**Online Appendix**
**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 23, 2020

## B. Additional Theoretical Results

In Appendix B.1, we provide formal justifications for the predictions in Section I.B. Appendix B.2 shows that Predictions (A)–(D) hold if we model rational inattention as coming from a constraint on information processing rather than a cost. Appendix B.3 considers these predictions with alternate cost functions. In Appendix B.4 we provide examples that scaled payoffs need not be monotone in leverage, which provides more intuition about Prediction (C).

### B.1. Proofs

*Proof of Prediction (A).* Prediction (A) states that if $w_s > w_t$ for states $s$ and $t$, then $p_s^f > p_t^f$. This is a direct consequence of (4) with the observation that $\mathcal{L}(p) \equiv \log(p) - \log(1-p)$ is increasing in $p$. $\square$

*Proof of Prediction (B).* Taking differences of (4) for states $s$ and $t$ yields

$$\left(\log p_s^f - \log\left(1 - p_s^f\right)\right) - \left(\log p_t^f - \log\left(1 - p_t^f\right)\right) = \frac{w_s - w_t}{\lambda},$$

so decreasing $\lambda$ increases the difference $\mathcal{L}\left(p_s^f\right) - \mathcal{L}\left(p_t^f\right)$. $\square$

*Proof of Prediction (C).* The equilibrium in state $t$ is defined by

$$p_t^f = \frac{p_0^f \exp(w_t/\lambda)}{p_0^f \exp(w_t/\lambda) + 1 - p_0^f} \tag{B.1}$$

$$w_t = q_t^f L[\tilde{w}(f,f) - \tilde{w}(n,f)] + (1 - q_t^f) L[\tilde{w}(f,n) - \tilde{w}(n,n)] \tag{B.2}$$

CONTACT. vivek.bhattacharya@northwestern.edu and glhoward@illinois.edu.

$$q_t^f = \frac{q_0^f \exp(v_t/\lambda)}{q_0^f \exp(v_t/\lambda) + 1 - q_0^f} \tag{B.3}$$

$$v_t = p_t^f L[-\tilde{w}(f,f) + \tilde{w}(f,n)] + (1 - p_t^f)L[-\tilde{w}(n,f) + \tilde{w}(n,n)], \tag{B.4}$$

where $p_0^f$ and $q_0^f$ are average probabilities of $f$ for the pitcher and the batter, respectively. Define $p^*$ and $q^*$ to be the Nash strategies with payoffs $\tilde{w}(\cdot,\cdot)$. By the payoff-equivalence definition of Nash equilibrium,

$$0 = q^*[\tilde{w}(f,f) - \tilde{w}(n,f)] + (1-q^*)[\tilde{w}(f,n) - \tilde{w}(n,n)] \tag{B.5}$$

$$0 = p^*[-\tilde{w}(f,f) + \tilde{w}(f,n)] + (1-p^*)[-\tilde{w}(n,f) + \tilde{w}(n,n)] \tag{B.6}$$

For notational convenience, define $\tilde{w} \equiv \tilde{w}(f,f) + \tilde{w}(n,n) - \tilde{w}(f,n) - \tilde{w}(n,f)$. Then, we can simplify equations (B.2), (B.4), (B.5), (B.6) to

$$w_t = L(q_t^f - q^*)\tilde{w} \tag{B.7}$$

$$v_t = -L(p_t^f - p^*)\tilde{w} \tag{B.8}$$

The equilibrium is described by equations (B.1), (B.3), (B.7), and (B.8). It exists and is unique. This is easy to see because plugging (B.8) into (B.3) into (B.7) into (B.1) gives one equation to solve for $p_t^f$. They must intersect because both are continuous and at $p_t^f = 0$, the left hand side is smaller and at $p_t^f = 1$, the right-hand side is smaller, because the right-hand side is always strictly between 0 and 1. The left-hand side is increasing in $p_t^f$, and the right-hand side is decreasing, so the intersection is unique.

We want to prove that as $L$ approaches $\infty$, $p_t^f$ converges to $p^*$. That means for any $\epsilon$, there exists a $\bar{L}$ such that $L > \bar{L}$ implies that $|p_t - p^*| \leq \epsilon$. We pick a specific $\bar{L}$,

$$\bar{L} = \max\left\{ \frac{\lambda}{\phi\tilde{w}} \log\left( \frac{q^*}{1-q^*} \frac{1-q_0^f}{q_0^f} \right), \frac{\lambda}{\epsilon\tilde{w}} \log\left( \frac{p^*+\phi}{1-p^*-\phi} \frac{1-p_0^f}{p_0^f} \right), \frac{\lambda}{\epsilon\tilde{w}} \log\left( \frac{p^*-\phi}{1-p^*+\phi} \frac{1-p_0^f}{p_0^f} \right) \right\}, \tag{B.9}$$

where $\phi$ is a fixed number that is less than $p^*$ and $1-p^*$. Our assumptions on $\tilde{w}$ guarantee that $p^*$ is bounded away from zero, so such a $\phi$ does exist.

From here, we proceed by contradiction. First, suppose that $q_t^f > q^* + \epsilon$. Then $w_t > L\epsilon\tilde{w}$. From (B.1) and $L > \frac{\lambda}{\epsilon\tilde{w}} \log\left( \frac{p^*+\phi}{1-p^*-\phi} \frac{1-p_0^f}{p_0^f} \right)$, we have us that $p_t^f > p^* + \phi$. That implies that $v < -L\phi\tilde{w}$. With that condition, $L > \frac{\lambda}{\phi\tilde{w}} \log\left( \frac{q^*}{1-q^*} \frac{1-q_0^f}{q_0^f} \right)$ guarantees that $q_t^f < q^*$. This is a contradiction with our supposition that $q_t^f > q^* + \epsilon$. So it must be that $q_t^f \leq q^* + \epsilon$. Second, suppose that $q_t^f < q^* - \epsilon$. Then $w_t < -L\epsilon\tilde{w}$. Equation (B.1) and $L > \frac{\lambda}{\epsilon\tilde{w}} \log\left(\frac{p^*-\phi}{1-p^*+\phi}\frac{1-p_0^f}{p_0^f}\right)$ imply that $p_t^f < p_t^* - \phi$. That implies $v_t > L\phi\tilde{w}$. That condition and the fact that $L > \frac{\lambda}{\phi\tilde{w}} \log\left(\frac{q^*}{1-q^*}\frac{1-q_0^f}{q_0^f}\right)$ means that $q_t^f > q^*$. This is a contradiction with our supposition that $q_t^f < q^* - \epsilon$. So it must be that $q_t^f \geq q^* - \epsilon$.

Together, this means that $|q_t^f - q^*| < \epsilon$. Therefore, $\lim_{L\to\infty} q_t^f = q^*$. The proof for $p_t^f$ is analogous. $\square$

*Proof of Prediction (D).* If $p_0^f \in (0, 1)$ then (4) must hold for all states. Rearrange, take exponents, and take the expectation over states to arrive at

$$\begin{aligned}
\mathbb{E}\left[\exp(w_s/\lambda)\right] &= \mathbb{E}\left[\exp\left(\mathcal{L}(p_s^f) - \mathcal{L}(p_0^f)\right)\right] \\
&= \mathbb{E}\left[\frac{p_s^f}{1-p_s^f}\cdot\frac{1-p_0^f}{p_0^f}\right] = \mathbb{E}\left[\frac{p_s^f}{1-p_s^f}\right]\cdot\frac{1-p_0^f}{p_0^f} \\
&\geq \frac{\mathbb{E}p_s^f}{1-\mathbb{E}p_s^f}\cdot\frac{1-p_0^f}{p_0^f} = 1,
\end{aligned}$$

where the inequality follows from the convexity of $x/(1-x)$, and the final equality follows from $p_0^f \equiv \mathbb{E}p_s^f$. As long as it is not the case that $p_s^f = p_0^f$ for all states $s$, the inequality is strict. To prove that $\mathbb{E}\exp(-w_s/\lambda) > 1$, start from

$$\mathbb{E}\left[\exp(-w_s/\lambda)\right] = \mathbb{E}\left[\exp\left(\mathcal{L}(p_0^f) - \mathcal{L}(p_s^f)\right)\right]$$

and repeat the steps, noting that $(1-x)/x$ is also convex.[1] $\square$

### B.2. Rational Inattention with an Information Processing Capacity Constraint

In this paper, we model rationally inattentive agents as having an effectively utility function that involves a cost of processing information parameterized by $\lambda$, following Matějka and McKay (2015). Another approach would be to model rationally inattentive agents as possessing a maximum information processing capacity $\kappa$. In this model, an agent would solve

$$\begin{aligned}
&\max_{p_s^f} \mathbb{E}_s\left[p_s^f w_s^*\right] \\
&\text{s.t. } \mathbb{E}_s\left[p_s^f \log p_s^f + \left(1-p_s^f\right)\cdot\log\left(1-p_s^f\right)\right] - p_0^f \log p_0^f - \left(1-p_0^f\right)\log\left(1-p_0^f\right) \leq \kappa,
\end{aligned} \tag{B.10}$$

