The Dynamics of Inattention in the (Baseball) Field*

James Archsmith†
University of Maryland, College Park

Anthony Heyes‡
University of Birmingham

Matthew Neidell§
Columbia University

Bhaven Sampat¶
Arizona State University

Draft Date: January 21, 2025

Preliminary and incomplete

Please do not circulate or cite without permission of the Author.

Please review and cite the most recent version of this paper available at
https://econjim.com/WP2101

*We are grateful to Stephen Coussens, Jonathan Guryan, Sandeep Kapur, Adam Leive, Devin Pope, Stephen Salant, four reviewers from this journal and seminar participants at the University of Virginia, University of Massachusetts and University of Connecticut for helpful comments and suggestions.

†Department of Agricultural and Resource Economics, University of Maryland, College Park, MD 20742 Email: archsmit@umd.edu, URL: https://econjim.com, ORCHID: 0000-0002-6052-0302

‡Department of Economics, University of Birmingham, Birmingham, UK Email: a.g.heyes@bham.ac.uk ORCHID: 0000-0003-1847-9374 Heyes acknowledges financial support from the Canada Research Chair program.

§Department of Health Policy and Management Columbia University, New York, NY 10032, NBER and IZA Email: mn2191@columbia.edu

¶School for the Future of Innovation in Society, Arizona State University, 1800 I Street NW, Washington DC 20016 and NBER. Email: bhaven.sampat@asu.edu ORCHID: 0000-0001-7387-6706

Abstract

Recent theoretical and empirical work characterizes attention as a costly 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 or the psychological theory of ego depletion. There is no such interdependence across the breaks that occur 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 the future induces reduced attention to current decisions, consistent with a forward-looking agent allocating his budget strategically across a sequence of decisions of varying importance. 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.

JEL: D83, D91

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

The Online Appendix is available at https://econjim.com/WP2101a

1 Introduction

Attention is perhaps the most important cognitive process; it allows humans to process the world around them to make better decisions. Yet, in daily and professional life, many decisions compete for our attention, 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, then amounts of attention applied across decisions in a sequence can become dynamically interdependent. That is, attentional effort at one point in time can be expected to both 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, is sensitive to attention devoted to previous ones. Not only would we expect decision quality to decline with the number of decisions made (as evidenced in the related literature on decision fatigue), but also the complexity of these decisions, and the effort exerted in their execution.

Moreover, if attentional effort is a depletable resource, a forward-looking agent might be expected to manage that resource strategically. Agents may exert less effort on an early, low-stakes decision to conserve effort for 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 athletes' management of physical energy. It is well-understood that athletes may not exert full effort (play at less than "full throttle") in low-stakes situations early in games, conserving physical energy for more critical situations later.

In this paper, we empirically explore such 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). Though they work in the sports industry, our subjects are not professional athletes. They are well-paid, skilled workers who we observe in their place of work, making high-stakes decisions under stress. As such, we believe our results may have more general implications for a much broader set of professional contexts. We ask three primary questions:

1. How does attention umpires allocate to a decision depend on the importance of that decision?
2. Does the attention umpires applied to past decisions deplete attention available for current decisions, and are there mechanisms through which it can be replenished?
3. How does an umpire's attention to a decision depend on the expectation of the importance of future decisions?

Our setting has several features that make it particularly well-suited for research on attention. 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. In other words, we assume that a correct decision, all else equal, implies that the umpires spent more attentional effort on that decision.

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 game's outcome.[1] Leverage varies substantially during a game – as in most sports, some calls are crucial to the outcome of the contest while others are largely irrelevant – evolving as a function of player actions and chance events. The quasi-random variation in the value from paying attention allows us to assess how attentional effort varies with stakes. If the importance of a decision increases the value of paying attention, and umpires are allocating attention strategically, 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, and the varying stakes associated with each, allow us to make progress in disentangling dynamic interdependence, even with potentially small effect sizes. We can see how attention to one decision depends upon the importance of that decision, the complexity of prior decisions, and a rational expectation of the importance of future decisions. Moreover, we explore whether attention is 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 as to 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. Baseball is a rare field setting where the econometrician can infer workers' rational 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 26,536 games between 2008 and 2018, enabling us to 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 game date. We can control flexibly for time elapsed, allowing us to separate the effects of cognitive fatigue from physical fatigue associated purely with the 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), allowing us to include detailed controls 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, showing a statistically significant increase in accuracy as the decision importance increases. This finding supports a central static prediction of rational inattention theories: umpires allocate more attention when the benefits from doing so increase.

[1]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.

We also find that periods of higher leverage in the past lead to less contemporaneous attention, even with controls for current leverage. Increased leverage in recent prior decisions leads to a statistically significance decrease in the probability of a correct decision. This finding is consistent with a model of a depletable budget of decision resources, such that more attention devoted to one decision depletes availability for subsequent decisions.

However, short respites in the decision series – provided by the break that the structure of the game gives the umpire between each half-inning – reset the process.[2] 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 breaks might impact performance in mentally challenging work tasks. Our results suggest that short breaks of just a few minutes allow for the replenishment of attentional capital.

Finally, we find evidence of forward-looking behavior by umpires. A rational expectation of facing higher leverage (more important) decisions later in the decision series leads to reduced attention to the current decision. More concretely, as the expectation of leverage in future decisions increases, we see a statistically significant decrease in the probability of the umpire making a correct decision.

These findings significantly extend existing evidence on rational inattention (see Maćkowiak, Matějka, and Wiederholt (2023) for an excellent overview) in several ways. The central *static* prediction of rational inattention models is that agents allocate more attention to more important decisions, which we confirm. Laboratory and field studies find that the 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 Pollet (2009)), consumer reactions to tax rates (Chetty, Looney, and Kroft (2009)), hiring decisions (Acharya and Wee (2020)) and information acquisition in the rental housing market (Bartoš 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. Our results support 'budget-of-attention'-type models (for example, Dragone (2009); 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. In this framework, 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 both prior high-leverage decisions and rationally anticipated future high leverage reduce attention to current decisions.[3] These results are inconsistent with other models of strategic inattention in which attention is costly (and agents optimize with respect

[2] 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.

[3] 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.

to how much attention to apply) but where 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).[4]

We proceed as follows. In the following section, we provide background on MLB umpiring and discuss how our work relates to existing research on attention. Section 3 outlines the data we assemble; describes how we operationalize "leverage," our measure of decision importance or stakes; and provides examples of how leverage is calculated. Section 4 reports descriptive statistics. Section 5 describes our econometric approach. Section 6 reports the main results and a battery of robustness checks. Section 7 concludes.

2 Background and related work

2.1 Baseball umpiring

Baseball umpiring is a skilled job, requiring sustained mental effort and decision-making under stress. We study professional umpires operating at the game's highest level, Major League Baseball.[5] The ballparks where they work are dispersed across many of the major cities of the United States, plus Toronto. MLB employs around 100 umpires in any given season, organized into "crews" of four, 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 post-season assignments and writing and speaking engagements for the high performers can supplement.

The home plate umpire's most significant task is "calling" the game: deciding which pitches are balls and which are strikes. A pitch should be called a strike if any baseball portion passes through the strike zone (see Figure 1) and a ball otherwise.[6] 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

[4] 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.

[5] For readers unfamiliar with the game of baseball, we provide a summary of rules relevant to our analysis and a glossary of terms in Appendix C.

[6] 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").

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 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" (J. W. Kim and King (2014)), the gambler's fallacy (Chen, Moskowitz, and Shue (2016)), how decision quality is affected by exposure to air pollution (Archsmith, Heyes, and Saberian (2018)) 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.[7]

2.2 Related research on inattention

Several economic studies explore the dynamics of inattention. The closest laboratory experimental evidence may be 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 are 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 the 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. In a recent application to online chess games, Howard (2024) studies players' allocation of a time budget across moves. He finds that high-skilled players are better at allocating time strategically in such a way as to more closely equate the marginal cost and benefit of additional seconds devoted to each move than their low-skilled counterparts. Bhattacharya and Howard (2022), also using baseball data, provide evidence consistent with rational allocation of attentional effort by baseball players. While they find significant deviations from Nash behavior

[7]While the rate of termination of umpires is low, the opportunities for playoff games and other high profile events create strong incentives for umpires to maintain high levels of accuracy.

