Rational Inattention in the Infield

VIVEK BHATTACHARYA[1] and GREG HOWARD[2]

[1]Department of Economics, Northwestern University and NBER
[2]Department of Economics, University of Illinois

December 11, 2020

ABSTRACT. This paper provides evidence of rational inattention by experienced professionals in strategic interactions. We add rational inattention to a game of matching pennies with state-dependent payoffs. Unlike the full-information mixed-strategy Nash equilibrium, payoffs of different actions need not be equated state-by-state. Moreover, players respond partially to payoff differences, this responsiveness is stronger when attention costs are lower, strategies converge to full-information Nash as stakes increase, and average payoffs across all states are approximately equal across actions. We test these predictions using data on millions of pitches from Major League Baseball, where we observe strategies, payoffs, and proxies for attention costs.

KEYWORDS. rational inattention, behavioral economics, cognitive costs.

JEL CODES. C72, D83, D91.

CONTACT. vivek.bhattacharya@northwestern.edu and glhoward@illinois.edu. We would like to thank Alex Bartik, Glenn Ellison, Daniel Ershov, Gastón Illanes, Nellie Liang, Jonathan Libgober, Matt Lowe, Alex MacKay, Adam Smith, and Farhan Zaidi for useful discussion as well as participants at the Illinois Applied Micro Lunch, the Strategy, Competition, and Markets Conference, and the Reading Online Sports Economics Seminar. All errors are our own.

You can't think and hit at the same time.

Yogi Berra, New York Yankees

Given the complexity of decision-making in most situations, inattention is an intuitively appealing modification to standard models of rationality. Rational inattention—that agents choose to pay attention to certain information and not others—germinated in the macroeconomics literature with Sims (2003) and follow-on work. The idea has since percolated into finance, industrial organization, and other branches of theory.[1] Despite the important theoretical contributions, empirical analysis of the relevance of inattention in the field has been comparatively limited, with most evidence coming from controlled laboratory settings. Maćkowiak et al. (2018) and Gabaix (2019), both recent surveys of the literature, provide a call to arms to evaluate the importance of inattention in real-world settings. As Maćkowiak et al. (2018) write,

> there is very high value to testing RI [rational inattention] in the field, since decision situations in the lab do not always fit the setup for which RI is meant to be applied to. Often, in the lab the agent does not have enough time or proper motivation to design the optimal strategy (as in repeated decisions in real life) and moreover, information is also often available in many fewer forms inside of the lab.

In this paper, we provide the first evidence that inattention affects the behavior of professionals in a high-stakes strategic setting. Our empirical setting is Major League Baseball, and we focus on the strategies of pitchers and batters pitch-by-pitch. A number of aspects of this setting make it an ideal one to test for rational inattention. Rich information about the *state* is in principle available to the players: they know exactly who they are facing in each interaction and can identify the exact situation of the game—including the score difference, the positions of the runners, the count, the number of outs, the inning, and many other aspects—but processing such a complex state space is likely difficult. Nevertheless, these professionals are highly trained and have been exposed to similar situations many times: unlike in one-shot laboratory settings, these agents would have had the chance to hone their information acquisition strategies through practice, introspection, or analysis ahead of time.[2] Finally, because of the extensive training and high stakes—industry analysts estimate the market value of a win to be about $10 million (Swartz, 2017)—there is a long tradition in economics of using sports as a model for understanding strategic behavior in the field. Evidence of inattention here may suggest that it is also prevalent in other strategic settings.

Baseball also has features that allow us to test specific implications of rational inattention. We can propose natural determinants of the cost of attention. The game naturally has variable

[1] There are many papers exploring theoretical implications. Examples include Woodford (2008) and Maćkowiak and Wiederholt (2015) in macroeconomics; Kacperczyk et al. (2016) and Van Nieuwerburgh and Veldkamp (2009) in finance; Matějka (2015) and Martin (2017) in firm pricing and industrial organization, and Hellwig and Veldkamp (2009), Yang (2015), Dessein et al. (2016), and Denti (2020) in theory. See Maćkowiak et al. (2018) for a survey.

[2] Maćkowiak et al. (2018) provide the examples of households making repeated consumption decisions and human resources managers learning how to quickly scan CVs of applicants after practice.

stakes, as certain situations and aspects of the state are more important than others—which leads to predictions on behavior through the rational inattention model. In addition to predictions from the single agent decision problems with a rationally inattentive agent, the strategic nature of the setting allows us to derive and test other predictions. Perhaps most importantly, the game is especially transparent and many aspects are easily measurable. Even though each pitch occurs under different and complex circumstances, the econometrician observes very good proxies for the state. The econometrician also observes good proxies for the realized payoff rather than having to infer payoffs from choice behavior, which provides tests of the model.

We model the environment as a matching pennies game between the pitcher and the batter. Payoffs of this game are driven by a random state of nature, corresponding to the different situations that might arise in baseball. Throughout the paper, we consider pitch choice as the main strategic decision for the pitcher.[3] In a Nash equilibrium, strategies would be such that payoffs are equalized across pitch types *state-by-state*. We first show that, unlike in other settings where professionals play mixed strategies, payoffs differ significantly across pitch types. We verify these predictions using a variety of payoff metrics, including both a team-level payoff—the win probability added—and a player-level payoff—field-independent pitching.

The previous results identify systematic patterns in the departures from Nash equilibrium, and we argue that these are consistent with rational inattention. Within the matching pennies framework, we embed the assumption that players are not fully attentive about the state but instead receive a noisy signal about it through a costly information acquisition strategy. Critically, this assumption allows for departures from full information Nash equilibrium, but it still places restrictions on the relationship between strategies and equilibrium payoffs. Using this model, we derive four predictions to help distinguish rational inattention from other departures from Nash. We then test these predictions in the baseball data.

Consider the situation from the perspective of a single agent, the pitcher, best-responding to a distribution of payoffs in each state. Results from Matějka and McKay (2015) show that the rational inattention model with the Shannon cost of information—which we focus on in this paper—resembles a logit discrete choice problem, where the magnitude of the error term is parameterized by the cost of attention. This implies that (A) pitchers will pitch more fastballs in states in which fastballs have higher *equilibrium* payoffs, and (B) pitchers with lower costs of attention will be more responsive to the payoff difference between fastballs and other pitches. We use a variety of proxies for the cost of attention. First, we use the emphasis that a team places on analytics, which can capture benefits that an analytics department may have on players' ability to respond appropriately to different states. Second, we use the player's age, which can

[3] In baseline specifications, we focus on the most salient decision facing a pitcher—a fastball vs. other pitches. However, we consider robustness to finer classifications of pitches throughout.

capture variation in inherent information processing ability. Finally, we use "quasi-experimental" variation generated by temperature and the timing of the game that could affect players' cognitive performance. We test these predictions and discuss the interpretation of the various proxies.

The strategic nature of the setting leads to additional implications of rational inattention. A third prediction is that (C) in states with larger underlying payoffs, the strategies pursued by the players approach the strategies they would choose in the full-information Nash equilibrium state-by-state. To test this prediction, we use *leverage*—variation in win probability added—as a proxy for such high-stakes situations. Essentially, if a particular game scenario could lead to large swings in the win probabilities for the two teams, then these scenarios are especially important. We provide indirect evidence that strategies approach the full-information Nash equilibrium. This indirect evidence comes from showing that proxies for payoffs that *do not* scale with leverage are equated across pitch types in high-leverage situations but not in others.

While the previous set of predictions are consequences of rational inattention, they can also be rationalized by unobserved pitch-level heterogeneity (or "errors"), especially if there is reason to believe that the magnitude of these errors are correlated with correlates for attention (for (B)) and independent of the leverage of the state (for (C)). For instance, quantal response equilibrium (QRE) may be a candidate explanation for these observations. We finally show that an often-overlooked implication of rational inattention is a cross-state restriction: (D) the fact that pitchers are throwing both fastballs and other pitches with positive probability implies that payoffs averaged across all states must be "approximately" equal, in a manner that we formalize. If payoffs for fastballs averaged across states were much larger than those of other pitches, then the pitcher would do better by just pitching fastballs in all states and not exert effort acquiring information about the state. While an analogous prediction would hold in a single-agent setting, the strategic setting provides conditions so that the pitcher must mix, making (D) a necessary condition.[4] We test this prediction directly and show payoffs do satisfy the approximate zero condition we derive. Moreover, this approximate zero condition is tight, so it does provide a strong test to reject rational inattention.

We further use the data to compute the cost of attention—not only to compute the power of the test described above but also to provide another way to interpret the magnitude of the behavioral friction through the lens of the model. The best response function of the pitcher serves as an estimating equation to quantify the cost of attention. Using this estimated cost of attention, the observed strategies, and the popular estimate for the market value of a win, we estimate that the information acquisition strategy that fully reveals the state would cost the equivalent of approximately $9,770 per pitch to implement on average—under the assumption

[4] In a single-agent case, high attention costs would simply mean agents use fewer actions in equilibrium (e.g. only throw fastballs even if non-fastballs are better occasionally). Here, the agent continue to consider all options, and the payoffs adjust in equilibrium.

that equilibrium payoffs do not change, so that other teams do not respond to such a unilateral deviation in information acquisition. By contrast, the current strategy heavily economizes on these attention costs, amounting to only $234 per pitch. These estimates are consistent with anecdotal evidence that processing information in the heat of the moment is demanding even for professional players: thinking and pitching at the same time is indeed difficult. Moreover, we find that the model imposing Shannon entropy exhibits good fit quantitatively as well.

In Section I, we propose a model of baseball and derive testable implications of the rational inattention framework, and we discuss the data in Section II. Section III argues that the data is not consistent with Nash equilibrium play. Sections IV and V verify the predictions of the rational inattention model and take it directly to the data. Appendix A provides a description of baseball that is sufficient to understand this paper, even for an uninitiated reader. The remaining appendices collect proofs and provide further theoretical and empirical results.

Related Literature. Rational inattention traces back to the theoretical work of Sims (2003), but recent work has been important in developing empirical implementations of these models. Matějka and McKay (2015) shows that rational inattention with the standard "Shannon" information cost leads to a logit-like discrete choice model, and Fosgerau et al. (2020) extends this to more general costs. Caplin et al. (2016) connects this intuition to inference from market share data. Caplin et al. (2019) shows a connection between rational inattention and consideration sets, which is similar to a prediction used in this paper for the payoffs of actions being approximately equal.

A large body of work connects these theoretical observations to data from the lab. Caplin and Dean (2015) derives a characterization of rationally inattentive behavior for general cost functions and uses this characterization to test behavior in lab settings. Dean and Neligh (2019) and Dewan and Neligh (2020) use lab experiments to estimate these cost functions and show some evidence of departures from Shannon entropy. This literature is too extensive to cite thoroughly, and Section 4.4 of Maćkowiak et al. (2020) provides a review.

On the other hand, tests of the model and estimates of the cost function remain considerably more limited outside the laboratory, and this paper contributes to this smaller but growing literature on rational inattention in a real-world setting. Cohen and Frazzini (2008), Menzly and Ozbas (2010), and Kacperczyk et al. (2016) show evidence from firm returns and mutual funds consistent with inattention. Mondria et al. (2010) and Sicherman et al. (2016) use online search and login behavior as a measure of attention in finance. Bartoš et al. (2016) use response to a correspondence study of landlords as a measure of attention. Taubinsky and Rees-Jones (2018) and Morrison and Taubinsky (2020) look at inattention to sales tax in online shopping experiments. A strand of papers measure and manipulate beliefs, mostly in the context of forecasting, to test rational inattention (Coibion et al., 2018; Gaglianone et al., 2020). Joo (2020) was the first to take

Matějka and McKay (2015) to demand estimation, and contemporaneous work considers health insurance (Brown and Jeon, 2020) and migration (Porcher, 2020). We are unaware of much work in real-world strategic settings.[5] We provide further evidence that professionals are rationally inattentive and extend tests to a strategic setting, which allows us to propose further tests—such as the convergence to Nash and the approximate payoff equivalence—that have not been tested before.

Gabaix (2019) argues that inattention more broadly can unify many behavioral biases. Through this interpretation, there is more empirical evidence on inattention. Agents are inattentive to types of prices (see Chetty et al. (2009) and Ellison and Ellison (2009) for taxes and Brown et al. (2010) for shipping costs), product characteristics (see Handel and Kolstad (2015) for health plans and Lacetera et al. (2012) for car sales), and even financial information (DellaVigna and Pollet, 2009). Even so, there is considerably less evidence in the context of strategic settings.[6]

Finally, there is a long tradition in economics of using sports as a laboratory to test predictions of game theory. Multiple studies have suggested that professionals do randomize mixed strategies consistent with theory: Walker and Wooders (2001) and Hsu et al. (2007) show this in the context of serves in tennis, and Chiappori et al. (2002) and Palacios-Huerta (2003) offer evidence in penalty kicks in soccer. In arguably more complex settings, results have not been as clean. Romer (2006) shows that football coaches are playing nonequilibrium strategies in the decision to punt on a fourth down. Pope and Schweitzer (2011) shows evidence that professional golfers exhibit loss aversion, even though eliminating this bias would be profitable. Green and Daniels (2018) shows that baseball umpires follow heuristics that are comparable to Bayesian updating. Phillips (2017) argues managers exhibit left-digit bias when substituting pitchers and discusses a connection with inattention by exploiting variation in stakes. Relative to these papers, we use a rich dataset to show evidence of departure from Nash equilibrium play in games, but we identify patterns in these departures consistent with a behavioral theory of inattention.[7,8]

[5]See Afrouzi (2020) for an example using survey evidence to analyze firms' inattention to aggregate shocks.

[6]See Brocas et al. (2014) for an exception in the context of level-k thinking in games. The limited evidence does not reflect a lack of theoretical interest in strategic settings with some element of inattention (Matějka and McKay, 2012; Bordalo et al., 2016). DellaVigna (2009) discusses inattention more broadly.

[7]A related and interesting paper is Kovash and Levitt (2009), who use a similar design to argue that fastballs are lower payoff on average. While we are able to replicate their results using their payoff metric (on-base percentage plus slugging) and state controls, we find our result is qualitatively more robust to different payoff metrics and finer definitions of the state. Moreover, we provide an explanation of our results through the lens of a behavioral model. Appendix C.1 replicates their results on our cut of the dataset and discusses differences in the research design.

[8]Another strand of the literature (Palacios-Huerta and Volij, 2008; Wooders, 2010; Levitt et al., 2010) has broadly found that the same professionals who play minimax in their professional settings do not play the Nash equilibrium in isomorphic lab settings. The results in this paper might suggest that the heuristics from sports are context-specific.

I. Model and Predictions

We model baseball as a game of generalized matching pennies with state-dependent zero-sum payoffs. Matching pennies is a natural model for the strategic situation. The pitcher (with the help of the catcher) covertly chooses a pitch and the batter prepares for a pitch, neither knowing the the action of the other. The batter would like to prepare for the pitch the pitcher will throw. The game is zero-sum, so the pitcher would prefer to throw a pitch the batter is not prepared for. Moreover, payoffs for different types of pitches are dependent on the state of the game—the count, situation in the inning, score difference, etc. This difference is because fastballs, one of the two main categories of pitches, are more likely to put a ball into play, which could be differentially beneficial depending on the state.[9]

In this paper, we add information acquisition to this otherwise standard model of baseball (Kovash and Levitt, 2009). The state space is especially complex, and neither agent has the time or the mental capacity to process it fully before each pitch. Nevertheless, they spend considerable effort planning strategies ahead of the game so that decision making can be quick on the spot. Another quote from Yogi Berra, a Hall of Fame player and manager for the New York Yankees, highlights this aspect:

> You do your thinking before you get up to bat. We used to spend a lot of time before the games talking about certain pitchers, what they threw, and what was the best way to hit them in certain situations. We did a lot of talking and a lot of thinking about hitting. We just didn't stand there thinking when we were up to bat. (Barra, 2010)

This combination of the ability to generate an information acquisition strategy, but limited information processing capacity when executing the strategy, suggests that rational inattention may be a natural model.

I.A. Setup

Each pitch in baseball is a simultaneous game between a pitcher and a batter that occurs in some state s.[10] The pitcher can pick an action $a_P \in \{f, n\}$ for pitching a fastball and not pitching a fastball, respectively. The batter also picks a strategy $a_B \in \{f, n\}$, corresponding to readying himself for a fastball or for a different pitch. The payoff to the pitcher from the profile (a_P, a_B) is $w_s(a_P, a_B)$, which importantly depends on the state s. The game is zero-sum, so the payoff to the batter is $-w_s(a_P, a_B)$. The batter would prefer to "coordinate" with the pitcher—play $a_P = a_B$—whereas the pitcher would not. Thus, we assume that $w_s(f, f) < w_s(f, n)$, $w_s(n, n) < w_s(n, f)$,

[9] For instance, a fly out—which requires the ball to be put into play—can be less beneficial to the pitcher than a strikeout when a runner is on third base, since the runner has a chance to score. With no runners on base, a fly out and a strikeout would be equivalent.