[1]One can also prove this statement directly starting from (5).

where, like in the body, $p_0^f \equiv \mathbb{E}_s p_s^f$ and $w_s^*$ is the equilibrium relative payoff of a fastball in state $s$ given the opponent's payoffs. Of course, if we let $\lambda$ be the multiplier on the information constraint, we can write the optimization problem as

$$\max_{p_s^f} \mathbb{E}_s\left[p_s^f w_s^*\right] + \lambda\left\{\kappa - \left[\mathbb{E}_s\left[p_s^f \log p_s^f + \left(1-p_s^f\right)\cdot\log\left(1-p_s^f\right)\right]\right.\right.$$
$$\left.\left. - p_0^f \log p_0^f - \left(1-p_0^f\right)\log\left(1-p_0^f\right)\right]\right\}, \quad \text{(B.11)}$$

which is the same problem as the one from combining (2) and (3). The first-order condition of (B.11) is

$$w_s^* = \lambda^* \cdot \left[\log p_s^f - \log\left(1-p_s^f\right) - \left(\log p_0^f - \log\left(1-p_0^f\right)\right)\right], \quad \text{(B.12)}$$

which coincides with (4), with the Lagrange multiplier $\lambda^* \geq 0$ at the optimum corresponding to the cost of attention in the baseline formulation. From (B.12), it is immediate that Prediction (A) holds. The proof of Prediction (C) in Appendix B.1 goes through identically, with $\lambda$ replaced by the equilibrium $\lambda^*$ for each agent. Averaging (B.12) over all states shows that the same proof for Prediction (D) applies as well.

The analogue of Prediction (B) is that agents larger information processing capacities $\kappa$ are more sensitive to payoff differences. Thus, it suffices to show that $\lambda^*$ is decreasing in $\kappa$, fixing $w_s^*$. We assume the constraint is binding (since otherwise $\lambda^* = 0$ and the problem is trivial). We will show that if $\lambda$ were the primitive, then the attention paid by the agent $\kappa(\lambda)$ would be decreasing in $\lambda$. Since $(\kappa, \lambda, p_0^f, p_s^f)$ solve the system of equations given by the constraint and the first-order conditions, this will show that the equilibrium multiplier $\lambda^*(\kappa)$ is decreasing in $\kappa$.

Rearranging the first-order condition (4), we have

$$p_s^{f*} = \frac{p_0^{f*}\exp(w_s^*/\lambda^*)}{1-p_0^{f*}+p_0^{f*}\exp(w_s^*/\lambda^*)}. \quad \text{(B.13)}$$

Substituting B.13 into the constraint (B.10) and rearranging, we have

$$\mathbb{E}_s\left[\frac{p_0^{f*}\exp(w_s^*/\lambda^*)}{1-p_0^{f*}+p_0^{f*}\exp(w_s^*/\lambda^*)}\cdot\frac{w_s^*}{\lambda^*} - \log\left(1-p_0^{f*}+p_0^{f*}\exp(w_s^*/\lambda^*)\right)\right] = \kappa^*, \quad \text{(B.14)}$$

where

$$p_0^{f*} = \arg\max_p \mathbb{E}_s\left[\log\left(p\exp(w_s^*/\lambda^*) + 1 - p\right)\right]. \quad \text{(B.15)}$$

The asterisks represent that the first-order conditions and value of the constraint hold at the equilibrium $(\kappa^*, \lambda^*, p_s^{f*}, p_0^{f*})$, but henceforth we simply view then as solutions to the system consisting of (B.14), (B.13), and the first-order condition associated with (B.15). Dropping the asterisks for simplicity, we can write

$$\frac{d\kappa}{d(1/\lambda)} = \frac{\partial\kappa}{\partial(1/\lambda)} + \frac{\partial\kappa}{\partial p_0^f}\frac{dp_0^f}{d(1/\lambda)}. \tag{B.16}$$

Partially differentiating (B.13), we have

$$\frac{\partial\kappa}{\partial(1/\lambda)} = \frac{1}{\lambda}\mathbb{E}_s\left[p_s^f\left(1-p_s^f\right)(w_s^*)^2\right] \geq 0. \tag{B.17}$$

We can also partially differentiate (B.13) with respect to $p_0^f$ and substitute in the first-order condition from (B.15), which is part of the system we are solving, to arrive at

$$\frac{\partial\kappa}{\partial p_0^f} = \frac{1}{\lambda}\frac{1}{p_0^f\left(1-p_0^f\right)}\mathbb{E}_s\left[p_s^f\left(1-p_s^f\right)w_s^*\right]. \tag{B.18}$$

The cross-partial with respect to $p_0^f$ and $1/\lambda$ in the maximand of (B.15) is

$$\frac{1}{p_0^f\left(1-p_0^f\right)}\mathbb{E}_s\left[p_s^f\left(1-p_s^f\right)w_s^*\right],$$

which means it shares the same sign as (B.18). Given standard results in monotone comparative statics, $dp_0^f/d(1/\lambda)$ has the same sign as the cross partial, meaning it has the same sign as $\partial\kappa/\partial p_0^f$. Thus, (B.17) establishes that the first term of (B.16) is positive. We established that the two components of the second term share a sign, so the product is positive. This means that $d\kappa/d(1/\lambda) \geq 0$. This, if $\lambda$ increases, $\kappa$ decreases, and vice versa. This shows that the analogous prediction to (B) holds in this case as well.[2]

Overall, the predictions in the body of the paper would hold even if we modeled rational inattention as players being subject to a capacity constraint on how much information they can process.

### B.3. Alternate Models of Information Acquisition

While this paper is concerned primarily with the common assumption of Shannon entropy as the cost of information acquisition, it is instructive to ask whether the predictions in this paper are specific to Shannon entropy. In this appendix, we consider two generalizations of Shannon

[2] Note that our proof is conditional on $w_s^*$. This is because our empirical test treats $w_s^*$ as observable, as we can estimate it directly from the data. If $\kappa$ changes for both parties, then in equilibrium $w_s^*$ would change as well, but the first-order conditions conditional on $w_s^*$, which we use for the prediction and for this argument, would still hold.

entropy—generalized entropies and posterior-separable cost functions—and show that while many of the predictions carry through for generalized entropies, the predictions are not true for all posterior-separable cost functions. Our goal in this appendix is not to identify more general predictions or characterize the set of cost functions consistent with each prediction but rather to note that these tests should not be viewed as predictions of general models of rational inattention. In particular, some of these predictions can be viewed as one-sided tests for particular cost functions.

