

Discrete Adjustment to a Changing Environment: Experimental Evidence*

Mel Win Khaw
Columbia University

Luminita Stevens
University of Maryland

Michael Woodford
Columbia University

August 2017

Abstract

A laboratory experiment illustrates cognitive limitations in decision-making that may be relevant for modeling price-setting. Our subjects systematically depart from the optimal Bayesian response in several respects. Their responses are random, even conditioning on available information. Subjects adjust in discrete jumps rather than after each new piece of information, and by both large and small amounts, contrary to the predictions of an “Ss” model of optimal adjustment subject to a fixed cost. And they prefer to report “round numbers,” even when that requires additional effort. A model of inattentive adjustment is found to quantitatively outperform popular alternatives.

*We would like to thank Fernando Alvarez, Marios Angeletos, Marty Eichenbaum, Eduardo Engel, Benjamin Hebert, Cosmin Ilut, Filip Matějka, Rosemarie Nagel, Erkut Ozbay, Isabel Trevino, Laura Veldkamp, Alex Wolman, and conference and seminar participants at the Barcelona GSE Summer Forum, the NBER Summer Institute, the University of Maryland, the Cowles Summer Conference on Macroeconomics, and the JME-SNB-SCG conference for useful comments, and the National Science Foundation for support of this research.

1 Introduction

A central problem in macroeconomics is understanding the ways in which households and firms respond to changing market conditions, with a particular emphasis on how the immediate (or relatively short-run) effects of new developments differ from the adjustments that eventually occur. When behavior is observed at the micro level, continuous decision variables (such as the price that a firm charges for a given product) are often observed to change only at discrete points in time, even though relevant market conditions are changing continuously; this is often attributed to fixed costs of adjustment (“menu costs” in the case of prices), though there is often little direct evidence about the magnitude of such costs. Here we present evidence for an alternative view, under which such discrete adjustment economizes on the cognitive resources of decision-makers.

This paper attempts to shed light on the nature of discrete adjustment dynamics using a laboratory experiment. While there are obvious questions about the similarities between the task faced by our subjects and those faced in settings of relevance for macroeconomic modeling (such as firms’ pricing decisions), a laboratory experiment also has important advantages. We can treat the decision-maker’s objective as known, to the extent that we assume that our subjects care only about maximizing their monetary payment from the experiment, and we can ensure that many complications sometimes supposed to be relevant for firms’ pricing decisions are *not* determinants of our subjects’ behavior. We can also be quite certain about exactly what information the decision-maker has at each point in time; not only do we know everything that the experimental subject has had an opportunity to observe, but we know what she ought to understand about the data-generating process, and hence what inferences could rationally be drawn from the information presented. Finally, the laboratory setting allows us to perform multiple replications of the same environment and to vary the environment to draw causal inferences.

Our experiment is a forecasting exercise in which one of two outcomes can occur

on each trial, and the subject must estimate the probability of a particular outcome. The underlying probability shifts from time to time, but it is likely to remain the same for many successive trials, offering an opportunity to estimate the current probability from observation of past draws. This kind of exercise allows us to test alternative theories of expectation formation, including the familiar benchmark of “rational expectations.” At the same time, the subjects’ problem can be viewed as an example of a more general class of problems, in which there is a continuous decision variable (here, the subject’s announced probability estimate), with the payoff from a given action depending on the current value of continuous state variable (here, the true probability of an outcome), which varies over time (so that it is necessary to continue monitoring), but is sufficiently persistent to make attempts to monitor the changing state variable worthwhile. Viewed in this way, it is an example of the same basic kind of decision problem faced by a firm that must set a price for its product, where the profits obtained by charging a given price depend on other state variables (demand conditions and factor costs) that fluctuate over time.

In fact, despite the differences in the setting, our experimental data exhibit some notable (and puzzling) features of data on individual prices, as we discuss further below. In particular, our subjects usually leave their decision variable unchanged for periods of time, despite the receipt of many new pieces of information in the meantime, and despite having a continuum of possible responses at each point in time. Since there are no true (external) adjustment costs in our problem, and subjects have (and we believe, are able to understand) all of the information needed to obtain the optimal estimate at each point in time, we conclude that their failure to track the optimal Bayesian estimate more closely reflects imperfect attention, limited memory or related cognitive limitations.¹

¹Our experimental design differs importantly from that of Magnani, Gorry and Oprea (2016), who study how well an “*Ss*” model fits laboratory data on the timing of adjustments. Their experimental setup imposes an external fixed cost of adjustment, in order to ensure that adjustment will be discrete, but considers whether adjustments are optimally timed; we are instead interested in observing discrete adjustment even when subjects are free to adjust at any time. Their setup also requires that when adjustment occurs, the decision variable is moved to exactly the currently optimal state, whereas we are interested in where the decision variable will be set when subjects are

We further demonstrate that our data are consistent with the predictions of a particular quantitative model of inattentive choice in a number of important respects. This model generalizes the model of inattentive discrete adjustment developed in Woodford (2009), most importantly by allowing not only the timing of adjustments but also the choice of where to set the decision variable conditional on adjustment to be inattentive. We compare the quantitative fit of our model with other models of random discrete adjustment, such as the “generalized Ss model” of Caballero and Engel (1993, 2007) and the optimizing models with random fixed costs of adjustment proposed by Dotsey, King and Wolman (1999) and Dotsey and King (2005). While the latter models describe the adjustment dynamics that we observe better than a simple “ Ss model,” we find that the rational inattention model outperforms these alternatives significantly.

Section 2 describes our experiment. Section 3 gives an overview of some of the salient features of the behavior that we observe, focusing on the ways in which subjects’ behavior resembles or differs from the predictions of the ideal Bayesian (or rational expectations) benchmark. Section 4 discusses the extent to which various models of discrete adjustment that have been common in the macroeconomic literature, especially the literature on price adjustment, can account for our data. Section 5 offers a concluding discussion.

2 The Experimental Design

Our laboratory experiment follows the setup of Gallistel et al. (2014), who study how well subjects can predict probabilities.² The subjects’ task is to estimate the probability that the next draw is a green ring from a box with both green and red rings. The subjects draw rings with replacement from the box and indicate their draw-by-draw estimate. Figure 1 shows a screenshot of the experiment. The screen displays the box with a hidden number of red and green rings on the left. The slider at

²We modify their experiment in a number of respects, because of the different focus of our study, as discussed in the online appendix.

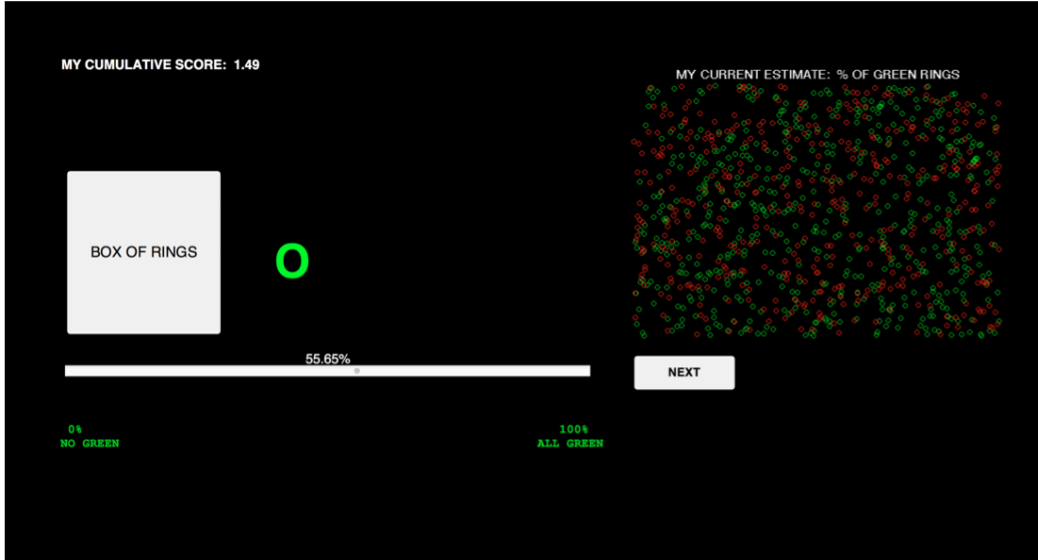


Figure 1: Experiment screenshot. The box of rings contains an unknown number of green and red rings. Subjects use the mouse to adjust the slider to their current estimate of the probability that the next draw from the box is a green ring. Their estimate is displayed numerically above the slider and visually both using the slider and in the box on the right side of the screen. Subjects click the NEXT button to draw another ring from the box of rings on the left. After each draw, the subject’s cumulative score is updated and displayed at the top of the screen. Each session consists of 1,000 ring draws.

the bottom indicates the subject’s estimate on the current trial, \hat{p}_t , with the number in percentage points displayed above the bar. The box on the right side of the screen displays 1,000 rings that also depict the subject’s estimate visually. Whenever the subject moves the slider, this box is adjusted in real time to reflect the probability indicated by the slider position. The subject begins with a guess and adjusts the position of the slider to indicate his or her estimate. Each time the subject clicks the NEXT button, a new ring is drawn randomly from the box and displayed on the screen. The subject may then adjust his or her forecast or leave it unchanged. After each ring draw, the subject’s cumulative score is updated and displayed at the top of the screen. This process is repeated until the session ends, after 1,000 ring draws.

