 2 * abs(p_{ub}(z_{1i}, z_{2i}) - 0.5)$. For certain balls or strikes ($p_{ub}(z_{1i}, z_{2i})$ in $\{0,1\}$), $\omega_i = 0$; for locations in which balls and strikes are equally probable ($p_{ub}(z_{1i}, z_{2i}) = 0.5$), $\omega_i = 1$. Given this definition of ω, the interpretation of β_j is the percentage point change in the probability of a strike call for state j from the baseline when the baseline is 0.5—i.e., the state-induced bias on a borderline pitch.

Table 1: Semiparametric regression on strike call. Coefficients of weighted linear component reported. Coefficient is percentage point change on baseline probability for a close call under the associated game state. Standard errors clustered by umpire|batter handedness.

3 ball count	0.081***	(14.29)
2 strike count	-0.19***	(-37.30)
Full count	-0.025	(-2.45)
Ball on last pitch in at-bat	0.0049	(1.55)
Called strike on last pitch in at-bat	-0.17***	(-42.01)
N	1036335	

t statistics in parentheses
* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

Table 1 reports estimates for β.[11] The semi-parametric estimates generally echo the effects depicted non-parametrically in Figure 2. In 3-ball counts, borderline calls are called strikes 58% of the time;[12] in 2-strike counts, borderline calls are called strikes only 31% of the time. In full counts, the probability of a strike decreases by about 13 percentage points $(0.081 - 0.19 - 0.025)$—a moderation of the 2-strike bias. The effect of a ball on the previous pitch in the at-bat is near zero: the 95% confidence interval is $(-.001, 0.011)$. However, calling a strike on the last pitch in the at-bat decreases the probability of a strike call on a borderline pitch by 17 percentage points. Note that while we estimate these effects jointly, causal identification rests on an assumption of exogeneity with respect to omitted states.

B   Structural estimation

We estimate strike thresholds using a utility framework. By assumption, umpire u receives a utility of 1 for making a correct call, a disutility of $\gamma_{s,u}$ for calling a strike incorrectly, and a disutility of $\gamma_{b,u}$ for calling a ball incorrectly. Allowing for non-symmetric valuations of positive and negative outcomes follows Kahneman & Tversky [6]. We model umpires as if they are risk-neutral and know the probability with which a strike is the correct call.[13] This known strike probability is the baseline probability from Appendix A—the umpire-specific likelihood of a strike

---

[11] We estimate β in three steps: first, by estimating p non-parametrically; second, by subtracting p from y; and third by regression this difference on $\omega x^T \beta$. Standard errors are clustered by umpire—batter handedness.
[12] In results not shown, this bias towards strikes is only present on the first 3-ball count. If the hitter takes a strike or swings on the first three ball count, subsequent calls in the at-bat are not biased in favor of strikes.
[13] An alternate framework could allow for biased beliefs instead.

based solely on pitch location and estimated on observations in the baseline state—which we denote here as p. We assign the following state-specific utilities to strike (s) and ball (b) calls:

$$U(\text{call} = s | p, \mathbf{x}; \gamma_{s,u}) = p - (1-p)\mathbf{x}^T \gamma_{s,u} + \nu_s$$

$$U(\text{call} = b | p, \mathbf{x}; \gamma_{b,u}) = (1-p) - p\mathbf{x}^T \gamma_{b,u} + \nu_b,$$

where $\mathbf{x}$ is a 4×1 vector of state indicators, and ν_s and ν_b are mean-zero error terms. If ν is an IID draw from a type I extreme value distribution, then the probability of a strike call, $P[U(\text{call} = s) > U(\text{call} = b)]$ is the logistic transformation of $U(\text{call} = s) - U(\text{call} = b)$, which we fit to the data using maximum likelihood. We recover estimates of $\gamma_{s,u}$ and $\gamma_{b,u}$. We then simulate the strike thresholds for each state j by finding p_j at which $E[U(\text{call} = s) = U(\text{call} = b)]$, or:

$$p_j = \frac{1 + \hat{\gamma}_{s,u}^{(j)}}{2 + \hat{\gamma}_{s,u}^{(j)} + \hat{\gamma}_{b,u}^{(j)}}$$

If $\gamma^{(j)}_{s,u} = \gamma^{(j)}_{b,u}$, then $p_j = 0.5$, and the state does not induce bias. When $\gamma^{(j)}_{s,u} < \gamma^{(j)}_{b,u}$, then $p_j < 0.5$, and the state induces a strike bias. When $\gamma^{(j)}_{s,u} > \gamma^{(j)}_{b,u}$, then $p_j > 0.5$, and the state induces a ball bias.