We first consider a generalization of Shannon entropy: Fosgerau et al. (2020) introduce a class of cost functions from the *Bregman information* associated with *generalized entropies*.[3] This cost function takes the form

$$c_{\mathbf{S}}(\mathbf{p}(\cdot);\lambda) = -\lambda \cdot \mathbf{p}_0^f \cdot \log \mathbf{S}(\mathbf{p}_0^f) + \lambda \cdot \mathbb{E}_s \left[\mathbf{p}_s^f \cdot \log \mathbf{S}(\mathbf{p}_s^f)\right], \tag{B.19}$$

where $\mathbf{p}_s^f \equiv \left(p_s^f, 1 - p_s^f\right)$ is the probability of the two actions in state $s$ and $\mathbf{S}(\cdot)$ is a vector-valued function on $\mathbb{R}^2$ (the number of actions per state) called a *scaled inverse demand function*. With this cost function, Proposition 6(ii) of Fosgerau et al. (2020) shows that the probability of each action is given by

$$\mathbf{p}_s^f = \frac{\mathbf{T}\left(\exp\left(\mathbf{w}_s + \lambda \cdot \log \mathbf{S}(\mathbf{p}_0^f)\right)\right)}{\sum_{j=1}^2 T_j \left(\exp\left(\mathbf{w}_s + \lambda \cdot \log \mathbf{S}(\mathbf{p}_0^f)\right)\right)}, \tag{B.20}$$

where $\mathbf{T} = \mathbf{S}^{-1}$ and $\mathbf{w}_s \equiv (w_s^f, w_s^n)$.[4] Rearranging (B.20), applying homogeneity, and subtracting the equation for fastballs in state $s$ from the equation for non-fastballs in the same state $s$, we arrive at

$$\begin{aligned} w_s/\lambda = {} & \log S_1\left(p_s^f, 1 - p_s^f\right) - \log S_2\left(p_s^f, 1 - p_s^f\right) \\ & - \left[\log S_1\left(p_0^f, 1 - p_0^f\right) - \log S_2\left(p_0^f, 1 - p_0^f\right)\right]. \end{aligned} \tag{B.21}$$

Since each component of $\mathbf{S}$ is increasing its corresponding argument, $\log S_1\left(p_s^f, 1 - p_s^f\right) - \log S_2\left(p_s^f, 1 - p_s^f\right)$ is increasing in $p_s^f$. Thus, (B.21) shows that if $w_s > w_{s'}$ for states $s$ and $s'$, then $p_s^f > p_{s'}^f$, which is Prediction (A). We also have an analogue of Prediction (B), with $\mathcal{L}(p)$

[3]Fosgerau et al. (2020) show an equivalence relationship between rational inattention with generalized entropies and discrete choice models: in this sense, this is the natural generalization of the intuition from Matějka and McKay (2015), which forms the basis of the predictions in the body of the paper.

[4]In Fosgerau et al. (2020), this results is only stated for $\lambda = 1$. However, is it easy to check that scaling the cost function (B.19) by $\lambda$ is equivalent to transforming $\mathbf{S}(\cdot) \mapsto \mathbf{S}(\cdot)^\lambda$.

replaced by $\mathcal{L}_{\mathbf{S}}(p) \equiv \log S_1(p, 1-p) - \log S_2(p, 1-p)$. Note that the general intuition that probabilities of fastballs are less sensitive to payoff differences as $\lambda$ increases still holds: as $\lambda$ increases, the left-hand side of (B.21) moves to zero. Since $\mathcal{L}_{\mathbf{S}}(p)$ is monotone is its argument, it must be that $p_s^f$ moves to $p_0^f$ for all states, so the probability of a fastball in each state moves closer to the average probability of a fastball.

Prediction (C) also continues to hold. Intuitively, the attention costs still become relatively small as leverage approaches infinity, leading to Nash strategies. To modify the proof, simply replace the value for $\bar{L}$ in (B.9) with

$$\begin{aligned}\bar{L} = \max \Bigg\{ & \frac{\lambda}{\phi \tilde{w}} \log \left( \frac{S_1(q^*, 1-q^*)}{S_2(q^*, 1-q^*)} \frac{S_2(q_0^f, 1-q_0^f)}{S_1(q_0^f, 1-q_0^f)} \right), \\ & \frac{\lambda}{\epsilon \tilde{w}} \log \left( \frac{S_1(p^*+\phi, 1-p^*-\phi)}{S_2(p^*+\phi, 1-p^*-\phi)} \frac{S_2(p_0^f, 1-p_0^f)}{S_1(p_0^f, 1-p_0^f)} \right), \\ & \frac{\lambda}{\epsilon \tilde{w}} \log \left( \frac{S_1(p^*-\phi, 1-p^*+\phi)}{S_2(p^*-\phi, 1-p^*+\phi)} \frac{S_2(p_0^f, 1-p_0^f)}{S_1(p_0^f, 1-p_0^f)} \right) \Bigg\}, \end{aligned} \tag{B.22}$$

The one subtlety to note is that we know $S_1(p, 1-p)/S_2(p, 1-p)$ is increasing in $p$ because $S_1(p, 1-p)$ is increasing in $p$ and $S_2(p, 1-p)$ is increasing in $1-p$. Otherwise, the proof is identical.

Finally, we can write (B.21) as

$$\exp\left(\frac{w_s}{\lambda}\right) = \frac{S_1\left(p_s^f, 1-p_s^f\right)}{S_2\left(p_s^f, 1-p_s^f\right)} \cdot \frac{S_2\left(p_0^f, 1-p_0^f\right)}{S_1\left(p_0^f, 1-p_0^f\right)}.$$

A sufficient condition for Prediction (D) to hold as stated is for $S_1(p, 1-p)/S_2(p, 1-p)$ and its inverse to be convex in $p$, in which case the argument from Appendix B.1 applies directly. While we are not aware of counterexamples, we have not been able to establish that the properties of an inverse scaled demand function $\mathbf{S}(\cdot)$ are sufficient for convexity of this function. Of course, note that a weaker version of Prediction (D) must hold: it cannot be the case that $w_s > 0$ for all $s$ if we observe mixing, since then the pitcher would simply play $p_s^f = 1$ and not acquire information; how disperse $w_s$ can be across states is a function of the specific cost function.

We then move to the even more general class of *posterior-separable* cost functions (Caplin et al., 2019; Denti, 2020a), where we show that fewer predictions must hold. If $\pi_s$ is the prior probability of state $s$, Lemma 4 of Denti (2020a) notes that we can write the objective function as

$$\mathbb{E}_s\left[p_s^f w_s\right]-\lambda\cdot\left[p_0^f\cdot\psi\left(\frac{p_1^f\pi_1}{p_0^f},\cdots,\frac{p_S^f\pi_S}{p_0}\right)+(1-p_0)\cdot\psi\left(\frac{(1-p_1^f)\pi_1}{1-p_0^f},\cdots,\frac{(1-p_S^f)\pi_S}{1-p_0^f}\right)\right] \tag{B.23}$$

if there are $S$ states and $\psi:\Delta^S\to\mathbb{R}$ is convex. Differentiating (B.23) with respect to $p_s$ shows that in general, the probability of each state enters the first-order condition. As an illustration, in the simpler case where

$$\psi\left(x_1,\cdots,x_S\right)=\sum\tilde{\psi}\left(x_s\right),$$

with $\tilde{\psi}(\cdot)$ convex, we would have the first-order condition

$$\frac{w_s}{\lambda}=\tilde{\psi}'\left(\frac{p_s^f\pi_s}{p_0^f}\right)-\tilde{\psi}'\left(\frac{(1-p_s^f)\pi_s}{1-p_0^f}\right). \tag{B.24}$$

Inspecting (B.24), it is easy to see that since $\pi_s$ is still in the expression, it need not be the case for general $\tilde{\psi}(\cdot)$ that if $w_s>w_t$ then $p_s^f>p_t^f$. Thus, Prediction (A) need not hold, as the action in a particular state depends on the payoff as well as the ex ante probability of that state. Moreover, no prediction directly analogous to (B) holds either, again because $\pi_s$ enters into the first-order condition (B.24) directly. However, (B.24) still shows that as $\lambda\to\infty$, $p_s^f\to p_0^f$: if attention becomes especially costly, then the probability of a fastball in each state converges to the average probability. While we conjecture that a result analogous to (C) would exist in this case—players would have an incentive to play close to $p_s^f\in\{0,1\}$ for states with sufficiently large payoff difference as long as they are not especially uncommon—the statement in the text does not apply exactly.[5] Similarly, there is no exact analogue of Prediction (D); however, as mentioned above in the context of generalized entropies, it cannot be that $w_s$ is uniquely signed for all $s$, since if so then $p_s^f$ would either be 0 or 1.

Finally, note that our estimates of attention costs and the evaluation of the fit of the model in Sections V.B to V.E of the paper use the specific functional form of Shannon entropy. The fact that payoffs and strategies have a reasonably linear relationship (Figure 4) provides evidence in favor of this functional form for the cost function. Overall, the discussion above suggests aside from the specific exercise of fitting the model, the predictions do generalize at least partially to cost functions other than Shannon entropy.

Another assumption we maintain is that information acquisition is uncorrelated across players. In a coordination game, Denti (2020b) notes that agents would prefer information acquisition to not be independent so that actions will inherit some of the correlation from the information.