(for example, the type of pitch a pitcher chooses to throw), play converges to Nash equilibrium as the stakes in any particular game situation get higher.

Our work also connects to recent psychology literature. There are numerous studies on decision fatigue, the idea that the character of decisions is a function of the cumulative number of decisions made in a sequence. Settings include healthcare (Linder et al. (2014); Philpot et al. (2018); Chan, Hartley, and Spiegel (2009); R. H. Kim et al. (2018)), 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 (Orasanu 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 amount 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 on the intensity of effort directed to them by the decisionmaker.

Our analysis relates closely to the psychological theory of 'ego depletion' proposed by Baumeister et al. (1998). That paper, entitled "Ego Depletion: Is the Active Self a Limited Resource?" proposed one of the most influential 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). The authors 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 is interpreted broadly as effort applied to self-control or resisting temptation. In our setting, the umpires need to resist the temptation to relax or give less-than-full attention to a particular decision, and so doing depletes their stock going forward. The contention of Gailliot et al. (2007) that exercising self-control or mental effort requires and depletes body glucose also lays the basis for a possible physiological basis for such depletion and suggests a closer analogy between attention fatigue and muscle fatigue. 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 optimization applied to a cake-eating problem.

Despite its popularity, ego depletion as a theory has fallen out of favor recently due to repeated replication failures (see, for example, Hagger et al. (2016) and citations therein). The replication failures are 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) published a cover article headlined "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 the application of attentional effort depletes the budget subsequently available. We believe that context is essential for uncovering these effects: boredom experienced by subjects in a laboratory setting may substantially impact inattention. Moreover, our forward-looking results point to the *anticipation* of ego depletion playing an essential role in the decision-making process, in the spirit of the assumptions embedded in the model of Ozdenoren, Salant, and Silverman (2012). Not only does the limited resource deplete, but our subjects act as if aware that their attentional capital is depletable and seek to allocate its use strategically across time, for example, conserving it for later decisions, where those future decisions are expected to be more important. This finding 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.

3 Data

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

3.1 MLB Pitch Data

Following Archsmith, Heyes, and Saberian (2018), we compiled data on the details of every pitch in all MLB games from 2008 to 2018 from the MLB website. This data is 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 focus our samples on only the pitches in the regular innings.[8]

3.2 Leverage: A measure of decision importance

The static element of our empirical analysis investigates whether MLB umpires apply more effort to making 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 commonly used in baseball to refer to the importance of a game situation to the outcome of the game. 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. The most

[8] In Appendix Table A.6 we estimate our main specifications on a sample based only on extra innings data. This delivers estimates very similar to our main results.

commonly used measure of leverage in professional baseball is the Tango (2006) "Leverage Index" (LI). This index is defined at the "at-bat" level, and there are between zero and six decisions by the umpire per at-bat. To focus on umpires' decisions, we extend Tango's measure to compute leverage at the pitch-level.

We define leverage for a given pitch as the absolute difference in what we will call the "win expectation," which is the probability the home team wins the game, between the states after the umpire calls a "ball" and the state where the umpire calls a "strike." [9] The stakes can change 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 game's current state, but future events impacting the outcome (many beyond the umpire's control) are unknown. Thus, computing this leverage measure requires determining two win expectations, the probability the batting team wins given a "strike" call and the probability they win given a "ball" call, conditional on the current state in the game. In each case, similar to Tango (2006), we assume events in a baseball game follow a Markov process with state A_t encompassing the game state variables at pitch t.

Lacking a repository of win expectations in MLB games incorporating the current ball/strike count into the state, we estimate them empirically. In the simplest case, for each game state one would count the number of times the home team wins conditional on the game entering that state and divide by the total number of times the state occurred. However, this state space is quite large, with over 100,000 possible combinations. Even with the large volume of data available for MLB games (over 7 million pitches during our sample period), some game states occur at low frequency, and the win expectation estimates are imprecise. Instead, we estimate these expectations 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 B for details):

1. We define a limited game state as by the inning part (top or bottom), the score difference between the home and away teams (-10 to 10), the number of outs (0-2), baserunner positions (occupied or not for 1st, 2nd, and 3rd bases), and current ball (0-3) and strike (0-2) count. This removes the inning from the state and assumes games will evolve similarly across innings.

2. Using actual MLB data, for each possible limited state we compute the probability of transitioning to all potential new limited states.

3. Using these probabilities, we simulate 5 million MLB games from start to finish, collecting the full game state observed for each pitch and the eventual winner. This information is used to compute the win expectation conditional on a given state.

[9] 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 state A is simply one minus the probability the home team wins given that state. 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.

4. Using these computed expectations, we calculate our leverage measure for each state as the difference in win expectations between the case where the state evolves by one strike versus evolving by one ball.

Appendix A.2 also examines the 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 each past pitch during the current half-inning, whether or not the umpire was required to make a ball/strike decision.[10] Expected future leverage is also computed from simulated baseball games.[11] Using the same set of simulated games, we assign the leverage value for each pitch corresponding to the game state. Future leverage in each simulated game is the sum of leverage on all remaining pitches in the half-inning. We compute expected future leverage in actual games as the mean future leverage for a game state across all times that state occurred in the simulations.

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 3.1 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) and the final regression sample (column 2). 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 negligible effect on the game's 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 with 2 outs, 0 balls, 1 strike, runners on 2nd and 3rd base, with the home team leading by 5 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

[10] In the appendix (A1.3) we separate past leverage into pitches that are called by the umpires and those that are not called.

[11] Here simulated games are essential since even states which are overall unlikely, and so 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.

calls a "ball" as compared to when the umpire calls a "strike." The table also provides examples of leverage variation within the same half-inning since, as described below, we include half-inning fixed effects.

[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 state 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 presents a scatter plot of each leverage measure (past, current, and future) against the other. Each suggests substantial independent variation in the measures of leverage, such that multicollinearity should not be a substantial 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 throughout the game and tends to peak as the game approaches critical junctures.

[Figure 3 About Here]

5 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:[12]

$$\mathbf{1}\left(C_p^* = C_p\right) = \beta^C \underbrace{L\left(A_p\right)}_{\substack{\text{Current}\\ \text{Leverage}}} + \beta^P \underbrace{\sum_{t=1}^{p-1} L(A_{p-t})}_{\substack{\text{Past}\\ \text{Leverage}}} + \beta^F \underbrace{\mathbf{E_p}\left[\sum_{s=1}^{\infty} L(A_{p+s})\right]}_{\substack{\text{Expected Future}\\ \text{Leverage}}} + \delta_{\mathbf{p}}^{\mathbf{I}} + \delta_g^G + \epsilon_{pg}$$

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 game state A_p where pitch p is thrown, $\delta_{\mathbf{p}}^{\mathbf{I}}$ are fixed effect for each inning and a dummy variable to indicate the top-half of an inning, δ_g^G is a game fixed effect, and ϵ_{pg} is an idiosyncratic error potentially correlated within games. 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 p. Given our definition of leverage, we

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

interpret β as the effect of a change in win probability for the home team, conditional on the game state, 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, an umpire bias toward one team (Sacheti, Gregory-Smith, and Paton (2015)), 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 (δ_p^I) to control for physical fatigue as the game progresses, enabling us to separately identify the effects of decision fatigue.[13] We also explore specifications that include pitch attributes as controls. However, our main specification excludes them because they may reflect "bad controls" (Angrist and Pischke, 2008), both discussed in more detail in Section 6.2.1 To define past and future leverage, we accumulate leverage measures within the same half-inning. For example, imagine we are at the 5$^{\text{th}}$ pitch in the inning for a game in the top of the third 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 6$^{\text{th}}$ pitch, past and future leverage updates to include the 5$^{\text{th}}$ pitch. Given that the effects from past leverage may extend beyond the current half-inning, we also explore the role of leverage from previous innings by including lags of past leverage. The existence of a short (two-minute) break between half-innings enables us to explore whether a short respite replenishes the umpire's stock of attention.

6 Results

6.1 Main Results

Table 4 shows our main results. The first column reports results from our estimating equation in which our measure of past leverage is from the current half-inning only. The next two columns add lags of past leverage. We exclude data from the first inning to keep the sample of pitches we explore fixed.[14] All coefficients are multiplied by 100 to improve readability.

