The Dynamics of Inattention in the (Baseball) Field[1]

James Archsmith (University of Maryland)

Anthony Heyes (University of Birmingham)

Matthew Neidell (Columbia University, NBER and IZA)

Bhaven Sampat (Columbia University and NBER)

Abstract

Recent theoretical and empirical work characterizes attention as a limited resource that decision-makers allocate strategically. There has been less research on the dynamic interdependence of attention: how paying attention now may affect performance later. In this paper, we exploit high-frequency data on decision-making by Major League Baseball umpires to examine this. We find that umpires apply greater effort to higher-stakes decisions, but also that effort applied to earlier decisions increases errors later. These findings are consistent with the umpire being endowed with a depletable 'budget' of attention. There is no such dynamic interdependence after breaks during the game (at the end of each half-inning) suggesting that even short rest periods can replenish attention budgets. An expectation of higher stakes decisions in future induces reduced attention to current decisions, consistent with forward-looking behavior by an agent aware of attention scarcity. We believe this to be the first large-scale empirical demonstration, from economics or psychology, that individuals may manage the stock of attention in anticipation of future use.

Keywords: Rational inattention – dynamic decision-making – cognitive capital – decision fatigue – theory of ego depletion – bounded rationality – behavioral economics

[1] We are grateful to Stephen Coussens, Jonathan Guryan, Adam Leive, Devin Pope, Stephen Salant and seminar participants at University of Virginia, University of Massachusetts, and University of Connecticut insightful comments. Heyes acknowledges financial support from the Canada Research Chair program.

1. Introduction

Attention is perhaps the most important cognitive process; it allows humans to process the world around them to make better decisions. Yet many choices demand our attention over the course of the day -- we make approximately 35,000 decisions each day -- making it impossible to pay full attention to each.

A simple prediction of almost any model of costly attention would be that, other things equal, a rational agent would apply more attentional effort to an important, or high stakes, decision than they would to one with lower stakes. Beyond this, however, if attention is a depletable stock and is rationally managed, then amounts of attention applied across decisions in a sequence themselves become dynamically interdependent. That is, attentional effort at one point in time can be expected to affect, and be affected by, attention paid at other points in time. Attention paid to a current decision, and therefore the quality of that decision, are sensitive to attention devoted to previous one. Therefore, not only will decision quality decline with the number of decisions made (as is evidenced in the related literature on decision fatigue) but also to the complexity of these decisions, and the effort exerted in their execution.

Insofar as attentional effort is a depletable resource, an agent anticipating such inter-dependence could be expected to take forward-looking actions to conserve attentional capital. Agents may exert less effort on an early, low stake decision, to allow for greater attention later – consistent in spirit with why, for example, Mark Zuckerberg and Barack Obama are reported to wear the same clothes every day (Baer, 2015). This management of mental energy is analogous to the management of physical energy by athletes. Athletes may routinely play at less than "full throttle" in low stakes situations early in games to conserve energy if they expect to face higher value situations later.

In this paper, we explore empirically these dynamics of inattention in a data-rich, high-stakes, field environment by focusing on the ball and strike calls of home plate umpires from Major League Baseball (MLB) and connect what we find to economic models of strategic inattention, and the psychological theory of ego depletion. Our setting has several features that make it particularly well-suited for this research. First, we have insight into the attention applied at a given moment by observing not just an umpire's decision, but also a measure of the quality of the decision. Data from camera-technology in MLB stadia provide information on the objectively correct ruling for each decision. Under the assumption that greater attention devoted to a decision increases the probability of a correct outcome, we can infer how umpires vary the effort applied to specific decisions. That is, a correct decision, all else equal, implies that greater attention was paid. We are also able to control for a wide set of variables that influence the difficulty of each decision problem.

Second, the importance of decisions varies over time. We call the importance of each decision "leverage," defined as how pivotal a specific umpiring decision is in influencing the outcome of the game.[2] Leverage varies substantially during a game – as in most sports some calls are crucial to the outcome of the contest, some are largely irrelevant – evolving as a function of player actions and chance events. By providing a quasi-random source of variation in the value from paying attention, this allows us to explore how attentional effort varies with stakes. If the importance of a decision increases the value of paying attention, and umpires were allocating attention strategically, other things being equal we would expect umpires to make more correct calls as leverage increases.

Third, in a typical game, a home-plate umpire makes around 120 distinct decisions in approximately three hours. The number of decisions, along with the varying stakes associated with each, allows us to make progress disentangling dynamic interdependence, even with potentially small effect sizes. We can see how attention on one decision depends upon the importance of that decision, complexity and effort applied to prior decisions, and a rational expectation of the importance of future decisions. Moreover, we explore whether attention can be replenished by short periods of mental rest.

Next, the evolution of a baseball game may be decomposed into a sequence of discrete states with limited possibilities how the game may proceed from one state to the next. We can estimate the transition probabilities between states from actual game data, allowing straightforward computation of the empirical distribution of possible future states – and the corresponding (rational) expectation of future leverage – at any point in a game. This may be one of very few field settings where the econometrician may observe workers' expectations of future effort allowing us to explore how anticipation may impact the allocation of attention.

Finally, rich data are available. We exploit data on the more than 3 million decisions made by 127 home-plate umpires in 26,523 games between 2008 and 2018, enabling us to

[2] The term "leverage" is commonly used in statistical analysis of baseball to capture how important a particular moment is to the outcome of the game.

control for a wide set of potential confounders. The large sample allows for precise estimates even from econometric specifications that include game fixed effects that control for time-invariant characteristics of the umpire, the teams involved in the game, and the date of the game. We flexibly control for time elapsed, allowing us to separate the effects of cognitive fatigue from physical fatigue associated purely with passage of time. In addition to decision accuracy, our data includes an array of characteristics for each pitch thrown (pitch speed, type, location, and movement) that allow us to control in detail for the complexity or difficulty an umpire faces at any moment.

Contrary to conventional models of decision-making that predict that errors are random and therefore uncorrelated with leverage, our results reject the prediction that umpires exert equal effort to all decisions. We find that umpires adjust the attention paid to a decision in response to the importance of the decision; a one standard deviation increase in the leverage of a decision increases the likelihood of a correct call by 0.61% -- equivalent to improving the accuracy of the median umpire to that of the 73rd percentile umpire. This supports a central static prediction of rational inattention theories: umpires allocate more attention when the benefits from doing so increase.

We also find that periods of higher leverage in the past lead to less contemporaneous attention even with controls for current leverage. A one standard deviation increase in prior leverage reduces the probability of a correct decision, other things equal, by 0.32%, equivalent to reducing the accuracy of the median umpire to that of the 45th percentile umpire. This finding is consistent with a model of a depletable budget of decision resource, such that more attention devoted to one decision depletes availability for subsequent decisions, i.e., increases its marginal cost.

However, short respites in the decision series – which are provided by the break that the structure of the game gives the umpire between each half-inning – reset the process.[3] The effect of higher leverage in *previous* half-innings on current decision quality is a precisely estimated zero. While it is intuitive that rest would increase productivity in a physical work setting, for example because of muscle fatigue, it is less obvious how the design of shift patterns and work

[3] In baseball an inning is the basic unit of play and a game comprises nine scheduled innings. Each inning is divided into two half-innings. In the "top" half the visiting team bats until three outs are made. In the "bottom" half the home team bats until three outs are made. The umpire receives a scheduled break between each half-inning as the teams reset their positions.

breaks might impact performance in mentally-challenging work tasks. Our results suggest breaks allow for replenishment of attentional capital.

Finally, we find evidence of forward-looking behavior by umpires. A rational expectation of having to deal with more important (higher leverage) decisions in the future of the decision series leads to reduced attention to the current. More concretely, an increase of one standard deviation in expected future leverage in the half-inning reduces the probability of a correct decision by 0.49% – equivalent to reducing the accuracy of the median umpire to that of the 39th percentile umpire.

Taken together, these findings significantly extend existing evidence on rational inattention in several ways. The central *static* prediction of rational inattention models is that agents allocate more attention to more important decisions, and this we confirm. Laboratory and field studies find that intensity of attention is increasing in the importance or stakes associated with a decision across a range of settings. Examples in the field include consumer purchases of durable goods (Allcott and Wozny, 2014; Levav et al, 2010), portfolio investment decisions (Dellavigna and Pollett, 2009), consumer reactions to tax rates (Chetty et al 2009), hiring decisions (Acharya and Wee, 2020) and information acquisition in the rental housing market (Bartos et al, 2016) to name a few.

Our results on the dynamics of attention are more novel. As already noted, an appealing feature of the setting that we exploit is that we observe our subjects making a long sequence of decisions within a contained period. This allows us to explore both backward- and forward-looking responses in a field setting for the first time, to the best of our knowledge.[4]

Our results support 'budget-of-attention'-type models (for examples Dragone, 2019; Gabaix et al., 2006) in which there is a link between decisions in a series through the (endogenous) evolution of the remaining stock of attention. Effort exerted at one decision moment is expected to influence optimal attention allocation in a subsequent decision, conditional on the importance of that later decision. Consistent with this framework, we find

[4] The closest laboratory experimental evidence is that presented by Gabaix et al (2006). In their set-up subjects face an open-ended series of choices between sets of eight different goods and given a fixed time budget (25 minutes). Collecting information (by clicking on boxes using the 'Mouselab' technology) allows for a better decision in that round but depletes budget of time available to devote to future rounds. "Our experimental design also allows us to evaluate how subjects allocate scarce search time *between* games …" (Gabaix et al, 2006: 1062). They find that subjects devote more effort to higher value rounds, but also show increased propensity to stop analyzing the current game as the remaining budget of decision time diminishes.

both prior high leverage decisions and rationally-anticipated future high leverage reduce attention to current decisions.[5] It is inconsistent with other models of strategic inattention in which attention is costly, and agents optimise with respect to how much attention to apply, but there is no linkage between decisions in a sequence. Our findings also contrast with the experimental assumptions of, for example, Levav et al (2010), who interpret consumer choices in the sequence of mentally-taxing decisions required to configure an automobile under the assumption that "consumers are partially myopic in their allocation of mental resources. Instead of distributing their mental effort efficiently across the configuration process … (they) behave as if the current decision in a sequence is practically their last, despite that in our experiments it is obvious that subsequent decisions will follow" (page 276).[6]

1.1 Relation to existing psychology and related research

There are numerous studies on decision fatigue – that the character of decisions is function of the cumulative number of decisions made in a sequence. Settings include healthcare (Linder et al, 2014; Philpot et al 2018; Chan et al 2009; Kim et al 2015), financial forecasting (Hirshleifer et al, 2019), voting (Augenblick and Nicholson, 2016), consumer science (Bruyneel et al, 2006)), manuscript evaluation (Kwan et al, 2016) and air traffic control (Orasunu et al, 2012). Only in a few of these studies is the *quality* of decision observed. Further, to the best of our knowledge there is no study in which a proxy exists for the quantity of effort exerted on prior decisions. Our findings therefore point to the depletion of attention capital stock as depending not only on the cumulative number of decisions made but also the intensity of effort directed to them by the decisionmaker.