[5] This is in part due to the fact that the formulation we follow for posterior-separable information costs from Denti (2020a) requires finitely many states, and the statement of (C) references a sequence of states. We have not pursued a formal treatment of analogues to (C) in this more general case.

While a formal analysis is beyond the scope of this paper, we conjecture that this is less of a concern in a game of strategic substitution. In our setting, imagine the batter and pitcher acquired correlated information, leading to positively correlated actions. This benefits the batter, who is trying to match the pitcher. Given our Shannon information setup, though, the pitcher could acquire uncorrelated information that led to uncorrelated actions for the same processing cost. He would do so, because he can get higher payoffs for the same cost. Similarly, if the correlated information benefited the pitcher, the batter could choose a different signal for the same cost that was uncorrelated. Therefore, we conjecture the only information structures that could survive in equilibrium will lead to uncorrelated actions, which suggests that that the independence assumption may not have been consequential.[6]

### B.4. Example of Strategies and Payoffs as a Function of Leverage

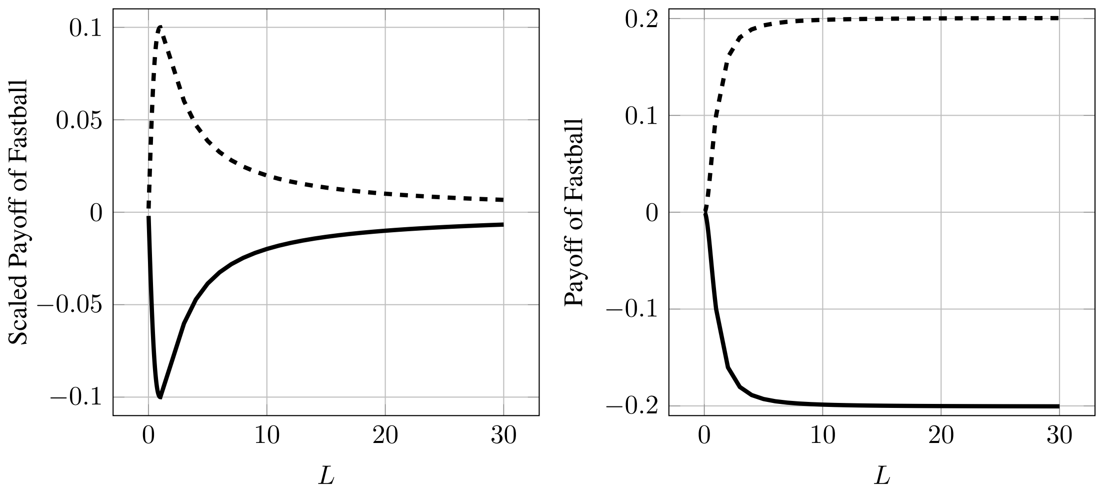
In this section, we solve a numerical example to show that (scaled) payoffs need not be monotonically decreasing in leverage. Consider the following state-dependent matching pennies game between a pitcher and a batter, where the first term indexed the payoffs of the pitcher.

| | | Batter | |
|---|---|---|---|
| | | Fastball | Other |
| Pitcher | Fastball | $-L(1+\epsilon), L(1+\epsilon)$ | $L(1+\epsilon), -L(1+\epsilon)$ |
| | Other | $L(1-\epsilon), -L(1-\epsilon)$ | $-L(1-\epsilon), L(1-\epsilon)$ |

States are indexed by $(\epsilon, L)$, which includes the relative payoff $\epsilon$ of a fastball for the pitcher as well as the leverage $L$, which scales payoffs. The rational inattention equilibrium involves a probability $p(\epsilon, L)$ and $q(\epsilon, L)$ that each agent plays a fastball as a function of the state. Prediction (D) notes that fixing $\epsilon$, as $L \to \infty$, strategies should converge to Nash equilibrium. We test this in the data by noting that payoff differences, scaled down by $L$, should be decreasing to zero—but only for sufficiently high $L$. For low-leverage situations, even the scaled payoff differences could be increase in $L$, as we show in this appendix.

Suppose that $\epsilon$ has mean zero and is symmetric about 0, and the distribution of $L$ is independent of that of $\epsilon$. Then, we claim that the average probabilities of a fastball across all states, $p_0^f$ and $q_0^f$ for the two players, are equal to 1/2. To see this, note that if $(p_0^f, q_0^f)$ were the average probabilities for an equilibrium of this game, then $(1-p_0^f, 1-q_0^f)$ would be the average probabilities of a fastball in the game where $\epsilon$ is replaced by $-\epsilon$ (by independence of $\epsilon$ with $L$). But, given the symmetry of $\epsilon$, this modified game is equivalent to the original one. Thus, $p_0^f = 1 - p_0^f$, so

[6] Of course, in more general setups besides Shannon entropy information costs, we would hesitate to speculate what might happen. For example, if acquiring the same information saved on information processing costs, then it might be optimal to pursue such a strategy.




(a) Normalized Payoff Differences

(b) Unnormalized Payoff Differences

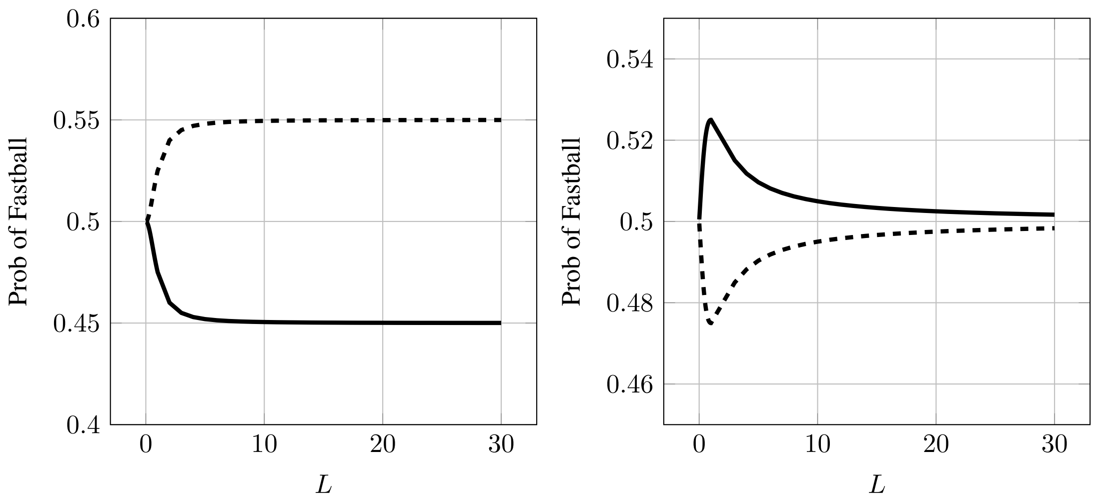
Figure B.1: Payoff differences for pitcher as a function of leverage $L$. (a) does not multiply the payoffs by $L$ and (b) does. The solid line is for $\epsilon = 0.1$ and the dashed one is for $\epsilon = -0.1$.

$p_0^f = 1/2$, and similarly for $q_0^f$.[7]

Fixing the average probabilities $p_0^f$ and $q_0$ of a fastball across all states, we can numerically compute the probabilities in a particular state using the system (B.1)–(B.4). This then lets us compute the both the true payoffs as well as the scaled payoffs (ignoring the $L$ in the payoff matrix) as a function of $L$. Figure B.1 shows the normalized and unnormalized payoffs differences for the pitcher as a function of leverage, for $\epsilon \in \{0.1, -0.1\}$ and $\lambda = 1$. As Panel (a) does not multiply the payoffs by $L$ where as Panel (b) does, we can think of the payoffs as representing FIP and WPA, respectively, thus making this figure the stylized analogue of Figures 5 and C.5. The scaled payoff differences do converge to zero, which is in line with Prediction (C). However, they exhibit a non-monotonicity as they move away from the default and towards the Nash equilibrium ones: for intermediate levels of $L$, payoffs show a departure from payoff equivalence. Moreover, note that the true payoff differences in Panel (b) do not converge towards zero even though the strategies approach Nash equilibrium. In this numerical example, they converge to points away from zero, as the convergence of the strategies and the increase in $L$ cancel out. This is again justification for why we do not look for convergence in WPA when testing Prediction (C), although finding convergence would be consistent with this prediction.