[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 37.158 indicates

[13] We use half-inning fixed effects over controls for time-elapsed since the latter is endogenously determined by leverage, though results are comparable using this alternative measure.

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

that increasing leverage from 0 to 0.014, the mean leverage in our sample, increases the probability that the umpire makes the correct call by 0.0051, a 0.611% percent increase. This estimate is statistically significant, with a 95% confidence interval of [0.551%,0.671%].

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 -1.497 indicates that moving accumulated past leverage from 0 to 0.137, the mean of past leverage, decreases current call accuracy by -0.2055 percentage points, a -0.245% percent change. This estimate is also statistically significant with a 95% confidence interval of [-0.287%, -0.202%].

Next, we focus on future leverage. Our estimates align with our theoretical prediction: higher future leverage decreases current attention. Our estimate of -3.550 indicates that changing future leverage from 0 to 0.137, the mean of future leverage, increases an umpire's likelihood of a mistake by -0.7383 percentage points, a -0.879% percent change.

Overall, we find that greater contemporaneous leverage increases umpires' attention, while past and future leverage decreases it. The inning dummy variables, which control at least imperfectly for physical fatigue over the course of the game, also indicate an interesting pattern. Except for the last inning, we see only small changes in umpire performance as the game progresses. Changes in a correct call vary between -0.05 to 0.12 percentage points compared to the second inning (the reference category), though with no clear pattern of physical fatigue through most of the game.[15] In other words the overall quality of the calls made by the umpire is quite level over the first eight innings of a game. However, in the last inning umpire performance drops by 0.38 percentage points. Since we control for leverage, this drop at the end of the game does not reflect the erosion of the importance of calls later in games but could reflect that close to the conclusion of a game the focus of an umpire turns to other things, leading to more mistakes.[16] We probe this potential 9th-inning result in more detail below and in Appendix A.1.

Two patterns emerge as we include leverage from the previous half-inning to our specification. First, the coefficient estimates for past, current and future leverage of the current half-inning are unaffected. Second, the effect of past half-inning total leverage is very small, coming in several orders of magnitude smaller than the current half-inning, and statistically insignificant. Both of these patterns hold whether we include leverage measures from the previous one or two previous half-innings. These results imply that umpires refocus their attention after a short break, suggesting that while attention is scarce, budgets can replenish quickly. This has potentially important implications, as we discuss in more detail in the conclusion.

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

[16] A recent paper by J. T. Dean (2024) provides experimental evidence that noise can reduce cognitive function, raising the possibility that umpire performance is negatively affected by crowd noise. The inclusion of game fixed effects should mean that any factors invariant within a game, for example noise levels that carry throughout a game, would not impact our conclusions. However, a reader might be concerned that the noise level from a large crowd might vary in a way that is correlated with leverage, confounding inference. In Appendix Table A.7 we re-estimate our main specification but adding regressors that interact attendance at a game (as a proxy for noise) with our leverage measures. The estimated coefficients on these interactions are small, and statistically insignificant at conventional levels. Further their inclusion does not disturb other primary estimates meaningfully, so we do not think they are an important source of confounding.

6.2 Additional Results

In Table 5, we explore how the effect of leverage varies within the game by estimating the effects separately by inning.[17] 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.

[Table 5 About Here]

As the game progresses, the effect of current leverage on umpire attention steadily decreases, though it always remains positive and statistically significant. By the 9$^{\text{th}}$ inning, the effect of contemporaneous leverage is nearly 30% the size of the effect in the 3$^{\text{rd}}$ inning. The difference between these two estimates is statistically significant[18]; 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.

With respect to 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.

6.2.1 Addressing Pitcher and Batter Decisions

Umpires are not the only participants in baseball games. Decisions made by players in the game may impact the difficulty of the decisions faced by an umpire. In particular, the pitcher, after observing leverage, may decide what kind of pitch to throw. Furthermore, the batter chooses whether to swing after observing leverage and some attributes of the thrown ball. If the batter swings, it relieves the umpire of the need to make a ball/strike decision and causes that pitch to be excluded from the sample. Therefore, the difficulty of an umpire's decision may depend on characteristics of the pitch in flight, such as velocity, spin, or trajectory and, for pitches in sample, the batter's decision as to whether to swing.

Directly addressing these issues in our econometric model raises a potential complication. Pitch characteristics, for example, are not exogenous but rather reflect the pitcher's actions. It is reasonable to think that the pitcher's behavior itself depends upon leverage, deciding where to pitch the ball and what kind of pitch to throw. As such pitch attributes could reasonably be treated as alternative outcome variables and their inclusion as controls in our specification would run into the "bad controls" critique (Angrist and Pischke (2009)) or overcontrol bias (Cinelli, Forney, and Pearl (2024)). On this basis, we excluded controls for pitch attributes from our primary specification. In this section, we investigate the impact of including pitcher or batter decisions, noting this caveat in interpreting any changes in estimates as bias.

Accounting for Pitch Location Pitchers might impact an umpire's decision by attempting to throw pitches in locations that vary in how difficult they are to correctly adjudicate. We consider whether particularly difficult decisions by the umpire may be associated with high leverage situations and the umpire's

[17] 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.

[18] A Wald test rejects these parameters are equal with an $F(1,26533)$ statistic of 22.72 and a p-value less than 0.001.

propensity to make the correct decision. There is no comprehensive measure of the difficulty of a particular ball/strike call and including any such measure would effectively condition our analysis on the outcome variable. Instead, under the assumption that "close" calls are challenging to call correctly, we split our sample into bins based on how far the pitch was from the strike zone boundary.

The results of this exercise are shown in Table 6. Column 1 repeats the primary specification for reference. In Columns 2 – 4 we consider pitches falling within successive three-inch bins – approximately the width of a baseball -- around the strike zone boundary. Overall, the accuracy of umpires increases as the pitch passes farther from the strike zone boundary, with 68.7% accuracy within 3 inches and 99.0% within 6-9 inches. In each case, the leverage effects persist, though they diminish in magnitude the further the pitch is from the strike zone boundary. Note that baseline accuracy in the more distant bins is already high, leaving little room for the leverage effects to manifest.

Controlling for Pitch Attributes We benefit from a wealth of data for each pitch encompassing a pitcher's decision of how to throw the ball, including the pitch speed, spin, break and so on, as shown in Table 2. We show several results that control for these attributes in Table 7. Column 1 repeats our primary specification. Column 2 adds linear controls to that regression for all pitch attributes, and our results change minimally when doing so. One may be concerned that these controls insufficiently capture the more complex impact of pitch attributes. We more flexibly control for this nonlinear relationship by forming Kronecker products of the pitch attributes and reducing the space of candidate attribute controls using the Post-Double Selection Lasso procedure of Belloni et al. (2012).[19] We estimate these models for 1st through 3rd-order interactions of the pitch attribute variables in Columns 3-6, respectively. Alternatively, we eschew any assumption on the specific functional form of the impact of pitch attributes by allowing them to enter as nuisance function flexibly estimated via gradient-boosted trees using Double-Debiased Machine Learning approach of Chernozhukov et al. (2018), with results shown in Column 7.[20] We find that including first-order terms has minimal impacts on our estimates. Including higher order terms, however, attenuates the current and future leverage estimates, though past leverage is largely unaffected. Importantly, all leverage estimates remain statistically significant. It is difficult to discern from this exercise if the changes in estimates indicate that further omitted variable bias remains, especially since we include endogenous variables that may reflect bad controls. Since we observe an important set of pitch attributes that only impact estimates when we control for them flexibly, the fact that our estimates remain significant gives us comfort in interpreting our estimates as causal impacts of leverage on umpire attention.

Accounting for Batter Decisions We address the batter's decision to swing, which affects whether or not the umpire has a decision to make, by treating this as a sample selection problem. We first address sample selection using the Heckman (1979) two-step control function approach (Heckit model), which separates the selection problem into two stages: 1) a participation equation estimated as a discrete choice of whether

[19]We use the oracle formula to choose the value of the regularization parameter in each Lasso model, as suggested by Belloni et al. The PDS-Lasso procedure estimates a traditional OLS model of our outcome, variables of interest, and the union of all variables in the nuisance function with non-zero parameters across all of the Lasso regressions.