Our findings also relate to the psychological theory of 'ego depletion' proposed by Baumeister et al (1998). That paper, entitled "Ego Depletion: Is the Active Self a Limited

[5] If an umpire anticipates that paying high attention to one decision negatively impacts subsequent, perhaps more important, decisions, he will optimally conserve attention by strategically 'allowing' more errors in the present.
[6] An important feature of our setting shared by Levav et al (2010) is that they have a proxy for decision complexity, namely the number of options among which the consumer is required to choose at any stage of the customization process. They show, contrary to predictions of conventional choice theory under unlimited attention, that the vehicle ultimately configured by the consumer is sensitive to the experimentally manipulated order in which decisions over options are required. If complex decisions are posed early in the sequence, a consumer is more likely to revert to reliance on heuristics, such as accepting the default, later. Apart from being a large-sample field setting, the environment that we study provides an exact and objective measure of decision quality, whether the 'right' decision is made, whereas the utility-maximising specification of vehicle is not observed.

Resource?" proposed one of the most influential (10,322 cites on Google Scholar at time of writing) and controversial theories in social psychology: "The core idea behind ego depletion is that the self's acts of volition draw on some limited resource, akin to strength or energy and that, therefore, one act of volition will have a detrimental impact on subsequent volition" (page 1252). They present a series of experimental results consistent with that, with diminished performance both in subsequent tasks and in contemporaneous but unrelated tasks and cite extensive other literature that "… suggest that exertions of self-control carry a psychic cost and deplete some scarce resource" (page 1253). Volition here is interpreted broadly as effort applied to self-control or resisting a temptation. In our setting, the umpire needs to resist the temptation to relax or give less-than-full attention to a particular decision, and so doing depletes his stock going forward. Ozdenoren, Salant and Silverman (2012) provide a theoretical economic model of ego depletion, in which an agent chooses how to allocate a budget of self-control over time, using the standard tools of dynamic optimisation applied to a cake-eating problem.

Despite its popularity, in recent years ego depletion as a theory has fallen out of favour with repeated replication failures (see, for example, Hagger et al (2016) and citations therein). This is consistent with false positives derived in small sample, low-powered laboratory studies, combined with publication bias in favor of non-null results. The popular magazine Psychology Today (28 November 2020) was prompted to the cover article headline "How Willpower Wasn't: The Truth About Ego Depletion – Why an Effect Found in Hundreds of Studies Didn't Replicate."

Contrary to the failure of the psychological literature to make a persuasive empirical case for ego depletion, in our large sample, field setting we find what we believe to be compelling evidence in support of its central contention, that exertion of self-control through application of attentional effort depletes the budget subsequently available. We believe the context is essential for uncovering these effects; boredom experienced by subjects in a laboratory setting can have substantial impacts on inattention. Moreover, our forward-looking results point to the *anticipation* of ego depletion playing an important role in the decision-making process, in the spirit of the assumptions embedded in the model of Ozdenoren, Salant and Silverman (2012). It is not just that the limited resource depletes, but people act as if aware that their attentional capital is depletable and seek to allocate its use accordingly, for example conserving it for later decisions, where those future decisions are expected to be more important. This provides the first evidence of a further level of rationality and sophistication with which agents manage their expenditure of attention when faced with a series of mental tasks of varying challenge and importance.

We proceed as follows. In next section we provide background on MLB umpires, including what they do during a game, the incentives they face, and how we measure attention. Section 3 outlines the data that we assemble and describe how we operationalize "leverage," our measure of decision importance or stake. Section 4 describes our central econometric approach. Section 5 reports main results and a battery of robustness checks. Section 6 concludes.

2. Background

Baseball umpiring is a skilled job, requiring sustained mental effort. We study professional umpires operating at the highest level of the game, Major League Baseball. The ballparks where they work are dispersed across many of the major cities of the United States, plus Toronto. MLB as an organization employs around 100 umpires in any given season, organized into "crews" of four, with each serving as the home plate umpire every fourth game. Umpiring at this level is a lucrative and competitive career, with an experienced umpire commanding a base salary of $350,000 per season, which can be supplemented by post-season assignments and writing and speaking engagements for the high performers.

The most significant task that the home plate umpire faces in his working day is "calling" the game: deciding which pitches are balls and which are strikes. A pitch should be called a strike if any portion of the baseball passes through the strike zone (see Figure 1), and a ball otherwise.[7] The accuracy of the adjudications is fundamental to the game. We use whether a call is correct as our measure of decision quality.

In an average game, an umpire makes calls on around 120 pitches. We observe the umpire's decision as well as the objectively correct call. We obtained the latter from a high-precision pitch-tracking technology called PITCHf/x which has been in operation at every MLB ballpark since 2008. The output of the PITCHf/x camera system will be familiar to baseball watchers since it forms the basis for the real-time on-screen pitch location graphic used in

[7] If the pitcher throws three strikes, the batter is considered out (a "strike out"). If the pitcher throws four balls, the batter advances to first base (a "walk", or "base on balls").

television broadcasts of games. Researchers have used the same data as a testbed for other hypotheses, including racial discrimination (Parsons et al, 2011), the effect of status on evaluations and the so-called "Matthew Effect" (Kim and King, 2014), the gambler's fallacy (Chen et al, 2016), how decision quality is affected by exposure to air pollution (Archsmith et al, 2018) and as a test of models of strategic interaction (Bhattacharya and Howard, 2022) and belief formation (Green and Daniels, 2021).

Figure 1 presents a spatial scatterplot of the true locations of pitches upon which the umpire had to make a call in one game, as generated by PITCHf/x. Correct and incorrect calls are the hollow and solid black shapes, respectively. Umpires make both Type 1 and Type 2 errors. A solid triangle in the plot denotes a pitch that passed outside the zone that an umpire erroneously called a strike. A solid circle indicates that a pitch passed through the zone, but the umpire called it a ball. Not surprisingly, only pitches close to the strike zone boundary are called incorrectly.

[Figure 1 About Here]

The incentive for umpires to make correct calls is substantial. MLB operates a stringent system of monitoring and incentives for its umpires, called the Supervisor Umpire Review and Evaluation (SURE) system. This system uses various sources, including evaluations and on-site supervisors, to track the performance of umpires (Drellich, 2012). More generally, in the PITCHf/x era umpire errors are easily observed by a wider audience. As such, an umpire's reputation is plausibly highly sensitive to how often he makes mistakes, especially at those important (high leverage) moments in games when players, fans and the media are paying the most attention.

3. Data

We compile data to reconstruct the decision environment and outcomes faced by MLB umpires during professional baseball games. Our primary data are based on detailed information from actual games. We augment these data with a calculation of decision leverage and the outcomes of additional simulated games. Below, we describe each data source in detail.

3.1 MLB Pitch Data

Following Archsmith et al (2018) we compile data on the details of every pitch in all MLB games from 2008 to 2018 from the MLB website. These data are reported as part of MLB's PITCHf/x tracking system. Game-level data include variables for the home and away team, venue, the umpires and their position on the field, starting time, starting weather conditions, game attendance, and total runs scored by each team. Pitch-level data include identity of the players on the field and their position (including the pitcher, batter and catcher), attributes of the game situation (current runs by each team, half-inning, baserunner positions, outs, balls, and strikes), attributes of the pitched ball, the location of the pitch as it crosses home plate, and the result of the play after the pitch, including the umpire's ball/strike call if one was made.

Given that the ending of a baseball game is endogenous – a game can go into "extra innings" if the score is tied at the end of "regular innings" – we choose to focus our samples on only the pitches in the regular innings.

3.2 A measure of decision importance

Our empirical analysis investigates whether MLB umpires expend more effort to make correct decisions when the stakes of that decision are large. Doing so requires an objective measure of the stakes of each decision. To this end, we adapt the concept of *leverage* (a term already used in baseball to refer to the importance of a game situation) to specific pitch-level ball/strike decisions of MLB umpires.

Leverage is a scalar metric that assigns large values to important or pivotal moments in sporting events. For example, a decision or action that breaks a tie late in a game will have a much larger impact on the probability of winning than breaking a tie early in the game because the opposing team has fewer chances to equalize the score. We define leverage for a given pitch as the absolute difference in the probability the home team wins in the situation where the umpire calls a "ball" and the situation where the umpire calls a "strike". [8] The stakes can change

[8] Unlike many other sports, the rules of regular- and post-season MLB games prohibit games that end in draws. Therefore, the probability the away team wins conditional on some game situation *A* is simply one minus the probability the home team wins given that situation. Likewise, umpire decisions of "ball" versus "strike" are mutually exclusive and collectively exhaustive when an umpire adjudicates a pitch. As such changing the team for which we compute leverage or whether "ball" is subtracted from "strike" would result in an identical leverage metric.

substantially from pitch to pitch, and umpires can make independent decisions at each pitch over the level of effort to expend on adjudicating it correctly.

This leverage metric captures the umpire's state of incomplete information at the time each decision is made. The umpire knows the current situation of the game, but future events impacting the outcome of the game (many beyond the umpire's control) are unknown. Thus, computing this leverage measure requires determining two expected probabilities: the probability the batting team wins given a "strike" call and the probability they win given a "ball" call, conditional on the current situation in the game. In each case, we assume events in a baseball game follow a Markov process with state A_t encompassing the game situation at pitch t.

We estimate leverage using probabilities derived from simulated MLB games. By simulating the evolution of a large number of games, we can compute win probabilities for states that occur infrequently in the available history of baseball games at the cost of additional assumptions over the evolution of game states. There are four basic steps in the simulation (see the appendix for details):

1. We define the state of the game by the number of outs, baserunner positions, strikes, and balls.

2. Using actual MLB data, we compute the probability of transitioning from given states to new states plus new runs scored. We compute these state transitions within each half-inning.

3. Using these probabilities, we simulate 5 million MLB games from start to finish, collecting the states observed in each half-inning of a game and the eventual winner. This information is used to compute the probability the home team wins conditional on a given state.

4. Using these win probabilities, we compute a leverage measure for each situation using the same method as the measure based on actual game data.

The appendix also examines robustness of our results to an alternative approach to calculating leverage, using actual game outcomes.

3.3 Past and Future Leverage

We also consider the impact of accumulated past and expected future leverage. Past leverage is simply the sum of the leverage measure for all past pitches during the current half-inning. Expected future leverage is computed from simulated baseball games.[9] For each possible game situation and across all simulated games, we compute the leverage for all future pitches in that half-inning. Expected future leverage is the mean future leverage across all times that situation occurred.

3.4 Summary Statistics

Table 1 shows game-level summary statistics. We have data on 26,536 games, with an average of 291 pitches per game. Of these, about 120 pitches are "called," meaning they are subject to umpire discretion. This leaves about 3.2 million observations where the umpire makes a call about a ball in flight. Table 2 shows summary statistics for these pitches on the full sample (column 1), the final regression sample (column 2), and then further limits the latter to decisions in the first inning (column 3) and ninth inning (column 4). On average, umpires call 84 percent of pitches correctly.

The main explanatory variable in our analyses is leverage. In theory, leverage ranges from 0 to 1, but the average leverage at any point in the game is low (0.014) because any single pitch generally has a small effect on the outcome of the game. Note that columns 3 and 4 show that average leverage increases throughout the game, as later decisions are more impactful on the final outcome. However, even the 99th percentile value of leverage is small (.08). Past and future leverage are higher than current leverage because they capture the accumulation of leverage within a half-inning.

[Table 1 about here]

[Table 2 about here]

To illustrate our leverage measures, Table 3 provides specific examples of current and future leverage. For example, the 50th percentile of current leverage, measured as .0097, corresponds to a situation where there are 2 outs, 0 balls, 2 strikes, a runner on 3rd base, with the

[9] Here simulated games are essential since even situations which are overall unlikely –and not frequently observed in the 11 years of available data – may have a relatively large probability of occurring conditional on the current state and thus influence the umpire's expectation over future leverage.

home team leading by 3 runs in the bottom of the 6th inning. The value of .0097 is the difference in the probability the home team wins if the umpire calls a "ball" as compared to when the umpire calls a "strike."