Figure B.2 shows the strategies of the pitcher and the batter. While convergence of strategies to Nash equilibrium is not directly a test we can run, investigating them in the numerical simulation

[7]We leverage symmetry for simplicity here, but it is easy to compute $p_0^f$ and $q_0^f$ numerically for games that do not exhibit the symmetry property.




(a) Pitcher's Probability of Fastball (b) Batter's Probability of Fastball

Figure B.2: Strategies for the (a) pitcher and (b) batter as a function of leverage $L$. The solid line is for $\epsilon = 0.1$ and the dashed one is for $\epsilon = -0.1$.

provides some intuition for the elements of the model. For low values of $L$, the players play the default strategy of mixing 50/50. In the state-by-state game, the Nash equilibrium is for batters to mix 50/50 but for pitchers to play fastballs with probability 0.55 when $\epsilon = -0.1$ and 0.45 when $\epsilon = 0.1$. The results in the figure show that the strategies do indeed converge to Nash as $L$ increases, and the batter's strategy exhibits a non-monotonicity as a function of $L$ (as the default strategy is the same as Nash for the batter). Moreover, comparing Figures B.1 and B.2 shows that the relative sign of the payoff differences match the strategies: pitchers throw more fastballs when the payoff of a fastball is better than average.

We should finally note that nothing in the model forces there to be a non-monotonicity in payoffs as a function of $L$. Indeed, the batter's scaled payoffs (not shown) are monotonically decreasing in magnitude as a function of $L$. In this example, this is because at low values of $L$, the pitcher is playing a strategy that does not correspond to the Nash equilibrium. The batter's true payoffs are non-monotone in $L$: for low $L$ they are mechanically small and for these parameters for high $L$ they converge to 0 as well. This example shows that the robust conclusion of the model—and indeed the one highlighted in Prediction (C)—is that the scaled payoffs should converge to 0 since the strategies would converge to Nash.

| | WPA $\times$ 1000 | OPS$^{-}$ $\times$ 100 | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Fastball | 0.0273 | 0.551 | 1.586 | 4.093 |
| | (0.0849) | (0.200) | (0.127) | (0.104) |
| $N$ | 2,707,048 | 2,707,048 | 5,442,316 | 6,881,063 |
| Controls | Full | Full | Partial | None |

Table C.1: Comparison of baseline result to Kovash and Levitt (2009), who use on-base plus slugging. Note that a lower OPS is better for the pitcher (as indicated by the $^{-}$). Full state controls include the standard list used in our paper. Partial state controls include inning, outs, number of runners, and pitcher-batter-count effects, but not interacted.

## C. Additional Empirical Results

### C.1. Comparison to Kovash and Levitt (2009)

Kovash and Levitt (2009) investigate a similar question to see if baseball players were playing Nash equilibrium and propose a different answer. In particular, they compute the *average* payoff difference across states and find it is different from zero. The two main differences are that (1) they use a measure of outcomes based on on-base-percentage plus slugging percentage (OPS) and that (2) they have fewer controls for the state space than we do. OPS is defined to be 0 for an out, 1 for a walk, 2 for a single, 3 for a double, 4 for a triple, and 5 for a home run. In this section, we run their regressions on our sample to replicate their result, and we then discuss our choice of payoffs.

Table C.1 compares our results to theirs by changing one thing at a time. Column (1) repeats our baseline regression, where fastballs do not have an effect on win probability added. Column (2) makes one change: it switches the outcome measure to OPS. Using OPS suggests that fastballs are worse: a higher OPS is bad for the pitcher. In Columns (3) and (4), we attempt to replicate some of the regressions run in Kovash and Levitt (2009). Our Column (3), labeled "Partial" is intended to replicate their regression with the most controls, where they control for inning, outs, number of runners, and pitcher-by-batter-by-count fixed effects, but not interacted. This regression appears in Column (5) of Table 3 in their paper. They also have indicators for sliders and curveballs, so their coefficient compares fastballs to pitches besides those, making the results slightly hard to compare.[8] Their estimated coefficient is 0.073, with standard error 0.008. This is larger than our replication, but still comparable. In Column (4), we run the regression with no controls, to correspond to Column (1) of their Table 3. Their estimated coefficient is 0.094, with standard error 0.004. We seem to be in the same ballpark, though slightly smaller. Any remaining

[8]On the other hand, their results for sliders and curveballs are fairly close to zero, so the magnitude should be not especially different.

difference is likely due to the fact that we use data from 2008 to 2017, while they use data from 2002 to 2006.

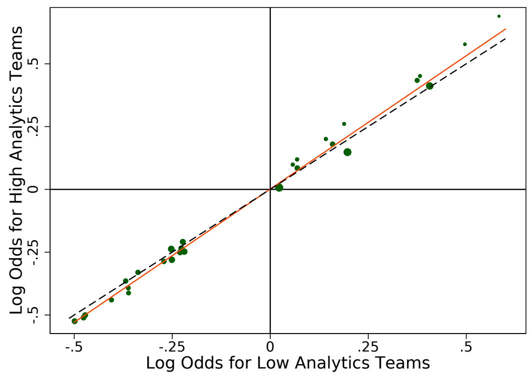
Thus, the difference in sample period is not causing us to draw a different conclusion from Kovash and Levitt (2009). However, there is reason to think that OPS may be misleading. Using OPS, an out is recorded as a 0, a walk is a 1, and a single is 2. However, in most situations, a walk and a single lead to the same subsequent state, differing only when there are runners on base that advance a different amount. Indeed, in our data, the average win probability added of a single is 0.071 worse than an out and the average win probability of a walk is 0.055 worse than an out, considerably less than the two-to-one ratio. We also find in our data that fastballs lead to more singles and fewer walks. Because of the difference in how OPS measures payoffs, it likely overstates the losses of fastballs compared to other pitches. Moreover, since a battery of other outcomes—win probability, FIP, wins, and runs remaining—broadly suggest similar results, and we provided further discussion in Section III.B about other payoffs, we do not include OPS as a main outcome in the paper.

### C.2. Responsiveness of the Pitching Team

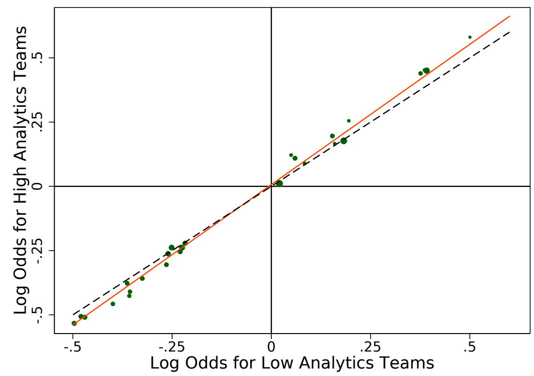
In this appendix, we provide further results related to the calculations in Section IV.B.

We first provide more details about the regressions underlying Table 3. Figure C.1 shows scatterplots of the log odds of a fastball in a particular state (relative the log odds of a fastball on average), with each axis corresponding to a particular group. The horizontal axis plots this quantity for the "baseline" category, which is omitted from the regressions in Table 2, and the vertical axis is the category for the interaction. Panels (a) and (b) use the two measures of analytics, where the vertical axis corresponds to teams with a higher emphasis on analytics (lower attention cost), and in the remaining panels the vertical axis refer to the group of observations with higher attention costs. Each datapoint is a count-out, although the graphs restrict to a subset of the data that has the better-populated states to make the dots more salient. Note that the two components of each data point are computed on entirely separate datasets. Nevertheless, the regression line (solid red) is steeper than the $45°$ line (dashed black) in Panels (a) and (b), and it is shallower in the other panels. This is consistent with Table 3.