For each session, the subject is rewarded with a fixed payment of \$10 plus a variable payment, equal to $2\text{¢} \times$ the subject’s cumulative score. For each trial t , if the subject sets the slider at \hat{p}_t and the ring drawn is s_t (equal to 1 for a green ring and 0 for a red ring), then the reward is $r(\hat{p}_t; s_t) = 1 - (s_t - \hat{p}_t)^2$. The total score is the

sum of the rewards for all trials in a session. This payoff structure generates a non-negative score that is increasing over the life of the session, to avoid loss aversion. If one believes that a green ring will be drawn with probability p_t (however this forecast might be formed), then the expected reward, $p[1 - (1 - \hat{p})^2] + (1 - p)[1 - \hat{p}^2]$, is maximized by setting the slider at $\hat{p}_t = p_t$.³ The monetary reward is only a function of the history of the subject’s behavior and the sequence of draws thus far. Before the start of the sessions, we explained to the subjects both the reward function and the optimal decision rule, using both written and verbal instructions.

From time to time, the box of rings is replaced by a new box with a different probability of drawing a green ring. The probability is initially drawn from a uniform distribution between 0 and 1. After each ring draw, there is a constant probability $\delta = 0.5\%$ of a regime change. If there is a regime change, the new percentage of green rings in the box is an independent draw from the uniform distribution on the unit interval. The subjects are not told when a change in the box occurs, but they are told in advance that the box might change, and they know the probability of a regime change and the distribution from which the new box is drawn. The frequency of regime changes is small enough relative to the number of noisy observations that subjects receive, to enable them to learn the underlying probability. But it is nonetheless changing, so that the subjects should continuously take into account the fact that there may have been a regime change.

It is important to note that the timing of slider adjustments is completely up to the subject. We also allow the subjects to control the speed at which they draw rings so that they may choose how much attention to give to the task, rather than forcing them to make decisions within an externally imposed time limit. This setup allows us

³We assume that the subjects are risk neutral for the small monetary rewards involved in the experiment. This kind of quadratic scoring rule is often used in experiments intended to elicit probability beliefs, for the reason given here (e.g., McKelvey and Page, 1990). In our case, we are not primarily interested in eliciting subjects’ beliefs about the probability, but simply in their performance on a tracking problem. However, choosing a task for which the optimal strategy is to report one’s estimate of the probability makes it easy for us to explain the task to subjects, and to make sure that any departures from the rational benchmark do not result from subjects’ misunderstanding of the task.

to investigate the extent to which the subjects choose to update their forecast every time they obtain a new piece of information. There is no explicit penalty for constant adjustment and new information arrives frequently, which puts the finding of discrete adjustment at odds with a large set of models, as we discuss in the next section.

Our setup is designed to minimize adjustment costs and to eliminate reasons for habit preferences, such that there would be no reason to not have constant adjustment to the currently optimal forecast in the absence of cognitive limitations. Additionally, we control fully the information that subjects have at each point in time, which allows us to test to what extent subjects are processing the information that is available to them. Finally, by controlling the true data generating process and what our subjects know about it, we control what the correct prior is. Thus we do not need to make conjectures about either the decision-makers' objectives or the information that is available to them, and we can focus instead on what our results indicate about potential cognitive limitations. Since we explained all features of the setup to the subjects, they have all the information needed to form the optimal rational expectations forecast, and we can compare their performance to that of the ideal Bayesian agent.

An important question that arises our experimental study is that of external validity: To what degree can we expect the behavioral regularities that we uncover in the experimental data to persist in other economic environments? The answer to this question depends on the degree to which the conditions that are relevant remain relatively unchanged across environments. In our particular setting, even though we place our subjects in a simple environment, the task that they face is an example of a general class of *tracking problems*. Formally, it is analogous to the problem of a pricing manager who sets the price of a product and whose task is to track the profit-maximizing price as closely as possible, subject to any frictions and limitations that might impede adjustment. There are some key features of the pricing problem that also apply to our experimental task: it features a continuous decision variable, the resulting profits depend on other state variables (demand, costs) that fluctuate continuously over time, and finally, the manager making the pricing decision receives

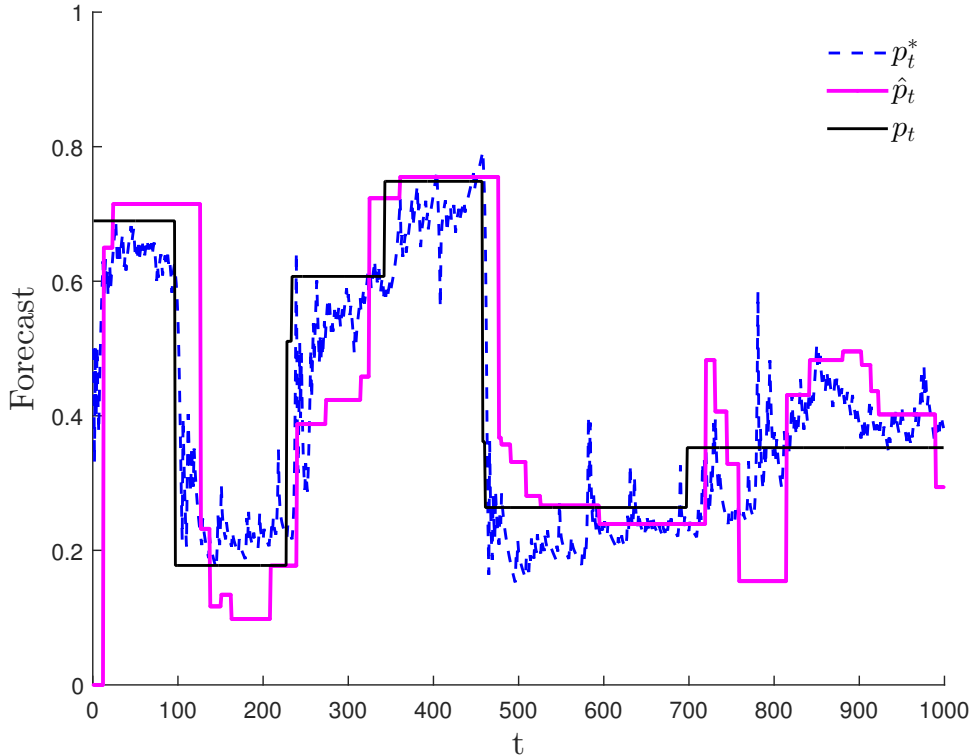


Figure 2: Experiment data for subject 11, session 10. The benchmark Bayesian forecast (blue dashed line) and the subjective estimate (magenta solid) are plotted against the true hidden probability (in black).

a steady stream of information about product demand, competitive pressures, cost conditions, the macroeconomy, etc., and incorporates these factors in the price of the product.

3 Features of Observed Behavior

Figure 2 shows an example of a typical session for a particular subject. The figure plots the true hidden probability and the subject’s forecast, compared to that of a Bayesian decision-maker who makes full use of all the information available.⁴ The figure illustrates key features of the data that we document more systematically below. Both the subject and the Bayesian decision-maker track the hidden probability quite well, and they detect changes in this probability relatively quickly. The subject outperforms the no-information rational benchmark (which would set the slider at

⁴The online appendix presents the Bayesian inference problem.