[Table 3 about here]

A potential concern with our measure of leverage in time is that it lacks independent variation from current leverage. That is, if current leverage reflects not only the current situation but also how the game has evolved to its current point or future possibilities, then past and future leverage may be highly correlated with current. Figure 2, which presents a scatter plot of each leverage measure (past, current and future) against the other. Each shows ample independent variation in the measures of leverage, suggesting multicollinearity will not be an issue for our analyses.

[Figure 2 about here]

Likewise, Figure 3 shows the evolution of the leverage metrics over time through one particular MLB game. Current leverage is highest toward the ends of games with close scores, particularly in crucial situations. Accumulated leverage is highest after these critical situations, even if the current leverage is low. Expected future leverage generally increases over the course of the game and tends to peak as the game approaches critical junctures.

[Figure 3 about here]

4. Methods

Our goal is to investigate the relationship between the effort an umpire expends on correctly adjudicating a decision and the leverage of the decision. We estimate this relationship using a linear regression for each pitch *p* in game *g* as follows:[10]

$$\mathbf{1}\left(C_p^* = C_p\right) = \beta^C L\left(A_p\right) + \beta^P \sum_{\tau=1}^{p-1} L(A_\tau) + \beta^F \boldsymbol{E_p}\left[\sum_{\tau=p+1}^{\infty} L(A_\tau)\right] + \beta X_p + \boldsymbol{\delta_p^I} + \delta_g^G + \epsilon_{pg}$$

Where C_p^* is the decision of the umpire, C_p is the correct call given the point at which the pitch crossed home plate, $L\left(A_p\right)$ is the leverage in the situation where pitch *p* is thrown, X_p is a

[10] We estimate these regressions using the `reghdfe` package from Correia (2014).

vector of continuous (such as velocity or rate of spin) and discrete (such as an indicator for fastballs or curveballs) controls for pitch attributes, $\boldsymbol{\delta_p^I}$ are fixed effect for each half-inning, δ_g^G is a game fixed effect, and ϵ_{pg} is an idiosyncratic error potentially correlated within games.[11] The parameters of interest are β^C, the coefficient on the current pitch leverage, β^P the coefficient on accumulated past leverage, and β^F the coefficient on future leverage expected prior to adjudicating the current pitch. Given our definition of leverage, we interpret β as the effect of a change in win probability for the home team, conditional on the game situation, on the probability of the umpire making a correct call.

Our use of game fixed effects controls for many unobserved factors. Specifically, these fixed effects control for all time invariant characteristics of the umpire, the teams that are playing, the venue of the game, and the date and time of the game. This will control for features like a game between two rivals, a venue more amenable to home runs, or a hot day. With game fixed effects, we are exploiting how leverage within a game affects correct calling within the same game. This enhances our ability to interpret β^c, β^p, and β^f as causal parameters.

Although this approach controls for many time invariant components, there may be factors varying within the game that affect umpires' focus, such as physical fatigue and player changes. We include half-inning fixed-effects ($\boldsymbol{\delta_p^I}$) to control for physical fatigue as the game wears on, enabling us to separately identify the effects of decision fatigue. These fixed effects also control for any differences between home and away teams, both across the entire game and that change throughout the game, such as umpire bias to the home team (Sacheti et al., 2014).

Factors may also change within innings. Teams may change pitchers at crucial moments if the situation calls for a particular pitcher. These relief pitchers often have different pitching styles than the pitcher they replace, and these different styles may affect the umpires' ability to make correct calls.[12] To account for such factors, in some specifications we control for various pitch attributes in the vector X. A potential concern with this approach is that pitch attributes

[11] Following Archsmith et al (2018) and Kim and King (2014) we control for all pitch attributes reported in PITCHf/x using linear controls for continuous attributes and fixed effects for discrete attributes. Unlike previous work, there is substantial within-game variation in our variable of interest and we can identify our parameters of interest relying on only within-game variation using game fixed effects. However, variation in leverage is driven by the game situation, so we do not control for game situation variables which define the state space for our leverage metric.

[12] For example, relief pitchers are likely to throw more fastballs relative to breaking balls, and fastballs are easier for an umpire to adjudicate.

may be endogenous, or “bad controls” (Angrist and Pishke, 2008). Because of this, our preferred specification excludes controls for pitch attributes, though we report results from a specification including pitch controls in Section 5.3.

For defining past and future leverage, we accumulate leverage measures within the same half- inning. For example, for a game in the top of the third inning, imagine we are at the 5th pitch in the inning. Past leverage is the sum of the contemporaneous leverage from the first four pitches in the half-inning. Future leverage is the expected sum of leverage for all remaining pitches in the half-inning. As we move forward to the 6th pitch, past and future leverage update to include the 5th pitch. Given that the effects from past leverage may extend beyond the current half-inning, we also include past leverage from previous innings. The existence of a two minute break between half-innings enables us to explore whether the umpires stock of attention is replenished by a short respite.

5. Results

5.1 Main Results

Our main results are shown in Table 4. The first column reports results from our estimating equation in which our measure of past leverage is from the current half-inning only, with the next column adding the lag of past leverage. We exclude data from the first inning to keep the sample of pitches we explore fixed.[13] All coefficients are multiplied by 100 to improve readability. In general, we contemporaneous leverage increases umpires’ attention, while past and future leverage decreases it.

[Table 4 about here]

Focusing on the effect of contemporaneous leverage on umpires’ attention, we find estimates consistent with our hypothesis that higher leverage increases umpire attention. Our estimate of 38.223 indicates that increasing leverage from 0 to 0.013, the mean leverage in our sample, increases the probability that the umpires makes the correct call by .0051, a 0.61 percent increase. This estimate is highly statistically significant, with a 95% confidence interval of

[13] The sample size changes slightly due to missing or inconsistent game situation data from PITCHf/x that affects our ability to calculate leverage.

[0.55%,0.67%]. Adding additional lags of past leverage scarcely affects the coefficient on contemporaneous leverage.

Turning to the effect of past leverage, we find evidence consistent with a hypothesis of a depleted attention budget. As umpires face more leverage earlier in the inning, this decreases their attention on the current call. The estimate of -3.26 indicates that moving accumulated past leverage from 0 to 0.064, the mean of past leverage, decreases current call accuracy by 0.209 percentage points, which is a 0.248 percent change. This estimate is also highly statistically significant with a 95% confidence interval of [-0.289%,-0.207%].

As we include lags of leverage to our specification, two patterns emerge. One, the estimate on the past leverage of the current half-inning is unaffected. Two, the effect of past half-inning leverage is very small, coming in several orders of magnitude smaller than the current half-inning, and statistically insignificant. (These patterns hold true when we include additional lags of leverage.) These results imply that umpires refocus their attention after a short respite, suggesting that while our attention is scarce, our budget can replenish quickly.

Next, we focus on future leverage. Our estimates are again in line with our theoretical prediction: higher future leverage decreases current attention. Our estimate of -3.575 indicates that changing future leverage from 0 to 0.167, the mean of future leverage, increases an umpires' likelihood of a mistake by .060 percentage points, a 0.71 percent change. This estimate is also insensitive to further controls for lagged leverage and highly statistically significant, with a 95% confidence interval of [-0.792%,-0.628%].[14]

The inning dummy variables, which at least partially control for physical fatigue over the course of the game, also indicate an interesting pattern. Except for the last inning, we see fairly modest changes in umpire performance as the game wears on. Changes in a correct call vary between 0.05 to .1 percentage points compared to the second inning (the reference category), though with no clear pattern of physical fatigue throughout the majority of the game.[15] This suggests umpires are fairly consistent in their performance over the course of the game. However, in the last inning umpire performance drops by 0.39 percentage points. Since we control for leverage, this drop at the end of the game does not reflect the erosion of the

[14] Estimating this and all other models excluding data from the ninth innings does not substantially disturb any of our results.

[15] We omit inning 1 to allow inclusion of a lagged measure of past leverage.

importance of calls later in games. One possible explanation could be that, as we near the end of the game, umpires focus their attention elsewhere, leading to more mistakes.

5.2 Additional Results

In the next table, we explore how the effect of leverage varies within the game by estimating the effects separately by inning.[16] We find the same general pattern of results in every inning for our three measures of leverage, but find some interesting trends within the game. We did not derive specific hypotheses for these patterns, so we only speculate about potential explanations.

[Table 5 about here]

As the game wears on, the effect of current leverage on umpire attention steadily decreases, though it always remains positive. By the 9th inning, the effect of contemporaneous leverage is nearly 30% the size of the effect in the 3rd inning. The difference between these two estimates is statistically significant; further, the general decrease in the coefficient suggests an important trend. A possible explanation is that the umpire fatigues as the game goes on and is less able to regain focus for an equally important call later in the game.

In terms of the dynamics of leverage, we consistently find negative and statistically significant effects for past and future leverage, with the effect size steadily diminishing over time. While there are some trends across innings, in general the results by inning support our main results.

5.3 Robustness

In Table 6 we explore whether our results are robust to a range of alternative regression controls. First, in column 1 we repeat our preferred specification. The difficulty of an umpire's decision may depend on characteristics of the pitch in flight, such as velocity, spin, or trajectory. One concern is these attributes are a result of what type of pitch a pitcher decides to throw, which is decided after leverage is determined for a given situation. Thus, pitch attributes are effectively simultaneously determined with leverage, and could qualify as "bad controls"

[16] We estimate leverage effects by inning within a joint regression framework over the full sample to constrain the game fixed effects to be identical across innings.

(Angrist and Pishke, 2008). Despite this concern, when we add controls for all pitch characteristics[17] to our primary specification, shown in column 2, our estimates move only slightly.[18]

Our primary specification, by including game fixed effects, identifies the effect of leverage on decision accuracy from within-game variation. In columns 3 and 4 of Table 5, we allow for identification across games, while still accounting for potentially confounding heterogeneity, by replacing game fixed effects with umpire, home team, away team, and date fixed effects (column 3) and umpire, home team, away team, year, month-of-year, and day of week fixed effects (column 4). Estimated effects of leverage are essentially unchanged in this specification.

Finally, it is possible individual players may be more likely to be included in high-leverage situations and may take actions which increase the difficulty of an umpire's decisions. In column 5 we add fixed effects for players in each of the three positions that directly participate in this component of the game: the pitcher, the batter, and the catcher. Again, the empirical estimates are very similar to those from our primary specification. A further result in Section A1.3 of the Appendix demonstrates the same patter of effects persists regardless of whether the batting team is winning, losing, or tied with their opponent.

[Table 6 about here]

Our preferred specification assumes a linear relationship between each of our leverage measures and its impact on the probability of a correct call. To relax this assumption, we estimate a model that more flexibly controls for past, current, and future leverage. For each of the leverage measures, we divide the observed values into quintiles of equal size and replace our linear leverage measures with indicators for these quintiles.[19] Results for each leverage measure are shown in Figure 4. Results from this more flexible specification reveals an approximately

[17] We use continuous, linear controls for all continuous pitch attributes and indicator variables for all discrete attributes in the PITCHf/x data. Specifically, the continuous attributes are the position of the ball when released by the pitcher (along horizontal, vertical, and distance from home-plate axes), the ball's direction of spin, rate of spin, angle of break and distance of break (measured in both horizontal and vertical axes) and the velocity of the ball when it crosses home plate. The discrete attributes are indicators for the pitch type.