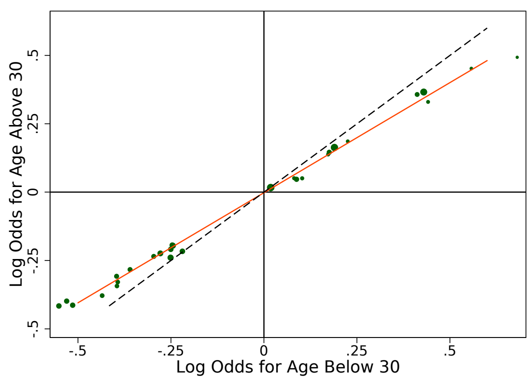
However, the scatter shows that it is largely the case that even *state-by-state*, strategies for the situations with lower attention costs are more responsive. Figure C.1 shows this by noting that in each panel, almost all dots are on the correct side of the 45° line. Figure C.2 makes this even clearer by plotting the difference between the points and the 45° line. In Panels (a) and (b), we would expect the majority of the points to lie in the first and third quadrants, and we would expect the points to lie in the second and fourth in the remaining panels. This seems to be the case. We should further note that these observations about the scatter plot remain robust if we



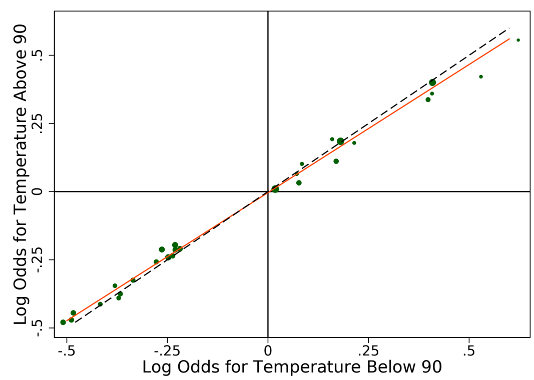

(a) Defensive shifts



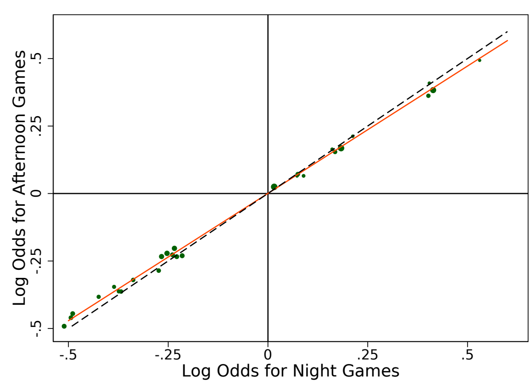

(b) ESPN




(c) Age




(d) Temperature




(e) Game Timing

Figure C.1: Strategies in situations with different attention costs. The horizontal axis plots the relative log odds of a fastball at the count-out level for particular situations: teams with low analytics intensity in (a) and (b), younger pitchers in (c), cooler temperatures in (d), and night games in (e). The vertical axis plots relative log odds in the complementary situations. In Panels (a) and (b), the horizontal axis corresponds to teams with higher attention costs, and in the remaining panels it corresponds to teams with lower attention costs. The dashed line is $45°$ and the solid (orange) line is the line of best fit.

move to the even finer categories of count-out-base position, although interacting with inning further increases sampling error substantially.

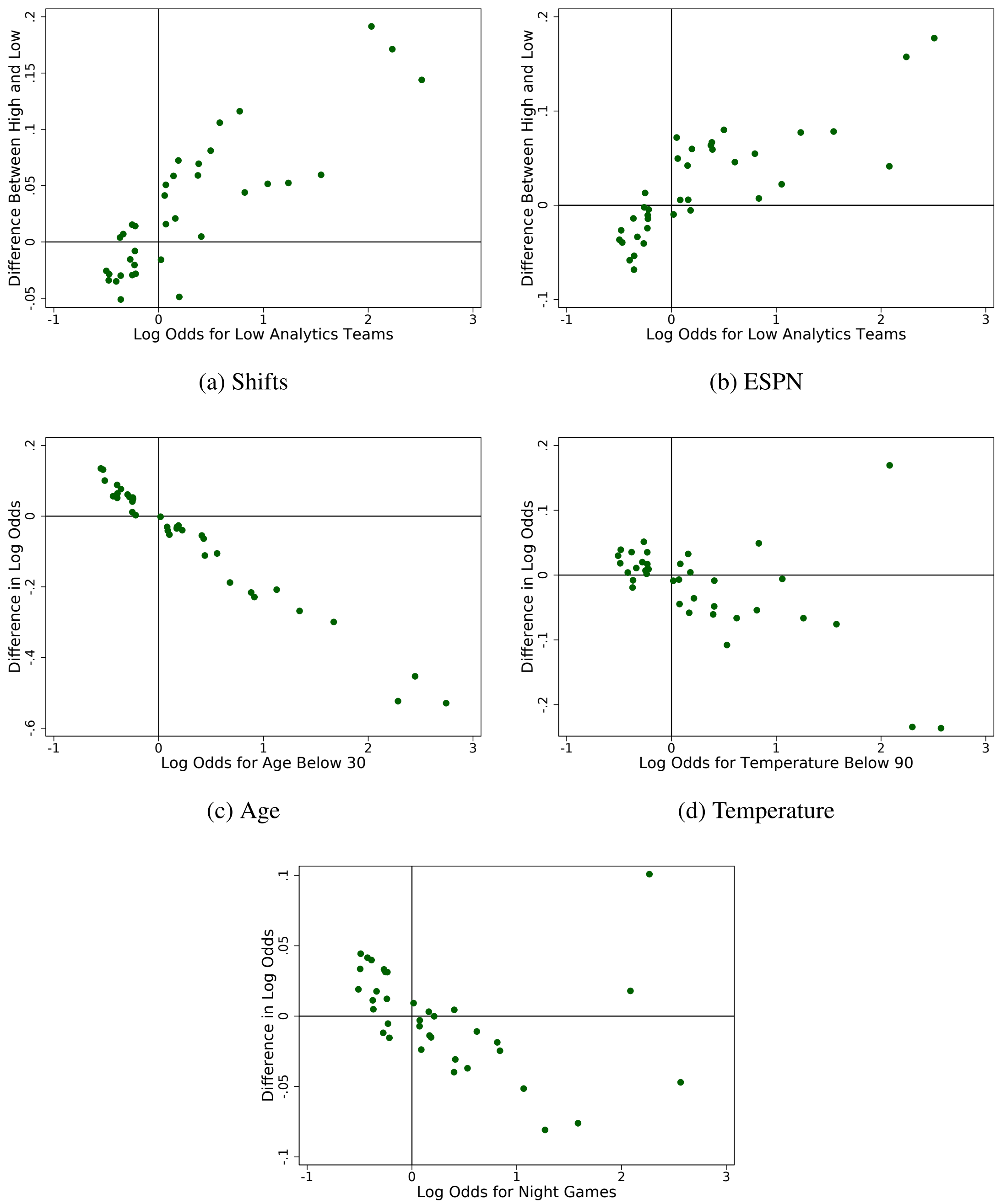
We next investigate a richer specification for the responsiveness of strategies to zero- and two-strike counts in Table 2. In particular, we run the same regression but instead of interacting strikes with a binary classification of age or temperature, we interact with dummies for age or bins of 10 degrees for temperature. The coefficients of the interaction terms are plotted in Figure C.3. Panel (a) shows an almost monotone relationship between responsiveness and age, peaking around age 20 and then declining beyond that. This is consistent with continuous decline in cognition related to information processing with age. Panel (b) shows a monotone effect of temperature on the responsiveness in zero-strike counts; for two-strike counts the effect of temperature on fastball probability is relatively flat for intermediate temperatures but does change for especially high ones. Taken together, Figure C.3 suggests that the results in Table 2 were not an artifact of the binning procedure.

To check the concern that payoff difference across states may differ by groups, we run the analogue of Section IV.A interacted with proxies for attention. We estimate

$$\begin{aligned} Y_p = \sum_{\sigma} \Big\{ & \beta_\sigma \cdot \mathbb{1}[\text{Fastball}]_p \cdot \mathbb{1}[\sigma \text{ strikes in the count}]_p \cdot \text{Attention}_p \\ & + \gamma_\sigma \cdot \mathbb{1}[\text{Fastball}]_p \cdot \mathbb{1}[\sigma \text{ strikes in the count}]_p \Big\} + \text{State FEs} + \epsilon_p, \quad \text{(C.1)} \end{aligned}$$

where $Y_p$ is the win probability added of a pitch. Results from this analysis are in Table C.2. They largely do not seem to indicate differences in payoffs that would change the conclusions from this section. The point estimates in Column (2) suggest that high analytics teams face *lower* payoffs from fastballs in 0-strike counts and higher in 2-strike counts, which would suggest that such teams are even more responsive than the previous results would imply by themselves. Column (3) suggests that older pitchers earn higher payoffs in both 0- and 2-strike counts—perhaps due to increased experience—although the net difference in payoffs between the two counts may in fact be slightly lower. Column (4) suggests that on hot days, payoffs to a fastball in 0-strike counts are higher and those to 2-strike counts are lower, which would exacerbate the limited responsiveness documented above. It is true that the point estimates in Column (5) suggest more muted payoff differences in afternoon games, and the estimates in Column (1) suggest that teams with high analytics (as measured through shifts) actually face slightly starker payoff differences. However, the point estimates in Column (1) are very small, and all point estimates are especially noisy. Thus, while we cannot definitively rule out differences in payoffs that would counteract the responsiveness documented above, the general pattern in Table C.2 makes this seem like a less likely explanation.