0.5 on all trials), but underperforms the Bayesian benchmark. This pattern is true for all our subjects (albeit to varying degrees): on average, the subjects’ scores are 2.5% lower than those of the Bayesian observers. Overall, the monetary gains from exerting effort and processing information about the task are fairly low: on average, the subjects would earn \$26.60 per session under complete information (knowing the hidden state) and \$25.00 per session under the no-information benchmark. Nevertheless, the subjects are engaging with the task and exerting effort, closing a significant fraction of the gap between the no-information and the complete-information extreme benchmarks. In the experiment, the subjects’ average reward is \$26.00 per session.⁵

One major distinction between our subject and the Bayesian observer is that while the Bayesian forecast changes by a small amount after each ring draw, the subject keeps his estimate unchanged for variable periods of time, and often adjusts by large amounts. On average, the subject makes much larger and more infrequent adjustments than the Bayesian decision-maker. This discreteness arises despite the fact that there is no discreteness in the set of available actions, nor any explicit cost of adjusting the slider in response to the stream of new information.

With these observations in mind regarding lumpiness in both the timing and the size of adjustments, we now look more closely at some of the ways in which our subjects’ forecasts differ from the predictions of the rational benchmark. We then use these observations to motivate our subsequent consideration of a variety of adjustment models.

3.1 Stochasticity

Our subjects’ deviations from the Bayesian benchmark are not simply due to the use of some boundedly rational but deterministic rule (as for example in typical “adaptive expectations” models). Instead, there is a random element in the subjects’ responses to any given history of evidence. This stochasticity is at odds with tra-

⁵The dollar amounts include a \$10 fixed “show-up” fee. The online appendix presents different measures of forecast bias as well as the cumulative scores for all subjects and for the corresponding Bayesian observers, relative to the no-information and the complete-information benchmarks.



Figure 3: Experiment data, subject 1 forecasts, all repeated sessions.

ditional economic models in which optimization rules out randomization, as well as with many models of behavioral heuristics.

In 20 out of our 110 sessions, we showed the subjects the same sequence of rings. These repeated sessions were never conducted on the same day for the same subject, and there is no evidence that subjects were able to recognize the repetition. We find substantial heterogeneity in different subjects' responses to the given ring sequence, but also in the responses of the same subject across the repeated sessions. Figure 3 shows the forecasts of subject 1 across the repeated sessions. The three forecast series are correlated, since they are tracking the same sequence of realized rings, but there is also considerable dispersion. The top panel of Figure 4 shows the cross-sectional dispersion across all 20 repeated sessions. Throughout the session, there is a wide range of forecasts that depart – at times substantially – from the rational expectations benchmark. Cross-sectional dispersion has also been documented in the pricing literature (see Kaplan and Menzio, 2015, for a recent addition to this literature). While the price setting literature has largely focused on external frictions (limits to arbitrage such as trade barriers and search costs) as the source of such dispersion, our experimental data generates this dispersion in the absence of any such frictions.

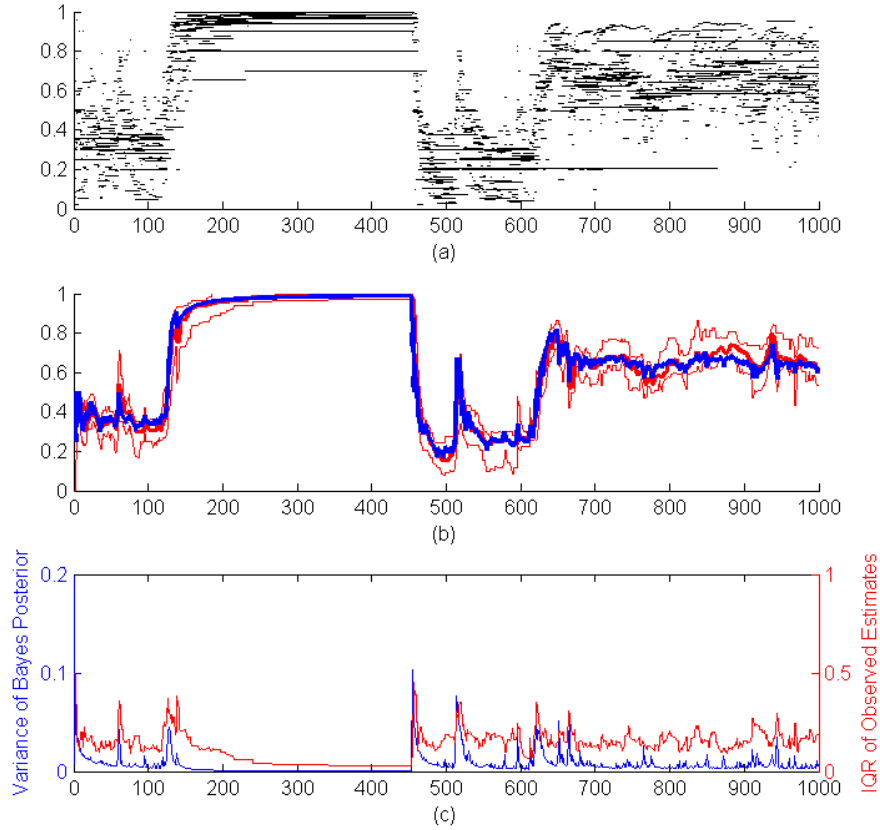


Figure 4: Experiment data, all subjects, all repeated sessions. Trials are on the x-axis. (a) Individual forecast series for all repeated sessions, (b) Bayesian posterior mean in blue, median subject forecast and interquartile range in red, (c) Bayesian posterior variance in blue (left axis) and subjective forecasts interquartile range in red (right axis).

Despite the considerable dispersion, the aggregate behavior over the 20 repeated sessions tracks the RE prediction fairly well, as shown in the middle panel of the figure, which plots the Bayesian forecast against the median slider position. This evidence, while at odds with the classic definition of rational expectations (Lucas (1972)), does provide some support for Muth's (1961) hypothesis that while individual behavior may deviate even substantially from rational expectations, it averages out to the truth. It is important to emphasize, however, that even though the aggregate appears to track the RE forecast, this finding does not mean that the RE forecast suffices for macroeconomic analysis. The heterogeneity of responses implies greater population dispersion in beliefs, and therefore actions, than would be predicted by the

RE model. In turn, this dispersion often matters for welfare, as well as for statistics like the volume of credit or trade.

We also uncover evidence of endogenous dispersion. The heterogeneity in forecasts does not simply reflect exogenous noise in decisions, independent of the state. Instead, the dispersion of forecasts is higher when objective uncertainty is higher, as measured by the Bayesian posterior variance. We show this relationship in the bottom panel of the figure: when the Bayesian posterior variance spikes, the cross-sectional dispersion also rises. This evidence suggests that variation in cross-sectional dispersion may be a reasonable proxy for uncertainty shocks, as it is sometimes used in the literature on uncertainty.

3.2 Discrete Adjustment

In our experiment, subjects should adjust their forecast after each new ring draw, since each draw provides additional information about the hidden probability that subjects seek to track. Instead, our subjects adjust with various lags, even though there are no explicit costs of adjustment. As shown in Figure 5, subjects exhibit a wide range of wait times between adjustments, and there is no evident pattern of time-dependence. There is also considerable variation in the size of the slider adjustments conditional on adjustment, as shown by the distribution of adjustment sizes in the lower left panel of the figure. Both of these features of our data diverge from the predictions of the Bayesian benchmark. At the same time, the distribution of adjustment sizes is also inconsistent with the bimodal distribution that would be predicted by a simple “ Ss model” of optimal adjustment. Such a model would predict no small adjustments and also no adjustments larger than a certain magnitude, since one would never drift that far away from the optimal forecast.