[18] In Appendix Section A1.4 we include additional specifications that include more flexible functions of pitch attributes, estimated using the Post-Double Selection Lasso of Belloni et. al (2012) and a partly linear regression with gradient boosted trees with Double/Debiased Machine Learning of Chernozhukov et al. (2018). Including very flexible functions of pitch attributes attenuates our estimates, consistent with pitch attributes being bad controls. However, in each specification, our parameters of interest are of identical sign, similar magnitude and maintain statistical significance.

[19] For each leverage measure we treat the first quintile as the omitted category.

linear relationship with effect sizes that are similar, if not slightly larger in magnitude, to the parametric specification from Table 4.[20]

[Figure 4 about here]

5.4 Heterogeneous Effects

We further explore heterogeneity in how individual umpires allocate effort by estimating umpire-specific leverage effects. Extending the regression specification from Table 4 Column 1, we interact each leverage measure with indicators for every umpire in a single regression.[21] The estimated effects for each umpire are shown in Figure 5. Panels (a) – (c) show the effects for past, current, and future leverage, respectively. In each panel, umpire specific effects are ordered from largest at the top and smallest at the bottom. For nearly all umpires, the estimated treatment effects are of the same sign as our main results, with a large portion being statistically significant at the 95% level despite the large increase in model parameters. These results suggest that very few, if any, umpires deviate from our main findings about dynamic inattention.

Panel (d) of Figure 5 combines the past, current, and future leverage estimates for each umpire into a single figure. Here, effects for a given umpire are aligned horizontally and ordered by the that umpire's estimated current leverage effect. For comparability, each leverage effect is divided by its standard deviation across all umpires. Lines show the moving average of each leverage effect across the 10 umpires above and below each observation. Umpire-specific effects for past and future leverage are negatively correlated with the current leverage effect.[22] Some individuals are highly responsive to high-leverage situations, appearing to expend substantial effort when decision stakes are high. These same individuals tend to exhibit larger decreases in accuracy from accumulated past and expected future leverage. This result is broadly consistent with umpires maintaining a budget for attention; umpires who expend more effort on high-leverage decisions will need to conserve more effort in other situations to maintain their budget.

We next explore whether there are factors that might explain the heterogeneity in umpires' inattention. To do so, we regress each of the umpire specific effects just obtained

[20] We observe similar results if we increase or decrease the number of quantiles used.

[21] To improve precision, we exclude umpires who are observed to serve as home plate umpire in fewer than 20 games during our sample. The median umpire served as home plate umpire in 240 games during this period. This restriction removes 14 of the 127 umpires or 132 of the over 26,000 games from the sample.

[22] The correlations between an umpire's past or future leverage effect and the current leverage effect are -0.401 and -0.485, respectively. Both correlations are significant at the 1% level.

against two variables: the age of the umpire and the average rate of correct calls by that umpire. We view age as a measure of both chronological age and umpire experience in an effort to assess whether increasing training or advanced age may affect the attention budget.[23] Results from this regression are shown in Table 7.

The results produce two findings. First, age appears to have little impact on how past, present, or expected future leverage influence an umpire's accuracy. Second, umpires who are more accurate on average tend to be less sensitive to leverage: their accuracy improves less when then current pitch is high-leverage. Conversely, the decrease in accuracy with high-leverage past or future pitches is less pronounced for umpires who tend to be correct more often.

We can place some of the estimates from Table 7 in context. Over the sample, the 5th percentile umpire has a 79.2% accuracy rate, the mean umpire has a 84.3% accuracy[24], and the 95th percentile umpire is 88.7% accurate. Umpires in the 95th percentile of accuracy are just under 40% less responsive to the leverage of the current pitch than the average umpire. Their accuracy is also less impacted by past (54% the effect of mean umpire) or expected future (64% of the effect of the mean umpire) high-leverage decisions. Effects for the 5th percentile umpire are slightly larger in magnitude and of the opposite sign. Their accuracy increases substantially relative to the mean umpire (44% more) when the current pitch is high-leverage, but decreases substantially when past decisions are high-leverage (51% larger decrease), or expected future decisions are (38% larger decrease).

These effects are consistent with more accurate umpires more evenly distributing their attention capital across all decisions and less accurate umpires focusing much more on the pivotal decisions. If attention is a depletable resource and umpires face pressure to get important calls correct, this observed pattern is rational. Less accurate umpires will invest more effort in getting the important calls correct to avoid making mistakes in critical situations, at the expense of having a smaller stock of attention capital available for lower-stakes decisions.

[23] Ideally, we would like to control for both age and experience because we believe the two may have opposing effects – one would expect the cognitive effort to achieve a given level of accuracy would decrease with experience but increase with advanced age – but the two are highly correlated in our data.
[24] This differs slightly from the average rate across all umpires since umpires are not observed to adjudicate identical numbers of pitches in our sample and less accurate umpires tend to umpire fewer games.

6. Conclusions

Conventional economic models embody agents able to make perfect, optimising decisions. An important strand of recent efforts to increase the behavioral realism of models has been to acknowledge that attention is not costless---the effort required to attend to decisions and execute them well can be costly and cognitively tiring---and incorporate that in models. Models of "strategic inattention", predicated on rational agents adjusting their behavior to account for attention being either limited and/or costly, are increasingly mainstream (for examples Caplin and Dean, 2015; Sims, 2003; Falkinger, 2011).

While the idea of costly attention is intuitively appealing, rigorous evidence characterizing its implications in real settings remains limited and primarily focuses on static effects in cross-sectional data. This paper adds to and extends this evidence. Studying the quality of decisions made by a panel of professional decision-makers with strong incentives to get these decisions right, we show that MLB umpires systematically vary the effort they apply to individual decisions: applying greater attention to those associated with higher stakes. This is consistent with established theoretical models of strategic inattention. Our data-rich setting, in which the same umpire is called upon to issue a long series of decisions, allows for careful study of the *dynamics* of inattention and delivers our most novel results. First, high effort applied early in a sequence of decisions reduces effort applied later in the series. Second, umpires act as if they anticipate high stakes decisions to come later, and conserve cognitive effort. Both results fit closely to the predictions from a model in which the umpire has a depletable stock of attention. These dynamics render inter-dependent otherwise separable decision problems. The short and exogenously mandated break that the umpire receives between half-innings appears sufficient to replenish his stock of attention, since there is no evidence of inter-dependence across those breaks. If repeated in other work settings such evidence could point to the utility of short breaks built into the working day in many cognitively-demanding professions (Gino, 2006).[25] The results prove robust to an array of alternative specifications and robustness tests.

---

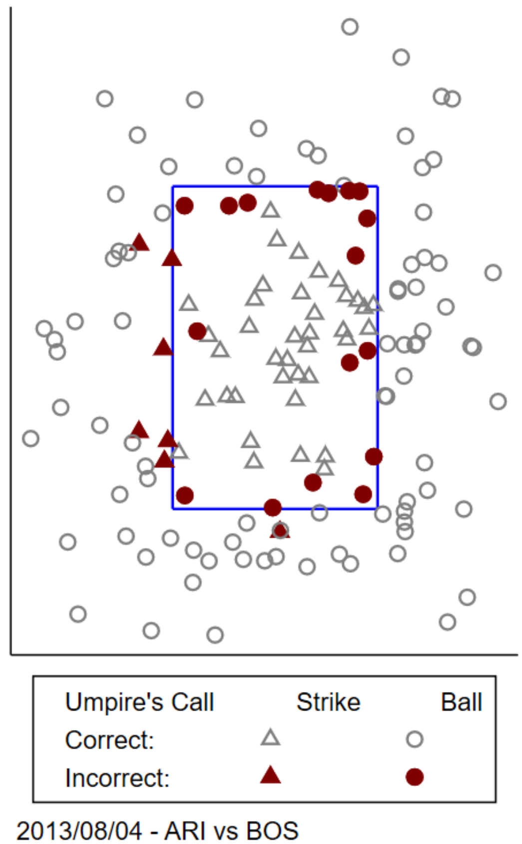
[25] Sievertsen et al (2016) found that the performance of Danish children in standardized tests declined as the time of the test became later in the day ("… because over the course of a regular day, students' mental resources get taxed." (p. 2621). They also found, however, that a twenty-minute break from mental work restored performance.

This is not just a paper about baseball. The richness of the data in a field setting affords a unique opportunity to explore the much broader issue of strategic inattention in novel ways.[26] Moreover, although umpires work in the sports industry, our subjects are not professional athletes, but rather professional decision-makers. Umpires attend specialized training schools, acquire 7-10 years of experience prior to achieving MLB status, are highly paid, and their work highly scrutinized, making their role much closer to a judge than an athlete. As with studies of any industry or profession, there may be concerns about how generalizable the results are. Examining whether similar dynamics of attention are seen in other settings is thus an important next step.

[26] Concerns over external validity from using sports data are more limited here than for some other influential research that have used sports data, for example testing tournament theory using player choices in professional golf tournaments (Ehrenberg and Bognanno, 1990).

Tables and Figures

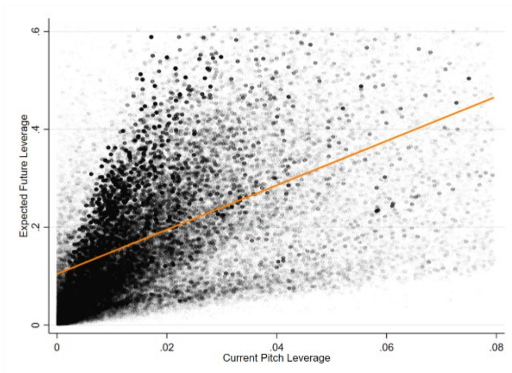
Figure 1 - Example Pitch Location and Umpire Decisions



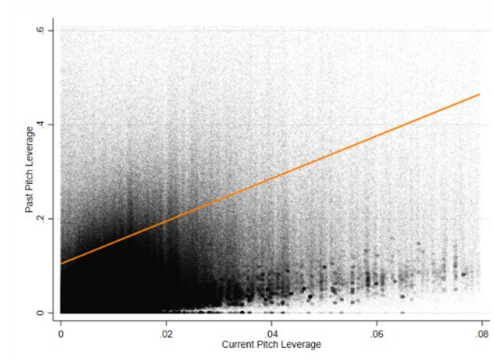
Visualization of pitch locations and umpire decisions from a typical MLB game between Arizona and Boston on August 4th, 2013. This game was selected because the number of total pitches and umpire error rate are close to the sample means. Circles denote pitches called balls and triangles denote called strikes. Filled shapes are incorrect decisions by the umpire. Pitch locations are normalized so boundary of the strike zone, shown as a rectangle, is identical for each pitch. Pitches far from the strike zone (all of which the umpire adjudicated correctly) are excluded from this visualization