[10] The state would include aspects of the game that affect the payoffs of different actions. For example, it would include the number of balls and strikes and the score. We discuss in detail how we interpret the state in Section II.

$w_s(f, f) < w_s(n, f)$, and $w_s(n, n) < w_s(f, n)$. In this sense, the game resembles matching pennies, and the full-information Nash equilibrium would be in mixed strategies.

However, pitchers and batters do *not* observe the state s costlessly, and they are rationally inattentive to it. Before playing an action, each agent decides on a flexible information acquisition strategy $F(\mu, s)$, which corresponds to a joint distribution of signals μ and potential states s. This joint distribution must be consistent with the prior $G(s)$ of the probability of each state s occurring. Throughout this paper, we assume that the cost of this strategy is

$$\lambda \cdot \left(H(G) - \mathbb{E}_\mu H\left(F(\cdot|\mu)\right)\right), \tag{1}$$

where $H(\cdot)$ is the entropy of a distribution. This so-called Shannon information cost has been used in many models of rational inattention, including Sims (2003), Matějka and McKay (2015), and Yang (2015). The interpretation of this cost function is that it is costlier to choose strategies that the agent expects will decrease the entropy significantly: since entropy captures a notion of imprecision of information, strategies that yield more precise information about states are costlier.[11] We follow the literature on coordination games with flexible information acquisition and assume that agents do not see each others' information acquisition strategies at any point (Yang, 2015). Indeed, the model we present here is similar to the one in Yang (2015), but adapted to a game of strategic substitutes rather than strategic complementarities.

In the most general formulation of the game, the strategy space of each player i consists of a set of signal realization M_i, the information acquisition strategy $F(\mu|s)$ is a probability distribution over M_i, and a strategy σ_i maps signals M_i to actions. However, as Woodford (2009) and Yang (2015) argue, a property of the Shannon entropy cost is that M_i will take on only two values here (see Cover and Thomas (1991) for a discussion of information inequalities). Interpreting these as "messages" to play a fastball or not, a strategy simply specifies a probability p_s^f that the pitcher picks a fastball in state s. Let the strategies of the batter be given by q_s^f. Conditioning on these strategies, the payoff from the game to the pitcher will be

$$\begin{aligned} U_P(p, q) = \mathbb{E}_s \Big[& p_s^f \cdot q_s^f \cdot w(f, f) + p_s^f \cdot \left(1 - q_s^f\right) \cdot w(f, n) \\ & + \left(1 - p_s^f\right) \cdot q_s^f \cdot w(n, f) + \left(1 - p_s^f\right) \cdot \left(1 - q_s^f\right) \cdot w(n, n) \Big]. \end{aligned} \tag{2}$$

[11] We make the assumption that agents incur a cost proportional to the Shannon entropy. Another way of modeling this is to assume that agents have a constraint on the expected Shannon entropy of their information acquisition. In Appendix B.2, we show this alternative formulation leads to the same predictions we derive in this section. Further, we explore the degree to which our predictions generalize beyond Shannon entropy in Appendix B.3.

Moreover, the cost is

$$c(p;\lambda) = \lambda \cdot \left\{ \mathbb{E}_s \left[p_s^f \log p_s^f + \left(1 - p_s^f\right) \cdot \log\left(1 - p_s^f\right) \right] - p_0^f \log p_0^f - \left(1 - p_0^f\right) \log\left(1 - p_0^f\right) \right\}, \tag{3}$$

where $p_0^f \equiv \mathbb{E}_s p_s^f$, the overall probability of throwing a fastball across all states. Combining (2) and (3), the pitcher maximizes $U_P(p,q) - c(p;\lambda)$. The batter maximizes the analogous equation.

The best response of the pitcher solves

$$\log p_s^f - \log\left(1 - p_s^f\right) = \log p_0^f - \log\left(1 - p_0^f\right) + \frac{1}{\lambda}\left(w_s^f - w_s^n\right), \tag{4}$$

where $w_s^f \equiv q_s^f \cdot w(f,f) + \left(1 - q_s^f\right) \cdot w(f,n)$ is the expected payoff of f given the batter's equilibrium strategy q_s^f.[12] Importantly, we are later able to econometrically measure $w_s^f - w_s^n$, i.e. the difference in equilibrium payoffs of throwing a fastball or not. The result in (4) is an important observation for our analysis, and it has been established in multiple papers: see Matějka and McKay (2015) for a single-agent discussion and Proposition 2 in Yang (2015) for the case of the game.[13]

Equation (4) establishes that there is a tradeoff between the cost of picking a state-specific strategy p_s^f too far from the average strategy p_0^f and the benefit of throwing a higher-payoff pitch. The pitcher can also choose the average strategy p_0^f, and he does so in a utility-maximizing way. Lemma 2 in Matějka and McKay (2015) shows that p_0^f solves

$$\max_{p_0^f} \mathbb{E}_s \log\left[p_0^f \exp(w_s/\lambda) + 1 - p_0^f\right], \tag{5}$$

where $w_s \equiv w_s^f - w_s^n$. This formulation helps understand comparative statics on p_0^f and could be used to derive testable conditions on payoffs given $p_0^f \in (0,1)$.

I.B. Empirical Predictions

The most basic property of the rational inattention model is that payoffs need not be equalized across strategies within state.

(0) It is not necessarily the case that $w_s = 0$.

[12] Note that (4) is true only if p_0^f is strictly between 0 and 1. If p_0^f is 0 or 1, then $p_s^f = p_0^f$ in all states. Further, note that (4) shows that a pitcher's strategy depends on its perceived payoffs w_s^f and w_s^n to fastballs and non-fastballs in each state. In a setting where a pitcher faces uncertainty over his opponent's attention cost, the heterogeneity in perceptions of payoffs in each state could come from beliefs over the attention cost as well.

[13] Note also that (4) is a best response characterization and thus a necessary condition of equilibria. As such, the predictions below will not require uniqueness of equilibria.

Prediction (0) differentiates the model from state-by-state Nash equilibrium. Given this equilibrium distribution of payoff differences, the results in (4) and (5) treat the game as a single-agent decision problem and lead to the following two predictions, which are direct consequences of (4) and are proved in Appendix B.1.

(A) The probability of pitching a fastball is larger in states where the equilibrium benefit of throwing a fastball w_s is larger.

(B) Agents with lower costs of information λ are more sensitive to payoff differences: for states s and t as associated relative payoffs $w_s > w_t$ for fastballs, $\mathcal{L}(p_s^f) - \mathcal{L}(p_t^f)$ increases as λ decreases, where $\mathcal{L}(p) \equiv \log p - \log(1-p)$.

That is, rationally inattentive agents are at least somewhat responsive to variation in payoffs across states, although those with higher costs of attention are less so.[14]

A third prediction is a consequence of the strategic nature of the setting and does not merely rely on single-agent predictions.

(C) Suppose there is a sequence of states parameterized by $L > 0$ so that $w_t(\cdot,\cdot;L) = L\cdot\tilde{w}(\cdot,\cdot)$. Then, as $L \to \infty$, the strategies in the rationally inattentive equilibrium converge to the strategies in Nash equilibrium of the full-information game with payoffs $\tilde{w}(\cdot,\cdot)$.

As the payoffs in a particular state become more variable, rationally inattentive agents pay "more attention" to this state, and play converges to Nash. While Appendix B.1 provides a formal proof, intuition follows from two steps. First, (4) shows that as $L \to \infty$, $p_t^f(w_t)$ converges to being 1 if $w_t > 0$ and 0 if $w_t < 0$. However, w_t is determined in equilibrium based on the batter's strategies in this state as well. The analogous equation for the batter shows that the batter will play fastball with probability close to 1 if his perceived payoff to fastballs is larger than 0—and will play with probability close to 0 otherwise. These are the same best responses as in the full-information Nash equilibrium. Importantly, (C) is not a prediction about the magnitude of the payoff differences in equilibrium as $L \to \infty$. The payoff difference is a function of both leverage and the opponent's strategy. It might converge to a non-zero value. The empirical strategy involves finding payoff proxies that do not scale mechanically with leverage, and which will converge to zero.

A final prediction provides necessary conditions for the solution to the maximization problem in (5) to be interior, although this result can be derived from (4) too.[15]

[14]In the empirical section, we aggregate Predictions (A) and (B) over coarse categories of states, e.g. the number of strikes. In testing Prediction (A) we often use the log odds of the average fastball probability, rather than the average of the log odds of fastball probability. This aggregation implicitly appeals to the fact that $\mathcal{L}(p)$ is well-approximated by $4p-2$ over a wide range. Similarly, note that Prediction (B) is not formally about differences in probabilities of throwing fastballs but about differences in log odds. We sometimes use probabilities in Section II for interpretability, but a number of our tests use log odds directly as well.

[15]These are conditions provided in Proposition 1(c) and Assumption 1 in Yang (2015). Our point here is that this is

(D) If the average probability p_0^f of pitching a fastball is in $(0, 1)$, then

$$\mathbb{E}_s \exp(w_s/\lambda) > 1 \text{ and } \mathbb{E}_s \exp(-w_s/\lambda) > 1. \tag{6}$$

We can interpret (D) as a statement that the relative payoff of a fastball is "approximately" zero when averaged across all states. To see why, note that $\mathbb{E}_s w_s = 0$ is a sufficient condition for (6). By Jensen's inequality, $\mathbb{E}_s \exp(w_s/\lambda) \geq \exp(\mathbb{E}_s w_s/\lambda) = 1$ if $\mathbb{E}w_s = 0$; the other inequality is analogous. In the empirical analysis in Section V.A, we will test for this prediction, which does not depend on estimates of λ. We should, however, caveat this test: the implication of rational inattention is not necessarily that payoffs are identically equalized on average, but rather that (4) holds.[16] As such, in Section V, we go further and test the fit of (4) and then use our estimates of λ to show that the bounds that (D) places on $\mathbb{E}_s w_s$ are fairly tight. However, the "approximate" zero interpretation provides clear intuition: Prediction (D) is the rational inattention analogue of the result that payoffs are equalized across strategies when players are mixing. If the batter were playing a strategy that yielded high payoffs for fastballs when averaged across states, the pitcher would prefer to not exert effort trying to understand the state: he would not be "mixing" and would simply throw fastballs regardless of the state.

In Appendix B.2 we derive Predictions (A)–(D) using an alternate formulation of rational inattention in which agents do not have a direct cost of information processing but rather a maximum amount of information they can process, given by the mutual information cost. As we discuss above, the empirical analysis of this paper is based on the Shannon information cost. Appendix B.3 investigates the extent to which these predictions generalize to other more general cost functions: to the extent that not all predictions generalize, they provide one-sided tests for rational inattention with Shannon entropy.

I.C. Rejecting Alternative Explanations to Rational Inattention

In addition to being predictions of rational inattention, Predictions (0) and (A)–(D) also help us distinguish rational inattention from other reasons that baseball players might not choose the strategies corresponding to Nash equilibrium.

Prediction (0) is a rejection of Nash equilibrium, but it is also a rejection of any other theory that predicts payoff equivalence. One example is the "simple framework" for inattention proposed in Section 2 of Gabaix (2019); here, players would only be partially responsive to payoff differences, interpreting them as a convex combination of the true difference and some default

also a necessary condition for there to be an interior equilibrium. Given we see an interior equilibrium, Prediction (D) must hold.

[16]Of course, (4) is a consequence of the functional form of Shannon entropy, which we maintain throughout this paper. With other costs of attention, (4) would take different forms; Appendix B.3 discusses some extensions.

value independent of state. Under natural interpretations of this model, we would still expect to observe payoff equivalence, so we can reject such a simple model of inattention.[17]

Similarly, observing pitchers mix strategies also helps us distinguish rational inattention from theories that would not predict stochastic choice. For example, a level-k thinker does not employ mixed strategies.[18]

However, there are theories besides rational inattention that could produce both mixing and deviations from payoff-equivalence. One theory is that agents do not react to the state, possibly because they are completely inattentive. Prediction (A) rejects this theory: a rationally inattentive agent is going to change strategies in response to the state. Another set of theories assumes that agents observe the state but have difficulty adjusting their strategy. For instance, they may play the default strategy with exogenous probability, and otherwise reoptimize. In this case, either the solution is not interior, i.e. the agents do not mix when allowed to reoptimize, or payoff equivalence continues to hold. This would generate a stair-step relationship between probability of a fastball and relative payoff of a fastball: the probability of a fastball would discontinuously rise when the estimated relative payoff of a fastball becomes positive, and then remain flat. This is in contrast to the smoother relationship between probability and payoffs as predicted by Prediction (A) and seen in Figure 4.

Another theory could be that baseball players consistently mis-perceive the payoffs in some counts, and they mistakenly believe that payoffs are indeed equalized. Prediction (B) would require that these beliefs change at times when we might expect attention to be lower-cost, which seems unintuitive. Prediction (C) would require that they only make this mistake in low-stakes situations. Both these requirements would seem like coincidences in absence of further details of a model justifying why we may expect them.

Finally, some of our predictions—together with the departure from payoff equivalence—bear resemblance to those from quantal response equilibria (QRE) proposed by McKelvey and Palfrey (1995). QRE posits that players are selecting actions based on their expected utilities plus some noise. Such a theory would also predict a relationship between payoffs and probabilities i.e. Prediction (A). However, since it is not a solution concept that inherently microfounds the source of this noise, generating Predictions (B) and (C) is possible but requires more structure. Moreover,

[17]There are probably multiple ways to implement the "simple framework for modeling attention," but we think the most natural one is one in which the pitcher observes observes only a discounted value of w_s, anchored to 0, which would be a reasonable default in a game with mixing. (In the notation of Gabaix (2019), x would be w_s and the anchoring point for every x would be 0.) As in Nash, the only time there is mixing is when there is indifference between actions. Observing a fraction of the payoffs would still have the same sign as observing the payoff itself.

[18]Randomness in how many levels are considered by the pitcher could lead to what looks like mixed strategies. However, the observed probability of a fastball would primarily be driven by this randomness, and not by the equilibrium payoffs. The state s could matter in that it determines the best response to a level-0 opponent, if the level-0 opponent were mixing. However, the observed probabilities would be limited to P(k is odd) and P(k is even). In this sense, Prediction (A) would serve as a rejection.

it does not link states to each other, so it does not necessarily generate Prediction (D). Section IV.D discusses this further.

I.D. Overview of Empirical Strategy

To take the model to the data, we need to take a stance on what the actions are, whose probabilities p we measure, the specific payoffs w_s^i for pitch type i, and the states s. As in the model above, our baseline analysis will focus on the binary classification of fastballs or other pitches; however, throughout many different points of the paper, we subdivide non-fastballs into finer pitch types and check analogous predictions. Second, we take advantage of the fact that we observe a number of proxies for payoffs. Our preferred metric is a direct measure of how the outcome of a pitch affects the probability the team wins the game, but we discuss other concerns—including individual pitcher-level metrics as well as metrics that incorporate more information about dynamics.

The state space requires some more discussion. As discussed above, baseball has a rich set of states, and we would expect that payoffs (w_s^f, w_s^n) (using the notation from the binary pitch classification) would differ between them. Ideally, we would have precise estimates of these payoffs in especially fine states, but this is not possible despite data on millions of pitches. As such, in regressions below we will control for fine state fixed effects that affect payoffs of all pitches in a particular state: if a team is leading by a lot near the end of a game, no pitch will have an appreciable effect on the outcome. We take a stance of a baseline categorization of states and show that further refinements do not change estimates appreciably. The predictions of rational inattention, however, depend on differences in the *relative* payoff of a pitch ($w_s = w_s^f - w_s^n$) across states. We compute these relative differences on a coarser classification of states; throughout the paper, we denote coarser classifications using σ. Purely for expositional purposes, we always begin with the number of strikes as the coarse category of states so that we can present these results in tables. However, the various scatter plots throughout the body of the paper as well as the supplementary figures in Appendix C show that the predictions are generally robust to estimating these payoff differences on considerably finer categories.[19] Moreover, estimates of the cost of attention use w_σ that are computed on these categories.