The distribution of step sizes and that of spell lengths instead resemble the distributions of product-level price changes in micro price data. For comparison, the right panels of Figure 5 show the distribution of the size of price changes and of the periods between price changes constructed for products in the AC Nielsen Retail

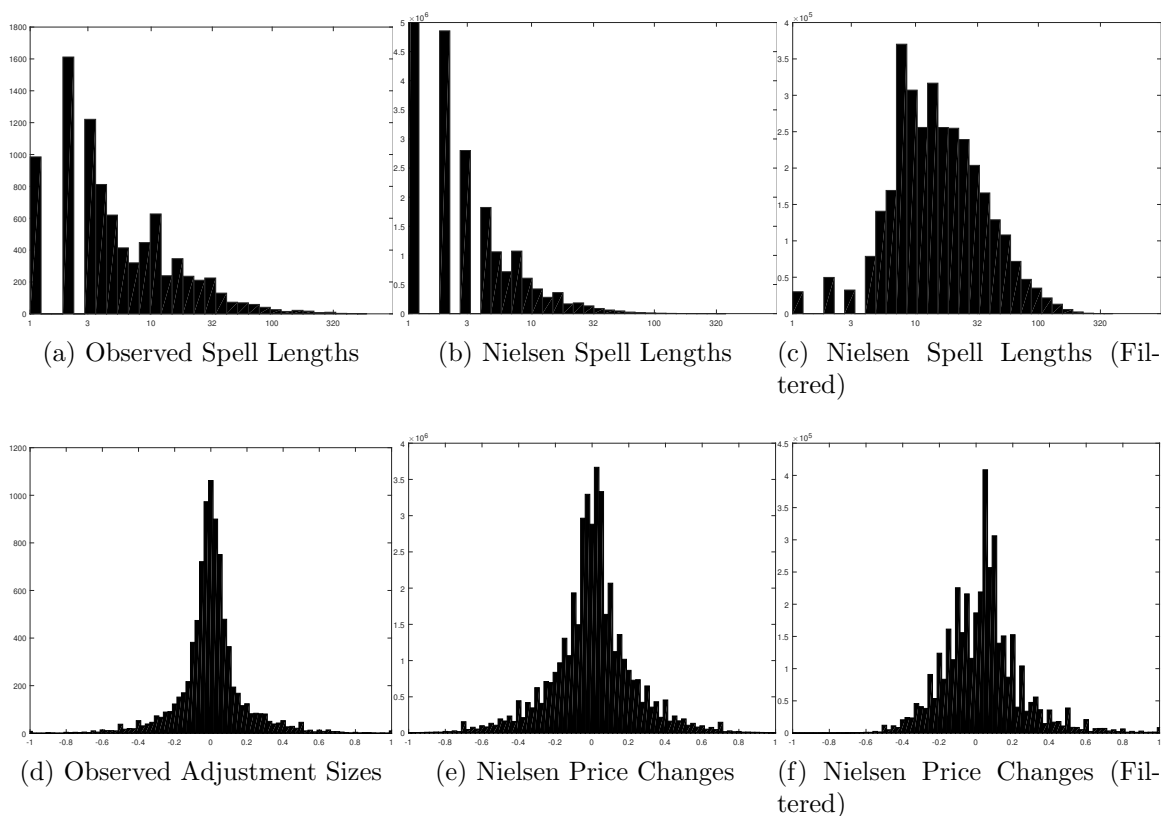


Figure 5: Discreteness. (a) Distribution of the number of ring draws between adjustments (all subjects and all sessions). (b) Distribution of the number of weeks between price changes (AC Nielsen Retail Scanner data; y-axis truncated). (c) The same distribution, but removing transitory price changes. (d) Distribution of the size of changes in the slider position, counting only trials on which the slider is moved (all subjects and all sessions). (e) Distribution of the size of price changes, conditional on adjustment (AC Nielsen Retail Scanner data). (f) The same distribution, but removing transitory price changes.

Scanner data, using both raw prices and regular prices, which exclude transitory price changes. Both our experimental data and the price data generate a distribution for the size of adjustment that is leptokurtic. However, in our experiment, this feature arises spontaneously, as a result of the subjects’ behavior, rather than for the reasons proposed in the pricing literature, such as leptokurtic shocks or opportunities to adjust that depend on other decisions (such as the repricing of other products).

Ricci and Gallistel (2017) document discreteness in a variant experiment in which the underlying probability changes continuously, rather than jumping on infrequent discrete occasions. They report a mean wait time between adjustments of approxi-

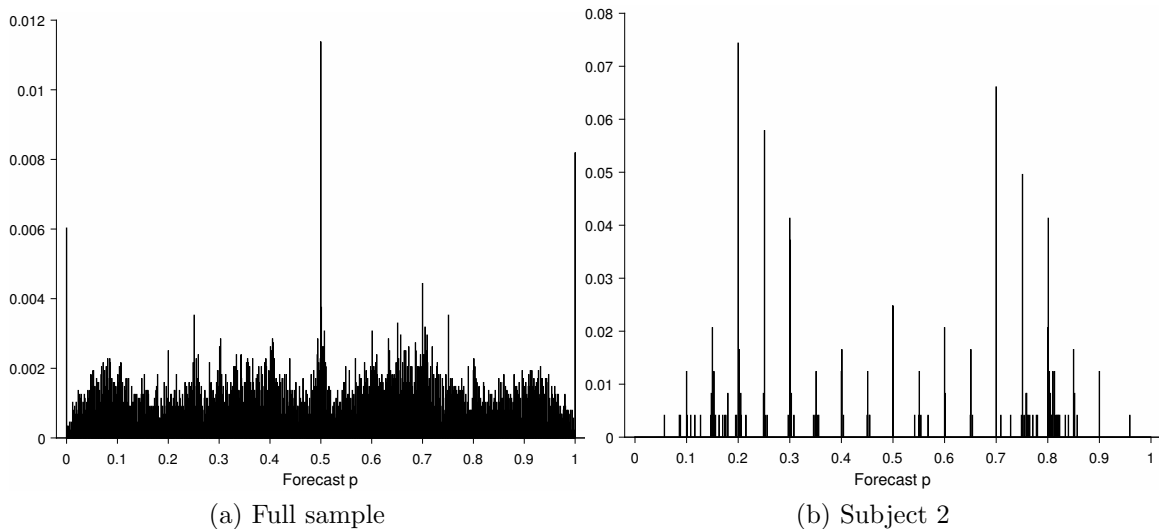


Figure 6: Experiment data. Distribution of the subjects’ slider position choices for (a) the full sample and (b) subject 2.

mately 30 trials. This suggests that the discrete slider adjustments that we observe are not an artifact of our subjects’ knowledge that the true underlying probability moves in discrete jumps.

3.3 Round-Number Bias

Our subjects also exhibit a bias towards selecting “round numbers” when choosing where to position the slider. As shown in Figure 6, subjects are more likely to estimate the hidden proportion of green rings using numbers that are multiples of 0.05. This bias is common to all of our subjects, to varying degrees. Subject 2 is a particularly stark example of this behavior, as illustrated in the right panel of the figure.

This discreteness in the chosen slider positions is another significant departure from the RE benchmark, which would predict a flat distribution. It occurs even though there is no discreteness in the distribution from which the hidden probability is drawn (which is continuous uniform between 0 and 1), or in the set of available actions (the mouse moves continuously⁶). There is also no sense in which the experiment makes it either easier or more profitable to choose certain slider positions over others.

⁶The step size of the mouse is 10^{-7} , hence much finer than even the estimate that is displayed on the subjects’ screen (which is on the order of 10^{-4})

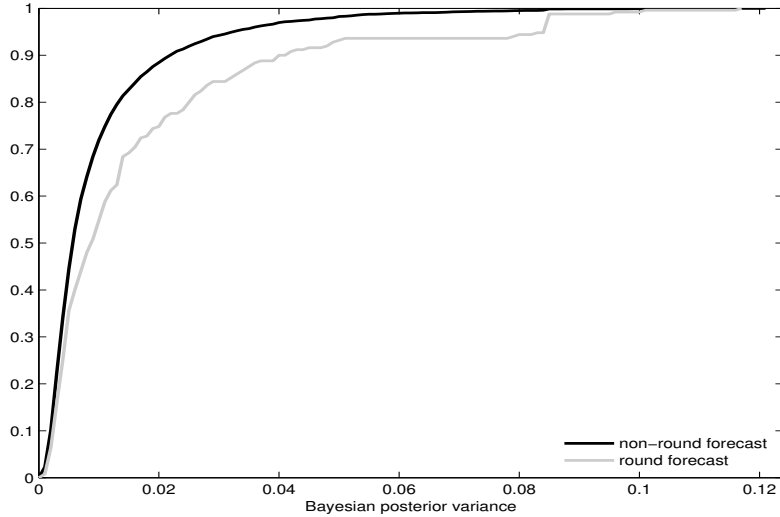


Figure 7: Experiment data. Cumulative distribution functions of the Bayesian posterior variance when the subjects’ forecasts are round numbers versus when they are not round, for trials on which hidden probability is inside the interval $[0.15, 0.85]$. Forecasts are defined as round if they are multiples of 0.05 when rounded to the third decimal.