Figure 2 - Current and Expected Future Leverage

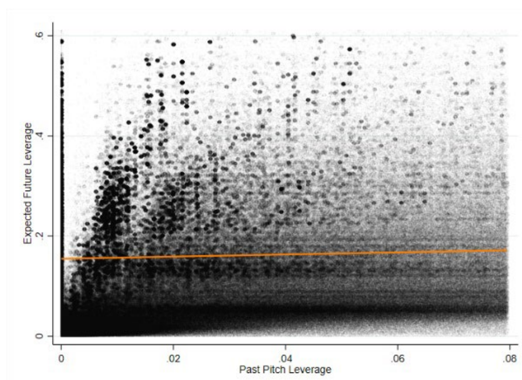
(a) Current v. Future Leverage



(b) Current v. Past Leverage

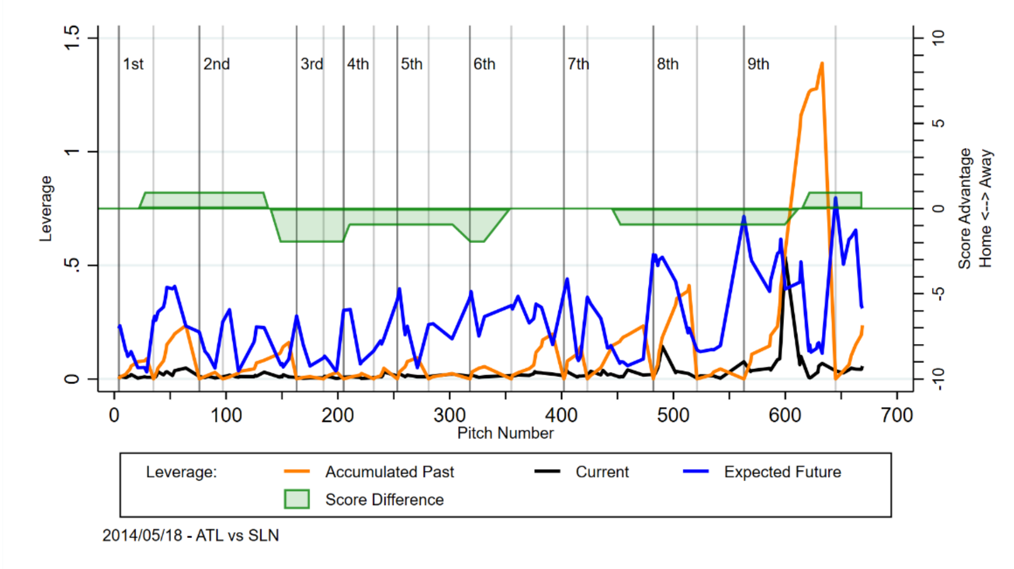


(c) Past v Future Leverage



A scatterplot showing the relationships between current, past, and expected future leverage for each pitch in MLB games during the sample period. Random noise uniformly distributed over 1% of the graph size has been added to each point for clarity. Data are limited to the 99th percentile values on each axis to remove infrequent, extreme values. The orange line is a best-fit regression line based on the full range of data.

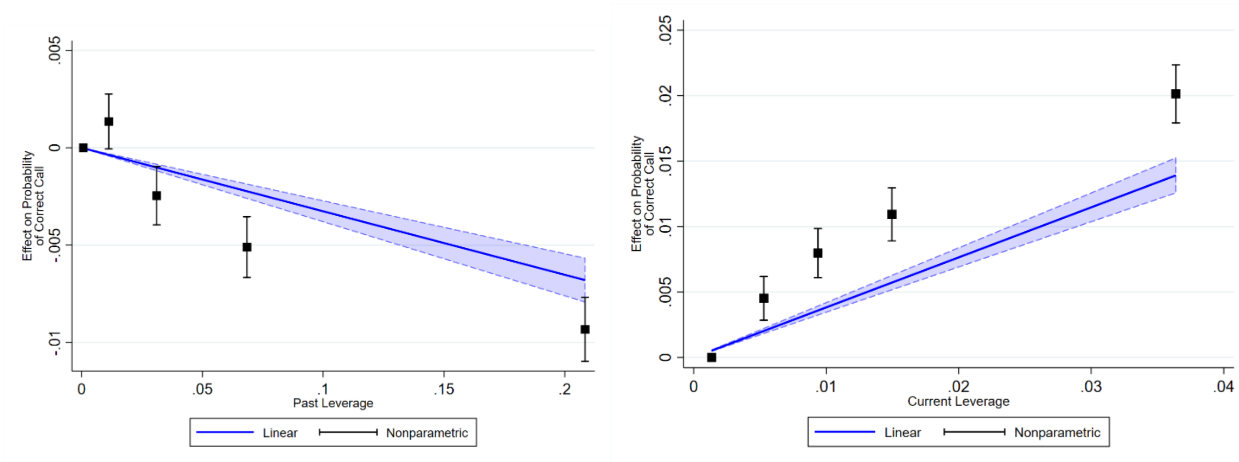
Figure 3 - Example of Evolution of Leverage Metrics



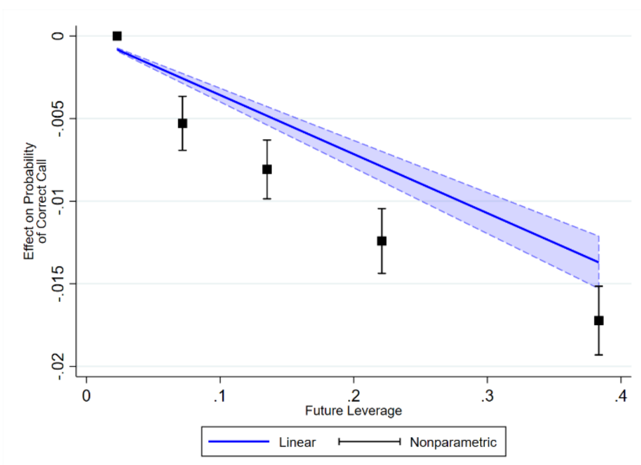
An example of the evolution of the leverage metrics through one MLB game on May 18th, 2014 between Atlanta and St. Louis. The horizontal axis represents each pitch in the game, regardless of whether the home plate umpire was required to make a ball/strike decision. Vertical black lines denote the first pitch of the top (black) or bottom (gray) of each half-inning. The black line represents leverage of the decision for the current pitch. The orange line represents accumulated leverage through the course of the current half-inning and the blue line denotes the expected cumulative leverage for the remainder of the half-inning. The score advantage/disadvantage of the away team is shown as the shaded green area.

Figure 4 – Comparison of Parametric and Nonparametric Effects

Panel (a) – Past Leverage   Panel (b) – Current Leverage



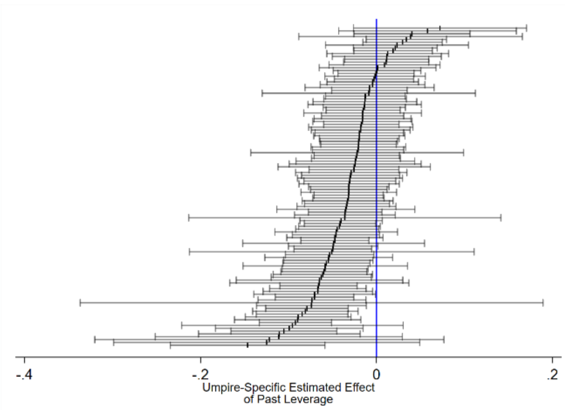
Panel (c) – Future Leverage



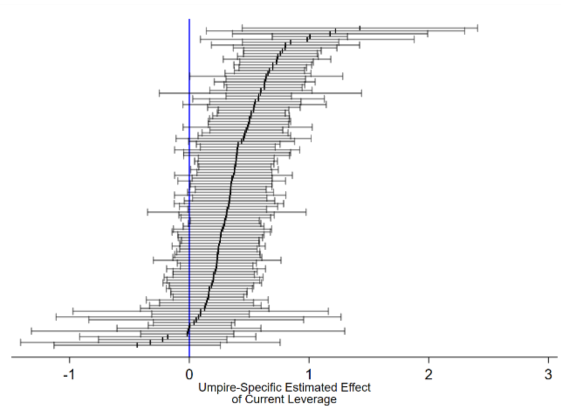
Comparison of the estimated total effects of past (Panel A), current (Panel B), and future (Panel C) leverage on the probability of a correct call from parametric and non-parametric specifications. Parametric estimates, using the specification from Table 4 Column 1, shown as the blue line with the shaded region representing pointwise 95% confidence intervals. Nonparametric estimates for quintiles of observed leverage values with 95% confidence intervals shown as point-and-whiskers. In each case, the first quintile is the omitted category.

Figure 5 – Heterogeneous Effects By Umpire

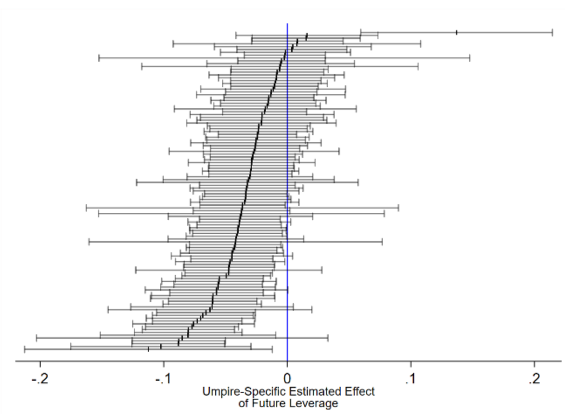
Panel (a) – Past Leverage



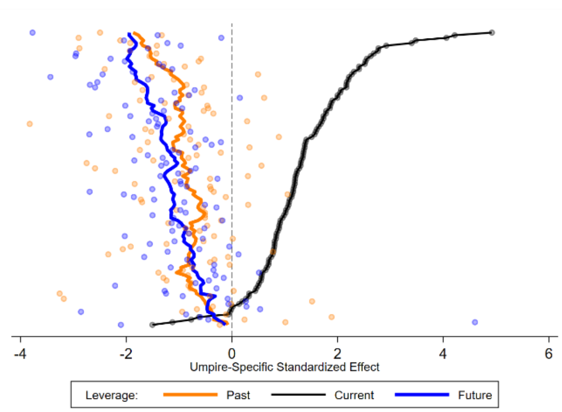
Panel (b) – Current Leverage



Panel (c) – Future Leverage



Panel (d) – All Leverage



Individual-specific estimates of past (Panel A), current (Panel B), and future (Panel C) leverage effects from a single regression. Regression controls are otherwise identical to Table 4 Column 1 and the sample is limited to umpires who are observed calling balls and strikes in at least 20 distinct games. 95% confidence intervals for each estimated effect shown as capped bars. Observations ordered by the estimated effect size. Panel D combines all three estimated effects, ordered by the magnitude of the current leverage effect. For ease of interpretation, effect sizes in Panel D are divided by the standard deviation across all umpires. Dots represent the estimated effect and lines are the moving average for the 10 individuals with larger and 10 individuals with smaller current leverage effects.

Table 1 - Summary Statistics by Game

	(1)
Final Home Team Score	4.338
	(3.070)
Final Away Team Score	4.299
	(3.116)
Game Total Pitches	290.951
	(40.012)
Game Total Called Pitches	119.254
	(19.629)
Game Total Leverage	1.648
	(0.736)
N Games	26,536
First Year	2008
Last Year	2018

Summary statistics for attributes that vary by game. Standard deviations shown in parenthesis.