Finally, to take (B) and (C) to the data, we need measures of attention costs and L. We use a number of distinct proxies that we believe capture the spirit of rational inattention—that λ is an information processing cost that determines how difficult it is to acquire information to execute a pre-specified strategy. Our proxies include technological decisions by a team that can help information processing (investment in analytics), a player characteristic that is related to cognition (age), and environmental factors that the psychological literature suggests is related to

[19]In fact, we can usually go even further than the refinements considered in this paper, although we quickly run into the issue that we cannot estimate w_s well on these categories.

information processing (temperature and game timing) that varies across games and players. We use a measure of the variation in w_s as our measure of L. All these decisions are discussed in more detail in the relevant subsection.

II. Data

We use data from Major League Baseball's Gameday dataset, from PITCHf/x. We observe every pitch in every MLB game from 2008 to 2017. For each pitch, we observe quantities that help us control for the state of the game: the pitcher, batter, score, count, base situation, number of outs, and inning. We also observe the type of pitch thrown and the outcome of the pitch—how the count, outs, base situation, etc. evolve after the pitch. In most of the analysis, we group pitches into fastballs and other pitches, although throughout the paper we study robustness to a finer classification of pitches. After restricting to regular season games and dropping extra innings,[20] we are left with 24,305 games and 6,881,063 pitches.[21]

One of our two primary measures of payoff is the *win probability added* (WPA), measured at the at-bat level. This is the probability of winning the baseball game, given the inning, the number of outs, the position of any runners on base, and the score difference. It is calculated as the fraction of historical games in that situation that ended as a win for the pitcher. We use a measure of win probability compiled by Greg Stoll,[22] although we have verified that computing an in-sample win probability does not change results. For each pitch, we calculate the win probability at the end of the at-bat, and subtract the win probability at the start of the pitch.[23,24]

For robustness, we also study *field-independent pitching* (FIP) as another outcome metric. FIP is a commonly-used summary measure of outcomes that are most related to the pitcher's performance. It is calculated as

$$\text{FIP} = 13 \cdot \text{Home Run} + 3 \cdot (\text{Walk + Hit By Pitch}) - 2 \cdot \text{Strikeout}.$$

[20]We drop pitches that are pitch-outs or are uncategorized as a type of pitch.

[21]By the time we partial out our state space many of these pitches will be dropped because the state only occurs once in our dataset.

[22]This metric uses baseball games from 1957 to 2017. See https://gregstoll.com/~gregstoll/baseball/probs.txt.

[23]An at-bat is a collection of consecutive pitches between a pitcher and a hitter. Typically, the important dimensions of win probability (outs, score, etc.) are more likely to change at the end of an at-bat, and it is rare for them to change in the middle. The fact that win probability added is calculated at the at-bat level rather than the individual pitch level is not a problem. If pitches have no effect on the probability of winning once the outcome of that pitch is realized, then win probability added, at the at-bat level, is a noisy measure of the change in the probability of winning at the pitch level. If there are dynamic effects within the at-bat (e.g. throwing a fastball sets up a curveball on the next pitch), this measure will capture those effects as well.

[24]Given our controls for the state fixed effects, the results would be identical by just using the win probability at the end of the at-bat. However, we find the concept of win probability added to be more intuitive.

It is also measured for the at-bat level, and lower magnitudes indicate success for the pitcher.[25] This outcome is related to WPA: the relative magnitudes of the coefficients are similar to how much a strikeout, walk, or home run affects the win probability (on average). There are several reasons to look at FIP. First, the realization of FIP, unlike WPA, does not depend on the actions of other members of the pitcher's team. This may assuage concerns that pitchers prioritize statistics they have control over, rather than just winning the game. It also leads to a less random measure of outcomes, enabling more precise estimates. A final benefit of FIP is its scaling. Win probability added has higher variance in some situations in the game, unlike FIP. Given FIP will not mechanically scale, it is useful to test (C).

Furthermore, in some tests we also consider *weighted on-base average* (wOBA), which is similar to FIP, as another robustness check. It assigns a year-specific score to walks, hit-by-pitches, singles, doubles, triples, and home runs based on how much they contribute to a win by the batter's team. A lower wOBA is better for the pitcher.[26]

We sometimes group states by their importance. We use *leverage*, meant to capture how important an at-bat is for the outcome of the game: a tie game late in the game will have high leverage, and an at-bat when one team is already winning by a lot will have low leverage. We initially measure leverage as the standard deviation of WPA, based on the inning and score-difference of the at-bat, although we also consider other quantities developed by sabermetricians. This quantity proxies for L in (C): high leverage situations will have both especially positive and especially negative WPA, which is consistent with elements of the payoff matrix in that state having large variance. Appendix C.3 verifies that while WPA scales with leverage, FIP does not.

Finally, to test (B), we use three different types of proxies for the cost of attention. The first is related to coaching that a team may provide its players about how to react to different situations. Maćkowiak et al. (2018) note that a reason to test rational inattention models in the field is that agents may have had the opportunity to put a lot of thought into their information acquisition strategy, relative to agents in the lab. Given that analytics departments may aid pitchers in improving pitch selection situation-by-situation, we use how heavily the team in invested in analytics as a proxy. We have two measures of analytics. The first is the total number of defensive shifts a team conducts over our sample period. A defensive shift involves adjusting the fielders from their normal positions during a particular situation in a game. Implementing this strategy involves effective analytics, and teams without a large analytics department employ them less.[27] We download defensive shift data from Baseball Reference. As an additional measure, we use

[25]When calculated for a pitcher, it is scaled by the number of innings thrown, and adjusted by a constant so that it is on a similar scale to a more familiar measure: earned run average. We do not make that adjustment because the scale of FIP is not the economically interesting outcome.

[26]The calculation of the measure can be found at https://library.fangraphs.com/offense/woba/.

[27]For a discussion of the defensive shift, see https://sloanreview.mit.edu/audio/what-would-happen-if-baseball-outlawed-the-shift/.

a ranking by Ben Baumer, an analyst at ESPN, based on qualitative analysis and interviews.[28] The second proxy is a pitcher-level one for information processing ability; we use pitcher age, which we obtain from Baseball Reference. The final proxy approximates quasi-experimental variation across pitchers and games in external factors that may affect information processing ability: we use temperature and game timing (whether the game is an afternoon game) as this variation. Game timing and temperature information are obtained from Bill Petti's website.[29]

Another measure of payoffs is the full outcome of the game—whether the pitching team wins or loses. While we show some results with this measure, we find it too noisy to say anything definitive. Another related measure—the number of runs scored by the other team in the rest of the inning—captures some dynamics involved without using the entire rest of the game. We also study the cumulative win probability added across multiple innings to incorporate an intermediate amount of dynamic effects.

III. Rejecting Nash Equilibrium Play

We first test Prediction (0), whether the data is consistent with full-information Nash equilibrium. In Section III.A, we discuss our empirical strategy and argue that we can reject Nash equilibrium play. In Section III.B we discuss alternate explanations.

III.A. Empirical Strategy and Main Results

Our empirical analysis exploits the fact that we observe multiple outcome metrics that are good proxies for the pitchers' payoffs. Thinking of each state as a different matching pennies game, a prediction of Nash equilibrium is that payoffs should be identical across all pitch types in *every* state of the game. Since we do not have the power to conduct such a test on fine states, we aggregate states into coarser categories of states and compute the average payoff difference across pitch types within these coarse categories—with the test for Nash play being that these differences are still zero. In particular, our regressions are of the form

$$Y_p = \sum_{\sigma=0}^{2} \beta_\sigma \cdot \mathbb{1}[\text{Fastball}]_p \cdot \mathbb{1}[\sigma \text{ strikes in the count}] + \text{State FEs} + \epsilon_p, \tag{7}$$

where Y_p is the payoff in a particular pitch. In (7), β_σ is the average payoff difference between fastballs and other pitches in counts with σ strikes, which correspond to the "coarse" categories.

[28] This ranking was created in 2015 and can be found at http://www.espn.com/espn/feature/story/_/id/12331388/the-great-analytics-rankings.

[29] See https://billpetti.github.io/baseball_tools_home/.

In Nash play, we would expect $\beta_\sigma = 0$ for all σ.[30,]

We control for a rich set of state fixed effects in (7).[31] These fixed effects should capture anything that affects the decision of what pitch to throw and the distribution of payoffs conditional on the pitch. The identification assumption is closely related: conditional on state fixed effects, there is nothing that affects both the probability of throwing a fastball and payoffs conditional on the pitch thrown. In practice, we use the full interaction of the balls, the strikes, the outs, the position of runners on base, the inning, the score differential, how many pitches have been thrown in that count,[32] the identity of the pitcher, and the handedness of the batter. This leads to more than one million different states in our data, though some are sparsely populated. We should stress that these state fixed effects help us use our large dataset to overcome the aggregation problem Chiappori et al. (2002) faced: the strategy in (7) allows payoff matrices to differ by state, and β_σ reports a (weighted) average of the within-state payoff differences. Section III.B evaluates alternate choices for the state space. Standard errors are clustered at the level that the payoff is realized: the at-bat for win probability added and FIP, the inning for runs, and the game for wins.

The results are presented in Table 1. Column (1) uses win probability added as the outcome metric and is our preferred specification. With zero strikes, the win probability added of a fastball is greater than other pitches, by about five-ten-thousandths of a win. With two strikes, the win probability added of a fastball is less than other pitches, by about nine-ten-thousandths of a win. Both are statistically significant at even the 0.1 percent level.

The magnitude of these coefficients is small, but the pitch decision occurs repeatedly. Switching twenty percent of non-fastballs for fastballs, in zero-strike counts, and ten percent of fastballs for non-fastballs in two-strike counts, with no equilibrium response, would lead to an additional win over the course of a season. Another way to get a sense of the magnitude is to use the monetary value of a win. The cost of a win on the free agent open market is approximately \$10 million (Swartz, 2017). On average, fastballs are thrown 63% of the time. Throwing fastballs with

[30]Given the estimated β_σ are averages of payoff differences in the states, rejecting $\beta_\sigma = 0$ would imply that at least some states must have payoff differences. Moreover, they are also averages across finer action spaces, such as types of non-fastballs. Rejecting $\beta_\sigma = 0$ would also reject payoff equivalence for these finer action spaces. Thus, we will focus on the binary pitch choice and strikes as the coarse category throughout this section. In the analysis conducted in Section IV, we show direct evidence that payoff equivalence is rejected for finer categories and finer actions.

[31]Using the notation from Section I, the state fixed effects capture w_s^n, the payoff associated with not throwing a fastball, so that β_σ can capture the average of $w_s^f - w_s^n$, the difference in payoffs between fastballs and non-fastballs in state s, averaged over the category σ.

[32]We include this in our state-space to eliminate a bias from multiple pitches in the same at-bat being compared to one another. It only applies to two-strike counts. With two strikes, if there is a foul ball, the state remains the same—except for this variable. If we did not include it, there would be a bias because the outcome measure is defined at the at-bat level. Suppose there is a state that occurs only once in our datatset, and it has two strikes. If the batter fouls off a pitch, and next gets a pitch of a different type, then the payoff difference between fastballs and non-fastballs will be estimated to be zero in that state, even though that is only because win probability is estimated at the at-bat level. Dropping this variable from the state-space has no effect on our estimates for zero or one strike, but moves our two-strike estimate toward zero, consistent with this bias.

	(1) WPA $\times$ 1000	(2) FIP$^-$ $\times$ 100	(3) wOBA$^-$ $\times$ 100	(4) Wins $\times$ 100	(5) Runs Rem$^-$ $\times$ 100
Fastball, 0 strikes	0.512***	-4.470***	-0.666***	0.169	-0.911***
	(0.134)	(0.658)	(0.133)	(0.115)	(0.229)
Fastball, 1 strike	-0.092	-0.847	0.367**	0.036	-0.101
	(0.135)	(0.698)	(0.146)	(0.135)	(0.248)
Fastball, 2 strikes	-0.902***	8.225***	1.46***	-0.012	0.756***
	(0.149)	(0.804)	(0.162)	(0.165)	(0.282)
Joint p-value	0.000	0.000	0.000	0.526	0.000
N	2,707,048	2,707,048	2,707,048	2,707,048	2,707,048

Table 1: Testing payoff equivalence, by strike. Each column shows a different payoff metric: (1) win probability added, (2) FIP, (3) weighted on-base average, (4) a dummy for whether the pitcher wins the game, and (5) the runs scored by the other team in the remainder of the inning. Variables with a $^-$ (Columns (2), (3), and (5)) indicate that lower values are better for the pitcher. We report the p-value of the joint test that all three estimates are zero. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

zero strikes and non-fastballs with two strikes, instead of playing 63% fastballs in all states, is worth about $2,370 per pitch, assuming no response from opponents (i.e., no change in payoffs).[33]

The above strategy is unlikely to be the partial-equilibrium payoff-maximizing strategy, since there may be some 0-strike states in which fastballs are lower-payoff and some 2-strike states in which fastballs are higher-payoff. We presume there are payoff-differences that depend on other dimensions of the state besides strike, but have shown the results for strike because we have strong empirical power when we split the states into only a few categories. We consider finer categories below, but for now we view this as a back-of-the-envelope partial-equilibrium lower bound, which illustrates the magnitude of the coefficients.

Column (2) shows that FIP is lower in zero-strike counts and larger in two-strike counts. Given higher FIP is worse for the batter, this is consistent with the WPA results in sign. A potential concern with WPA is that it uses historical data before our sample period. Column (3) shows that results are similar with wOBA, which is signed in the same manner as FIP; we also see a significant payoff difference for 1 strike counts (which is of course still consistent with departures from Nash equilibrium), but the ordering remains the same. Column (4) uses wins as an outcome metric, and while the point estimate is positive for no strikes and negative for two strikes, it is an especially noisy measure. Column (5) shows the runs remaining in the inning,

[33]This is calculated by taking the percent of counts with zero strikes (41%) times throwing 37% more fastballs times the gain from a fastball, 0.000512 wins times $10,000,000 per win. That is added to the percent of pitches with two strikes (28%) times 63% fewer fastballs times the gain from throwing a non-fastball, 0.000902 wins times $10,000,000 per win. This exercise should not be taken as a counterfactual. Any adjustment in the strategy would lead to a response by the batter. The exercise is solely meant to interpret the magnitude of the coefficients.

which is again a metric where lower values are preferred by the pitcher. It shows the same pattern as Columns (1), (2), and (3).[34]

In this section, we have focused on rejecting Nash equilibrium play, but we also think it is important to understand why these payoff differentials might occur. We think such payoff differences would be plausible under any theory that predicts an under-reaction by the pitcher, i.e., if the pitcher throws a fastball with probability less than 1 even when the equilibrium payoff of a fastball is positive. Throwing a non-fastball when the batter is ready for a fastball is likely to be a strike. With two strikes, that leads to a strikeout, one of the best possible outcomes for the pitcher. With fewer strikes, the at bat continues with one additional strike, a smaller payoff. If the pitcher under-reacts to these payoff changes, the batter will not necessarily adjust his strategy to fully close any equilibrium payoff differences. This example also explains our decision to group states by strikes in this analysis, as a reader familiar with baseball would expect strikes to be one of the most important dimensions of the state in determining payoffs.[35]

III.B. Alternate Explanations

Our analysis is predicated on the existence of equilibrium payoff differences across pitch types. Accordingly, before moving to the rational inattention model, we consider alternate explanations for the results above.

Incorrect Payoffs. One set of concerns is that pitchers may care about other outcomes, such as their individual statistics. Certain strategies could lead to outcomes advantageous to them but which are not captured by the probability of their team winning the game. Perhaps with the appropriate metric for payoffs, one would observe payoff equivalence in all states.

We provide some arguments against this concern. First, we already show results using FIP and wOBA, easily available individual statistics often used to evaluated pitchers; patterns are similar between FIP, wOBA, and WPA. However, perhaps the pitcher cares about one statistic over the others. For example, if we replace the left-hand side of our baseline regression with a dummy for the at-bat being a strikeout, then the coefficient on fastball is always negative, i.e. fastballs lead to fewer strikeouts regardless of how many strikes there are. Thus, if the pitcher's utility were increasing in both strikeouts and win probability, then a Nash equilibrium would have

[34] While weights from wOBA are computed differently year-by-year and wins and runs remaining and contemporaneous metrics, one may be concerned that WPA and FIP are computed using historical data. We have checked that an in-sample version of WPA—computed by averaging win probabilities within inning, score differential, outs, and base position—returns quantitatively and statistically similar numbers as Column (1). Furthermore, using these in-sample WPAs to compute an in-sample FIP returns similar results to Column (2).

[35] We should also stress that while one could also devise other explanations for the payoff differences, all tests for rational inattention but (C) simply rely on our estimates of equilibrium payoffs rather than their source.

fastballs always leading to higher win probability—not what we find.[36] Walks and home runs follow the same pattern as FIP—they are higher for fastballs with two strikes and lower with zero strikes. There is a linear combination of win probability and home runs that would be equalized, but it would require the pitcher prefers home runs to other outcomes, which would be against his own interest as well as the team's. The same argument applies to walks.