An alternate approach would be to estimate $\lambda$ following the procedure in Section V separately




(a) Shifts

(b) ESPN

(c) Age

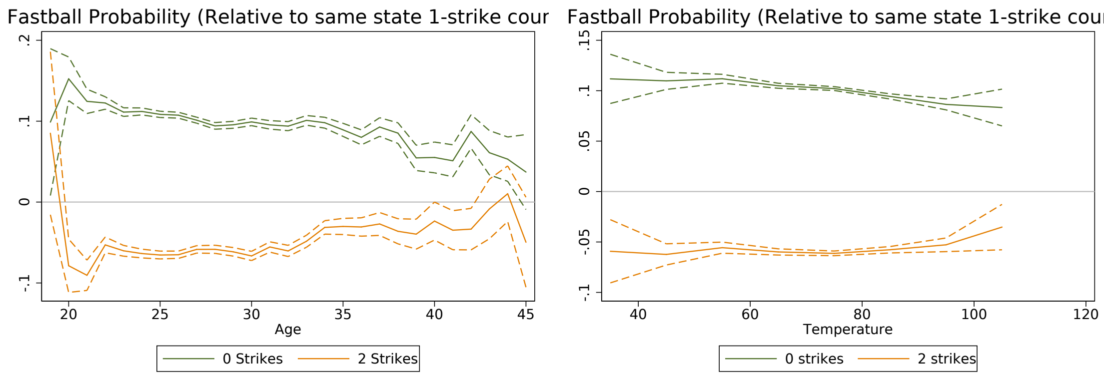
(d) Temperature

(e) Game Timing

Figure C.2: Strategies in situations with different attention costs. The plotted data is the same as Figure C.1 but the vertical axis in this figure is the difference between the vertical axis and the 45° line.

| | Shifts (1) | ESPN (2) | Age (3) | Temp (4) | Afternoon (5) |
|---|---|---|---|---|---|
| Fastball, 0 strikes | 0.535 | 0.591 | 0.439 | 0.465 | 0.574 |
| | (0.170) | (0.165) | (0.168) | (0.139) | (0.155) |
| Fastball, 1 strike | -0.002 | -0.124 | 0.018 | -0.081 | -0.003 |
| | (0.171) | (0.167) | (0.170) | (0.141) | (0.158) |
| Fastball, 2 strikes | -0.766 | -0.855 | -0.939 | -0.905 | -0.939 |
| | (0.191) | (0.185) | (0.190) | (0.155) | (0.177) |
| Fastball, 0 strikes | 0.042 | -0.158 | | | |
| $\times$ High Analytics | (0.297) | (0.309) | | | |
| Fastball, 1 strikes | -0.501 | -0.310 | | | |
| $\times$ High Analytics | (0.301) | (0.310) | | | |
| Fastball, 2 strikes | -0.024 | 0.222 | | | |
| $\times$ High Analytics | (0.337) | (0.348) | | | |
| Fastball, 0 strikes | | | 0.242 | | |
| $\times$ (Age $\geq$ 30) | | | (0.308) | | |
| Fastball, 1 strike | | | -0.265 | | |
| $\times$ (Age $\geq$ 30) | | | (0.312) | | |
| Fastball, 2 strikes | | | 0.429 | | |
| $\times$ (Age $\geq$ 30) | | | (0.354) | | |
| Fastball, 0 strikes | | | | 0.353 | |
| $\times$ (Temperature $\geq$ 90) | | | | (1.02) | |
| Fastball, 1 strike | | | | -2.74 | |
| $\times$ (Temperature $\geq$ 90) | | | | (1.27) | |
| Fastball, 2 strikes | | | | -0.460 | |
| $\times$ (Temperature $\geq$ 90) | | | | (1.30) | |
| Fastball, 0 strikes | | | | | -0.291 |
| $\times$ Afternoon | | | | | (0.451) |
| Fastball, 1 strike | | | | | 0.370 |
| $\times$ Afternoon | | | | | (0.478) |
| Fastball, 2 strikes | | | | | 0.449 |
| $\times$ Afternoon | | | | | (0.542) |
| $N$ | 2,479,402 | 2,492,807 | 2,443,911 | 2,536,315 | 2,266,237 |

Table C.2: Win probability added of a fastball, relative to a one-strike count. The specification in given in (C.1). Details of Table 2 apply.




(a) Age (b) Temperature

Figure C.3: Interaction coefficients for a regression of (8) where Attention$_p$ is replaced by (a) dummies for pitcher age or (b) dummies for bins of temperature. Dotted lines indicate 95% confidence intervals.

by attention cost. This would account for both differences in actions and any potential differences in payoffs. Results of this exercise are reported in Table C.3. Each element in this table is an estimate of $\lambda$ using a particular subset of the data: we drop all datapoints that do not belong in that bin (e.g., all data points for games played when the temperature is not high) and estimate (13). In Columns (1) and (2), the baseline category are teams with low analytics emphasis, so we would expect them to have higher $\lambda$. The baseline is young pitcher for Column (3), games in cool temperatures for Column (4), and night games for Column (5), and in all three would expect $\lambda$ for the baseline to be lower than $\lambda$ for the other category. The point estimates for age and temperature go in this direction, but the point estimate for $\lambda$ for afternoon games is not only smaller than the one for night games but negative. However, note that this estimate is especially noisy. In fact, all estimates are somewhat noisy and while the evidence usually points in the expected direction, we cannot reject that the costs of attention between the two groups are equal in any particular column.

### C.3. Scaling Properties of FIP

Section II notes that one benefit of FIP is that it does not scale with leverage. In fact, we use this property when testing Prediction (C), noting that FIP is approximately proportional to $L$ times win probability added. Here we provide evidence that this is the case.

Figure C.4 shows a binned scatter of the win probability added as a function of leverage—as defined by the standard deviation of the win probability added within inning-score difference combinations—for each of the components of FIP. The salient observation from this figure is that while win probability added scales with leverage, as expected, the relationship is linear within

| | (1) Shifts | (2) ESPN | (3) Age | (4) Temp | (5) Afternoon |
|---|---|---|---|---|---|
| $\lambda$ for Baseline | 1.445 | 1.431 | 1.132 | 1.308 | 1.330 |
| | (0.401) | (0.341) | (0.367) | (0.284) | (0.346) |
| $\lambda$ for Other | 1.195 | 1.080 | 1.536 | 1.927 | -0.130 |
| | (0.386) | (0.434) | (0.478) | (2.203) | (1.113) |

Table C.3: Estimates of $\lambda$ by bins for attention. The "Baseline" category is the excluded category in Table 2 and the "Other" category is the bin used in the interaction. Thus, for Columns (1) and (2) the baseline involves teams with low analytics, Column (3) is young pitchers, Column (4) is cooler games, and Column (5) is night games. Thus, we would expect higher $\lambda$ estimates for the baseline category in Columns (1) and (2) and higher for the other category in the remaining columns.