[20]We tune the learning rate and size of categorical variable bins using 10-fold cross validation grouped by game ID.

an observation is included in the sample, and 2) the outcome equation where one accounts for selection-into-sample using the inverse Mill's Ratio of residuals from the participation equation as a control function. We implement this model by constructing an instrument for participation using attributes of the current game participants (the pitcher and batter) and the thrown pitch -- information available to batters when they make the decision to swing. The rationale for this approach is that batters decide to swing using heuristics based on complicated interactions between components of the specific pitches thrown. We use machine learning to estimate a nonlinear function of player identities and pitch attributes that predicts a batter's decision to swing at that pitch.[21] We then use this predicted swing decision in the participation equation.

The results of these regressions are shown in Table 8. Column 1 repeats our primary specification. Since the Heckman procedure imposed sample and specification restrictions beyond our primary specification, column 2 repeats our main estimation results without accounting for selection but imposing these limits.[22] The results changed minimally from these limits. The results in column 3 show the impact of addressing selection estimating. The effect of current and future leverage nearly doubles in absolute value compared with our primary specification, while the effect of past leverage remains quite similar in magnitude. The estimates do not change much upon adding controls for pitch attributes to the outcome equation, as shown in Column 4. However, all parameter estimates are of identical sign and statistical significance to our primary specification.

As an extension of the Heckman procedure, we also estimate a partly linear model where leverage effects and the inverse Mills' ratio enter linearly but we allow pitch attributes to enter as a nuisance function, estimated using gradient-boosted trees via Double-Debiased Machine Learning.[23] This method allows for a more flexible selection effect beyond the inverse Mills' ratio. Estimates using the model, shown in Column 5, are similar to the previous estimates.

Overall, these estimates suggest that the batter's decision to swing may bias the impacts of leverage on umpire attention. However, as with our interpretation of the models with controls for pitcher decisions, we caveat this interpretation because the batter's decision to swing is endogenous. We are nonetheless assured that our estimates remain statistically significant and qualitatively comparable to our primary results, supporting the notion that sample selection does not eliminate the impact of leverage on umpire attention.

[21]We construct the predicted probability using gradient-boosted tree classifier on the full set of pitch data from 2008-2018. The outcome is a binary indicator for whether the batter swings at a given pitch. As candidate predictors, we include all pitch attributes and categorical variables for the identity of the batter, the identity of the pitcher, and the game ID. We tune the learning rate and the number of bins for categorical variables using 10-fold cross-validation grouped by game ID. For each pitch, we compute the propensity to swing as the predicted probability of a swing from the trained classifier. This model yielded strong predictive power, with a cross-validated pseudo-R2 of 0.37.

[22]In the two-step control function estimators, we modify our primary specification in two ways. First, since the first step is nonlinear, we are unable to use typical methods for accounting for high-dimensional game fixed effects. Instead, we use year, inning part, home team, away team, and umpire fixed effects. Second, we perform inference using bootstrap clustered by game. Some resamples would exclude umpires who appear in very few games, so we limit the sample to umpires appearing in at least 100 games from 2008 to 2018.

[23]We tune the learning rate and categorical variable bin size using 10-fold cross-validation grouped by game ID.

6.2.2 Other Robustness

In Table 9 we explore whether our results are robust to a range of alternative regression controls. Our primary specification, which includes game fixed effects, identifies the effect of leverage on decision accuracy from within-game variation. In Columns 2 and 3 of Table 9, 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). The estimated effects of leverage are essentially unchanged in these specifications.

Finally, it is possible individual players may be more likely to be involved in high-leverage situations and may take actions that increase the difficulty of an umpire's decisions. In column 4 we add fixed effects for players in each of the positions that directly participate in this component of the game: the pitcher and the batter. Again, the empirical estimates are very similar to those from our primary specification. A further result in Section A.4 of the Appendix demonstrates that the same pattern of effects persists regardless of whether the batting team is winning, losing, or tied with their opponent.

[Table 9 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.[24] Results for each leverage measure are shown in Figure 4. Results from this more flexible specification reveal an approximately linear relationship with effect sizes that are similar, if not slightly larger in magnitude, to the parametric specification from Table 4.[25]

[Figure 4 About Here]

6.2.3 Heterogeneous Effects

We further explore heterogeneity in how individual umpires allocate effort by estimating umpire-specific leverage effects. This aims to ensure that our main conclusions are not driven by a small number of outlier umpires, but rather reflect a pattern of behavior observed across many umpires.

Extending the regression specification from Table 4 Column 1, we interact each leverage measure with indicators for every umpire in a single regression.[26] 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

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

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

[26] 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.

all umpires, the estimated treatment effects have 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 that umpire's estimated current leverage effect. Each leverage effect is divided by its standard deviation across all umpires for comparability. 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.[27] 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 can place some of the estimates from Figure 5 in context. Over the sample, the 5th percentile umpire has 83.4% accuracy rate, the mean umpire has 87.4% accuracy, and the 95th percentile umpire is 90.9% accurate. Umpires in the 95th percentile of accuracy are 37.9% less responsive to the leverage of the current pitch than the average umpire. Their accuracy is also less impacted by past (36.6% of the effect of mean umpire) or expected future (41.5% 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 (42.8% more) when the current pitch is high-leverage, but decreases substantially when past decisions are high-leverage (41.3% larger decrease), or expected future decisions are (46.9% 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.

7 Conclusions

Conventional economic models embody agents that are able to make perfect, optimizing decisions. An important strand of recent efforts to increase the behavioral realism of models has acknowledged that attention is not costless---the effort required to attend to decisions and execute them well can be costly and/or 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 or costly, are increasingly mainstream (for examples see Caplin and M. Dean (2015); Sims (2003); Falkinger (2008)).

While the idea of costly attention is intuitively appealing, rigorous evidence characterizing its implica-

[27] 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.

tions in non-experimental 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, in particular allocating greater attention to those associated with higher stakes. This finding is consistent with established theoretical models of strategic inattention and several existing empirical studies.

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 interesting 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 with the predictions from a model in which the umpire has a depletable stock of attention. These dynamics render inter-dependent otherwise separable decision problems. Our forward-looking results are consistent with the stock of attention being depletable, as projected by the ego depletion theory in psychology. Moreover, our subjects seem to behave in a sophisticated way, conserving capacity in a manner sensitive to their expectations about the importance of future decisions they will face. We believe this is the first paper to demonstrate such behavior. While the precise magnitude of the effects depend somewhat on the precise empirical specification, the results prove broadly robust in sign and statistical significance across a wide array of alternative specifications and robustness tests.

The short and exogenously mandated breaks that the umpires receive between half-innings, as the players rotate between offense and defense, appear sufficient to replenish the 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 (2016)). It would provide a rationale for the advice that the hedge fund Voss Capital has given to stock traders that they should take short but frequent breaks during the work day (Wadhwa (2016)).[28]

Although we focus on baseball umpires, this is not just a paper about baseball. The richness of the data in a field setting affords a unique opportunity to explore the broader issue of strategic inattention in novel ways. 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. While we cannot speak directly to external validity, it is plausible that our qualitative results would apply in other contexts involving repeated decision-making under stress. Examining whether similar dynamics of attention are seen in different settings is thus an important next step.

[28]Sievertsen, Gino, and Piovesan (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.

References

Acharya, Sushant and Shu Lin Wee (2020). “Rational inattention in hiring decisions.” In: *American Economic Journal: Macroeconomics* 12 (1), pp. 1–40. ISSN: 19457715. DOI: 10.1257/mac.20180390.

Allcott, Hunt and Nathan Wozny (Dec. 2014). “Gasoline Prices, Fuel Economy, and the Energy Paradox.” In: *Review of Economics and Statistics* 96 (5), pp. 779–795. DOI: 10.1162/REST_a_00419.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press. ISBN: 978-0-691-12034-8.

Archsmith, James, Anthony Heyes, and Soodeh Saberian (2018). “Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation.” In: *Journal of the Association of Environmental and Resource Economists* 5 (4), pp. 827–863. ISSN: 23335963. DOI: 10.1086/698728.

Augenblick, Ned and Scott Nicholson (2016). “Ballot position, choice fatigue, and voter behaviour.” In: *Review of Economic Studies* 83 (2), pp. 460–480. ISSN: 1467937X. DOI: 10.1093/restud/rdv047.