This argument can in fact be interpreted as one alleviating concerns about risk aversion. Since no linear combination of strikeouts, walks, and home runs will explain the state-specific deviation of FIP, no convex transform of FIP can yield zero payoff differences. A similarly clean explanation does not apply to WPA, but we have checked that CARA transforms of these metrics are still not equalized across states, for reasonable risk aversion parameters.

A separate concern may be that the correct payoffs explicitly incorporate dynamics beyond the at bat.[37] For instance, one may worry that throwing a fastball in a particular state may be effective for that at-bat but may reveal information (say about the quality of one's fastball that day) that would reduce payoffs later in the game. If this were a concern, then we may expect the cumulative effect of a fastball to decay to zero. Columns (4) and (5) of Table 1 use outcome metrics that address this issue. Wins are in some sense the ideal cumulative metric, as they take into account everything that happens in the remainder of the game; estimates, however, are fairly noisy using wins as a metric. Runs remaining in the inning control for a limited amount of dynamics by taking into account what happens beyond an at-bat. Here, we can strongly reject payoff equivalence for zero- and two-strike counts. Figure 1 uses a metric that partially "interpolates" between these two extremes: the cumulative win probability added for the next N innings, with $N = 0$ denoting the cumulative win probability added for the contemporaneous inning.[38] Estimates become progressively noisier as N increases, but there does not seem to be any obvious evidence that dynamics induce a drift of the point estimates towards 0.[39] Another concern related to dynamics is that the order of pitches may matter, in terms of throwing a batter off-balance, say; we address such issues related to pitch sequencing in Figure 2(b) below. Moreover, omitting all pitches involving the first time through the batting lineup, where one may be worried about setting up pitches for the future, yields similar results (Figure 2(b)).

Finally, a concern may be unobserved payoffs not related to any outcome of the game. That is,

[36] A similar argument applies to the total number of balls put into play, if the pitcher has preferences over being "in control" of outcomes. Regressions indicate that fastballs lead to more balls in play regardless of the number of strikes.

[37] All our payoff measures are measured at the at bat or coarser level and so would account for any dynamic effects within the at bat.

[38] We set the WPA to be 0 for all innings that do not actually occur. For instance, if an at-bat occurs in the ninth inning, and the game ends in the ninth inning, the cumulative WPA for that at-bat will be the same for all $N \geq 0$.

[39] The point estimate for two strikes is closer to zero after six innings in Figure 1. Most pitchers will not be in the game six innings later anyway, so it is hard to imagine a mechanism in which this was a robust feature of the data and not sampling variance.

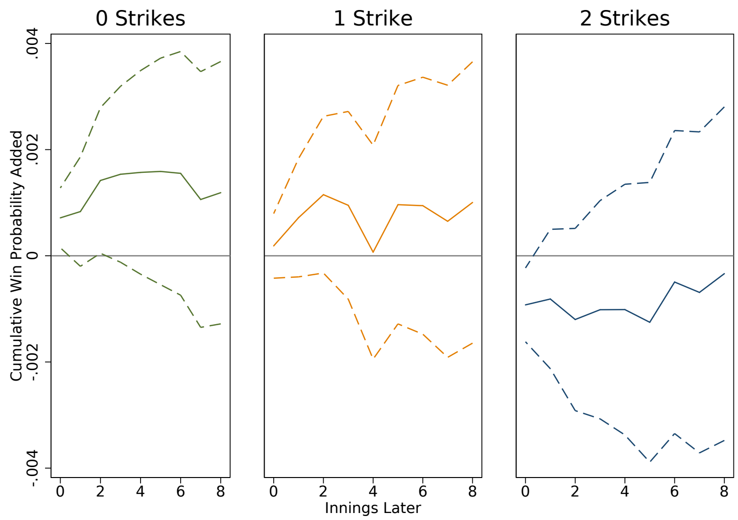
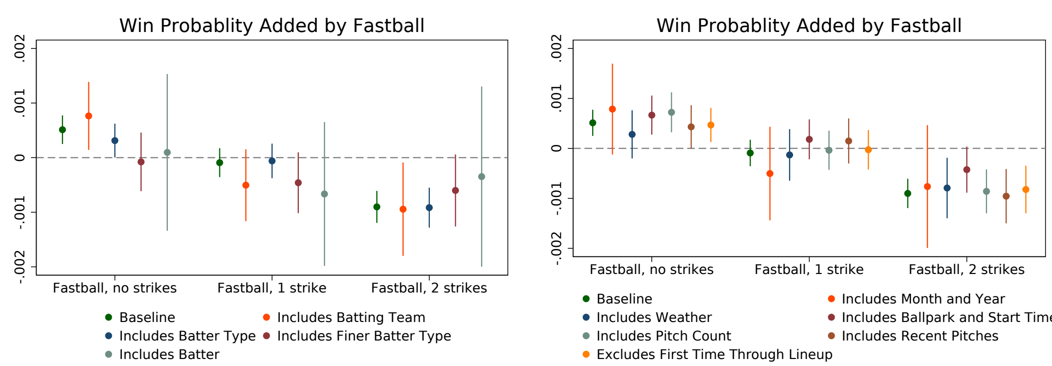


Figure 1: Cumulative win probability of a fastball for the next N innings, by strike for various N



(a) Batter

(b) Other Attributes

Figure 2: Robustness to finer state spaces

pitchers may also have direct utility from throwing a particular type of pitch; e.g., fastballs could be more tiring and lead to longer recovery periods, or they may cause fewer injuries since they put less strain on elbows. To achieve payoff equivalence, however, we would need the unobserved preference for throwing fastballs to be correlated with the strike: pitchers must have an unobserved preference for throwing fastballs in two-strike counts and unobserved preference for throwing other pitches in zero-strike counts. We find it difficult to generate reasonable examples of such payoffs in this setting.

Unobserved Heterogeneity in Payoffs. The ideal econometric setting would be to observe outcomes of different pitch types in otherwise identical situations. Given limited data, we compare different types of pitches in situations that are identical as defined by our state space. A concern would be that a variable omitted from our state space is influencing both the strategy and payoffs, potentially biasing our regression. We check to see if other aspects of the situation have any effect on our baseline result. We will look for any changes to the payoff deviations, by the number of strikes. Because any confounding variable would need to affect both the pitcher's strategy and payoffs, we focus on aspects that could plausibly affect payoffs. As we add elements to our state space, we fully interact them with all other variables already in the state space.

A potentially concerning aspect absent from our state space is the batter, as they have different absolute and relative abilities. In Figure 2(a), we run (7), interacting our state space with attributes of the batter. Controlling for who the batter is, the finest partition we can do, limits our observations and increases our standard errors. However, we can control for some attributes of the batter without losing so much power. The simplest addition to the state space is adding in the batting team. The standard errors increase and the point estimates are similar. To be more aggressive, we group batters into types via k-means clustering on the rates of the following outcomes for batters: strikeouts, home runs, walks, singles, doubles, triples, and other. Batters that have similar rates on these outcomes should be grouped together. We cluster into either three types or twenty "finer" types. With twenty types or the batter himself, the point estimates with zero strike become close to zero, but with significantly wider confidence intervals. For three types or with two strikes, the estimates are largely the same.

Concerns unrelated to the batter are shown in Figure 2(b). We include the month and year of the game, to compare two pitches thrown in the same state within one month of each other, to address changes in how well pitchers throw over the course of a season or over the course of their career.[40] The point estimate becomes negative with one strike, but results with zero or two strikes are unchanged. We include the weather, recorded in the box score and downloaded from Bill Petti's website,[41] since warmer weather is known to lead to more home runs. Baseball stadiums differ in sizes and architecture, which could affect visibility and the likelihood of home runs. We thus include the home stadium interacted with whether the game starts before 5 pm, to control for differences between day and night games. We also include the number of pitches the pitcher has thrown that game, grouped into bins of 5, to see if fatigue affects our results. Because of concerns about pitch sequencing, we include whether each of the previous three pitches was a

[40]For example, the fastball of a typical pitcher will get slower over the course of their career. Some pitchers may also get tired over the course of a season.

[41]We group "clear" and "sunny," "dome" and "roof closed," and "cloudy" and "overcast." We interact the qualitative description, which also includes things like "drizzle" and "partly cloudy," with the starting temperature, binned every ten degrees.

fastball.[42] Finally, we have a specification where we exclude the first at-bat for each batter when dynamic considerations would be particularly strong. Results do not change appreciably in any specification.

Overall, the additional state space controls have only minor effects on our estimated coefficients, adding credence to the fact that our baseline estimates do capture the different payoffs from throwing a fastball or not. We keep this baseline state space as our preferred specification throughout the remainder of the paper.[43]

IV. Testing Rational Inattention

In this section, we test Predictions (A)–(C) of our rational inattention model. We conclude with a brief discussion on the interpretation of these predictions through the lens of QRE, another canonical departure from Nash equilibrium.

IV.A. Partial Responsiveness to Payoffs

The first prediction of our model is that there should be partial adjustment of the pitcher's strategy to payoff deviations.[44] Given our previous results that payoffs to a fastball are high in zero-strike counts and low in two-strike counts, the most basic version of this question boils down to: do pitchers throw more fastballs with zero-strikes, and fewer fastballs with two strikes? The answer is yes. With zero strikes, fastballs are thrown 69% of the times, with one strike, 61% of the time, and with two strikes 56% of the time. This pattern confirms Prediction (A).[45]

A more stringent test of this prediction estimates the win probability and pitcher strategy by finer categories of states, instead of just by strike. Figure 3 presents WPA and the pitcher strategy by count, and we present the two quantities in separate figures to highlight that the pattern exploited in Table 1 and the previous paragraph survive when splitting strikes into counts. In Panel (a), we run the same regression as presented in Table 1, but split up estimates by count (combinations of balls and strikes), giving us twelve overall estimates.[46] In Panel (b), we show the probability of a fastball in each count. The two are clearly positively correlated, providing

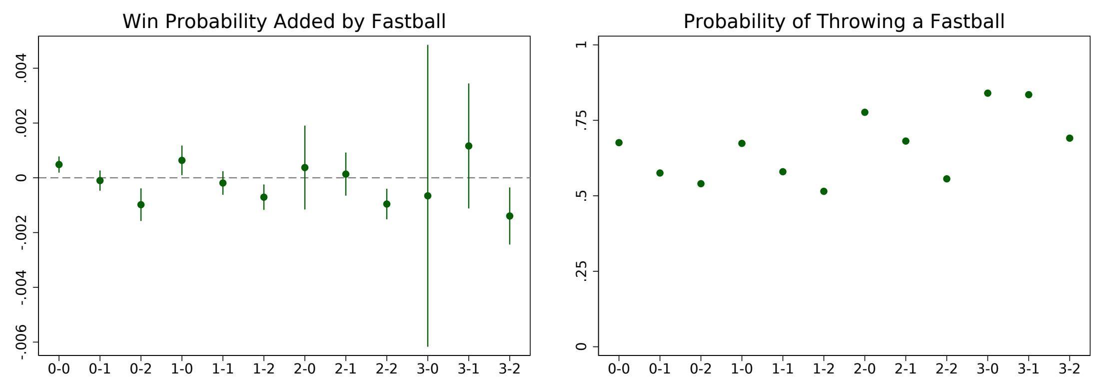
[42] Baseball players may get used to seeing a fastball and be better prepared to hit it after having seen it once.

[43] Splitting the state-space into more categories has the advantage of reducing any bias that might come from the omitted dimension. The downside is that the variance of our estimates increases. Given the robustness of our results, we think our baseline makes the appropriate bias-variance trade-off, especially since some of our predictions split the sample into even smaller parts.

[44] A requirement of the model for Prediction (A) to hold is that the pitchers are mixing. Implicitly, our regressions drop any states in which there is not at least one fastball and one non-fastball. So for the handful of pitchers that do not mix, they are effectively dropped from our dataset.

[45] We should note that this relationship is certainly not mechanical: a priori, we had no reason to believe that pitchers would be responding to payoff differences at all.

[46] The 3-0 count has large standard errors because it is a rare count and because pitchers throw a high percentage of fastballs, limiting the identifying variation.



(a) Payoff difference, by count

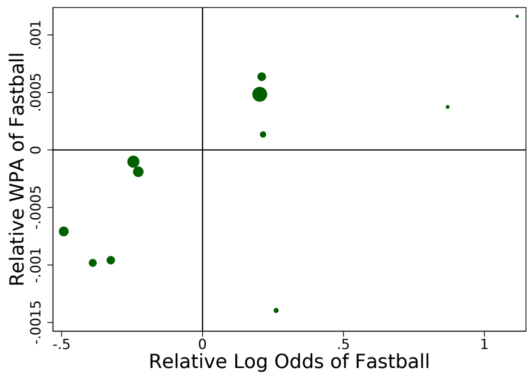
(b) Probability of fastball, by count

Figure 3: (a) Relative payoff and (b) probability of a fastball, by count. Intervals are 95% confidence intervals.

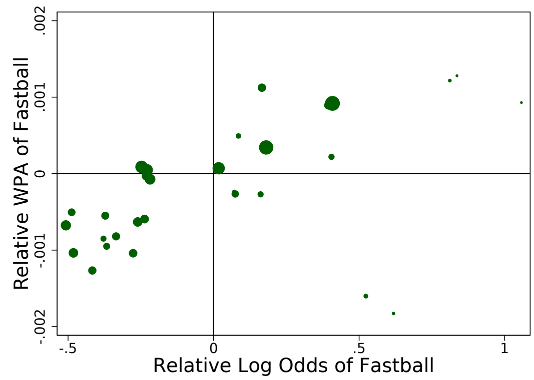
more empirical support for Prediction (A).

Figure 4(a) effectively plots the two subplots of Figure 3 against each other. We replace the probability of a fastball by the log odds of a fastballs to remain faithful to the structure of the rational inattention model (see Section V), but over the range of probabilities this does not lead to an appreciable difference. The size of the dot corresponds to the inverse standard error on the win probability added estimate. The interpretation of Prediction (A) is therefore that we should see a positive correlation between the two quantities, which we clearly do. Panel (b) replicates this analysis for an even finer division of categories—count-out—and finds a similarly strong positive relationship. We can subdivide further into smaller categories (such as count-out-base position) and still see robust positive relationships, although standard errors on the outcome metric become especially large. Panels (c) and (d) replicate this analysis with FIP as the outcome; noting that lower values of FIP correspond to better outcomes for the pitcher, the message is the same. To estimate λ, we verify in Section V that the visual relationships are statistically significant.

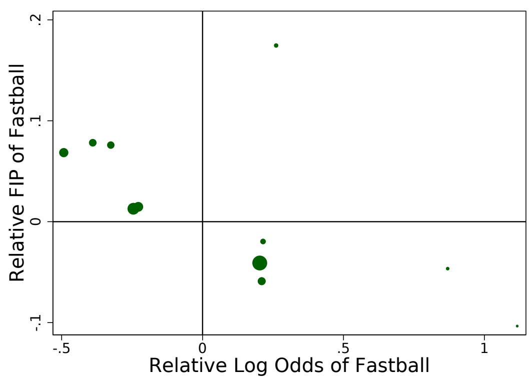
Yet another more stringent test is to break up the actions into finer ones and check that payoff differences across finer actions correlate positively with differences in the probabilities of playing these actions. We can subdivide non-fastballs into curveballs, changeups, and sliders, and for each pair of pitches we can correlate the relative probability of throwing one pitch versus another against the relative payoff of one pitch versus another. In the course of estimating λ in Section V, Figure 6 shows these correlations for a variety of combinations of pitches for both count and count-out, and the pattern suggests that the correlation remains stable for any pair of pitches. Thus, the test for (A) does not hinge on the binary classification into fastballs and other pitches.



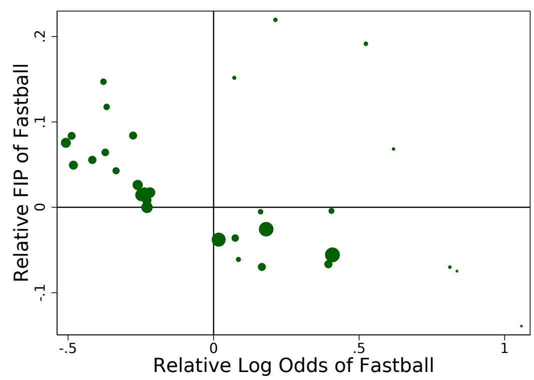
(a) Count, WPA



(b) Count-Out, WPA



(c) Count, FIP



(d) Count-Out, FIP

Figure 4: Scatterplots of payoff differences against relative log odds of fastball. Panels (a) and (b) use win probability added and (c) and (d) use FIP, for which lower values are beneficial to the pitcher. The size of the dot is proportional to the inverse variance of the estimate of the outcome metric for that state.