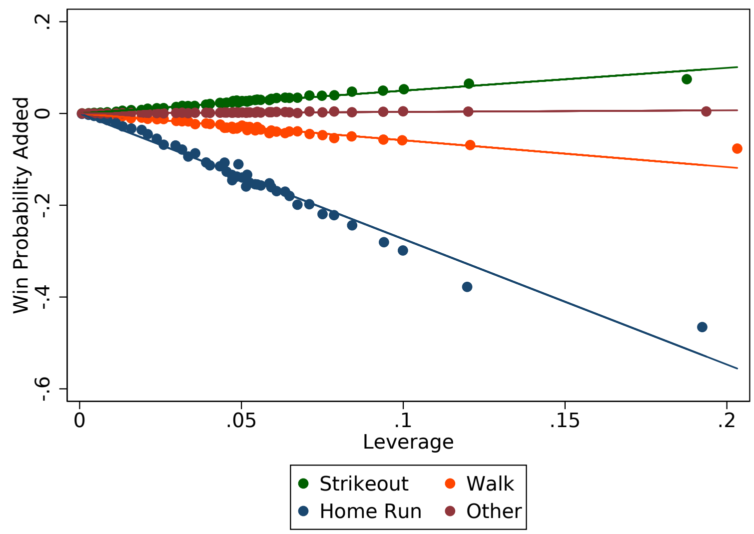
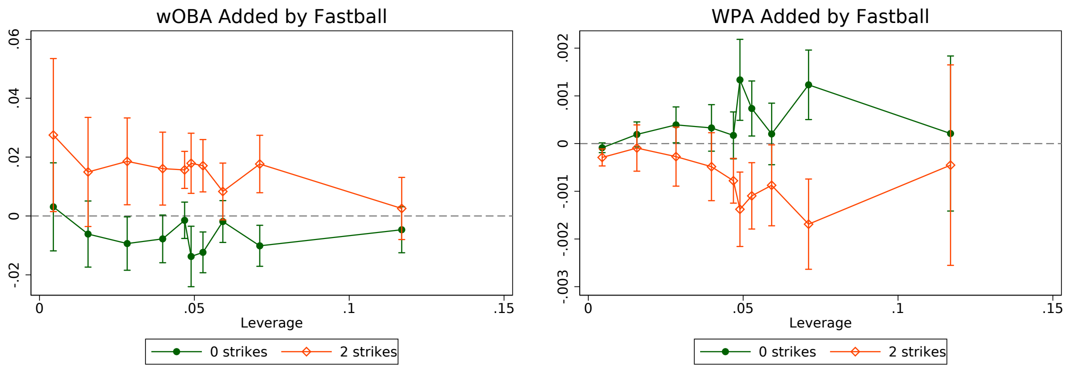



Figure C.4: Win probability added by elements of FIP, against leverage

outcome; e.g., win probability added by a home run is approximately $\alpha_{\text{Home Run}} \cdot L$. Furthermore, as mentioned in Section II, the coefficients in FIP are chosen so that they roughly mirror the effects of these outcomes on win probability added. To see this in the graph, we can compare the slopes of the lines—relative to the line for other—to each other. The ratio of the slope of home runs (minus other) to walks (minus other) is 4.4, whereas FIP uses $13/3 \approx 4.33$. That of home runs (minus other) to strikeouts (minus other) is 5.9, while FIP uses $13/2 = 6.5$. The ratio between walks (minus other) to strikeouts (minus other) is about 1.33 while the ratio used by FIP is 1.5. Putting this together, we can say that win probability added is approximately proportional to leverage times FIP.




(a) wOBA

(b) WPA

Figure C.5: (a) wOBA and (b) WPA by leverage for zero-strike and two-strike counts. Note that for wOBA, lower values are better for the pitcher.

### C.4. Convergence of wOBA and Win Probability Added by Leverage

Figure C.5 shows the results of (10) with (a) wOBA and (b) WPA as the left-hand side variable. Since wOBA does not scale with leverage, we would expect convergence in differences to 0 as leverage increases. We see some (albeit noisy) evidence that this is the case—especially for 2-strike counts. As discussed in Section IV.C, since WPA mechanically scales with leverage, we may not see payoff convergence in WPA as leverage increases even if strategies do converge to Nash. Appendix B.4 shows a numerical example of this lack of convergence even though strategies converge to Nash. Thus, evidence that the payoff differences measured in WPA do not explode as leverage increases can be suggestive evidence in favor of convergence to Nash. Indeed, Figure C.5 shows that while payoff differences do increase as a function of leverage for intermediate values of leverage, there is some evidence that they are not monotone in leverage despite the mechanical effect of leverage. In particular, the point estimates for WPA by a fastball in the highest decile of leverage are 0, although the standard errors are especially large.

### C.5. Approximate Payoff Equivalence for Dynamics

We conduct the analysis of approximate payoff equivalence in Section V.A but uses the cumulative win probability added in the next $N$ innings as the left-hand side variable. This analysis is to check whether payoffs are approximately equivalent on average even if the true payoffs were the ones that incorporated dynamic considerations. We use the weighting scheme discussed in Section V.A to weight by inverse variance and then further upweight states to deal with selection due to dropped observations.

Figure C.6 shows these results. We are unable to reject the the average is 0 for any inning,

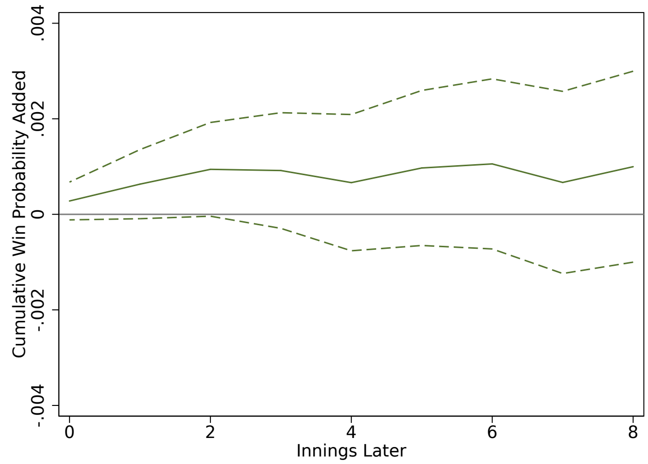



Figure C.6: Coefficient of (11) with cumulative win probability added in the next $N$ innings as the outcome. Weighting is as described in Section V.A. Dotted lines indicated 95% confidence intervals.

and point estimates are reasonably small. Of course, standard errors are larger than they are for other metrics—as can be expected given the variance of the outcome. However, there does not seem to be any systematic evidence that the difference between these dynamic payoffs across pitch dynamics are significantly different from zero.

### C.6. Interpreting the Model Fit

In this section, we interpret the $R^2$ values in Table 7. The exercise we run is in the spirit of a bootstrap: we ask whether we would expect an $R^2$ at this level *purely* due to measurement error if the true data generating process were from rational inattention. In particular, for each column we generate 500 bootstrap samples (indexed by $n$) by drawing

$$\beta_\sigma^n = \hat{\alpha} + \hat{\lambda} \cdot (\mathcal{L}(\hat{p}_\sigma^f) - \mathcal{L}(\hat{p}_0^f)) + \epsilon_\sigma^n,$$

where $\epsilon_\sigma^n$ is normal with standard deviation equal to the standard error of $\hat{\beta}_\sigma$. We then regress $\beta_\sigma^n$ on $\mathcal{L}(\hat{p}_\sigma^f) - \mathcal{L}(\hat{p}_0^f)$ and compute the $R^2$. We finally compute the proportion of such simulations where such an $R^2$ is less than the one observed in the data. Note that any departure from perfect linearity would lead to a low $R^2$. Since our focus throughout the paper is on Shannon entropy, we use this measure of model fit as it directly imposes the parameteric restrictions given by this choice.

This exercise does suggest that at the baseline standard errors, the $R^2$ that we estimate is low relative to what one might estimate if the true data generating process were (4): for count-out with WPA, the $p$-value is 0.25, but for all other specifications shown in Table 7, it is less than

0.06. However, further analysis suggests that the $R^2$—at least for win probability added—are in line with what one might expect with modestly more error. Inflating the standard errors by only 10% leads to a $p$-value of 0.12 for the count and 0.44 for count-out. The model fit seems slightly worse with FIP: one has to inflate the standard error by about 40% to achieve a $p$-value larger than 0.1 and 50% to achieve one larger than 0.2 for count-out with FIP as the metric.[9]

In summary, we take these results as reassuring for the model. While measurement error by itself would cause us to expect larger $R^2$ values than the (already rather large) ones we see, one does not have to add much further model misspecification to justify the fit.

## References


CAPLIN, A., M. DEAN, AND J. LEAHY (2019): "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," Tech. rep., Columbia.

DENTI, T. (2020a): "Posterior Separable Cost of Information," Tech. rep., Cornell.

——— (2020b): "Unrestricted Information Acquisition," Tech. rep., Cornell.

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.

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.

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


[9] We should also note that with finer categories, say count-out-base or count-out-inning, the $p$-values are considerably larger. However, for sufficiently fine categories, measurement error is large enough that this exercise itself may be underpowered.