Baer, Drake (2015). *The Scientific Reason why Barack Obama and Mark Zuckerberg wear the same Outfit Every Day*. URL: https://www.businessinsider.com/barack-obama-mark-zuckerberg-wear-the-same-outfit-2015-4.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka (2016). “Attention discrimination: Theory and field experiments with monitoring information acquisition.” In: *American Economic Review* 106 (6), pp. 1437–1475. ISSN: 00028282. DOI: 10.1257/aer.20140571.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice (1998). “Ego depletion: Is the active self a limited resource?” In: *Journal of Personality and Social Psychology* 74 (5), pp. 1252–1265. DOI: 10.4324/9781315175775.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012). “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain.” In: *Econometrica* 80 (6), pp. 2369–2429. ISSN: 0012-9682. DOI: 10.3982/ECTA9626.

Bhattacharya, Vivek and Greg Howard (2022). “Rational Inattention in the Infield.” In: *American Economic Journal: Microeconomics* 14 (4), pp. 348–393. ISSN: 1945-7669. DOI: 10.1257/mic.20200310.

Bruyneel, Sabrina, Siegfried Dewitte, Kathleen D. Vohs, and Luk Warlop (2006). “Repeated choosing increases susceptibility to affective product features.” In: *International Journal of Research in Marketing* 23 (2), pp. 215–225. ISSN: 01678116. DOI: 10.1016/j.ijresmar.2005.12.002.

Caplin, Andrew and Mark Dean (2015). “Revealed Preference, Rational Inattention, and Costly Information Acquisition.” In: *American Economic Review* 105 (7), pp. 2183–2203. DOI: 10.1257/aer.20140117.

Chan, Michael Y, Cohen Hartley, and Brennan M. R Spiegel (2009). “Fewer polyps detected by colonoscopy as the day progresses at a Veteran’s Administration teaching hospital.” In: *Clinical Gastroenterology and Hepatology* 7 (11), pp. 1217–1223. DOI: 10.1016/j.cgh.2009.07.013.

Chen, Daniel, Tobias J Moskowitz, and Kelly Shue (2016). “Decision-Making under the Gambler’s Fallacy : Evidence from Asylum Judges, Loan Officers, and Baseball Umpires.” In: *The Quarterly Journal of Economics* 131 (3), pp. 1181–1242. DOI: 10.1093/qje/qjw017.

Chernozhukov, Victor et al. (2018). "Double/debiased machine learning for treatment and structural parameters." In: *Econometrics Journal* 21 (1), pp. C1–C68. ISSN: 1368423X. DOI: 10.1111/ectj.12097.

Chetty, Raj, Adam Looney, and Kory Kroft (Sept. 2009). "Salience and Taxation: Theory and Evidence." In: *American Economic Review* 99 (4), pp. 1145–1177. DOI: 10.1257/aer.99.4.1145.

Cinelli, Carlos, Andrew Forney, and Judea Pearl (Aug. 2024). "A Crash Course in Good and Bad Controls." In: *Sociological Methods and Research* 53 (3), pp. 1071–1104. ISSN: 15528294. DOI: 10.1177/00491241221099552.

Correia, Sergio (2014). *REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects.* URL: https://ideas.repec.org/c/boc/bocode/s457874.html.

Dean, Joshua T. (Oct. 2024). "Noise, Cognitive Function, and Worker Productivity." In: *American Economic Journal: Applied Economics* 16 (4), pp. 322–360. ISSN: 1945-7782. DOI: 10.1257/app.20220532.

Dellavigna, Stefano and Joshua M. Pollet (2009). "Investor inattention and friday earnings announcements." In: *Journal of Finance* 64 (2), pp. 709–749. ISSN: 00221082. DOI: 10.1111/j.1540-6261.2009.01447.x.

Dragone, Davide (2009). "I am getting tired: Effort and fatigue in intertemporal decision-making." In: *Journal of Economic Psychology* 30 (4), pp. 552–562. ISSN: 01674870. DOI: 10.1016/j.joep.2009.03.008.

Drellich, Evan (Aug. 2012). *Complex system in place to evaluate umpires.* URL: http://mlb.mlb.com/news/print.jsp?ymd=20120828&content_id=37468304&vkey=news_mlb&c_id=mlb.

Falkinger, Josef (2008). "Limited Attention as a Scarce Resource in Information-Rich Economies." In: *The Economic Journal* 118 (May 2004), pp. 1596–1620. DOI: 10.1111/j.1468-0297.2008.02182.x.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006). "Costly information acquisition: Experimental analysis of a boundedly rational model." In: *American Economic Review* 96 (4), pp. 1043–1068. ISSN: 00028282. DOI: 10.1257/aer.96.4.1043.

Gailliot, Matthew T. et al. (Feb. 2007). "Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor." In: *Journal of Personality and Social Psychology* 92 (2), pp. 325–336. ISSN: 00223514. DOI: 10.1037/0022-3514.92.2.325.

Gino, Francesca (2016). *Don't Make Important Decisions Late in the Day.* URL: https://hbr.org/2016/02/dont-make-important-decisions-late-in-the-day.

Green, Etan A and David P Daniels (2021). "Bayesian Instinct." In: *Working Paper.* DOI: 10.2139/ssrn.2916929.

Hagger, Martin S. et al. (2016). "A Multilab Preregistered Replication of the Ego-Depletion Effect." In: *Perspectives on Psychological Science* 11 (4), pp. 546–573. ISSN: 17456924. DOI: 10.1177/1745691616652873.

Heckman, James J (1979). "Sample Selection Bias as a Specification Error." In: *Econometrica* 47 (1), pp. 153–161. DOI: 10.2307/1912352.

Hirshleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh (2019). "Decision fatigue and heuristic analyst forecasts." In: *Journal of Financial Economics* 133 (1), pp. 83–98. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2019.01.005.

Howard, Greg (2024). "A Check for Rational Inattention." In: *Journal of Political Economy Microeconomics* 2 (1). DOI: 10.1086/727556.

Kim, Jerry W. and Brayden G. King (2014). "Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring." In: *Management Science* 60 (11), pp. 2619–2644. ISSN: 15265501. DOI: 10.1287/mnsc.2014.1967.

Kim, Rebecca H. et al. (2018). "Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination." In: *JAMA network open* 1 (5), e181770. ISSN: 25743805. DOI: 10.1001/jamanetworkopen.2018.1770.

Kwan, James et al. (2016). "Does "Decision Fatigue" Impact Manuscript Acceptance? An Analysis of Editorial Decisions by the American Journal of Gastroenterology." In: *American Journal of Gastroenterology* 111 (11), pp. 1511–1512. DOI: 10.1038/ajg.2016.452.

Levav, Jonathan, Mark Heitmann, Andreas Herrmann, and Sheena Iyengar (2010). "Order in product customization decisions: Evidence from field experiments." In: *Journal of Political Economy* 118 (2), pp. 274–299. ISSN: 00223808. DOI: 10.1086/652463.

Linder, Jeffrey A. et al. (2014). "Time of day and the decision to prescribe antibiotics." In: *JAMA Internal Medicine* 174 (12), pp. 2029–2031. ISSN: 21686114. DOI: 10.1001/jamainternmed.2014.5225.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt (Mar. 2023). "Rational Inattention: A Review." In: *Journal of Economic Literature* 61 (1), pp. 226–273. ISSN: 00220515. DOI: 10.1257/jel.20211524.

Orasanu, Judith et al. (2012). "Evaluating the effectiveness of schedule changes for Air Traffic Services (ATS) providers: Controller Fatigue and Alertness Monitoring (C-FAM) study." In: (November), pp. 1–66. DOI: 10.13140/RG.2.1.4037.0643.

Ozdenoren, Emre, Stephen W. Salant, and Dan Silverman (2012). "Willpower and the optimal control of visceral urges." In: *Journal of the European Economic Association* 10 (2), pp. 342–368. ISSN: 15424766. DOI: 10.1111/j.1542-4774.2010.01050.x.

Parsons, Christopher A, Johan Sulaeman, Michael C Yates, and Daniel S. Hamermesh (2011). "Strike Three : Discrimination, Incentives, and Evaluation." In: *American Economic Review* 101 (4), pp. 1410–1435. DOI: 10.1257/aer.101.4.1410.