IV.B. Differential Sensitivity to Payoffs

Prediction (B) states that players with lower attention costs should have higher responses to differences in payoffs. Through (7) for testing Nash, we show that fastballs have higher payoffs in 0-strike counts and lower ones in 2-strike counts. Thus, we can check whether pitcher are more likely to pitch fastballs in 0-strike counts and less likely to do so in 2-strike counts in scenarios in which they have lower costs of attention. We show that these patterns hold for a finer categories of states as well, so that lower attention costs do correspond to more responsiveness. We finally verify that an alternate explanation for these responses—that payoffs are systematically different for the two groups—is not something for which we find evidence.

To test (B), we need proxies for the cost of attention. Here, we take three different approaches.

A challenge of testing rational inattention in the field is that we cannot experimentally alter the cost of attention. As such, none of these approaches is a silver bullet, but all of them try to capture the fundamental interpretation of λ as a cost of processing information. Taken together, we argue that various metrics that plausibly increase λ (for different reasons) seem to have the expected effects on strategies.

The first approach uses differences in technological investments across teams. As discussed in Section II, we use defensive shifts and a qualitative ranking generated by ESPN as measures of analytics intensity. In our time period, there was substantial variation across teams. Why is analytics a reasonable determinant of attention costs? To make a choice, the pitcher observes the state and tries to map that onto payoffs for his available actions. While there should be little uncertainty about the observable dimensions of the state, mapping the state onto payoffs is surely complicated, subject to mistakes, and hard to be attentive to. For example, precisely estimating the relevant payoffs state-by-state is challenging even with our massive dataset, but that means that it is difficult for the pitcher as well. A team highly invested in analytics is likely to have a more accurate mapping of these payoffs, may be better at communicating such a mapping to the pitcher, and may even be in the business of simplifying this mapping to make information processing easier in the heat of the moment. More analytics-intensive teams run analyses similar to ours with regularity, looking to improve a pitcher's pitch selection.[47]

The second proxy exploits potential differences in cognitive ability across players, which may directly affect information processing. There is psychological evidence that information processing speed declines starting from age 18 or 19.[48] Given a pitcher's career can last into the mid-thirties, we may expect that older pitchers have higher costs of attention. Of course, a concern with this proxy is that better pitchers would stay in the MLB for longer.[49]

The final approach exploits "quasi-experimental" variation generated by weather and game timing that may affect cognitive ability. There is some evidence that cognitive ability declines in especially high temperatures, so we use the daily temperature as a proxy for attention costs.[50]

[47] Our conversation with Farhan Zaidi, now President of Baseball Operations of the Giants, confirms they do these types of analyses.

[48] The effect of age on various forms of cognition is a very large field in psychology, and Merriam and Baumgartner (2020) provides a book-length summary. To our knowledge, the most closely related study on information processing is Hartshorne and Germine (2015), which provides such evidence from an analysis of a large set of prior studies along with a large-scale online study of cognitive tests. Craik and Bialystok (2006) also provides similar evidence.

[49] A natural hypothesis is that more attentive pitchers stay in the league for longer, which would dampen our observed effects. What would be a problem for us is if ability and strike-responsiveness were inversely correlated, and age was selecting on ability.

[50] Dai et al. (2016) and Laurent et al. (2018) use heat waves and variation in air conditioning to show that temperature has an effect on a number of cognitively demanding tests. Künn et al. (2019) finds limited effects of temperature on chess players' decisions in a tournament, but their range of temperature is well below what other studies—including this paper—consider especially high temperatures.

Moreover, there is a lot of evidence that lack of sleep can impair cognitive function.[51] Since most games are night games, a typical baseball player goes to sleep especially late,[52] and they are likely to get less sleep before an afternoon game. Of course, due to variation in other factors that may affect sleep, or differences in time zones that may lead to jet lag, game timing is admittedly a noisy proxy for sleep and thus attention.

Our primary analysis involves a linear probability model of whether the pitch is a fastball on the number of strikes, interacted with a dummy affecting the cost of attention,

$$\mathbb{1}[\text{Fastball}]_p = \sum_{\sigma=0,2} \big\{\beta_\sigma \cdot \mathbb{1}[\sigma \text{ strikes in the count}]_p \cdot \text{Attention}_p + \gamma_\sigma \cdot \mathbb{1}[\sigma \text{ strikes in the count}]_p\big\} + \text{State FEs} + \epsilon_p, \quad (8)$$

where Attention_p is a measure of attention and varies at the team level for analytics, pitcher-year-level for age, and game-level for temperature and timing.[53] We interact the state fixed effects with the attention measure to include the main effect of the attention measure; doing so makes strikes collinear with the state fixed effects, so we drop strikes from the state for these regressions and interpret the coefficients as averaged across states with strikes.

Results for the regression in (8) are found in Table 2. The first two rows show the observation from Section IV.A that fastballs are more likely in 0-strike counts and less in 2-strike counts, regardless of the further interactions included in the specification. Columns (1) and (2) use analytics-intensity of the team as the metric, using either the top third of defensive shifts (Column (1)) or the highest category in the ESPN categorization (Column (2)) as a proxy for low cost of attention. Such teams are more likely to throw fastballs in zero-strike counts, although the magnitudes are somewhat small—0.2 to 0.7 pp off a baseline of 10 pp, depending on the classification used. These teams also throw fewer fastballs in two-strike counts, and these responses are considerably larger—1.1 to 1.7 pp off a baseline of around 6 pp. Thus, more analytics-intensive teams are reacting more to the payoff differences.

Column (3) interacts the strike count with a dummy for whether the pitcher is at least 30

[51]Bessone et al. (2020) conduct lab experiments of sleep on economic outcomes (as well as cognition) in India, and the references within that paper point to a very rich history in psychology of studies of the relationship between sleep at cognition. Moreover, there is evidence specifically in sports that lack of sleep (measured through things as varied as game timing or late-night tweets) affects some metrics of performance (Mah et al., 2017; Song et al., 2017; Jones et al., 2019).

[52]See, for instance, news reports that document that baseball players leave the clubhouse at 11 pm after a night game and do not fall asleep for many more hours due to adrenaline (Schlact, 2012; Wagner, 2019). Wagner (2019) explicitly mentions the difficulty of fitting afternoon games into a sleep schedule.

[53]While a logit would be faithful to the functional form of Shannon entropy, running a logit with high-dimensional fixed effects is computationally much more cumbersome than the linear regression. Given that probabilities of fastballs are far from 0 and 1, we anticipate the results would be qualitatively similar to a logit. Moreover, we run other regressions below that do impose functional forms implied by Shannon entropy.

	Shifts (1)	ESPN (2)	Age (3)	Temp (4)	Afternoon (5)
0 Strikes	0.101***	0.100***	0.106***	0.103***	0.096***
	(0.0006)	(0.0006)	(0.0006)	(0.0005)	(0.0007)
2 Strikes	-0.059***	-0.058***	-0.067***	-0.064***	-0.059***
	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0009)
0 Strikes	0.0026**	0.0077***			
$\times$ High Analytics	(0.0010)	(0.0011)			
2 Strikes	-0.0109***	-0.0166***			
$\times$ High Analytics	(0.0013)	(0.0013)			
0 Strikes			-0.0125***		
$\times$ (Age $\geq$ 30)			(0.0010)		
2 Strikes			0.0142***		
$\times$ (Age $\geq$ 30)			(0.0013)		
0 Strikes				-0.0164***	
$\times$ (Temp $\geq$ 90)				(0.0020)	
2 Strikes				0.0121***	
$\times$ (Temp $\geq$ 90)				(0.0026)	
0 Strikes					-0.0012
$\times$ Afternoon					(0.0014)
2 Strikes					0.0033*
$\times$ Afternoon					(0.0017)
Joint p-value	0.000	0.000	0.000	0.000	0.044
N	5,101,768	5,107,156	5,089,574	5,111,591	3,146,846

Table 2: Fastball rate, versus a one-strike count. Column (1) classifies a team as high analytics if the number of defensive shifts in conducts is in the top third of all teams. Column (2) uses the highest category in ESPN's classification. All regressions include state fixed effects without interacting with strike but with further interactions with the measure of attention used in the column, which leads to small variation in sample size due to different sets of singleton states. Column (5) drops games in the 2008 and 2009 seasons as the data does not include start time. We report the p-value of the joint test that the interaction of 0 and 2 strikes with the category are both zero. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

years old. We would hypothesize that such pitchers have *higher* costs of attention given natural cognitive decline in information processing ability. Indeed, older pitchers tend to react less to strikes, reducing their probability of a fastball in 0-strike counts by 1.3 pp and increasing it in 2-strike counts by 1.4 pp. While Table 2 bins age into a coarse bin, this result is not the artifact of the binning decision. Figure C.3(a) in Appendix C.2 shows coefficients of a regression that interacts with dummies for age rather than a single bin, and there is a clear decrease in responsiveness as age increases, especially after the early twenties. This is consistent with reductions in cognitive processing ability and may be indicative of higher costs of attention.

Columns (4) and (5) move away from team- or pitcher-level properties and show results using the variation induced by temperature and game timing. Column (4) interacts strikes with a dummy for a game played in especially hot weather—over 90°F. We see a marked decrease in responsiveness, which is consistent with increased difficulty in processing information corresponding to higher λ. Figure C.3(b) shows the continuous pattern underlying this regression and indicates a continuous decrease in responsiveness, although the slope becomes more noticeable at higher temperatures. Finally, the point estimates in Column (5) suggest that responsiveness is also lower in afternoon games, but we hesitate to read too much into the estimates since they are especially small. Moreover, the coefficient on 0 strikes is not significantly different from 0 at any conventional level, although we reject the joint hypothesis that both interaction terms are zero at the 5% level. As discussed above, afternoon games are at best a noisy proxy for alertness of the players, so we may expect some attenuation bias.

We admit that none of the previous results is the ideal test of the effects of changing attention costs. Furthermore, we recognize that we have not checked an exhaustive list of the determinants of rational inattention. Indeed, an unfortunate drawback of not having preregistered our empirical analysis is that we cannot credibly show *ex post* that we did not cherry-pick ones among a larger list of candidates for proxies. Nonetheless, we believe that taken together, and as one piece of evidence amongst the several predictions in this paper, they provide support for rational inattention.

The second, related, test for the responsiveness of agents is to directly compare $\mathcal{L}(p^f_\sigma)$ state-by-state for agents with arguably different costs of attention. For a group of states σ, we run

$$\mathcal{L}\left(\hat{p}^{f2}_\sigma\right) - \mathcal{L}\left(\hat{p}^{f2}_0\right) = \alpha + \beta \cdot \left[\mathcal{L}\left(\hat{p}^{f1}_\sigma\right) - \mathcal{L}\left(\hat{p}^{f1}_0\right)\right] + \epsilon_\sigma, \tag{9}$$

where $\hat{p}^{fi}_\sigma$ is the estimated probability in state s of a fastball for agents in Group i, and $\hat{p}^{fi}_0$ is the average probability of a fastball for such agents. If these agents are facing the same payoffs in each state, then if agents in Group 1 have higher costs of attention than those in Group 2, we would expect $\beta > 1$ and $\alpha = 0$ in (9). Group 2 agents would play more fastballs than Group 1 agents if (and only if) a fastball has a higher payoff. If the opposite were the case, then we would expect $\beta < 1$ and $\alpha = 0$.[54]

Table 3 provides results from the regression in (9), using count-outs as the definition of state.[55] In each column, the group on the right-hand side (Group 1 in the notation above) is the baseline category from the corresponding column of Table 2, and the one on the left-hand side is

[54]We use $\mathcal{L}(p)$ instead of p to directly mirror (4). If payoff differences are equal for the two groups, then lower attention costs would directly imply $\beta > 1$.

[55]While estimates of the log odds are especially precise, we still weight by the inverse variance of the left-hand side, as some states are (relatively) more sparsely populated and the log odds of a fastball are estimated (slightly) less precisely for them. Figure C.1 in Appendix C.2 shows the underlying scatters and visually suggests that this weighting is not relevant for the qualitative message, which is also something we verify.

	Shifts (1)	ESPN (2)	Age (3)	Temp (4)	Afternoon (5)
$\mathcal{L}(p^f_\sigma) - \mathcal{L}(p^f_0)$ for Baseline Category	1.059***	1.092***	0.805***	0.939***	0.943***
	(0.018)	(0.012)	(0.011)	(0.013)	(0.007)
Constant	0.0010	0.0066	-0.0022	-0.0035	0.0006
	(0.0075)	(0.0049)	(0.0041)	(0.0048)	(0.0028)
N	36	36	36	36	36
R^2	0.993	0.996	0.995	0.994	0.998

Table 3: Regression of log odds of fastball for the excluded category on the log odds for the baseline category. In Columns (1) and (2), the baseline category is teams with low analytics and the excluded category (the dependent variable) are teams with high analytics. In Column (3) the baseline category are pitchers younger than 30 years of age, in Column (4) games with temperatures below 90° F, and in Column (5) games played at night. Thus, the baseline category involves *high* attention costs in Columns (1) and (2) and *low* attention costs in Columns (3)–(5). Each observation is a count-out. Significance for first row is measured against a null hypothesis of 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

one included in the explicit interaction. Thus, Group 1 has plausibly higher costs of attention in Columns (1) and (2) and lower costs of attention in Columns (3)–(5). Indeed, we see that β is significantly larger than 1 in Columns (1) and (2): pitchers on teams with high analytics intensity adjust the log odds of throwing a fastball by 6–9%. The estimates in Columns (3)–(5) are all significantly lower than 1. Older pitchers react about 20% less than younger ones, and pitchers tend to about about 6% less responsive on hotter days or afternoon games. These estimates provide a quantification of how different aspects of teams, players, and situations affect attention costs.

Of course, if we believe that Group 1 is more responsive than Group 2, a prediction of rational inattention is that they would be more responsive *state-by-state*, while the regressions in Table 3 show average responsiveness. In Appendix C.2, we show the scatters underlying the regressions in Figures C.1 and C.2. The scatters verify that the regressions are not masking important heterogeneity and that the responsiveness does happen almost state-by-state.

Finally, a natural question is whether the results above are instead due to differences in payoffs. One may speculate that teams with stronger analytics departments, or younger pitchers, are facing different equilibrium payoffs as a function of states.[56] To check, we run the analogue of the analysis in Section IV.A but interact strikes with proxies for attention to see whether payoffs are different by strike in different bins of attention. Appendix C.2 discusses this further, and Table C.2 shows the results. While the estimates are noisy, there largely does not seem to be any

[56]As an example, pitchers farther along in their careers are likely more effective on average—perhaps due to experience or due to the selection effect of higher-ability pitchers staying in the league. Given the interaction of the state with the measure of attention, this by itself is not a concern. The concern would be that older pitchers are relatively more effective with fastballs in 0-strike counts and less so in 2-strike counts, say.

evidence that payoff differences across bins of attention differ in a way that would undermine the results in this section. We then go further and estimate λ by bin using the procedure in Section V, which would take into account both differences in responsiveness and differences in payoffs. While results are imprecise, the point estimates in Table C.3 consistently confirm the hypothesized direction.[57] Thus, evidence largely indicates that payoff differences across bins are not the source of differential responsiveness documented here and the patterns are consistent with rational inattention.

IV.C. Convergence to Nash in High-Leverage States

Prediction (C) states that in high-leverage situations, the strategies pursued by the batter and pitcher should converge to their Nash strategies. Our test is indirect. First, we identify states with large L by classifying states according to their *leverage*. Given that leverage is a metric used in sabermetrics to quantify the importance of a state—and is directly related to the variability in outcomes—it is a natural proxy for L. Since each instance of a payoff we do observe is an element of the payoff matrix for the state, increasing L by a factor of two, say, would increase the standard deviation of the observed payoffs by two. We thus use the standard deviation of observed payoffs within groups of states as our basic measure of leverage. As before, we need to group states into coarser categories to compute this metric. Intuitively, the two largest determinants of the stakes are the inning and the differences in scores. For each combination of the two,[58] we take the standard deviation of win probability added across all at-bats in our dataset.

Second, we will test for convergence of strategies to full-information Nash equilibrium by testing for payoff equivalence between pitches *within* state. To do so, we need to identify a payoff metric that is positively related to our preferred metric of win probability added but does not scale with L. For instance, WPA mechanically scales with leverage, which would counteract the convergence of payoff difference to zero as strategies approach Nash. As such, our preferred specification for testing (C) is based on FIP, whose scale does *not* change with leverage and thus

[57] The exception is game timing, where point estimates go in a different direction but are especially noisy.