Nevertheless, our subjects appear to exert considerable effort to set the slider at or near these preferred numbers. We conclude that these round decimal numbers offer some cognitive advantage that compensates for the effort needed to select them.

For trials on which the hidden probability is not too close to 0 or 1, the use of round numbers is associated with higher uncertainty, as shown in Figure 7, which plots the cumulative distribution function for the Bayes posterior variance for trials on which subjects report round forecasts versus those on which they report non-round forecasts, for values of the hidden probability inside the interval $[0.15, 0.85]$. Round forecasts are multiples of 0.05 when rounded to the third decimal. This finding confirms research in cognitive psychology and linguistics that has associated the use of round numbers (such as multiples of five or ten) with conveying uncertainty about the quantity being expressed (e.g. Albers and Albers, 1983). It also supports the method of Binder (2017), who exploits rounding behavior in survey data to develop a novel measure of forecast uncertainty.

The evidence for round number bias also mirrors the patterns found in product-level retail price data, where prices alternate among a small set of values for fairly

long periods of time (Eichenbaum et al., 2011, and Stevens, 2015), and where 9 is the most frequently used price-ending for the penny, dime, dollar and the ten-dollar digits, and where multiples of dimes, dollars and ten-dollars are the most common price changes (Levy et al, 2011).

4 Modeling Discrete Adjustment

We study the degree to which our experimental data conform to the predictions of a variety of different possible models of discrete adjustment of the decision variable, with a particular focus on types of models that have been popular in the macroeconomic literature on price adjustment. Given the evidence that subjects do not adjust continuously, it is useful to decompose a model of discrete adjustment into two parts: first, a criterion that determines *when* the slider will be adjusted, and second, a criterion that determines *where* the slider position will be set, conditional on adjustment. We consider a variety of alternative models of each of these decisions.

4.1 “Gap-Based” Models of Adjustment

Under the Bayes-optimal benchmark, no “Bayesian gap” $\Delta_t \equiv p_t^* - \hat{p}_t$ should ever exist. This is plainly not the case in our experiment, as we have documented in Section 3. One might however still hypothesize that the size of the “Bayesian gap” should determine whether there is sufficient reason to adjust the slider. For example, we might hypothesize that adjustment should occur if and only if the gap is either above some upper threshold or below some lower threshold, as in an “*Ss* model.” More generally, we might suppose, as in the “generalized *Ss* models” of Caballero and Engel (1993, 2007), that there are no sharp thresholds, but that instead the probability of adjustment of the slider in any period is given by some continuous function of the Bayesian gap $\Lambda(\Delta_t)$.

Figure 8 shows the fraction of trials on which the slider is adjusted, if the data are sorted into bins according to the value of the Bayesian gap at the time of the decision whether to adjust the position. The figure is consistent with the existence

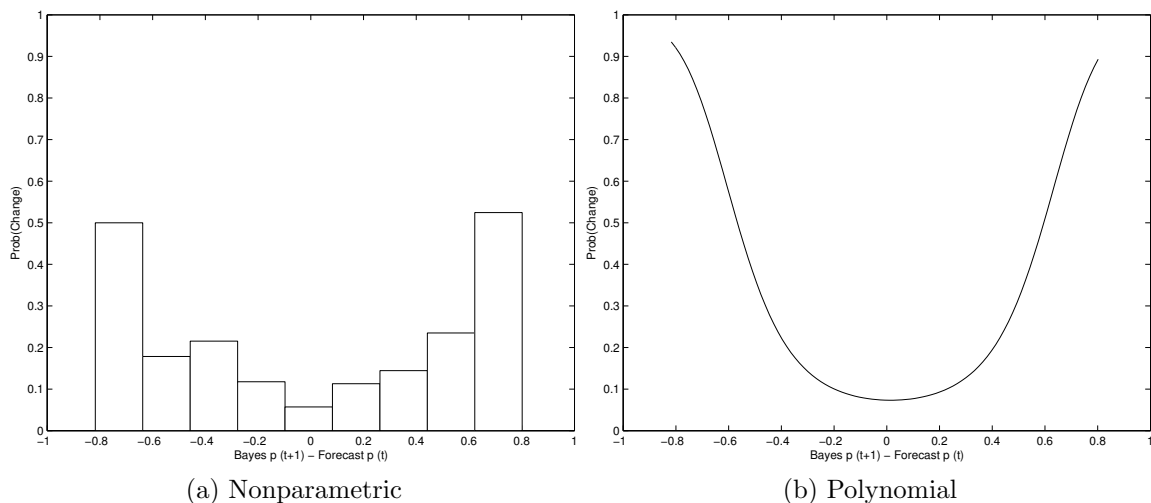


Figure 8: Hazard functions for the slider adjustment probability as a function of the Bayesian gap. (a) Nonparametric empirical hazard plotting the fraction of trials on which adjustment occurs, for each range of values of the gap, with equal-width bins. (b) The best-fitting logistic-polynomial hazard function. (Coefficients given in Table 1.)

of a non-uniform hazard function, which is furthermore decreasing for $\Delta < 0$ and increasing for $\Delta > 0$, as assumed by Caballero and Engel. We note, however, that our data do not appear to support the further common assumption that the hazard falls to zero at least when $\Delta = 0$, if not over some larger “zone of inaction.”

Thus there is at least some degree of “state-dependence” of the decision to adjust the slider. We can quantify the degree of state-dependence by computing how much the likelihood of our data is increased by allowing the adjustment hazard to depend on the Bayesian gap. We consider parametric families of possible hazard functions $\Lambda_k(\Delta; \theta)$, where θ is a vector of k parameters, and we choose both the family of models (the value of k) and a specific model within that family (the value of the parameter vector θ) so as to minimize the Bayes Information Criterion (BIC), $BIC \equiv -2LL + k \log N$, where LL is the log likelihood of the data and N is the number of observations.

Suppose that the log odds of adjustment, $\lambda(\Delta) \equiv \log \frac{\Lambda(\Delta)}{1-\Lambda(\Delta)}$, is a polynomial

function of Δ , with $k - 1 \geq 0$ the order of the polynomial,

$$\lambda_k(\Delta; \theta) = \sum_{j=0}^{k-1} \theta_j \Delta^j. \quad (1)$$

In the case of our data, the BIC is minimized when $\lambda(\Delta)$ is a quadratic function, with coefficients $\{\theta_j\}$ given in the top panel of Table 1. The implied hazard function $\Lambda(\Delta)$ is shown in the right panel of Figure 8. The best-fitting hazard function reaches its minimum near $\Delta = 0$, and approaches 1 for large enough positive or negative values of the gap; but even at its minimum, the probability of adjustment is positive.

The degree to which the data support these conclusions can be determined by comparing the minimum values of the BIC statistic for more restrictive families of models. For example, if we assume a constant hazard function (as in the Calvo (1983) model of adjustment), then as indicated in Table 1, the minimum BIC statistic is higher by 1,318 natural log points. Thus the data increases the relative posterior probability that the quadratic log odds model is correct by a factor greater than 10^{286} , relative to the hypothesis of no dependence of the adjustment decision on the Bayesian gap.⁷

As another example, we consider the degree to which the data support the conclusion that the hazard function is asymmetric, by considering the more restrictive family of only symmetric polynomials. We again find that the best-fitting polynomial is quadratic, but the BIC statistic is only 0.7 log points higher than in the unconstrained case, indicating only weak evidence in favor of an asymmetric hazard function.

While we are able to reject the hypothesis of no dependence on the Bayesian gap (the Calvo model), our data are also inconsistent with the strong degree of state-dependence assumed in a simple S_s model, in which adjustment never occurs for gaps within some fixed thresholds, and occurs with certainty for gaps outside them. Any such model will have an unboundedly large BIC statistic. This exact version of an

⁷See the appendix for discussion of our use of the BIC statistic to determine the relative posterior probabilities of alternative models.