Philpot, Lindsey M., Bushra A. Khokhar, Daniel L. Roellinger, Priya Ramar, and Jon O. Ebbert (2018). "Time of Day is Associated with Opioid Prescribing for Low Back Pain in Primary Care." In: *Journal of General Internal Medicine* 33 (11), pp. 1828–1830. ISSN: 15251497. DOI: 10.1007/s11606-018-4521-8.

Sacheti, Abhinav, Ian Gregory-Smith, and David Paton (2015). "Home bias in officiating : evidence from international cricket." In: *Journal of the Royal Statistical Society. Series A: Statistics in Society* 178 (3), pp. 741–755. DOI: 10.1111/rssa.12086.

Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan (2016). "Cognitive fatigue influences students' performance on standardized tests." In: *Proceedings of the National Academy of Sciences of the United States of America* 113 (10), pp. 2621–2624. ISSN: 10916490. DOI: 10.1073/pnas.1516947113.

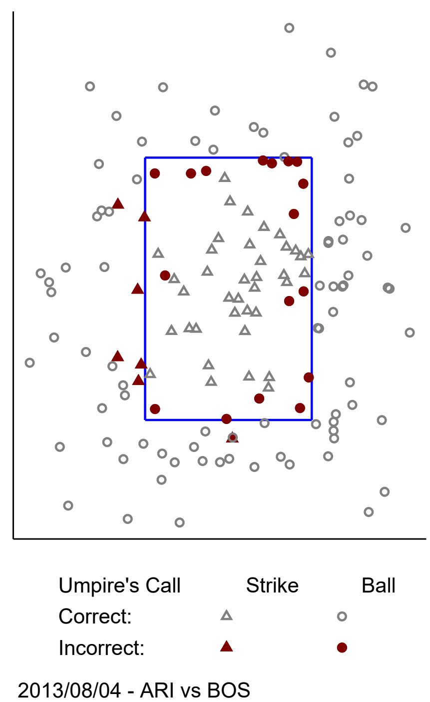
Sims, Christopher A. (2003). "Implications of rational inattention." In: *Journal of Monetary Economics* 50 (3), pp. 665–690. ISSN: 03043932. DOI: 10.1016/S0304-3932(03)00029-1.

Tango, Tom M (May 2006). *Critical Situations*. URL: https://tht.fangraphs.com/crucial-situations/.

Wadhwa, Tina (2016). *A Hedge Fund Wrote a Letter to Investors Examining Why They Should Read a Classic Book about Cognitive Biases.*

Tables and Figures

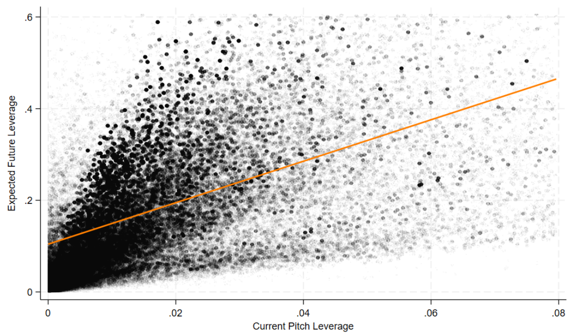
Figure 1: Example Pitch Location and Umpire Decisions



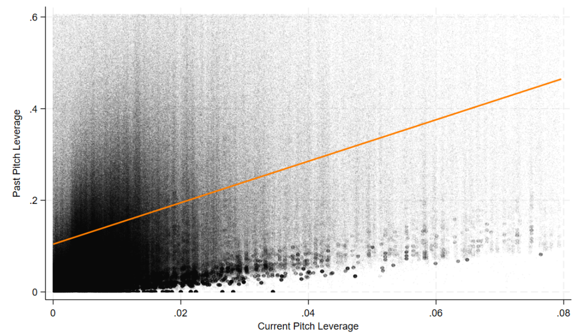
Visualization of pitch locations and umpire decisions from a typical MLB game between Arizona and Boston on August 4th, 2013. This game was selected because the number of total pitches and umpire error rate are close to the sample means. Circles denote pitches called balls and triangles denote called 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: Relationship between Measures of Leverage

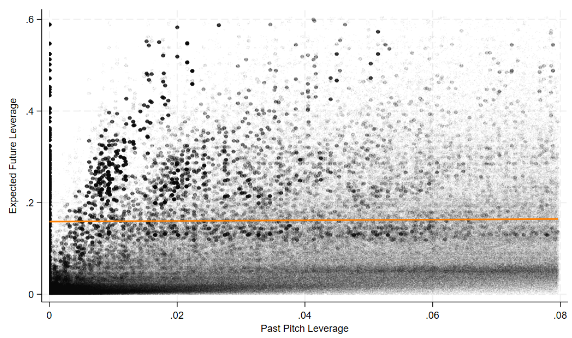
(a) Current v. Future Leverage



(b) Current v. Past Leverage

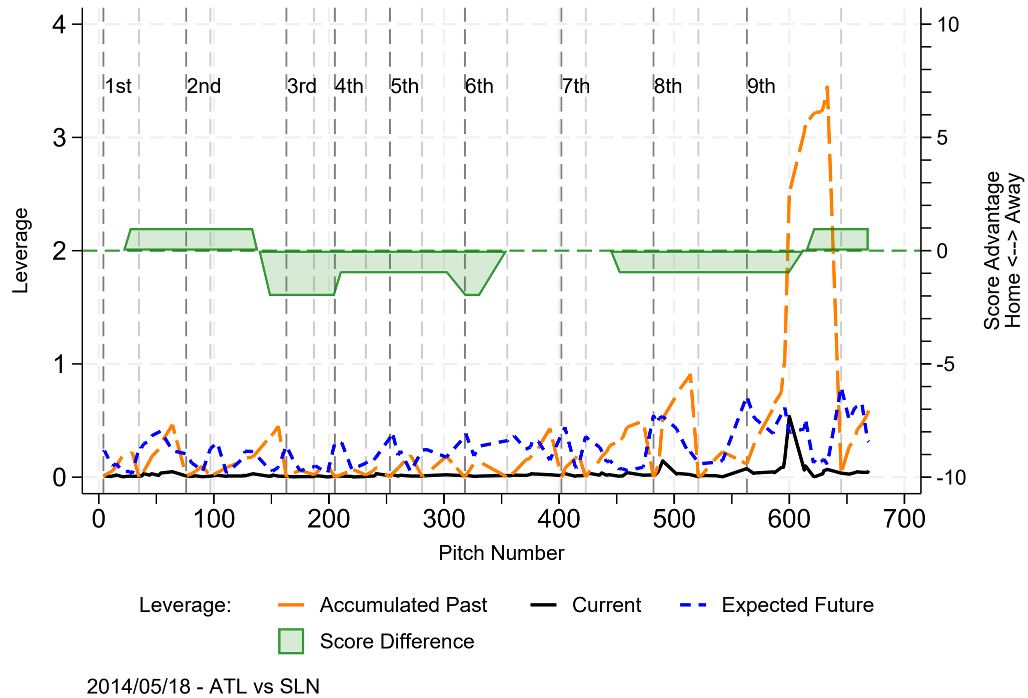


(c) Past v. Future Leverage



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.