[58] We require that there be at least 100 instances of that score-inning combination to calculate the leverage, and we also condition on whether it is the top or bottom of the inning.

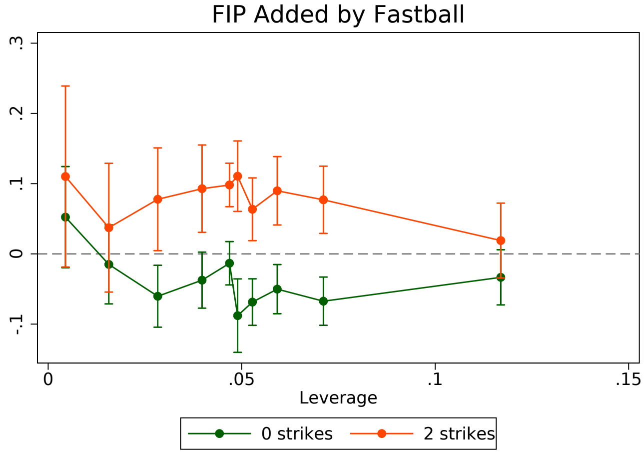


Figure 5: FIP by leverage for zero-strike and two-strike counts. Lower FIP is better for the pitcher.

can be a proxy for $\tilde{w}(\cdot,\cdot)$.[59] We use the specification

$$Y_p = \sum_b \alpha_b \cdot \mathbb{1}[\text{Fastball}]_p \cdot \xi_{b(p)} + \text{State FEs} + \epsilon_p, \tag{10}$$

where $\xi_{b(p)}$ is a dummy that is 1 if the pitch is thrown in a state with leverage in bin b. We run (10) separately for zero- and two-strike counts. We would expect payoff differences to converge to zero as leverage increases. Note, however, that (C) does *not* state that payoffs must monotonically converge across all levels of leverage; under rational inattention, payoff differences could still increase in leverage for low values of leverage. See Appendix B.4 for a simple numerical example in which strategies are not monotonic in leverage.

In a first pass, we look at ten equally-sized bins of leverage and look at FIP-differential of fastball in zero- and two-strike counts, plotting α_b from (10). These are shown in Figure 5, which shows a bow shape for FIP as a function of leverage. With some deciles as exceptions, FIP differences tend to increase in magnitude until approximate the median leverage, and they then generally show a downward trend—although the pattern is not especially smooth, especially for 0-strike counts. In the highest decile of leverage, the payoff differences are statistically indistinguishable from zero in zero- and two-strike counts with smaller point estimates than

[59]The map from a home run to win probability added, say, changes with leverage. In high leverage situations, a home run will decrease the win probability (of the pitching team) by a significant amount whereas a strikeout will increase it by a lot. In low leverage situations, the win probability added will be small in both outcomes. However, in both situations, a home run will earn 13 points for the FIP metric. A related note is that if one were to believe that FIP is the true objective of the pitcher, leverage would be similar in all settings, and (C) would be difficult to test. Appendix C.3 verifies that win probability added can indeed be thought of as proportional to FIP times leverage, which makes it a candidate for $\tilde{w}(\cdot,\cdot)$ in (C).

	(1)	(2)	(3)	(4)
Fastball, 0 strikes	-0.0752***	-0.1172***	-0.0690***	-0.0677***
	(0.0167)	(0.0222)	(0.0162)	(0.0143)
Fastball, 2 strikes	0.1227***	0.1162***	0.0882***	0.1207***
	(0.0200)	(0.0264)	(0.0191)	(0.0144)
Fastball, 0 strikes × Leverage	0.4320*	0.8179***		
	(0.2523)	(0.2924)		
Fastball, 2 strikes × Leverage	-0.6688**	-0.6193*		
	(0.3091)	(0.3507)		
Fastball, 0 strikes × Leverage (Tango)			0.0181	
			(0.0122)	
Fastball, 2 strikes × Leverage (Tango)			-0.0104	
			(0.0156)	
Fastball, 0 strikes × Inning				0.0053*
				(0.0028*)
Fastball, 2 strikes × Inning				-0.0104***
				(0.0031)
Cutoff Percentile	25	50	25	–
Joint p-value	0.019	0.003	0.253	0.0005
N	1,692,802	1,045,289	1,449,186	1,879,452

Table 4: Effect of leverage on FIP, by strike. Column (1) uses the top 75% leverage situations. Column (2) uses the top 50%. Column (3) also uses the top 75% but use the leverage measure computed by Tom Tango. Column (4) uses all the data and uses inning as the proxy of leverage. We report the p-value of the joint test that both the interactions of 0 and 2 strikes with leverage are 0. Lower FIP is better for the pitcher. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

previous deciles. Figure 2(a) in Appendix C.4 shows similar results with wOBA.

In Appendix C.4 we also run a version of this regression with WPA as the outcome. This is not our preferred metric to test this prediction since, as discussed above, WPA mechanically scales with leverage. As such, there is a tendency for WPA to diverge as L increases even if strategies are converging to Nash. Figure C.5 shows that for low values of L, WPA differences are mechanically close to zero. However, the figure is suggestive that payoff differences do not diverge as leverage increases and may in fact be converging back to zero (although estimates are noisy) despite the mechanical effect. This provides further evidence in favor of (C).

The pattern in the graph is admittedly noisy, so we run a statistical test corresponding to (C). In particular, we test whether FIP differences are increasing for zero-strikes and decreasing for two-strikes in the right half of the graph. We run regressions to answer this question in Table 4. Here we run the analogue of (8) but replace Attention_p with Leverage_s and use only half or three-quarters of the data which is high-leverage. We control for the state space, and we interact our fastball coefficient with leverage, to see if the effect of a fastball is changing as leverage increases. While the cutoffs of this regression are arbitrary, the idea is to test if the relationship we see in Figure 5(a)—i.e., the convergence of FIP for sufficiently high-leverage situations—is a robust feature of the data. We can further test the robustness of this relationship by using alternative measures of leverage. First, we use a leverage index commonly used in sabermetrics created by Tom Tango.[60] We also trim this at the bottom 25%.[61] Second, we also use a very simple measure, which is the inning of the game, with the idea that later innings tend to have higher-leverage situations. While the results with Tango's measure are noisy (especially for 2 strikes), the estimates paint a consistent picture that a fastball's effects on FIP are positive and decreasing with zero strikes and negative and increasing with two strikes, consistent with them converging towards zero.

IV.D. Comparison to Quantal Response Equilibrium

While we have derived our empirical predictions from a model of rational inattention with Shannon entropy, it is useful to discuss another canonical model that can generate departures from Nash equilibrium: QRE of McKelvey and Palfrey (1995). In this solution concept, players choose strategies based on the expected payoff from the strategy plus some noise. In the context of baseball, we can think of this as unobservable random preference over throwing a fastball or not; on some pitches, a player may just feel more confident throwing a fastball for entirely idiosyncratic reasons.

In a QRE, payoff needs not be equalized across actions. As such, it can explain the results in Section III. Moreover, since the probability of playing an action is positively related to the expected payoffs from it in a QRE by design, (A) would also be rationalized. Replicating (B) would require an assumption on how the unobserved random preferences vary across players: in particular, our measures of attention costs must be correlated to the dispersion of the unobserved random preferences. Replicating (C) would require an assumption on how the unobserved random preferences vary across states: a sufficient condition would be that the dispersion does not vary across states based on leverage.

Without a microfoundation for the errors in QRE, it is difficult to gauge how reasonable these

[60] Tom Tango is a respected sabermetrician, author of *The Book: Playing the Percentages in Baseball*, and currently at MLB Advanced Media. See http://www.insidethebook.com/li.shtml for the leverage data. This metric assumes fixed likelihoods of a limited set of outcomes (home run, single, etc.) and uses the mean absolute deviation of win probability. It is computed at a finer level than our baseline one: at the inning-score differential-out-base position level. This metric has been used in the academic literature by Phillips (2017) as well. Note that Tango scales his measure of leverage differently than ours, making the magnitudes of the coefficients not directly comparable across specifications.

[61] Again, in the absence of preregistration, we recognize that choosing cutoffs gives us wide latitude to manipulate the statistical significance of the results in Table 4. This is partly the reason we present the nonparameteric analysis in Figure 5. Together, we hope this provides plenty of information related to Prediction (C) for readers to judge the evidence for themselves.

restrictions are.[62] Admittedly, some models for the error could still justify parts of (B). One may argue that teams with high emphasis on analytics do not improve information acquisition strategies but rather educate players about payoffs—which may allow for less room for idiosyncratic beliefs about payoffs, which might be what the error term in QRE captures. This argument could be plausibly extended to age, if older pitchers have strong state-independent priors of the benefit of a fastball and younger pitchers are more open to listening to their manager or their catcher. However, we have a difficult time justifying why the error term in a QRE would be larger for hot games or for afternoon games. Thus, while we cannot formally reject that the agents have unobserved stochastic payoffs with a structure that happens to yield (B) and (C), it does require an auxiliary interpretation of the error term, while one exists naturally in rational inattention.

The key difference between QRE and rational inattention is that the former does not impose any restrictions on payoffs across states. In this sense, (D) provides a more interesting restriction of rational inattention and one that need not be satisfied by QRE. Doing so would effectively require a restriction on error terms across states, and unlike for (C), we are not aware of interpretable restrictions that would guarantee (D) in a QRE. We test (D) below and show that it is indeed a strong restriction of rational inattention with Shannon entropy. That is, rational inattention provides very tight conditions on the average deviations from payoff equivalence; this average could be satisfied by QRE, but it would have to be a coincidence.

V. Approximate Payoff Equivalence and the Cost of Attention

A rational inattention model allows for payoff differences between pitches state-by-state. As discussed above, while other models also allow for departures from payoff equivalence, the rational inattention model imposes a further constraint on payoffs that need not be satisfied in other models: Prediction (D) shows that the payoff differences should be "approximately" zero when averaged across all states. Otherwise the pitcher would find it too costly to acquire information and thus would not mix between pitches. In Section V.A, we test the sufficient (but not necessary) condition that the payoff of a fastball averaged across states is the same as that of a different pitch. Despite narrow confidence intervals, we largely cannot reject this null hypothesis.

The approximate payoff equivalence requirement comes from the relationship between payoffs and actions that a rational inattention model imposes. In the remainder of the section, we go further and impose this structure on the data. This structure is a consequence of the Shannon entropy assumption that we have maintained throughout this paper, and this exercise has a number of benefits. First, we test a stronger necessary condition that implies approximate payoff equivalence. Second, we quantitatively gauge the fit of the model. Third, we revisit the "approximate" nature

[62] QRE does not provide a natural interpretation of the errors, and Goeree et al. (2016) note that endogenizing the error in QRE is an important area for research. See Friedman (2019) for one approach.

of payoff equivalence and show that rational inattention does in fact impose rather tight bounds on the average difference in payoffs between actions, and approximate payoff equivalence is in fact a powerful test for rational attention. Finally, we can estimate a cost of attention; this adds another data point to the growing endeavor of estimating such costs in different economic settings.

V.A. Testing Approximate Payoff Equivalence

In Section I.B, Prediction (D) showed that a necessary condition for rational inattention was that, on average, payoff equivalence should approximately hold when mixing occurred. If one action were substantially better on average, even if it were not better in all states, then high enough attention costs would lead the agent to always choose that action. An empirical test that showed that payoffs were on-average equal would be sufficient to satisfy this condition. So if we are unable to reject $\mathbb{E}_s w_s = 0$, then we are unable to reject (D).

We run the same regression as (7), but without splitting the sample by strikes:

$$Y_p = \beta \cdot \mathbb{1}[\text{Fastball}]_p + \text{State FEs} + \epsilon_p. \tag{11}$$

However, without proper weights, $\hat{\beta}$ does not identify $\mathbb{E}_s w_s$. There are two reasons. First, regression implicitly weights the expectation by the variance of the fastball dummy within states (Angrist and Pischke, 2008).[63] So to estimate the average effect, we weight by the inverse variance of the fastball dummy within states, undoing the implicit weights. Second, there are states in which we cannot estimate a w_s because they are so infrequent we observe only fastballs or only non-fastballs in them. In the regression, they are dropped. These states are not random: they are more likely to occur with two strikes than with zero strikes. This would lead $\hat{\beta}$ to be biased upward, since w_s tends to be positive with zero strikes and negative with two strikes. However, if we assume that conditional on strike, the unidentified states are random, we can weight the observations to correct for this bias. For a pitch with 2 strikes, we upweight by the total number of pitches with 2 strikes divided by the total number of pitches with 2 strikes in states where we can identify w_s. We do the same for 0 and 1 strikes.

Table 5 shows our results. In Column (1), we see that the magnitude for win probability added is both economically and statistically insignificant. Column (2) shows that the average FIP difference is considerably smaller than the magnitudes observed in Table 1 by strike, and it is also not significant. Column (3) shows that the effect on wOBA is marginally significant 10%, but much smaller in magnitude that the effects for zero- or two-strikes. The effect on wins is noisy, as expected. The effect for runs remaining in Column (5) is significant, but still smaller than the

[63] A state in which fastballs are thrown half of the time (variance of 0.25) gets more weight than a state in fastballs are thrown 75 or 25 percent of the time (variance of 0.1875).

	(1) WPA × 1000	(2) FIP^{-} × 100	(3) wOBA^{-} × 100	(4) Wins × 100	(5) Runs Rem^{-} × 100
			A. Weighted by Inverse Variance		
Fastball	0.0145	0.217	0.155*	0.147	-0.336**
	(0.0871)	(0.458)	(0.094)	(0.118)	(0.171)
N	1,959,222	1,959,222	1,959,222	1,959,222	1,959,222
			B. Unweighted		
Fastball	0.0273	-0.154	0.104	0.088	-0.308**
	(0.0849)	(0.427)	(0.0087)	(0.107)	(0.153)
N	2,707,048	2,707,048	2,707,048	2,707,048	2,707,048

Table 5: Testing payoff equivalence on average. Panel A weighs observations by the inverse variance of a fastball in that state, and Panel B is unweighted. Observations are dropped in Panel A if the state does not have sufficiently many data points to compute a variance. States are upweighted in Panel A to maintain the same proportion of strikes, as described in the text. For variables labeled with a $^{-}$ (columns 2, 3 and 5), lower values are better for the pitcher. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

effect we found in either zero- or two- strike counts.[64] Note that Columns (4) and (5) speak to dynamic considerations; Figure C.6 in Appendix C.5 shows that payoffs are approximately equal on average even when using the intermediate metrics of cumulative win probability added.

For reference, Panel B gives results from unweighted regressions. This regression would not give an estimate the quantity $\mathbb{E}_s w_s$ of interest, because unweighted regression implicitly weights more heavily the states with more variance in the fastball dummy. However, it is reassuring that the results are similar and that standard errors—already economically small—are not inflated appreciably by adding the weights.[65]

We also test (D) for other types of pitches in Table 6: in a model with finer actions, we would expect payoffs of all actions to be approximately equal on average. To do the weighting correctly, we test the pitches separately, with the excluded category being all other pitches. For example, in Column (2), we test if curveballs have the same expected payoff as all non-curveballs, using a weighting scheme based on the inverse variance of a curveball dummy, as well as the weighting by strike. For reference, Column (5) shows a joint regression in which we do not weight, where the omitted category is fastball. All point estimates are small, and with the exception of changeups,

[64]This result could be due to the fact that runs are not what the pitcher is likely trying to minimize. Throwing a fastball may have an especially large effect on decreasing the probability of a many-run outburst. In some situations, however, the last of the many runs will have a negligible impact on whether the team wins the game. Consistent with this interpretation, in Appendix C.5 Figure C.6, we see that there is no effect on the win probability added over the course of the entire inning.

[65]Making the inverse-variance adjustment but not the strike adjustment also gives similar coefficients, and runs remaining is the only statistically-significant variable.

	(1)	(2)	(3)	(4)	(5)
Fastball	0.0145				
	(0.0871)				
Curveball		-0.0684			-0.0636
		(0.147)			(0.152)
Slider			-0.0683		0.0705
			(0.137)		(0.142)
Changeup				-0.262*	-0.255*
				(0.140)	(0.143)
Joint p-value	0.868	0.641	0.618	0.061	0.270
N	1,959,222	829,536	1,129,063	1,062,161	2,025,999

Table 6: Testing approximate payoff equivalence for other pitches. The outcome metric is WPA, scaled by 1000. Columns (1)–(4) report (11) but with the fastball dummy replaced by a different pitch; they are weighted as described in the text. Column (5) uses the included dummies and is unweighted. The joint p-value is for the test that the coefficients on all included pitch types are zero. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

they are statistically insignificant. A joint test of all the pitch-types in Column (5) is insignificant.