TABLE 1: Best fitting models

Models	k	Parameters				BIC	
Gap Models		θ_0	θ_1	θ_2			
Constant hazard	1	-2.33	-	-		59,325	
Polynomial hazard	3	-2.53	-0.23	7.76		58,006.5	
Symmetric poly.	2	-2.53	-	7.82		58,007.3	
Gap Models with Errors		ϵ	$\underline{\Delta}$	$\bar{\Delta}$			
<i>Ss</i> with errors	3	0.089	-0.641	0.771		59,329	
Symmetric <i>Ss</i>	2	0.089	-0.923	0.923		59,336	
Value Models		θ_0	θ_1	ϵ_1	ϵ_2	$\bar{\Delta}$	
Constant hazard	1	-2.33	-	-	-	-	59,325
Polynomial hazard	2	-2.49	0.95	-	-	-	58,169
<i>Ss</i> with errors	2	-	-	0.089	0.911	4.03	59,285
Step function	3	-	-	0.083	0.207	0.60	58,663
Value Model with Inattention		θ_0	θ_1	ψ_1	$\tilde{\Lambda}$		
Rational inattention	2	-2.052	0.816	1.225	0.1139		57,598

Polynomial coefficients and BIC for the log odds of adjustment, $\lambda(\Delta) \equiv \log \frac{\Lambda(\Delta)}{1-\Lambda(\Delta)}$, a polynomial function of Δ , with $k - 1 \geq 0$ the order of the polynomial, and $\lambda_k(\Delta; \theta) = \sum_{j=0}^{k-1} \theta_j \Delta^j$. Top panel: Coefficients and BIC for the best-fitting models within the Bayesian gap class of models. Middle panel: Coefficients and BIC for the best fitting models that allow adjustment to depend on the expected value of adjustment, assuming optimization subject to a random fixed cost of adjustment. Bottom panel: Coefficients, BIC and implied model parameters for the model assuming that adjustment economizes on information costs.

Ss model may seem a straw man, so we also consider a generalization in which it is assumed that there will inevitably be some random error in the execution of any decision with regard to adjustment: that even when a decision to adjust is made, the slider will only be adjusted with probability $1 - \epsilon$, and when a decision not to adjust is made, the slider will nonetheless be adjusted with probability ϵ , for some $0 < \epsilon < 1/2$. But even allowing for errors of this kind, the minimum value of the BIC statistic remains 1,323 log points higher under a model with sharp thresholds than under the quadratic log odds model. If we assume that the thresholds must

be symmetric, reducing the number of free parameters by one, the BIC statistic further increases by 7 log points. Hence, we conclude that the adjustment decision is state-dependent — with the Bayesian gap providing at least an imperfect proxy for the relevant state — and furthermore, that the Caballero-Engel hypothesis of a continuous hazard function fits the data better than a model with sharp thresholds.

We turn next to the question of where the slider is moved when it is adjusted. Caballero and Engel (1993, 2007) assumed, as in the classic “*Ss*” model, that when the decision variable is adjusted, it is moved to its current (myopically) optimal value. In the present context, this would correspond to a hypothesis that \hat{p}_t is set equal to p_t^* . Panel (b) of Figure 10 shows a scatter plot of the values of \hat{p}_t and p_t^* for all of the trials on which the slider is adjusted. Under the simple hypothesis, all of these points should lie on the diagonal. We do observe some relationship between the two variables; in fact, a linear regression of \hat{p}_t on p_t^* yields a regression line quite close to the diagonal. Nonetheless, there is clearly a great deal of noise in the relationship, not predicted by the simple hypothesis that the “Bayesian gap” should be eliminated whenever the slider is adjusted.

4.2 Optimizing Models with a Fixed Cost of Adjustment

We next consider the extent to which our data are consistent with more explicitly optimizing (and more forward-looking) models in which the decision whether to adjust the slider is assumed to be optimal (expected-payoff maximizing) on each trial, subject to a fixed cost of adjustment. Note that while there is no cost in terms of points deducted from the subject’s score when the slider is moved, one might suppose that there is nonetheless a utility cost of having to move the mouse to adjust the slider position, a utility cost of the additional delay before proceeding to the next trial,⁸ or a cognitive cost of having to decide where to move the slider.

Consider first the classic “menu cost” models of price adjustment such as Sheshin-

⁸We note that there is a time cost of slider adjustment for our subjects: on average, subjects take only 1.4 seconds to request the next ring on trials where they do not move the slider, but 4.4 seconds on trials where they move the slider before requesting the next ring.

ski and Weiss (1977, 1983). Because the cost of adjustment is the same regardless of the new position chosen, whenever the slider is adjusted, the new position \hat{p}_t should be chosen to maximize $V_t(p)$, where for any possible slider position $0 \leq p \leq 1$, $V_t(p)$ is the expected cumulative payoff from all remaining trials (net of adjustment costs on trials after t) if the slider is set at p . The slider should then be adjusted on trial t if and only if the “value gap” $\Delta_t \equiv \max_p V_t(p) - V_t(\hat{p}_{t-1})$ is greater than or equal to κ , the fixed cost of adjustment. Note that the value function $V_t(p)$ will depend not only on the number of the trial t , but also on the (optimal Bayesian) posterior π_t at that time; the argument π_t has been suppressed to simplify the notation.

We can alternatively assume, as in the generalized model of state-dependent pricing proposed by Dotsey, King and Wolman (1999), that the fixed cost of adjustment is a random variable, drawn independently on each trial from some distribution with CDF $G(\kappa)$.⁹ The probability of adjustment on trial t should then be given by a non-decreasing hazard function $\Lambda(\Delta_t) = G(\Delta_t)$, where the “value gap” Δ_t is again defined as above. If the distribution of fixed costs is continuous (contains no atoms), the hazard function $\Lambda(\Delta)$ must be a continuous, non-decreasing function with $\Lambda(0) = 0$. It is possible, however, to have $\Lambda(0) > 0$ if we suppose that there is a positive probability of a zero fixed cost, as in the model of Calvo (1983).

The empirical content of either version of the model depends on evaluating the value function $V_t(p)$. We would like to compare alternative models of the adjustment decision on a single trial using a measure of the value gap that is (to the extent possible) independent of assumptions regarding behavior in later trials. As explained further in the appendix, we estimate an *atheoretical* statistical model of subjects’ behavior after any trial t conditional on the slider setting \hat{p}_t chosen on that trial and the posterior π_t at that point, and use this atheoretical model to compute an implied value function $V_t(p)$. We can then test the degree to which both the timing of slider adjustments and the new slider positions chosen are consistent with the predictions above, under any assumption about the distribution of fixed costs.

⁹Caballero and Engel (1999) propose a similar model of optimization with random fixed costs as a micro-foundation for the type of “generalized *Ss* model” proposed in Caballero and Engel (1993).

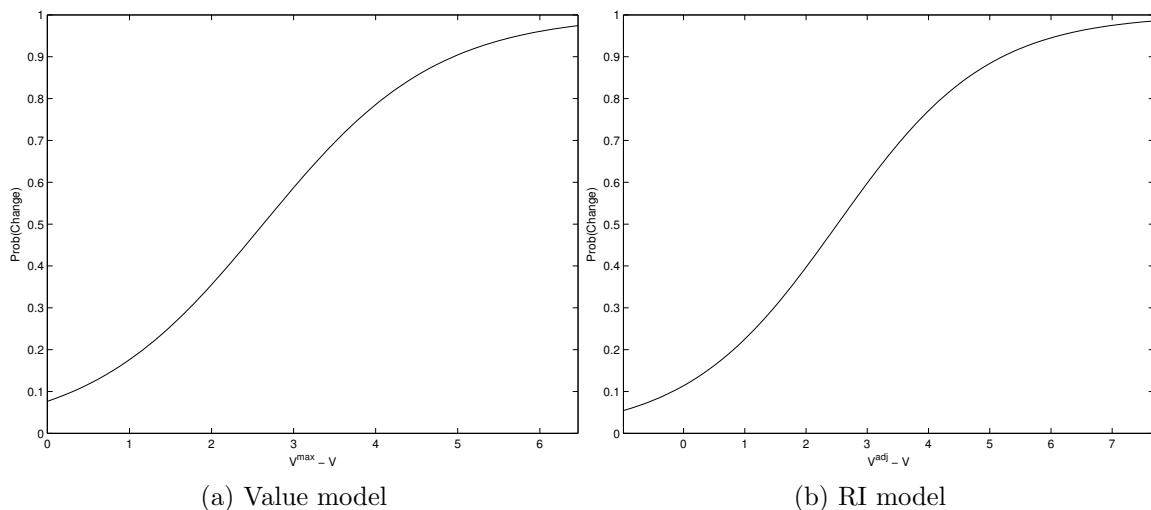


Figure 9: Adjustment probability as a function of the value gap: the best-fitting logistic-polynomial hazard function as a function of the value gap. (Coefficients given in Table 1.)(a) Value gap computed under an assumption that if the slider is adjusted, it will be moved to the currently optimal position. (b) Value gap implied by the rational inattention model.