Table 2 - Summary Statistics by Pitch

	Full Sample (1)	Regression Sample (2)	1st Inning (3)	9th Inning (4)
Correct Call	0.840	0.840	0.842	0.836
	(0.367)	(0.367)	(0.365)	(0.371)
Current Leverage	0.014	0.014	0.014	0.016
	(0.016)	(0.016)	(0.012)	(0.026)
Running sum leverage current half-inning	0.117	0.117	0.120	0.104
	(0.126)	(0.126)	(0.099)	(0.173)
Expected sum future leverage current half-inning	0.170	0.171	0.171	0.204
	(0.136)	(0.138)	(0.086)	(0.226)
Pitch release point (X-axis)	-2.130	-2.110	-2.480	-3.302
	(10.909)	(10.898)	(11.056)	(10.765)
Pitch release point (Y-axis)	27.007	26.987	27.249	28.216
	(4.533)	(4.544)	(4.351)	(4.670)
Pitch release point (Z-axis)	-22.254	-22.352	-20.697	-21.432
	(8.977)	(9.002)	(8.312)	(9.334)
Pitch spin direction (deg)	180.291	180.075	183.862	182.264
	(66.074)	(66.447)	(59.449)	(64.042)
Pitch spin rate (rpm)	1,795.912	1,789.485	1,894.113	1,842.551
	(670.399)	(672.427)	(632.757)	(676.447)
Pitch initial velocity (mph)	87.986	87.941	88.562	89.733
	(6.018)	(6.045)	(5.535)	(5.840)
Pitch break angle (deg)	5.313	5.268	6.099	8.315
	(24.862)	(24.787)	(26.027)	(24.975)
Pitch break length (in)	6.466	6.495	6.028	6.047
	(2.946)	(2.956)	(2.713)	(2.834)
Pitch break (Y-axis)	23.803	23.803	23.802	23.796
	(0.100)	(0.100)	(0.101)	(0.098)
Pitch final velocity (mph)	81.019	80.979	81.541	82.541
	(5.390)	(5.412)	(4.965)	(5.207)
N	3,164,525	2,971,642	200,174	265,885

Summary of attributes that vary across each pitch. Standard deviations shown in parenthesis. Column 1 summarizes all pitches in the data for which an umpire makes a ball/strike decision. Column 2 limits the sample to observations where all covariates from our primary regressions are non-missing.

Table 3 - Examples of Current and Future Leverage by Situation

(a) 25th Percentile Current Leverage

Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Bottom 5	2	0	1	2nd	9	.00485044	.00485303
Top 7	0	2	1	2nd 3rd	-3	.00484496	.57780694

(b) 50th Percentile Current Leverage

Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Bottom 6	2	0	2	3rd	3	.00973686	.02099236
Top 3	1	2	0	2nd 3rd	-3	.00972773	.36656121

(c) 75th Percentile Current Leverage

Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Top 3	2	3	2	1st	-5	.01773382	.04311057
Bottom 7	1	3	1	2nd 3rd	0	.0177395	.36116757

(d) 95th Percentile Current Leverage

Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Bottom 1	2	2	2	2nd	-3	.04212482	.0752991
Bottom 9	0	1	2	None	-2	.04213369	.52704869

(e) 99th Percentile Current Leverage

Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Top 6	0	1	2	2nd 3rd	1	.07954606	.29869399
Top 9	0	0	2	1st 2nd 3rd	-3	.07953886	.72586055

Examples of the difference in expected future leverage for situations with similar current-pitch leverage. Current leverage situations selected to be at the specified percentile of the observed leverage distribution in actual MLB games. The examples provided are the most extreme differences in expected future leverage for all situations with current leverage within 0.00001 of the percentile value. Score Diff. is the score advantage (positive) or deficit (negative) of the currently batting team. Current Leverage is the absolute change in the probability the batting team wins should the umpire call a strike versus a ball. Future Leverage is the expected sum of leverage on all pitches for the remainder of the half-inning.

Table 4 - Lag Leverage Only Innings 2-9

	No Lag (1)	1 Inning Lags (2)	2 Inning Lags (3)
Current Leverage	38.223	38.101	38.037
	(1.866)***	(1.868)***	(1.869)***
Past Leverage (Current Inning)	-3.260	-3.246	-3.239
	(0.274)***	(0.274)***	(0.275)***
Expected Future Leverage (Current Inning)	-3.575	-3.571	-3.573
	(0.210)***	(0.210)***	(0.211)***
Inning 3	0.102	0.098	0.099
	(0.086)	(0.086)	(0.086)
Inning 4	-0.061	-0.061	-0.062
	(0.088)	(0.088)	(0.088)
Inning 5	0.043	0.043	0.042
	(0.087)	(0.088)	(0.088)
Inning 6	0.023	0.023	0.023
	(0.088)	(0.088)	(0.088)
Inning 7	0.016	0.015	0.013
	(0.088)	(0.088)	(0.088)
Inning 8	-0.060	-0.061	-0.066
	(0.088)	(0.088)	(0.089)
Inning 9	-0.391	-0.389	-0.392
	(0.097)***	(0.097)***	(0.097)***
Lag Leverage Inning - 1		-0.011	-0.004
		(0.197)	(0.197)
Lag Leverage Inning - 2			0.035
			(0.204)
N	2,712,508	2,710,491	2,708,875
N Clusters	26,535	26,535	26,535
Mean Correct	0.84	0.84	0.84

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning. Lag leverage is the average of the leverage measure for all ball/strike decisions by the umpire during a previous half-inning. Regressions include game fixed effects, inning, and inning part fixed effects. Estimates limited to innings 2-9.

Table 5 - Leverage Effects by Inning

	Inning 2 (1)	Inning 3 (2)	Inning 4 (3)	Inning 5 (4)	Inning 6 (5)	Inning 7 (6)	Inning 8 (7)	Inning 9 (8)
Current Leverage	40.444	53.013	53.426	40.206	45.724	44.275	35.871	16.037
	(6.048)***	(6.088)***	(5.917)***	(5.844)***	(5.449)***	(4.987)***	(4.453)***	(4.283)***
Past Leverage (Current half-inning)	-4.117	-5.463	-6.425	-3.942	-2.129	-2.986	-2.140	-2.072
	(0.874)***	(0.934)***	(0.902)***	(0.880)***	(0.755)***	(0.729)***	(0.667)***	(0.628)***
Expected Future Leverage (Current half-nning)	-4.738	-5.013	-5.593	-3.893	-3.456	-3.661	-3.442	-2.011
	(0.776)***	(0.762)***	(0.713)***	(0.666)***	(0.595)***	(0.535)***	(0.476)***	(0.436)***
Inning Effect	0.000	0.061	0.045	-0.110	-0.383	-0.284	-0.339	-0.652
	(0.000)	(0.203)	(0.196)	(0.189)	(0.181)**	(0.176)	(0.172)**	(0.178)***
N	2,712,508	2,712,508	2,712,508	2,712,508	2,712,508	2,712,508	2,712,508	2,712,508
N Clusters	26,535	26,535	26,535	26,535	26,535	26,535	26,535	26,535
Mean Correct	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning. Lag leverage is the average of the leverage measure for all ball/strike decisions by the umpire during a previous half-inning. Regressions include game fixed effects. Estimates limited to innings 2-9.

Table 6 - Alternative Regression Controls

	Pref. Spec (1)	Pitch Controls (2)	Alternative FEs (3)	Coarse FEs (4)	Player FEs (5)
Current Leverage	38.223	37.454	37.842	37.702	39.809
	(1.866)***	(1.869)***	(1.853)***	(1.853)***	(1.873)***
Past Leverage	-3.260	-3.590	-3.089	-3.075	-3.258
(Current half-inning)	(0.274)***	(0.275)***	(0.266)***	(0.266)***	(0.274)***
Expected Future Leverage	-3.575	-3.256	-3.444	-3.443	-3.486
(Current half-inning)	(0.210)***	(0.211)***	(0.202)***	(0.202)***	(0.213)***
N	2,712,508	2,712,508	2,712,508	2,712,508	2,712,419
N Clusters	26,535	26,535	26,535	26,535	26,535
Mean Correct	0.84	0.84	0.84	0.84	0.84
Controls	Game FE	Game FE	Umpire, Date	Umpire, Yr, MOY,DOW	Game FE
		Pitch Attribs	Team FEs	Team FEs	Player FEs

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning prior to the current pitch. Expected future leverage is the expected sum of leverages for all future pitches this half-inning. Column (1) repeats the primary specification. Column (2) adds controls for the attributes of the pitch, including velocity, break, starting position, spin rate, and pitch type. Column (3) removes game fixed effects and replaces them with umpire, home team, away team, and date fixed effects. Column (4) further replaces game date fixed effects with year, month-of-year, and day-of-week fixed effects. Column (5) adds fixed effects for the identity of the current pitcher and batter to the primary specification.

Table 7 - Determinants of Umpire-Specific Leverage Effects

	Current Leverage (1)	Expected Future (2)	Accumulated Past (3)
Age (years)	0.106	0.000	-0.033
	(0.241)	(0.027)	(0.034)
Rate of Correct Calls	-334.822	33.565	27.061
	(63.035)***	(7.213)***	(10.982)**

Regression of umpire attributes on umpire-year-specific slope coefficients from our primary specification. Coefficients multiplied by 100 for legibility. Standard errors clustered by umpire shown in parentheses with *,**,*** denoting coefficients significant at the 10, 5, and 1% levels, respectively. Regressions weighted by the number of games during which the umpire worked at home plate during the corresponding season.

Appendix 1 Robustness

A1.1: Results Excluding the Ninth Inning

The leverage measure employed in our analyses is the effect an individual decision will have on the likelihood of a given team winning a game. By construction, this leverage measure will tend to be larger toward the end of games, where there are fewer decisions remaining for the umpire and a single decision can be pivotal to the outcome of a game. As test of robustness, we re-estimate our primary model, limiting our sample to the eighth inning and earlier. Results are shown in the tables below. Estimated parameters are not substantially different from models where we include the ninth inning.

Table 8 - Robustness: Omit 9th Inning

	Incl. Lag 0 innings (1)	Incl. Lag 1 innings (2)	Incl. Lag 2 innings (3)
Current Leverage	43.990 (2.068)***	43.942 (2.070)***	43.881 (2.071)***
Past Leverage (Current half-nning)	-3.506 (0.305)***	-3.494 (0.305)***	-3.488 (0.306)***
Expected Future Leverage (Current half-inning)	-3.946 (0.239)***	-3.944 (0.239)***	-3.949 (0.240)***
Lag Leverage half-inning - 1		0.065 (0.219)	0.072 (0.219)
Lag Leverage half-nning - 2			0.051 (0.228)
N	2,446,325	2,444,606	2,443,182
N Clusters	26,534	26,534	26,534
Mean Correct	0.84	0.84	0.84

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning. Lag leverage is the average of the leverage measure for all ball/strike decisions by the umpire during a previous half-inning. Regressions include game fixed effects. Estimates limited to innings 3-9.

Table 9 - Robustness: Omit 9th Inning, Inning Level Effects

	Inning 2 (1)	Inning 3 (2)	Inning 4 (3)	Inning 5 (4)	Inning 6 (5)	Inning 7 (6)	Inning 8 (7)
Current Leverage	40.609	53.023	53.165	40.464	45.555	44.167	35.652
	(6.047)***	(6.089)***	(5.924)***	(5.846)***	(5.442)***	(4.990)***	(4.455)***
Past Leverage	-4.136	-5.397	-6.386	-3.975	-2.175	-3.027	-2.086
	(0.874)***	(0.938)***	(0.905)***	(0.883)***	(0.756)***	(0.733)***	(0.667)***
Expected Future Leverage	-4.738	-4.949	-5.608	-3.894	-3.426	-3.634	-3.396
	(0.777)***	(0.764)***	(0.714)***	(0.667)***	(0.596)***	(0.537)***	(0.477)***
Inning Effect	0.000	0.045	0.050	-0.110	-0.386	-0.282	-0.344
	(0.000)	(0.203)	(0.196)	(0.190)	(0.181)**	(0.176)	(0.172)**
N	2,446,325	2,446,325	2,446,325	2,446,325	2,446,325	2,446,325	2,446,325
N Clusters	26,534	26,534	26,534	26,534	26,534	26,534	26,534
Mean Correct	0.84	0.84	0.84	0.84	0.84	0.84	0.84

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning. Lag leverage is the average of the leverage measure for all ball/strike decisions by the umpire during a previous half-inning. Regressions include game fixed effects. Estimates limited to innings 3-9.