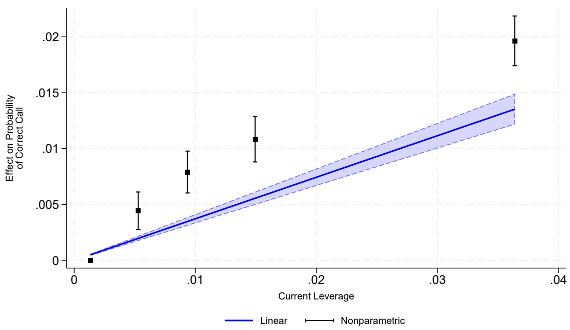
Figure 3: Example of Evolution of Leverage Metrics



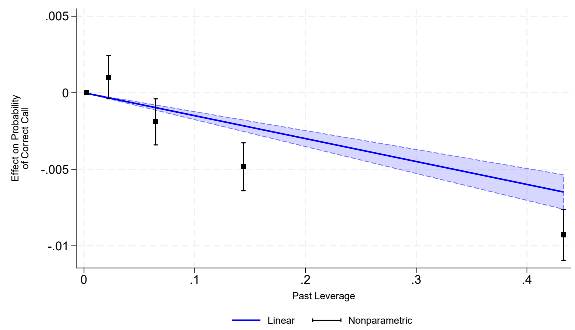
An example of the evolution of the leverage metrics through one MLB game on May 18th, 2014 between Atlanta and St. Louis. The horizontal axis represents each pitch in the game, regardless of whether the home plate umpire was required to make a ball/strike decision. Vertical black lines denote the first pitch of the top (black) or bottom (gray) of each half-inning. The black (solid) line represents leverage of the decision for the current pitch. The orange (long dash) line represents accumulated leverage through the course of the current half-inning and the blue (short dash) 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 of Leverage on Umpire Decision Accuracy

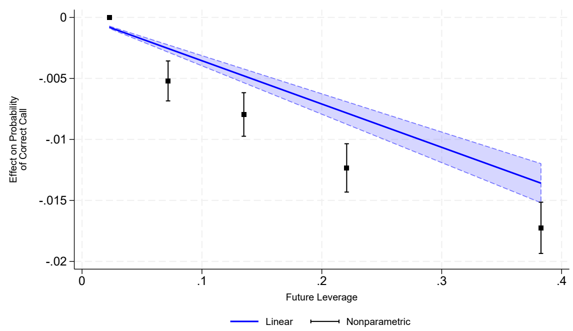
(a) Past Leverage



(b) Current Leverage

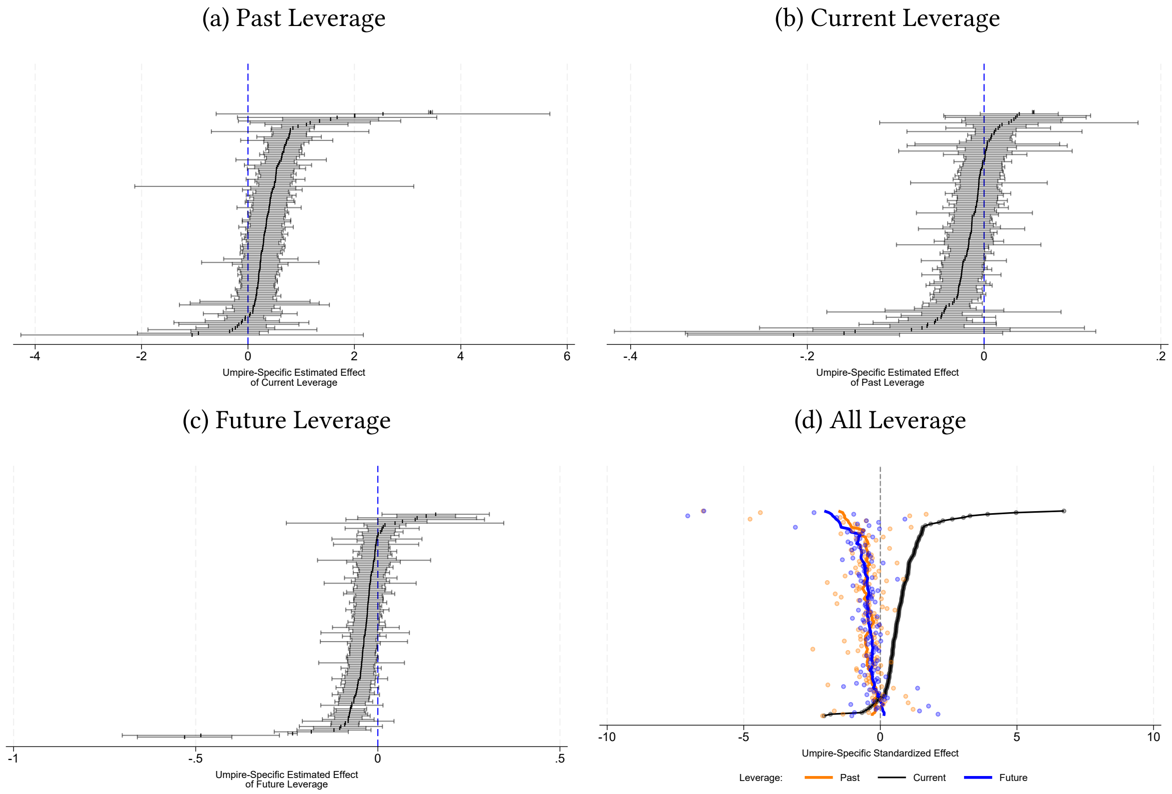


(c) Future Leverage



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: Comparison of Parametric and Nonparametric Effects of Leverage on Umpire Decision Accuracy



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	290.918
	(40.048)
Game Total Called Pitches	118.338
	(19.549)
Game Total Leverage	1.635
	(0.730)
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)
Correct Call	0.840	0.840
	(0.366)	(0.366)
Current Leverage	0.014	0.014
	(0.016)	(0.016)
Past Leverage (Current Inning)	0.137	0.137
	(0.200)	(0.203)
Expected Future Leverage (Current Inning)	0.170	0.171
	(0.136)	(0.138)
Lag Leverage Half Inning - 1	0.270	0.270
	(0.275)	(0.275)
Pitch release point (X-axis)	-2.127	-2.107
	(10.909)	(10.898)
Pitch release point (Y-axis)	27.006	26.987
	(4.532)	(4.543)
Pitch release point (Z-axis)	-22.257	-22.354
	(8.976)	(9.002)
Pitch spin direction (deg)	180.279	180.065
	(66.080)	(66.450)
Pitch spin rate (rpm)	1,795.590	1,789.173
	(670.546)	(672.578)
Pitch initial velocity (mph)	87.986	87.942
	(6.018)	(6.045)
Pitch break angle (deg)	5.306	5.263
	(24.862)	(24.787)
Pitch break length (in)	6.466	6.495
	(2.945)	(2.956)
Pitch break (Y-axis)	23.803	23.803
	(0.100)	(0.100)
Pitch final velocity (mph)	81.019	80.979
	(5.390)	(5.413)
N	3,140,224	2,951,251

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) 50th Percentile Current Leverage Full Game

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Bottom 6	2	0	1	2nd 3rd	5	.0097246	.018712
High	Top 7	0	1	0	3rd	-2	.0097233	.5181867

(b) 75th Percentile Current Leverage Full Game

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Top 4	2	1	2	2nd	3	.0177252	.0391887
High	Bottom 7	1	3	1	2nd 3rd	0	.0177395	.3611676

(c) 95th Percentile Current Leverage Full Game

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Bottom 1	2	2	2	2nd	-3	.0421248	.0752991
High	Bottom 9	0	1	2	None	-2	.0421337	.5270487

(d) 50th Percentile Current Leverage Single Inning

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Bottom 7	2	1	0	2nd 3rd	-5	.0097306	.1082078
High	Bottom 7	1	0	1	1st	1	.0097176	.1316078

(e) 75th Percentile Current Leverage Single Inning

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Bottom 7	0	2	1	1st 2nd	-8	.0177286	.1255942
High	Bottom 7	1	3	1	2nd 3rd	0	.0177395	.3611676

(f) 95th Percentile Current Leverage Single Inning

	Inning	Outs	Balls	Strikes	Baserunners	Score Diff.	Current Leverage	Future Leverage
Low	Bottom 8	2	1	2	1st 2nd 3rd	-4	.0420924	.1527406
High	Bottom 8	0	0	1	1st 2nd	-1	.0421085	.7395389

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: Effect of Leverage on Umpire Decision Accuracy, Innings 2-9