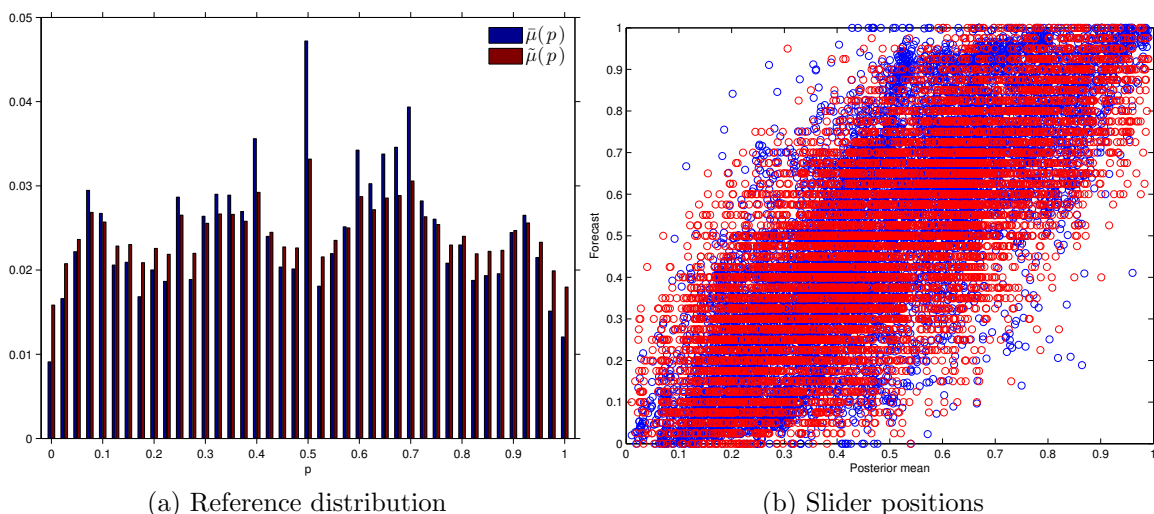
As in the case of the models based on the Bayesian gap, we can estimate a continuous hazard function using the BIC to penalize excessive flexibility in the family of functions considered. If we again consider the class of logistic-polynomial hazard functions, we find that the best-fitting model makes the log odds of adjustment a linear function of the value gap, with a positive intercept and positive slope, as reported in the middle panel of Table 1. The implied best-fitting value-based hazard function is shown in the first panel of Figure 9. This corresponds to a random-fixed-cost model in which the distribution of fixed costs has a CDF also given by the curve in that figure. The implied distribution has an atom at zero fixed cost (about 8 percent of the total probability), and a continuous distribution otherwise, with a positive probability density at zero and at all positive levels; the fixed cost can be unboundedly large, but is finite with probability one.

We again use the BIC to measure the degree to which the data favor this model. The best-fitting linear function $\lambda(\Delta)$ achieves a statistic that is lower by 1,156 log points than in the case of the best-fitting constant hazard. We also reject the hypothesis of a constant fixed cost (the optimizing S_s model). Even if we assume (as in the

previous section) that it is not possible to reduce the probability of adjustment below some $\epsilon > 0$ or to raise it above $1 - \epsilon$ (with ϵ a free parameter that can be chosen to fit the data), the lowest BIC statistic obtainable with such a model remains 1,116 log points higher than that obtained with our preferred random-menu-cost model. Alternatively, we can also consider the performance of a step-function variant of the flat hazard model, in which we suppose that for value gaps below a certain threshold, the probability of adjustment is $\epsilon_1 > 0$, while above that threshold, the probability rises to $\epsilon_2 > 0$ (not necessarily equal to $1 - \epsilon_1$). Allowing for this flexibility significantly improves the fit over the flat hazard model, but the BIC remains 413 log points higher than under the best polynomial-hazard model. Thus, the data strongly prefer a model in which the probability of adjustment increases continuously with the value gap.

While our value gap measure does have some ability to predict the trials on which subjects are more likely to adjust the slider, it remains a less successful predictor than the Bayesian gap. This failure of the optimizing model to do better, even when we allow for a very flexible possible distribution of fixed costs, may indicate that in fact our subjects use a simple rule of thumb rather than a fully optimal strategy, or perhaps that they are myopic optimizers. Alternatively, it may simply indicate that our numerical approximation to the value function $V_t(p)$ is not accurate enough; we should expect our numerical estimates of the Bayesian gap to be much more accurate, as it does not depend on any approximate characterization of observed behavior.

The model of optimization subject to a (possibly random) fixed cost also fails in a more obvious way. It implies that conditional on adjustment of the slider in period t , the new slider position \hat{p}_t should be exactly the p that maximizes $V_t(p)$. In the appendix, we show a scatter plot of the values of \hat{p}_t and $\arg \max_p V_t(p)$; these form a cloud similar to the one in panel (b) of Figure 10. While some of the noise might in this case reflect error in our estimation of the continuation value function, the noise in the relationship seems too great to be explained by the inaccuracy of our approximation method alone. It appears that the decision where to move the slider when it is adjusted, like the decision whether to adjust it, involves a significant degree



(a) Reference distribution (b) Slider positions

Figure 10: Predictions of the RI model. (a) The model-implied reference distribution $\tilde{\mu}$ of slider positions and the empirical distribution μ , conditional on adjustment. (b) The model-implied joint distribution of slider settings and Bayesian posterior mean, conditional on adjustment. Blue circles: scatter plot of the actual slider positions [vertical axis] against the Bayesian optimal setting. Red circles: scatter plot of the simulated RI slider positions [vertical axis] against the Bayesian optimal setting.

of randomness. The hypothesis of fixed costs of adjustment (even randomly varying fixed costs) cannot explain why this should be the case. Instead, the hypothesis of inattentiveness on the part of decision-makers can equally explain the randomness of both decisions, as we discuss next.

4.3 A Model of Inattentive Adjustment

We now consider a model of discrete adjustment in which both the decision when to adjust the slider and the decision where to move it conditional on adjustment are assumed to be optimal, but in terms of an objective that includes costs of paying closer attention to both decisions. The particular theory of attention costs that we consider is based on the theory of “rational inattention” proposed by Sims (2003, 2011) and applied to discrete adjustment problems by Khaw, Stevens and Woodford (2016).¹⁰ The theory generalizes the models of inattentive adjustment by Woodford

¹⁰See Cheremukhin *et al.* (2011) for an earlier application of the theory of rational inattention to the explanation of randomness in experimental data, albeit to a sequence of independent static decision problems rather than to a dynamic setting like that considered here.

(2009) and Stevens (2015) by relaxing the assumption that agents observe the true state with perfect precision conditional upon deciding to review the current setting (i.e., to adjust the slider), and by introducing the possibility of intrinsic preference for particular actions.

We specify a subject's behavioral rule by a sequence of functions $\Lambda_t(h_t)$, indicating the probability of adjusting the slider on trial t for each possible history h_t prior to that trial's ring draw, and a sequence of functions $\mu_t(h_t)$, specifying a probability measure μ_t over possible new locations for the slider (if adjusted on trial t) for each possible prior history h_t . In addition to assuming that it is costly for a subject to pay closer attention to the history h_t , we assume that there may also be intrinsic costs or convenience associated with particular actions. For example, there may be an effort cost associated with moving the mouse in order to adjust the slider position (as in the model with a constant fixed cost considered above). Or there may be a preference for choosing certain slider positions, independent of the monetary payoff expected from choosing those positions; our subjects' apparent preference for positions corresponding to round numbers suggests that this is the case.