A1.2: Results Using Actual Instead of Simulated Leverage

Our measure of the leverage at each pitch requires that we compute two probabilities in each game situation: the probability a given team wins in the event of a called ball and the probability they win if there is a called strike. Although we use simulated data, it is possible to empirically calculate these probabilities using observed outcomes in MLB games. However, while we have a wealth of data on which to base these estimates (over three million pitches in over 26,500 games), the space of possible situations is also large, and using actual game data to compute leverage could lead to substantial measurement error.[27] To address this concern, the leverage measure used as the basis of our primary specifications is derived from simulations of 5 million MLB games.

[27] Accounting for all possible combinations of balls, strikes, outs, baserunner positions, inning and inning part, and score differences between a 10-run advantage and a 10-run disadvantage, there are 108,864 possible states. Some states occur very frequently, e.g., every game starts in an identical state, and some states are not observed at all. A given state is observed, on average, around 30 times over the course of our data. If the probability of a team winning were 0.50, the estimated probability based on 30 observations would have a standard error of approximately 0.09. The standard error in leverage estimated this way would be even larger since it is the difference of two such probabilities measured with error. Computing the standard error of leverage requires knowledge of the covariance in the two estimated win probabilities. Applying the Cauchy-Schwarz inequality, we can bound the standard error of a leverage measure based on two outcomes observed thirty times each to [0.1286,0.1296].

A test of the robustness of using this simulated leverage metric would be to re-estimate our primary specifications using leverage computed using a leverage measure based on game data-only (GDO), with the understanding that it is poorly measured. Assuming classical measurement error, we would expect the parameter on leverage to be biased toward zero in these estimates.[28]

As a prerequisite, we first compute the degree of attenuation bias that can be expected given the measurement error in the game data-only leverage measure. Assuming the simulated leverage measure is the "true" measure of leverage on every pitch, attenuation bias is a function of the variance of the true measure divided by the sum of the variance of the true measure and the variance of the error in the noisy measure. Using estimates of these variances from our observed data, we can then compute the expected ratio of parameters from our preferred specification to those from a specification using the GDO leverage measure.

Measurement error in the game data-only measure is driven by sampling variance. Therefore, values of the leverage metric based on game states that are observed more often should be more precise. A natural approach to reducing measurement error would be to limit the sample to situations that are observed more frequently, with lower sampling variance. Table 9 recomputes the variance in the true leverage measure ("signal") and the sampling variance ("noise"), limiting the sample to observations where the game data-only measure is computed using at least 1, 100, 250, 750, 1000, and 2500 observations. Then, using these variances, we compute the ratio between the "true" parameter and the expected value of a parameter when the corresponding independent variable is measured with error.

The results of this table demonstrate the large impact measurement error might have on the estimated model parameters. Using the full sample, the true effect of leverage would be over 8 times value of the parameter one would expect to observe given the magnitude of measurement error in the GDO leverage measure. This bias decreases steadily as we limit the sample to situations observed more frequently, but is still over 3 times the value when considering situations occurring over 250 times in our sample.[29]

---

[28] If measurement error in the game data-only measure arises only from sampling variation in the estimated win probability in each state, then the measurement error is orthogonal to any other unobserved variables in our regression and meets the criteria of classical measurement error.

[29] As the minimum number of games increases over 750, the bias factor increases. This results from the fact that while increasing the number of games threshold reduces sampling variance (noise), it also reduces variance in the true leverage measure (signal) as the set of game situations in the sample decreases.

Table 10 - Estimated Bias from Measurement Error in Game Data Leverage

Minimum Num Games	Variance Signal	Variance Noise	Signal-to Noise Ratio	Bias Factor $b / \hat{b}$
1	0.0001764	0.0012982	0.1196263	835.90%
100	0.0001532	0.0004344	0.2607022	383.60%
250	0.0001390	0.0003351	0.2931118	341.20%
750	0.0001170	0.0002416	0.3262512	306.50%
1000	0.0000960	0.0002323	0.2924336	342.00%
2500	0.0000816	0.0002147	0.2753444	363.20%

Estimates of the magnitude of attenuation bias due to measurement error by using actual game, as opposed to simulated, outcomes to compute leverage. Assumes simulated leverage is the "true" measure of leverage. The "signal" is this true value (x). "Noise" is the difference between the GDO leverage measure and the simulated measure for a given pitch (u). Under classical measurement error, attenuation bias is proportional to the signal-to-noise ratio $\left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}\right)$ in the probability limit. The Bias Factor is the ratio of the true b (absent attenuation bias) and the estimated b when using the GDO leverage measure assuming classical measurement error. Minimum number of games denotes the minimum number of games on which the GDO leverage measure is based.

We follow by estimating the impact of leverage on the umpire making a correct call using the GDO leverage measure. That is, for each pitch in each of these games, we define the situation as the score differential[30], current inning[31], inning part, number and position of baserunners, number of outs, number of balls, and number of strikes. The estimated probability of the home team winning conditional on that situation is the proportion of games where that situation occurred to games where that situation occurred and the home team won.[32] We then compute the leverage in some situation A_t as the difference in win probability for the situation A_t incremented by a ball and the situation A_t incremented by a strike. This method of computing leverage requires minimal assumptions, only that events in a baseball game follow a Markov process with a state defined by the game situation variables. However, in spite of our large dataset consisting of over 26,000 individual games, the large state space leaves some relevant states unobserved or so infrequently observed that the win probabilities for these states are poorly estimated.[33]

[30] We limit to cases where the score difference between teams is 10 or less.
[31] MLB games that are tied after nine innings continue one inning at a time until the tie is broken at the end of the inning. Consistent with our assumption that states evolve as a Markov process, we treat any inning after the 9th inning as the 9th for the purposes of computing the state.
[32] If a given situation occurs multiple times in a game – which frequently occurs when a batter hits a foul ball with two strikes – it is only counted once for the purposes of this calculation.
[33] The state space consists of 21 possible run differences, nine innings, two inning parts (called half-innings), three outs, eight possible arrangements of runners on the bases, three strike states, and four ball states. This is a total of

To further investigate the role of measurement error, we again limit the sample to cases where the GDO measure is computed using data from at least 1, 100, 250, 750, 1000, and 2500 unique occurrences in MLB games. For comparison, we also estimate the leverage effect using the same sample of observations and our leverage measure derived from simulated games. Finally, we compute the bias factor dividing the coefficient on the simulated measure to the coefficient on the GDO measure. The results are shown in Table 10.

These estimates demonstrate two key advantages to using the simulated measure. First, the discrepancy between results using the simulated and GDO measure are broadly consistent with the magnitudes estimated in Table 9, declining to approximately 300% when limiting to cases where the GDO measure is based on at least 2500 observations. Second, limiting the sample to cases where the GDO measure is based on more observations increases the magnitude of the estimated coefficient in the simulated leverage regressions. Situations with few or many underlying observations on which to base the calculation of the GDO leverage measure are not randomly assigned and limiting the sample in this way can bias the estimated coefficients. Using the simulated leverage measure avoids both issues. While this table reveals discrepancies between estimates, the qualitative results still hold.

Table 11 - Comparison of Estimated Leverage Effects from Different Measures

	Leverage Effect		
Minimum Num Games	GDO	Simulated	Bias Factor $b / \hat{b}$
1	-0.381	15.149	-3979.20%
100	1.577	19.804	1255.40%
250	4.133	23.668	572.60%
750	10.250	47.772	466.10%
1000	7.558	43.940	581.30%
2500	15.384	44.675	290.40%

Estimates from linear probability models that the umpire makes the correct call of a given pitch. “Simulated Leverage” estimates computed using our preferred leverage measure from 5 million simulated MLB games. “GDO” (Game Data Only) estimates computed using a leverage measured derived only from actual game data. Standard errors clustered at the game level shown in parenthesis. Attenuation ratio shows the ratio of the estimated coefficients from each model. All coefficients and standard errors multiplied by 100 for legibility. Regressions include only contemporaneous leverage, game, inning, and inning part fixed effects. The first column uses all non-missing observations. Each subsequent column limits to observations where the GDO leverage measure is computed using a minimum of the number of games shown in the column header.

108,864 possible states. A typical game will pass through around 300 unique states. Given some states are more likely to occur than others (e.g., the state in the top of the first inning, tied game, zero base runners, balls, and strikes occurs in every game) there is incomplete coverage of the state space.

A1.3: Results By Current Score

One might be concerned that umpires may make mistakes that favor a team that is behind in a game, and this may be correlated with the leverage of a given series of pitches. In the table below we repeat our primary specification across three subsamples of the data: pitches where the batting team is leading in the score, behind, or tied. Leverage effects are weaker when the batting team is losing, but have identical sign and significance to estimates from the full sample.

Table 12 – Effects by Current Score Differential

	Primary Spec (1)	Batting Team Winning (2)	Batting Team Losing (3)	Teams Tied (4)
Current Leverage	68.866	61.235	30.847	46.420
	(1.509)***	(4.844)***	(2.572)***	(3.660)***
Past Leverage	-3.990	-2.083	-4.843	-3.728
(Current Inning)	(0.221)***	(0.472)***	(0.446)***	(0.628)***
Expected Future Leverage	-6.094	-5.769	-3.460	-5.252
(Current Inning)	(0.168)***	(0.705)***	(0.293)***	(0.580)***
N	3,278,210	1,043,466	1,129,494	539,042
N Clusters	26,540	26,294	26,358	19,064
Mean Correct	0.87	0.84	0.84	0.84

Estimates from linear probability model that the umpire makes the correct call for a given pitch. Standard errors clustered at the game level shown in parenthesis. All coefficients and standard errors multiplied by 100 for legibility. Past leverage is the total of current leverage in the current half-inning. Lag leverage is the average of the leverage measure for all ball/strike decisions by the umpire during a previous half-inning. Regressions include game fixed effects. Estimates limited to innings 3-9.

A1.4: Flexible Controls for Pitch Attributes

The PitchFX data include a variety of attributes of each pitch thrown (e.g., velocity, acceleration, spin rate, etc.). These attributes are the outcome of a pitcher's decision on the precise pitch to throw and is determined simultaneously with the umpire's level of effort to exert adjudicating the pitch in response to the current, past, and expected future leverage. In Section 5.3 we argue these pitch attributes are "bad controls" (Angrist and Pishke, 2008) and including them will likely bias our estimates toward zero. As a test of robustness, Table 6 Column 2 adds controls for these pitch attributes to our primary specification, with only small changes to our parameters of interest.

As a further test of robustness, we consider alternative specifications which allow for a flexibly-estimated nuisance function of pitch attributes in our specification. Specifically we replace our primary specification with

$$\mathbf{1}(C_p^* = C_p) = \beta^C L(A_p) + \beta^P \sum_{\tau=1}^{p-1} L(A_\tau) + \beta^F \boldsymbol{E_p}\left[\sum_{\tau=p+1}^{\infty} L(A_\tau)\right] + f(X_p) + \boldsymbol{\delta_p^I} + \delta_g^G + \epsilon_{pg}$$

Where $f(X_p)$ is a flexible function of pitch attributes. An initial path to flexibly estimating this function would be to include higher-order interactions of the pitch attribute controls in our specification. This approach, however, is infeasible. High-order interactions of some pitch attributes are mathematically equivalent to higher-order interactions of other attributes.[34] The PitchFX data include only 3 or 4 significant figures, so mechanical equivalence manifests as approximate, but not perfect collinearity and typical methods of removing collinear variables from a regression are ineffective. Given the large space of high-order interactions we were unable to identify all the mechanical relationships and any regression including second- or higher-order pitch attribute interactions had ill-conditioned variance-covariance estimates.