	Primary Spec (1)	Half Inning Lags (2)	Half Inning Lags (3)
Current Leverage	37.1577***	37.1561***	37.1560***
	(1.8696)	(1.8694)	(1.8694)
Past Leverage	-1.4971***	-1.4969***	-1.4957***
	(0.1322)	(0.1322)	(0.1324)
Expected Future Leverage	-3.5497***	-3.5494***	-3.5452***
	(0.2127)	(0.2127)	(0.2137)
Lag Leverage Inning − 1		0.0136	0.0134
		(0.0906)	(0.0906)
Lag Leverage Inning − 2			-0.0171
			(0.0950)
Inning 2	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Inning 3	0.1155	0.1157	0.1155
	(0.0865)	(0.0866)	(0.0866)
Inning 4	-0.0561	-0.0558	-0.0562
	(0.0877)	(0.0877)	(0.0877)
Inning 5	0.0455	0.0458	0.0455
	(0.0877)	(0.0877)	(0.0877)
Inning 6	0.0318	0.0321	0.0317
	(0.0881)	(0.0881)	(0.0882)
Inning 7	0.0313	0.0315	0.0312
	(0.0883)	(0.0883)	(0.0883)
Inning 8	-0.0435	-0.0432	-0.0435
	(0.0885)	(0.0885)	(0.0886)
Inning 9	-0.3804***	-0.3801***	-0.3799***
	(0.0969)	(0.0969)	(0.0969)
N Pitches	2,692,669	2,692,665	2,692,665
N Clusters	26,534	26,534	26,534
Correct Rate	0.840	0.840	0.840

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: The Effect of Leverage on Umpire Decision Accuracy 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	38.9426***	50.4411***	52.1035***	39.0338***	46.1528***	43.7753***	34.7163***	14.8489***
	(6.0379)	(6.1183)	(5.9047)	(5.8668)	(5.4557)	(4.9858)	(4.4919)	(4.2833)
Past Leverage	-1.8465***	-2.5174***	-3.0181***	-1.7468***	-1.1837***	-1.4791***	-0.9122***	-0.8812***
	(0.4336)	(0.4559)	(0.4311)	(0.4212)	(0.3656)	(0.3458)	(0.3239)	(0.3030)
Expected Future Leverage	-4.8566***	-4.8649***	-5.8007***	-3.9016***	-3.5557***	-3.6158***	-3.3346***	-1.9582***
	(0.7963)	(0.7801)	(0.7259)	(0.6767)	(0.6010)	(0.5399)	(0.4795)	(0.4399)
Inning	0.0000	0.0475	0.0730	-0.1304	-0.3655**	-0.2862	-0.3611**	-0.6721***
	(0.0000)	(0.2093)	(0.2017)	(0.1950)	(0.1863)	(0.1812)	(0.1763)	(0.1832)

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. Regressions include game fixed effects.

Table 6: The Effect of Leverage on Umpire Decision Accuracy by Distance to the Strike Zone Boundary

	Primary Spec	Pitch Distance to Boundary 0-3 in.	3-6 in.	6-9 in.
	(1)	(2)	(3)	(4)
Current Leverage	37.1577***	21.7171***	21.0724***	3.9433***
	(1.8696)	(3.7787)	(2.3514)	(1.0635)
Past Leverage	-1.4971***	-0.7325***	-1.6811***	-0.4046***
	(0.1322)	(0.2543)	(0.1696)	(0.0778)
Expected Future Leverage	-3.5497***	-2.0606***	-2.3271***	-0.4571***
	(0.2127)	(0.4162)	(0.2703)	(0.1233)
N Pitches	2,692,669	1,122,591	928,851	641,221
Correct Rate	0.840	0.687	0.922	0.990

Estimates from a linear probability model that the umpire makes a 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. Column 1 repeats the preferred specification. Columns 2 to 4 limit the sample to pitches passing within successive 3-inch bands around the strike zone boundary.

Table 7: The Effect of Leverage on Umpire Decision Accuracy Including Controls for Pitch Attributes

	Primary Spec (1)	Attrib Controls (2)	PDS-Lasso 1-Order (3)	PDS-Lasso 2-Order (4)	PDS-Lasso 3-Order (5)	DDML GBT (6)
Current Leverage	37.1577***	32.8137***	33.4155***	12.9966***	10.2305***	17.9911***
	(1.8696)	(1.8693)	(1.8726)	(1.8479)	(1.8432)	(1.7410)
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Past Leverage	-1.4971***	-1.7102***	-1.7297***	-1.4423***	-1.4044***	-1.9760***
	(0.1322)	(0.1326)	(0.1325)	(0.1310)	(0.1305)	(0.2459)
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Expected Future Leverage	-3.5497***	-2.7730***	-2.7646***	-1.0316***	-0.8612***	-1.6871***
	(0.2127)	(0.2125)	(0.2123)	(0.2110)	(0.2106)	(0.2564)
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N Pitches	2,692,669	2,692,665	2,692,665	2,692,665	2,692,665	2,692,665
Correct Rate	0.840	0.840	0.840	0.840	0.840	0.840
Specification Details						
Attribute Controls	No	Yes	Yes	Yes	Yes	Yes
Attribute Interaction Levels	N/A	1	1	2	3	N/A
Attribute Model Selection	None	None	PDS-Lasso	PDS-Lasso	PDS-Lasso	GBT

Estimates from a linear probability model that the umpire makes a 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. Included pitch attributes described in Section 6.2.1. For attribute interaction levels greater than 1 candidate attribute controls are the Kronecker product of all pitch attributes. DDML estimates are a partly linear model estimated using double-debiased machine learning, with linear controls for the parameters of interest and nonlinear nuisance function pitch attributes estimated using gradient boosted trees.

Table 8: The Effect of Leverage on Umpire Decision Accuracy Addressing Sample Selection

	Primary Spec (1)	No Selection /w Limits (2)	Heckit No Attributes (3)	Heckit With Attributes (4)	DDML GBT (5)
Current Leverage	37.1577***	40.2392***	79.7090***	76.7083***	83.5523***
	(1.8696)	(1.7630)	(1.9919)	(1.9285)	(2.0282)
Past Leverage	-1.4971***	-1.4819***	-1.7060***	-1.9593***	-1.2715***
	(0.1322)	(0.1216)	(0.1454)	(0.1460)	(0.1050)
Expected Future Leverage	-3.5497***	-3.8786***	-6.3381***	-5.5919***	-6.3472***
	(0.2127)	(0.1982)	(0.2512)	(0.2493)	(0.2576)
N Pitches	2,692,669	3,102,824	5,514,696	5,514,696	5,514,696
Specification Details					
Control Function Type	None	None	Probit	Probit	DDML/GBT
Limited Fixed-Effects	No	Yes	Yes	Yes	No
Limited Sample	No	Yes	Yes	Yes	No
Pitch Attribute Controls	No	No	No	No	Yes
N Bootstrap Reps	N/A	200	200	200	NA

Estimates from selection-adjusted linear probability model that the umpire makes a correct call for a given pitch. Column 2 repeats the primary specification without accounting for selecting, but imposing sample limits and fixed effects used in columns 3 and 4, and performs inference using the cluster bootstrap. Standard errors clustered at the game level shown in parenthesis and are computed using the cluster bootstrap in Columns 3 and 4. 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. Columns 3-4 account for selection using a control function estimated using a probit adding a machine-learning predicted probability the batter swings as an additional control. DDML estimates are a partly linear model estimated using double-debiased machine learning, with linear controls for the parameters of interest and inverse Mills' ratio and nonlinear nuisance function pitch attributes estimated using gradient boosted trees.

Table 9: The Effect of Leverage on Umpire Decision Accuracy Using Alternative Fixed Effects

	Primary Spec (1)	Alternative FEs (2)	Coarse FEs (3)	Player FEs (4)
Current Leverage	37.1577***	36.6493***	36.5422***	38.7357***
	(1.8696)	(1.8557)	(1.8557)	(1.8761)
Past Leverage	-1.4971***	-1.3829***	-1.3795***	-1.4788***
	(0.1322)	(0.1283)	(0.1284)	(0.1324)
Expected Future Leverage	-3.5497***	-3.4017***	-3.3984***	-3.4591***
	(0.2127)	(0.2040)	(0.2044)	(0.2161)
N Pitches	2,692,669	2,692,665	2,692,665	2,692,578
Correct Rate	0.840	0.840	0.840	0.840
Fixed Effects				
Game	X			X
Umpire	N/A	X	X	N/A
Date	N/A	X		N/A
Home and Away Team	N/A	X	X	N/A
Month, Day, Year	N/A	N/A	X	N/A
Starting Hour	N/A	X	X	N/A
Pitcher and Batter				X

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. N/A denotes a level of fixed effect that perfectly nests a more granular fixed effect and is therefore excluded.