Overall, we see that payoffs across different actions are approximately equal on average.

V.B. Empirical Specification

Since Matějka and McKay (2015) show that the rational inattention model with Shannon entropy has a logit formulation with a bias parameter that depends on the mean probability of picking an action across all states, one could estimate a combination of utility and the cost of attention purely from choice data via a maximum likelihood procedure. In our setting, we have access to interpretable metrics of payoffs as well, which will allow us to estimate costs of attention in units of this payoff metric. Rearranging (4), we have

$$w_s^f - w_s^n = \lambda \cdot \left[\log p_s^f - \log\left(1 - p_s^f\right) - \left(\log p_0^f - \log\left(1 - p_0^f\right)\right)\right]. \tag{12}$$

Essentially, if we estimate the differential payoff of a fastball for a set of states, the slope between the payoff and (a known function of) the probability a fastball is actually thrown in a particular state should return the cost of attention.

With a sufficiently large dataset for each state—the full combination of count, out, inning, base position, score differential, and batter that each pitcher might face—we could estimate both sides of the equation for each of these states. Despite our millions of observations, the state space is of the same order of magnitude, so we cannot implement this strategy directly. Instead, we use the coarser categories of counts and count-outs.[66] Within each of these aggregate categories σ, we estimate the payoff difference between fastballs and other pitches using the OLS regression

$$Y_p = \beta_\sigma^f \cdot \mathbb{1}[\text{Fastball}]_p + \text{State FE} + \epsilon_p,$$

where the Y_p is the preferred metric of the outcome of pitch p. This specification allows for differences in payoffs for *both* fastballs and other pitches across states, and β_σ^f is the average payoff difference between these two types of pitches to be the same across all states within σ. We can compute $\hat{p}_\sigma^f$ as the proportion of pitches within σ that are fastballs, and $\hat{p}_0^f$ as the proportion of pitches overall (across all states) that are fastballs. Thus, to move from (12) to an estimable equation, we run

$$\hat{\beta}_\sigma^f = \hat{\alpha} + \hat{\lambda} \cdot \left[\log \hat{p}_\sigma^f - \log\left(1 - \hat{p}_\sigma^f\right) - \left(\log \hat{p}_0^f - \log\left(1 - \hat{p}_0^f\right)\right)\right] + \epsilon_\sigma. \qquad (13)$$

Equation (13) is the analogue of the standard logit regression of (log) shares on characteristics to recover utility parameters. In a standard utility setting, the scale of the parameters is set by a normalization that the logit shock takes on a Type 1 Extreme Value distribution. Here, the normalization comes from the fact that we observe payoffs. Second, in a standard demand setting, one can justify departures from the logit share equation by including an unobserved component of demand. Here, we interpret ϵ_σ as measurement error in the left-hand side. As such, the standard deviation of ϵ_σ should be the standard error of the estimate of $\hat{\beta}_\sigma^f$, and we estimate (13) via weighted least squares, where the weight is the inverse variance of the estimate $\hat{\beta}_\sigma^f$.[67]

Our estimate for the cost of attention is $\hat{\lambda}$. This exercise also provides two related tests for the model. Since the estimating equation (4) must hold state-by-state, the ordered pairs of payoff differences and relative log odds for a fastball must lie on a line with positive slope. Second, this line must go through the origin, so that $\alpha = 0$. The latter prediction is indeed stronger than Prediction (D), which we will also revisit with our estimate of λ.

V.C. Estimating λ

Table 7 shows estimates of α and λ for two partitions of the states and two outcomes. Column (1) shows estimates simply using count and Column (2) uses combinations of count-out. Columns (3) and (4) repeat the exercise with FIP as the outcome. Recall that since lower FIP is preferred by the pitcher, we estimate a negative λ. Estimates are similar for the two partitions, and λ is

[66] We have checked for robustness with even finer categories, such as count-out-inning and count-out-base. While the results are similar, many payoffs are estimated imprecisely at these finer categories of states.

[67] Standard errors on $\hat{p}_\sigma^f$ are sufficiently small that we do not anticipate measurement error on the right-hand side to yield appreciable attenuation bias. This is the benefit of using the coarser categories of states.

	WPA × 1000		FIP^{-}	
	Count	Count-Out	Count	Count-Out
	(1)	(2)	(3)	(4)
λ	1.410***	1.405***	-0.129***	-0.116***
	(0.408)	(0.274)	(0.037)	(0.022)
α	0.014	0.002	-0.003	-0.000
	(0.128)	(0.093)	(0.011)	(0.007)
N	12	36	12	36
R^2	0.544	0.437	0.554	0.445
Bound	0.149	0.243	0.0151	0.0219
	(0.064)	(0.088)	(0.0051)	(0.0060)

Table 7: Estimates of the cost of attention. Columns correspond to different aggregates of states and different The number of observations corresponds to the number of aggregate states used in the estimate. The bound is the right-hand side of (14). Recall that low FIP is beneficial for the pitcher (indicated by the $^{-}$). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

always statistically significant and has the expected sign. The numbers are not interpretable by themselves, and we provide quantifications of the cost of attention in the next subsection.[68,69]

Before discussing the quantification, we discuss how this exercise helps gauge model fit. One restriction of the rational inattention model is that $\alpha = 0$. The second row of Table 7 reports these estimates, and we find that they are indeed precise zeros. We should emphasize that nothing in the estimation procedure mechanically forces them to be close to zero, and it is a consequence of the observed strategies and payoffs that it is. Indeed, since (D) can be derived from the estimating equation (4), the test that $\alpha = 0$ is stronger. We revisit this in more detail in the next subsection as well.

Another test is that if the rational inattention model—with Shannon entropy as the cost of information acquisition—were fully explaining the data, the relative payoff of a fastball would be exactly linearly related to the log odds of a fastball. We can return to Figure 4 in Section IV.A, which plots the scatter plots corresponding to the various regressions in Table 7. The size of the point is inversely related to the variance of the estimate of the relative payoff of a fastball. Visual inspection of the panels suggest that the data do exhibit a linear pattern, and the points that depart from the pattern tend to be imprecisely estimated.[70] Moreover, the estimated R^2 range from 0.44

[68]One might alternatively consider running regressions at the pitch level of the outcome on the log probability of a fastball in that state less the log of the mean probability of a fastball, with the state fixed effects, with standard errors clustered at the coarse category level. Doing so returns very similar estimates of λ with similar precision. We choose to run the estimation procedure in this manner for transparency with Figure 4.

[69]We also estimate a single λ for the entire population. Table C.3 in Appendix C.2 explores some dimensions of heterogeneity.

[70]We have also investigated finer categories of the state. The linear pattern is salient with interacting with the base positions, for instance, but moving to especially fine categories leads to very imprecisely estimated payoffs.

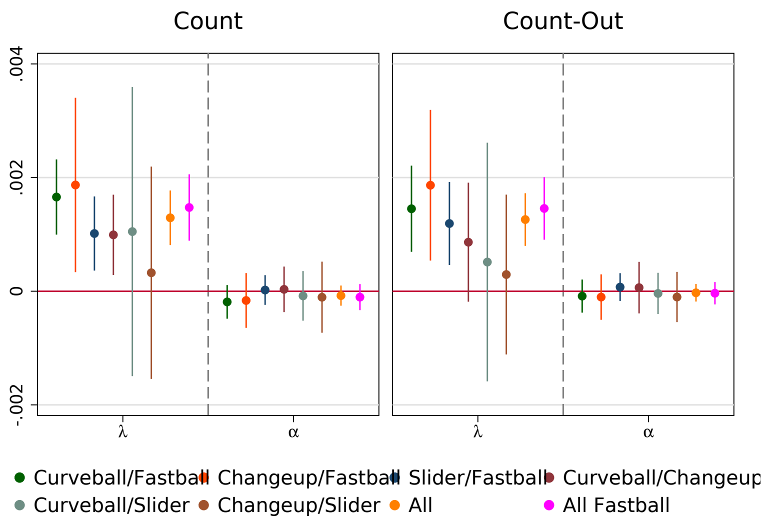


Figure 6: Cost of attention, computed using pairwise comparisons of finer pitch types and win probability added as the metric. "All" denotes using all pairwise combinations of pitch types, and "All Fastball" denotes just using the pairs involving fastballs. Errors bars show 95% confidence intervals.

to 0.55 depending on the state and outcome metric. These R^2 are only slightly smaller than one might expect based on the standard errors of the win probability added estimates. The smaller R^2 could reflect the restrictive nature of the Shannon entropy assumption.[71]

As with all other results in the body, the estimates of λ are done through a comparison of fastballs and other pitches. However, with a wider set of actions, such as finer pitch types, an analogous equation to (12) can be estimated by replacing "fastball" and "not" by any other pair. For pitch types a and b, we would have

$$w_\sigma^a - w_\sigma^b = \lambda_{ab} \cdot \left[\log p_\sigma^a - \log p_\sigma^b - \left(\log p_0^a - \log p_0^b \right) \right],$$

and we can write an estimating equation like (13). The rational inattention model with Shannon entropy would then imply that the estimate for λ_{ab} should be the same for all (a, b) pairs, and the constant in that regression should always be zero. Figure 6 plots this for win probability added for both count and count-out. We also include two specifications that include all pairwise combinations (labeled "All") and all pairwise combinations involving a fastball (labeled "All Fastball"). In most specifications, the estimate for the cost of attention remains relatively stable

[71] Appendix C.6 investigates the likelihood of observing the estimated R^2 were the true data generating process a rational inattention model plus measurement error. The conclusion is generally that modest increases in the measurement error, relative to the sampling error in this setting, can explain these values. Note that recent papers that estimate the shape of the attention cost function from lab experiments have found less support for Shannon entropy (Dean and Neligh, 2019; Dewan and Neligh, 2020).

and within the error bar of the point estimate from Table 7—which means the underlying sensitivity of actions to payoff differences across states does not depend on the comparison of pitches considered. The exception is the changeup/slider comparison, which is likely explained by the fact that those two are not typically used in the same states, so the relative WPA is hard to identify.[72] Moreover, the estimated constant is a precise zero in all specifications, which again is a stronger prediction than (D).[73]

V.D. How Tight is Prediction (D)?

In Section I.B we argued that $\mathbb{E}_s w_s = 0$ was a sufficient condition for (D) to hold. However, this condition is not necessary, and especially small values of λ could lead to (D) being satisfied by a wide range of average payoff differences. To test whether this is the case in our setting, we compute an approximate bound for $\mathbb{E}_s w_s$ based on λ. Taking the second-order expansion of the exponential, we arrive at

$$|\mathbb{E}_s[w_s]| \leq \frac{\mathbb{E}_s\left[w_s^2\right]}{2 \cdot |\lambda|}.^{74} \tag{14}$$

Using estimated payoffs and λ, we can estimate on the right-hand side of (14). The final row of Table 7 shows these bounds. They are indeed satisfied by the average payoff difference of a fastball (Table 5). However, they are almost an order of magnitude smaller than the payoff differences of a fastball for 0 and 2 strike counts (Table 1), suggesting they are economically rather tight. This provides additional verification that Prediction (D) was a priori falsifiable.

Figure 7 provides a graphical illustration of how tight (D) is for the estimated λ. In olive, we show the confidence interval for the average fastball payoff, which (D) said should be approximately zero. In orange is what we mean by "approximately," i.e. it is the range of counterfactual average WPA for which pitchers would still mix under rational inattention and our estimated value of λ, as given by (14). Given the overlap of the confidence interval for the average with the bound, we an conclude that it is likely that the average payoff to a fastball falls within the fairly narrow bounds implied by the rational inattention model. To highlight that the bound is tight, we provide for comparison confidence intervals for the payoff difference for 0 and 2 strikes. Even the edges of the confidence intervals do not overlap with the range for mixing.

[72] Many pitchers throw a mix including sliders but not change-ups to same-handed batters and change-ups but not sliders to opposite-handed batters

[73] We observe a similar degree of stability when using FIP as a metric, although the changeup/slider comparison yields an estimate of λ that is significantly different from the others.

[74] This result follows from noting that, to second order, both

$$1 + \frac{\mathbb{E}_s[w_s]}{\lambda} + \frac{1}{2}\frac{\mathbb{E}_s[w_s^2]}{\lambda^2} > 1 \text{ and } 1 - \frac{\mathbb{E}_s[w_s]}{\lambda} + \frac{1}{2}\frac{\mathbb{E}_s[w_s^2]}{\lambda^2} > 1.$$

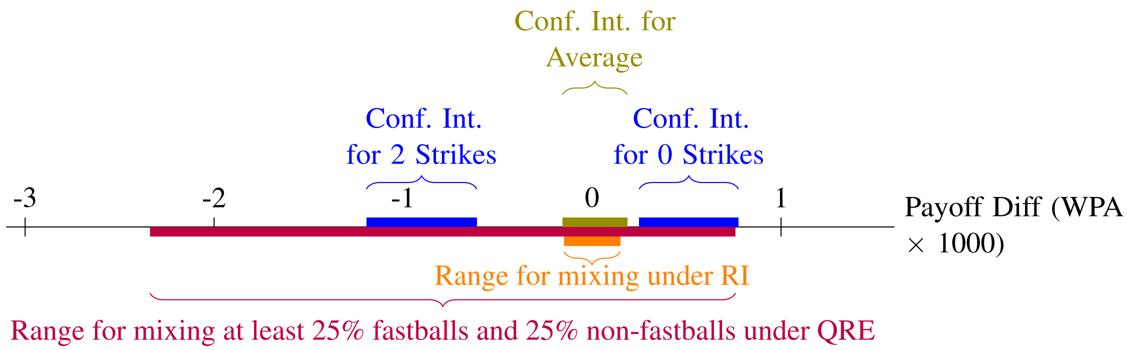


Figure 7: How Tight is Prediction (D)? The empirical confidence interval of the average payoff difference is shown above the line. The confidence intervals for 0 and 2 strikes are also provided for reference. Predictions from the RI and QRE models are below the line. "Range for mixing" refers to the bounds for the average payoff difference between fastballs and non-fastballs that would lead to a mixed strategy. In the case of QRE, there would be mixing at any average payoff difference, so we show the bounds for a significant amount of mixing, i.e. at least 25% of each action. Finding a zero for the average payoff difference is a possibility under either model, but it is a strong prediction of RI, while it is fairly unlikely under QRE.

Figure 7 also provides a back-of-the-envelope comparison to QRE, which could provide another explanation for state-by-state departures from payoff equivalence. In particular, we compute the range of payoff differences under which we would observe "substantial" mixing between fastballs and non-fastballs under a QRE with a logit shock with parameter λ. Since a QRE with an error term with infinite support would always predict mixing, we compute the average payoff differences such that a player who knows he is facing a random draw of a state will play both a fastball and a non-fastball with probability at least 25%.[75] This range is illustrated in purple, and it is clear that even substantial mixing can be justified by a very wide range of average payoff differences. Of course, this interval includes the range of payoff differences over which rational inattention justifies mixing, but given its width, observing an average WPA near zero would be only by coincidence. Thus, (D) provides a powerful one-sided test against rational inattention.

V.E. Quantifying the Cost of Attention

We can use the estimate of λ, together with the rough estimate of the value of a win from Swartz (2017), to place a dollar value on the cost of attention. One could interpret $\lambda / \log_2(e)$ as the

[75] To estimate this range, we note that in a QRE with a logit shock, we would expect the probability of a fastball in state s to be proportional to $\exp(q_s/\lambda)$, where q_s is the perceived payoff of a fastball in state s. If we let $q_s = \tilde{\alpha} + w_s$, where w_s is the WPA of a fastball and $\tilde{\alpha}$ is a constant term that denotes an unobserved component payoff of a fastball chosen to fit the average probability. Then, we can see that $\tilde{\alpha} = \alpha - \lambda \cdot (\log p_0^f - \log(1 - p_0^f))$ in the notation of (13). The interval we report is the range of $\mathbb{E} w_s$ such that $\exp((\tilde{\alpha} + \mathbb{E} w_s)/\lambda)/[1 + \exp((\tilde{\alpha} + \mathbb{E} w_s)/\lambda)]$ lies between 0.25 and 0.75.

cost of an additional "bit" of information—asking one additional binary question to acquire information about the state. With a value of $10 million on one win, such a question would cost $9,790—which is admittedly abstract.