As an alternative, we consider alternative methods for estimating the nuisance function $f(X_p)$ that are robust to potential mechanical relationships in high-order interactions. First, we follow the empirical design above, allowing for higher-order interactions of pitch attributes in $f(X_p)$, but reducing the space of included controls using lasso, following the post-double selection lasso (PDS-Lasso) algorithm of Belloni et al. (2012), which are shown in Table 13 below. We were able to estimate regressions including up to 4th-order interactions of the pitch attributes. Consistent with pitch attributes being "bad controls", including their higher-order interactions somewhat attenuates our parameter estimates, but they maintain sign and significance with our primary specification.

Table 13 – High-order Pitch Attribute Contols using PDS-Lasso

	Primary Spec. (1)	PDS with 1-order (2)	PDS with 2-order (3)	PDS with 3-order (4)	PDS with 4-order (5)
Current Leverage	38.223	37.483	26.099	22.480	21.794
	(1.866)***	(1.868)***	(1.859)***	(1.857)***	(1.855)***
Past Leverage	-3.260	-3.414	-3.040	-3.002	-3.003
(Current Inning)	(0.274)***	(0.275)***	(0.274)***	(0.273)***	(0.273)***
Expected Future Leverage	-3.575	-3.211	-2.325	-2.086	-2.032
(Current Inning)	(0.210)***	(0.210)***	(0.210)***	(0.210)***	(0.210)***
N	2,712,508	2,712,508	2,712,508	2,712,508	2,712,508
N Clusters	26,535	26,535	26,535	26,535	26,535
Mean Correct 1em	0.84	0.84	0.84	0.84	0.84
Attribute Interactions	1	1	2	3	4
Feature Selection	None	PDS	PDS	PDS	PDS

Regression of the probability a call is correct on leverage measures, game and inning fixed effects and pitch attribute controls. Column 1 repeats the primary specification. Column 2 uses the primary specification but instead chooses pitch attribute controls using the Post-double selection Lasso (PDS-Lasso) algorithm of Belloni et al (2012). Columns 3 to 5 interact pitch attribute controls with themselves and relevant controls are selected using the PDS-Lasso algorithm. Standard errors clustered by game shown in parentheses.

[34] E.g., The pitch attributes include pitch initial velocity and three variables representing the components pitch's initial velocity vector in each of the x, y, and z dimensions. While these variables are not collinear, the square of velocity is be definition equal to the sum of the squares of each component of the velocity vector, introducing collinearity in the second-order interactions.

Rather than limit our estimates to the linear functional form, we also estimate the nuisance function $f(X_p)$ using gradient-boosted trees, a flexible regression tree-based method, and our parameters of interest in a linear regression as a Partly Linear Regression (PLR) following the Debiased/Double Machine Learning (DDML) approach of Chernozhukov et al. (2018).[35] The estimates are shown in Table 14. Consistent with pitch attributes being bad controls, our estimates are attenuated toward zero, but maintain sign and significance.

Table 14 - High-order Pitch Attribute Contols using PLR and DDML

	Primary Spec. (1)	PLDDML Iteration (2)	PLDDML Iteration (3)	PLDDML Iteration (4)
Current Leverage	38.223	17.991	17.822	18.020
	(1.866)***	(1.741)***	(1.740)***	(1.737)***
Past Leverage	-3.260	-1.976	-1.928	-1.975
(Current Inning)	(0.274)***	(0.246)***	(0.246)***	(0.246)***
Expected Future Leverage	-3.575	-1.687	-1.663	-1.648
(Current Inning)	(0.210)***	(0.256)***	(0.256)***	(0.256)***
1em				
Nuisance Function via	None	xgboost	xgboost	xgboost
Feature Selection	None	DDML	DDML	DDML
RNG State	NA	123456	8675309	3141592654

Coefficients on leverage measures from Double/Debiased-ML (DDML) of Chernozhukov et al. (2018) for three initial states of the random number generator. The pitch attribute nuisance function is estimated using gradient-boosted trees. DDML implemented using 10 cross-validation folds clustered by game and 10 repetitions. Standard errors clustered by game shown in parentheses. DDML utilizes the random number generator (RNG) as part of its estimation process. Following best practice, we repeat estimation for several choices of the initial RNG state in columns (2) to (4).

[35] Specifically, we estimate the nuisance function using gradient-boosted trees. Growing 100 trees, using MSE as the loss function. We treat all categorical variables as unordered categorical data when partitioning leaves. Each branch partition may contain as many as 8 leaves. We employ the DDML procedure using 10 cross-validation folds and 10 repetitions.

References

Acharya, S. and S. L. Wee (2020). "Rational Inattention in Hiring Decisions." American Economic Journal: Macroeconomics 12(1): 1-40.

Allcott, H. and N. Wozny (2014). "Gasoline Prices, Fuel Economy and the Energy Paradox." Review of Economics and Statistics 96(5): 779-95.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Archsmith, J., A. Heyes and S. Saberian (2018). "Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focussed Occupation." Journal of the Association of Environmental and Resource Economists 5(4): 827-63.

Augenblick, N. and S. Nicholson (2016). "Ballot Position, Choice Fatigue, and Voter Behavior." Review of Economic Studies 83(2): 460-80.

Baer, D. 2015. "The Scientific Reason why Barack Obama and Mark Zuckerberg wear the same Outfit Every Day". Business Insider. (Retrieved 24 October 2018).

Bartos, V., M. Bauer, J. Chytilova and F. Metejka (2016). "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." American Economic Review 106(6): 1437-75.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). "Ego depletion: Is the Active Self a Limited Resource?" Journal of Personality and Social Psychology 74(5), 1252-65.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." Econometrica, 80(6), 2369–2429.

Bhattacharya, V. and Howard, G., 2022. "Rational Inattention in the Infield." American Economic Journal: Microeconomics. 14(4), 348-93.

Bruyneel, S. S. Dewitte and K. Vohs (2006). "Repeated Choosing Increases Susceptibility to Affective Product Features." International Journal of Research in Marketing 23(2): 215-25.

Caplin, A. and M. Dean (2015). "Revealed Preference, Rational Inattention and Costly Information Acquisition." American Economic Review 105(7): 2183-2203.

Chan, M. Y., H. Cohen and B. Spiegel (2009). “Fewer Polyps Detected by Colonoscopy as the Day Progresses.” Clinical Gastroenterology and Hepatology 7(11): 1217-23.

Chen, D. T. Moskovitz and K. Shue (2016). “Decision-making under the Gambler’s fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires.” Quarterly Journal of Economics 131(3): 1181-1242.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). “Double/debiased machine learning for treatment and structural parameters.” Econometrics Journal, 21(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chetty, R., A. Looney and K Kroft (2009). “Salience and Taxation: Theory and Evidence.” American Economic Review 99(4): 1145-1177.

Correia, S., (2014) "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects," Statistical Software Components S457874, Boston College Department of Economics, revised 18 Nov 2019.

DellaVigna, S. and J. Pollett (2009). “Investor Inattention and Friday Earnings Announcements.” Journal of Finance 64(2): 709-49.

Dragone, Davide (2009). “I Am Getting Tired: Effort and Fatigue in Intertemporal Decision-making.” Journal of Economic Psychology 30(4): 552-62.

Drellich, E. (2012). Complex system in place to evaluate umpires. https://www.mlb.com/news/c-37468304/print.

Ehrenberg, R. G. and M. Bognanno (1990). “Do Tournaments Have Incentive Effects?” Journal of Political Economy 98(6): 1307-24.

Erwin, M. (2019). “6 Reasons We Make Bad Decisions, and What to Do About Them.” Harvard Business Review.

Falkinger, J. (2011). “Limited Attention as a Scarce Resource in Information-Rich Economies.” Economic Journal 118(3): 1596-1620.

Gabaix, X. (2019). "Behavioral Inattention." Chapter 4 in Eds. D. Bernheim, S. Dellavigna and D. Laibson *Behavioral Economics: Application and Foundations* (Volume 1), Elsevier: North-Holland.

Gabaix, X., D. Laibson, G. Moloche and S. Weinberg (2006). "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." American Economic Review 96(4): 1043-1068.

Gino, F. (2016). "Don't Make Important Decisions Late in the Day." Harvard Business Review 23.

Green, E. A. and D. P. Daniels (2021). "Bayesian Instinct." DOI: 10.2139/ssrn.3916929

Hagger, M., N. L. Chatzisarantis, M. Zweinenberg et al (2016). "A Multilab Preregistered Replication of the Ego-Depletion Effect." Perspectives on Psychological Science 11(4): 546-73.

Hirshleifer, D. A., Y. Levi, B. Lourie and S. H. Teoh (2019). "Decision Fatigue and Heuristic Analyst Forecasts." Journal of Financial Economics 133(1): 83-98.

Kim, J. and B. King (2014). "Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring." Management Science 60(11): 2381-2617.

Kim R. H., S. Day, D. Small C., Snider, C. Rareshide and M. Patel (2018). "Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination." JAMA Network Open 1(5).

Kwan, J., L. Stein et al (2016). "Does Decision Fatigue Impact Manuscript Acceptance? An Analysis of Editorial Decisions by the American Journal of Gastroenterology." American Journal of Gastroenterology 111(11): 1511-12.

Levav, J., M. Heitmann, A. Herrmann and S. Iyengar (2010). "Order in Product Customization Decisions: Evidence from Field Experiments." Journal of Political Economy 118(3): 274-99.

Linder J. A., J. N. Doctor and A. W. Friedberg (2014). "Time of Day and the Decision to Prescribe Antibiotics". JAMA Internal Medicine 174(12): 2029 – 2031.

Orasunu, J. B. Parke, N. Kraft et al (2012). "Evaluating the Effectiveness of Schedule Changes for Air Traffic Service (ATS) Providers: Controller Alertness and Fatigue Monitoring Study." Available at: https://trid.trb.org/view/1364660.

Ozdenoren, E., S. W. Salant and D. Silverman (2012). "Willpower and the Optimal Control of Visceral Urges." Journal of the European Economic Association 10(2): 342 – 68.

Parsons, C., J. Suleman, M. Yates and D. Hamermesh (2011). "Strike Three: Discrimination, Incentives and Evaluation". American Economic Review 101(4): 1410 – 35.

Philpot L, B. A. Khokhar, D. L. Roellinger, P Ramar and J. O. Ebbert (2018). "Time of Day is Associated with Opioid Prescribing for Low Back Pain in Primary Care." Journal of General Internal Medicine 33(11): 1828-1830.

Sacheti, A, Gregory-Smith, I and Paton, D (2014), "Home Bias in Officiating: Evidence from International Cricket", Journal of the Royal Statistical Society: Series A (Statistics in Society).

Sievertsen, H. H., F. Gino and M. Piovesan (2016). "Cognitive Fatigue Influences Student Performance in Standardized Tests." Proceedings of the National Academy of Sciences of the United States of America (PNAS) 113(10): 2621-24.

Sims, C. (2003). "Implications of Rational Inattention". Journal of Monetary Economics 50(3): 665-90.

Zheng, B., E. Kwok, M. Taljaard (2020). "Decision Fatigue in the Emergency Department." The American Journal of Emergency Medicine 38(12): 2506-2510.