For a more concrete measure, we can use (3) to compute the information processing cost for any candidate strategy. Consider a strategy that would deliver a value of $p_s^f \in \{0,1\}$ for all states s; such a strategy would correspond to a pitcher perfectly observing the state—given the equilibrium payoffs fixing the batter's strategy—and knowing whether to pitch a fastball. From (3), the cost of such a strategy would be $-\lambda \cdot \left(p_0^f \cdot \log(p_0^f) + (1-p_0^f) \cdot \log(1-p_0^f)\right)$, where p_0^f is the average probability of a fastball in that strategy. Using estimates of the payoff in each state, we can compute p_0^f for a strategy where the pitcher throws fastballs only when it is beneficial, assuming no other changes to payoffs, including responses by opponents; this strategy would cost the pitcher approximately $9,770 per pitch. By contrast, the current strategy of the pitchers heavily economizes on attention costs: using the estimated λ and the observed strategies by state, we estimate that the current strategy only costs the pitchers approximately $234 per pitch.[76]

VI. Conclusion

We provide one of the first pieces of evidence that professionals are rationally inattentive to the state in a high-stakes setting. We use data from Major League Baseball, which qualitatively has the features where one would expect to find rational inattention: players have to make decisions quickly without fully processing information, they do so repeatedly, but they also have the time and incentives to devise strategies ahead of time. Moreover, some modern baseball franchises have embraced the "design" of strategies through analytics departments. One may have expected that in such a setting, Nash equilibrium would explain the data well. On the contrary, we find that Nash equilibrium play state-by-state does not explain the data, since payoff differences do exist across actions.

However, players do partially respond to the payoff differences in the state, and the response is larger in situations we may expect would correspond to a lower cost of attention. Moreover, play converges (albeit slowly) to Nash equilibrium as the stakes become higher. Finally, payoff differences do not exist in the aggregate. These results are all consistent with a model of rational inattention. From the responsiveness of strategies for payoffs, we use the structure of the model to estimate a cost of attention. We find that the fully revealing strategies would be prohibitively expensive, while the strategy employed in equilibrium economizes on attention costs substantially. These results suggest that attention costs are high even for professionals.

[76]These estimates are based on partitioning the state into count-out and using the win probability added measures. However, they are robust to finer categories of the states, such as interacting with bases or inning.

References

AFROUZI, H. (2020): "Strategic Inattention, Inflation Dynamics, and the Non-Neutrality of Money," Tech. rep., Columbia University.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

BARRA, A. (2010): *Yogi Berra: Eternal Yankee*, W. W. Norton and Company.

BARTOŠ, V., M. BAUER, J. CHYTILOVÁ, AND F. MATĚJKA (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106, 1437–75.

BESSONE, P., G. RAO, F. SCHILBACH, H. SCHOFIELD, AND M. TOMA (2020): "The Economic Consequences of Increasing Sleep Among the Urban Poor," Tech. rep., Harvard.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2016): "Competition for Attention," *Review of Economic Studies*, 83, 481–513.

BROCAS, I., J. D. CARRILLO, S. W. WANG, AND C. F. CAMERER (2014): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *Review of Economic Studies*, 81, 944–970.

BROWN, J., T. HOSSAIN, AND J. MORGAN (2010): "Shrouded Attributes and Information Suppression: Evidence from the Field," *Quarterly Journal of Economics*, 125, 859–876.

BROWN, Z. Y. AND J. JEON (2020): "Endogenous Information Acquisition and Insurance Choice," Tech. rep., University of Michigan.

CAPLIN, A. AND M. DEAN (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105, 2183–2203.

CAPLIN, A., M. DEAN, AND J. LEAHY (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 86, 1061–94.

CAPLIN, A., J. LEAHY, AND F. MATĚJKA (2016): "Rational Inattention and Inference from Market Share Data," Tech. rep., CERGE-EI.

CHETTY, R., A. LOONEY, AND K. KROFT (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–77.

CHIAPPORI, P.-A., S. LEVITT, AND T. GROSECLOSE (2002): "Testing Mixed-Strategy Equilibria When Players are Heterogeneous: The Case of Penalty Kicks in Soccer," *American Economic Review*, 92, 1138–1151.

COHEN, L. AND A. FRAZZINI (2008): "Economic Links and Predictable Returns," *Journal of Finance*, 63, 1977–2011.

COIBION, O., Y. GORODNICHENKO, AND S. KUMAR (2018): "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, 108, 2671–2713.

COVER, T. M. AND J. A. THOMAS (1991): *Elements of information theory*, John Wiley & Sons.

CRAIK, F. I. AND E. BIALYSTOK (2006): "Cognition Through the Lifespan: Mechanisms of Change," *Trends in Cognitive Sciences*, 10, 131–138.

DAI, L., I. KLOOG, B. A. COULL, D. SPARROW, A. SPIRO III, P. S. VOKONAS, AND J. D. SCHWARTZ (2016): "Cognitive Function and Short-Term Exposure to Residential Air Temperature: A Repeated Measures Study Based on Spatiotemporal Estimates of Temperature," *Environmental Research*, 150, 446–451.

DEAN, M. AND N. NELIGH (2019): "Experimental Tests of Rational Inattention," Tech. rep., Columbia.

DELLAVIGNA, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–72.

DELLAVIGNA, S. AND J. M. POLLET (2009): "Investor Inattention and Friday Earnings Announcements," *Journal of Finance*, 64, 709–749.

DENTI, T. (2020): "Unrestricted Information Acquisition," Tech. rep., Cornell.

DESSEIN, W., A. GALEOTTI, AND T. SANTOS (2016): "Rational Inattention and Organizational Focus," *American Economic Review*, 106, 1522–36.

DEWAN, A. AND N. NELIGH (2020): "Estimating Information Cost Functions in Models of Rational Inattention," *Journal of Economic Theory*, 187, 1–32.

ELLISON, G. AND S. F. ELLISON (2009): "Tax Sensitivity and Home State Preferences in Internet Purchasing," *American Economic Journal: Economic Policy*, 1, 53–71.

FOSGERAU, M., E. MELO, A. DE PALMA, AND M. SHUM (2020): "Discrete Choice and Rational Inattention: A General Equivalence Result," *International Economic Review*, 61, 1569–89.

FRIEDMAN, E. (2019): "Endogenous Quantal Response Equilibrium," Tech. rep., Columbia University.

GABAIX, X. (2019): "Behavioral Inattention," in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 2, 261–343.

GAGLIANONE, W. P., R. GIACOMINI, J. V. ISSLER, AND V. SKRETA (2020): "Incentive-Driven Inattention," *Journal of Econometrics*.

GOEREE, J. K., C. A. HOLT, AND T. R. PALFREY (2016): *Quantal Response Equilibria*, Springer.

GREEN, E. AND D. DANIELS (2018): "Bayesian Instinct," Tech. rep., Wharton.

HANDEL, B. R. AND J. T. KOLSTAD (2015): "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare," *American Economic Review*, 105, 2449–2500.

HARTSHORNE, J. K. AND L. T. GERMINE (2015): "When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span," *Psychological Science*, 26, 433–443.

HELLWIG, C. AND L. VELDKAMP (2009): "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies*, 76, 223–251.

HSU, S.-H., C.-Y. HUANG, AND C.-T. TANG (2007): "Minimax Play at Wimbledon: Comment," *American Economic Review*, 97, 517–523.

JONES, J. J., G. W. KIRSCHEN, S. KANCHARLA, AND L. HALE (2019): "Association Between Late-Night Tweeting and Next-Day Game Performance Among Professional Basketball Players," *Sleep Health*, 5, 68–71.

JOO, J. (2020): "Rational Inattention as an Empirical Framework: Application to the Welfare Effects of New-Product Introduction," Tech. rep., UT Dallas.

KACPERCZYK, M., S. VAN NIEUWERBURGH, AND L. VELDKAMP (2016): "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica*, 84, 571–626.

KOVASH, K. AND S. D. LEVITT (2009): "Professionals Do Not Play Minimax: Evidence From Major League Baseball and the National Football League," Tech. rep., National Bureau of Economic Research.

KÜNN, S., J. PALACIOS, AND N. PESTEL (2019): "The Impact of Indoor Climate on Human Cognition: Evidence from Chess Tournaments," Tech. rep., IZA.

LACETERA, N., D. G. POPE, AND J. R. SYDNOR (2012): "Heuristic Thinking and Limited Attention in the Car Market," *American Economic Review*, 102, 2206–36.

LAURENT, J. G. C., A. WILLIAMS, Y. OULHOTE, A. ZANOBETTI, J. G. ALLEN, AND J. D. SPENGLER (2018): "Reduced Cognitive Function During a Heat Wave Among Residents of Non-Air-Conditioned Buildings: An Observational Study of Young Adults in the Summer of 2016," *PLoS Medicine*, 15, e1002605.

LEVITT, S. D., J. A. LIST, AND D. H. REILEY (2010): "What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments," *Econometrica*, 78, 1413–1434.

MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2018): "Rational Inattention: A Disciplined Behavioral Model," Tech. rep., CERGE-EI.

——— (2020): "Rational Inattention: A Review," Tech. rep., CERGE-EI.

MAĆKOWIAK, B. AND M. WIEDERHOLT (2015): "Business Cycle Dynamics Under Rational Inattention," *Review of Economic Studies*, 82, 1502–1532.

MAH, C. D., J. A. ANGUERA, A. GAZZALEY, AND A. LUKE (2017): "Sleep Loading Improves Visual Search Response Time and Reduces Fatigue in Professional Baseball Players," in *SLEEP*.

MARTIN, D. (2017): "Strategic Pricing with Rational Inattention to Quality," *Games and Economic Behavior*, 104, 131–145.

MATĚJKA, F. (2015): "Rationally Inattentive Seller: Sales and Discrete Pricing," *Review of Economic Studies*, 83, 1125–1155.

MATĚJKA, F. AND A. MCKAY (2012): "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review*, 102, 24–29.

——— (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105, 272–98.

MCKELVEY, R. D. AND T. R. PALFREY (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

MENZLY, L. AND O. OZBAS (2010): "Market Segmentation and Cross-Predictability of Returns," *Journal of Finance*, 65, 1555–1580.

MERRIAM, S. B. AND L. M. BAUMGARTNER (2020): *Learning in Adulthood: A Comprehensive Guide*, John Wiley & Sons.

MONDRIA, J., T. WU, AND Y. ZHANG (2010): "The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data," *Journal of International Economics*, 82, 85–95.

MORRISON, W. AND D. TAUBINSKY (2020): "Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes," Tech. rep., UC Berkeley.

PALACIOS-HUERTA, I. (2003): "Professionals Play Minimax," *Review of Economic Studies*, 70, 395–415.

PALACIOS-HUERTA, I. AND O. VOLIJ (2008): "Experientia Docet: Professionals Play Minimax in Laboratory Experiments," *Econometrica*, 76, 71–115.

PHILLIPS, D. C. (2017): "Stopping on Nine: Evidence of Heuristic Managerial Decision-Making in Major League Baseball Pitcher Substitutions," *Southern Economic Journal*, 84, 577–599.

POPE, D. G. AND M. E. SCHWEITZER (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 101, 129–57.

PORCHER, C. (2020): "Migration with Costly Information," Tech. rep., Dartmouth University.

ROMER, D. (2006): "Do Firms Maximize? Evidence from Professional Football," *Journal of Political Economy*, 114, 340–365.

SCHLACT, M. (2012): "Pregame Routines: A Typical Day in the Life of a Professional Baseball Player," *Bleacher Report*.

SICHERMAN, N., G. LOEWENSTEIN, D. J. SEPPI, AND S. P. UTKUS (2016): "Financial Attention," *Review of Financial Studies*, 29, 863–897.

SIMS, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

SONG, A., T. SEVERINI, AND R. ALLADA (2017): "From the Home Team to Home Runs: How Jet Lag Impacts Major League Baseball," in *SLEEP*.

SWARTZ, M. (2017): "The Recent History of Free Agent Pricing," https://blogs.fangraphs.com/the-recent-history-of-free-agent-pricing/.

TAUBINSKY, D. AND A. REES-JONES (2018): "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Review of Economic Studies*, 85, 2462–2496.

VAN NIEUWERBURGH, S. AND L. VELDKAMP (2009): "Information Immobility and the Home Bias Puzzle," *Journal of Finance*, 64, 1187–1215.

WAGNER, J. (2019): "Justin Verlander: The Astros' Ace and Sleep Guru," *New York Times*.

WALKER, M. AND J. WOODERS (2001): "Minimax Play at Wimbledon," *American Economic Review*, 91, 1521–1538.

WOODERS, J. (2010): "Does Experience Teach? Professionals and Minimax Play in the Lab," *Econometrica*, 78, 1143–1154.

WOODFORD, M. (2008): "Inattention as a Source of Randomized Discrete Adjustment," Tech. rep., Columbia University.

——— (2009): "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 56, S100–S124.

YANG, M. (2015): "Coordination with Flexible Information Acquisition," *Journal of Economic Theory*, 158, 721–738.

A. A Description of Baseball

In this appendix, we present a brief description of baseball so that a reader unfamiliar with the sport can follow the discussion in the paper. A pitcher pitches a ball approximately 3 inches in diameter towards a batter, who stands approximately 60 feet away at *home plate*. The average pitch velocity is around 80–90 miles per hour, so the baseball takes only about half a second to reach the batter.[77] Pitches can be one of a few "types," which differ in target location, speed, and spin. This paper focuses on the distinction between a fastball, which is an especially fast pitch with some upward spin and is the most common type of pitch, and the rest (which generally tend to be more curved). Almost all pitchers mix between throwing fastballs and some subset of other pitches, and it is usually straightforward to distinguish a fastball from other types of pitches.[78] Given the especially short time horizon, most batters must prepare themselves for a particular pitch before the pitch is released.

In general, a pitcher stays throughout the majority of a game, and substitutions may take place near later parts of a game. An *at-bat* corresponds to a "game" between a pitcher and a single batter. If the batter does not swing after the pitcher pitches, one of two outcomes are possible. If the pitch is thrown relatively centrally, in an area know as the strike zone, a *strike* is called. (A swing that misses is automatically a strike as well.) If the pitch outside this zone, a *ball* is called. The *count* is the ordered pair of total balls and strikes in the current at-bat. The at-bat can end if there are four cumulative balls, in which case the batter moves to first base (and runners on base, a concept which will be discussed later, cycle through to the next base if needed). This event is called a *walk*. (If the batter is physically hit by the pitch, then that leads to an automatic walk too, regardless of the count.) The at-bat can also end if there are three cumulative strikes, an event known as a *strikeout*. In this situation, a new batter comes to the plate, and an *out* is recorded.

If the batter swings and hits the ball, he has the opportunity to run to the bases (described below). The defense on the pitching team has a chance to (i) catch the ball before it hits the ground, (ii) catch the ball after it hits the ground and tag the runner while he is running between bases, (iii) or catch the ball after it hits the ground and then throw it to first base before the batter

[77]This is comparable to penalty kicks in soccer, where speeds are around 70–80 mph and the distance is 12 yards.
[78]This is not true of all pitches, especially in terms of telling apart sliders and curveballs.

gets there (a "force-out"). All these events are marked as an out for the batting team. If the batter makes it onto a base and stops running, the batter is *safe* (no out is recorded) and stays on that base while the next batter comes to the plate. All these events will end the at-bat.

During an at-bat, runners—who were previously batters—may already be on different bases. Since the baseball field is shaped like a diamond, there are three bases. (The final corner of the diamond is home plate.) When a ball is in play after a successful hit by the batter, the runners can run to the next open base. The runner on third base can run to home plate and score a *run* for the batting team. As such, the positions of the runners on bases is an important part of the state space: having multiple runners involves the potential for multiple runs scored during an at-bat, and having a runner on third base can be especially dangerous for the pitching team.[79] These runners can also be tagged out or forced out at the next base (if all the bases before them are occupied) while the ball is in play, much like the batter can be tagged out while he is running; this would also contribute to outs.

Note that even though runs and outs (and thus the win probability added) depend on actions of other players—the defense, the runners, etc.—they also depend fundamentally on the outcome of the interaction between the pitcher and the batter. Even if the pitch is such that the batter hits the ball, the quality of the hit determines whether the ball is easy to catch, how long it gives runners to run to the next base, and so on. An especially strong hit can allow the batter to run to all four bases himself and thus score a run before being tagged out. This is called a *home run*.[80]

Once the batting team accumulates three outs, then the teams switch roles: the pitching team becomes the batter team. After both teams have had a chance to bat, the *inning* ends. At the end of nine innings, the team with the most runs wins. If the score is tied at the end of nine innings, innings are added successively until the score is no longer tied at the end. As a result, the score difference (difference in runs) and the inning are also important components of the state space.

[79] It is possible for positions of the batters on base to change within an at-bat due to stealing, but we ignore this in the description. The empirics control for such changes in the state since the state is computed at the pitch level.

[80] While home runs are automatic when the ball leaves the stadium, this is not a necessary condition.