Let the measurable function $c(p)$ denote the cost associated with choice of slider position p , borne only at times when the slider is adjusted; a negative value corresponds to a positive preference for that position choice. Similarly, let c^{adj} denote the cost associated with adjustment of the slider, and c^{non} the cost of non-adjustment (which quantities are also possibly negative).¹¹ Under the hypothesis of rational inattention, the subject's decision rule maximizes an objective of the form

$$E \left\{ \sum_{t=1}^T \left[r(\hat{p}_t; s_t) - \psi_1 I_1 - \psi_2 \Lambda_t I_2 - \Lambda_t c^{adj} - (1 - \Lambda_t) c^{non} - \Lambda_t \int c(p) d\mu_t(p) \right] \right\}, \quad (2)$$

where I_1 is a measure of the amount of information used in deciding on each trial whether to adjust the slider; I_2 is a corresponding measure of the amount of infor-

¹¹Note that the amount by which c^{adj} differs from c^{non} may reflect costs of having to decide about where to move the slider, whether or not it is actually moved, as in Alvarez *et al.* (2011). Here such costs, if found to exist, would be purely cognitive.

mation used in deciding where to set the slider, on those trials where it is adjusted; $\psi_1, \psi_2 > 0$ are attention cost parameters for the two types of information; and $E\{\cdot\}$ indicates an expectation over the possible sequences of realizations of ring draws, the possible outcomes of the random decision whether to adjust the slider each period, and the possible outcomes of the random decision as to where to set the slider when it is adjusted. Here, the history h_t consists of the complete history of ring draws through trial $t - 1$, together with the complete history of slider positions chosen through trial $t - 1$. It does *not* include the history of evolution of the underlying state $\{p_t\}$, as this is not visible to the subject, no matter how closely he may pay attention.

The solution to this problem is given by an optimal adjustment hazard

$$\log \frac{\Lambda_t}{1 - \Lambda_t} = \lambda(\Delta_t) \equiv \log \frac{\tilde{\Lambda}}{1 - \tilde{\Lambda}} + \psi_1^{-1} \Delta_t \quad (3)$$

and an optimal measure over new slider positions

$$\mu_t(p) = \frac{\exp\{\psi_2^{-1} V_t(p)\} \tilde{\mu}(p)}{\int \exp\{\psi_2^{-1} V_t(p')\} d\tilde{\mu}(p')}. \quad (4)$$

Here $V_t(p)$ is again the continuation value function, and the Δ_t in (3) is now the “RI value gap” defined as

$$\Delta_t \equiv \int V_t(p) d\mu_t(p) - V_t(\hat{p}_{t-1}). \quad (5)$$

The value of the reference adjustment probability $\tilde{\Lambda}$ depends on both the unconditional probability of adjustment and on the intrinsic cost of adjustment relative to non-adjustment (it would equal the unconditional probability if $c^{adj} = c^{non}$); similarly, the reference measure $\tilde{\mu}(p)$ depends on both the unconditional frequency distribution of slider location choices and on the intrinsic costs of different locations. In the absence of any independent evidence about the intrinsic costs, we can treat $\tilde{\Lambda}$ and $\tilde{\mu}(p)$ as free parameters to be estimated, from which we can then infer the implied intrinsic costs.

Note that (3) implies that the probability of adjusting is again given by a contin-

uous hazard function, and indeed that the hazard function must belong to a specific parametric family: the log odds of adjustment are predicted to be a linear function of Δ_t , as in the case of the best-fitting random-fixed-cost model discussed in section 4.2, though here the definition of Δ_t is different. Unlike the models considered above, (4) implies that the slider location decision is similarly randomized. Note however that the model reduces to the simple prediction that the new slider location will be the value of p that maximizes $V_t(p)$ in the limiting case $\psi_2 \rightarrow 0$.

In the work reported here, we do not optimize over a completely unrestricted reference distribution $\tilde{\mu}$, and instead consider only distributions of the parametric form $\tilde{\mu}(p) = A\bar{\mu}(p)^\gamma$ for some $\gamma \in [0, 1]$, where $\bar{\mu}$ is the unconditional frequency distribution of slider positions.¹² Within this family, the best-fitting reference distribution corresponds to $\gamma = 0.45$. The implied distribution $\tilde{\mu}(p)$ is shown in the left panel of Figure 10; note that that this distribution is significantly flatter than the empirical distribution of slider location choices, also shown in the figure. (The implied intrinsic cost function $c(p)$ is shown in the appendix.)

The degree to which this model of the slider location decision matches the experimental data when $\psi_2 = 0.64$ is shown in the right panel of the figure. Here we show a scatter plot of slider location choices against the value of p_t^* at the time of the choice (blue dots), with the corresponding scatter plot from a simulation of the rational inattention model overlaid on it (red dots). In this simulation, we assume the same sequences of ring draws (and hence the same evolution of the Bayesian posterior) as in the experimental sessions, and we suppose that the slider is adjusted on exactly the same trials as in the experimental sessions (so that the set of Bayesian posteriors associated with trials on which the slider is adjusted is the same as in the experimental data); but we suppose that when the slider is adjusted, the new position is drawn from the distribution μ_t given by equation (4). We find that the predicted joint distribution of the Bayesian posterior mean and the new slider position is reasonably

¹²Allowing a more flexible family would necessarily allow an even better fit to our data, at a possible risk of overfitting. This simple family nests both the cases of zero intrinsic costs ($\gamma = 1$) and a uniform reference distribution ($\gamma = 0$) as assumed in the “logit dynamics” model of Costain and Nakov (2014).

similar to the one observed in our data.

We turn next to the ability of the RI model to account for the timing of adjustments of the slider. We can test the model by estimating a polynomial log odds model for the hazard function, again assuming a flexible relationship as in equation (1) for the case in which the argument of the polynomial function is the “Rational Inattention value gap” defined in equation (5). The RI model makes a very specific prediction, that $\lambda(\Delta)$ should be an increasing linear function (equation (3)). We find that our data are best fit by a model according to which the log odds are indeed an increasing linear function of the gap, as predicted by the theory. The polynomial parameters and the associated BIC value are shown in the bottom panel of Table 1 and the polynomial hazard function is plotted in the second panel of Figure 9. The estimated polynomial parameters imply values for the reference frequency of adjustment $\tilde{\Lambda}$ and for the information cost ψ_1 , which we report in the bottom panel of Table 1 as well. The estimated reference probability of adjustment, $\tilde{\Lambda} = 11.4\%$, is larger than the empirical probability of adjustment, $\bar{\Lambda} = 8.9\%$. This difference corresponds to a relative cost $c^{adj} - c^{non} = -0.34$, meaning an intrinsic *preference for adjustment*, so that subjects adjust more frequently than would be optimal given the precision of their estimate of where to set the slider. In fact, our subjects often adjust when the RI value gap is negative. If instead, subjects adjusted only for non-negative the RI value gaps, the estimated relative adjustment cost would be zero and the reference probability of adjustment would be equal to the empirical probability of adjustment. Subjects would adjust even more infrequently than we observe in our data.

Interestingly, the information cost ψ_1 for the decision whether to adjust is estimated to be roughly twice as large as the information cost ψ_2 for the decision where to move the slider conditional on adjustment.¹³ The BIC statistic implies that the RI model best fits our data on the timing of slider adjustments, among all of the models considered. Even relative to the best-fitting gap-based model, the BIC of the

¹³This finding is consistent with the results of Stevens (2015), who finds that the parameterization that best fits micro price data is one in which the decision whether to review the firm’s pricing plan is more costly than the decision about which prices to charge.

RI model is lower by 489 log points, implying a relative posterior probability of the RI model greater than 10^{106} .

Hence the rational inattention model augmented with intrinsic costs succeeds in capturing the main features of our data, at least in terms of the pooled data, describing the average behavior of our population of subjects. It predicts both the timing of adjustments (which is stochastic, with an adjustment hazard that is a continuously increasing function of the expected value of adjusting) and the position to which the slider is moved if adjusted (which is also stochastic, with a probability of movement to each position that is a continuously increasing function of the expected value associated with that position). Moreover, the specific functional form (3) that is predicted for the adjustment hazard function fits our data better than any of the other models considered above. Finally, the model allows (though it does not require) the kind of “round-number bias” observed on the part of our subjects.

5 Discussion

We have used a controlled laboratory setting to study how successful our subjects are in tracking a changing environmental state, and adjusting their behavior so as to continue to maximize expected reward. While our experimental setting is very stylized, the pattern of discrete adjustments that we observe shares some notable features with data on the adjustment of individual goods prices by retailers. This suggests that our findings about adjustment behavior in our experiment may well be relevant to understanding discrete adjustment in economically relevant settings as well, such as price-setting behavior.

Moreover, we observe a pattern of discrete adjustment reminiscent of that observed in micro-level data on prices even in a setting where many factors that might be thought relevant to price-setting are clearly not present: here, for example, there are no costs of communicating a new policy to customers, no reasons to doubt whether customers will recognize and respond to a price change, nor any reasons to fear the emotional reaction of customers to a price change that they notice. The discrete-

ness thus must result from some kind of cognitive constraint on subjects' ability or willingness to more closely track the reward-maximizing slider setting. And while we might think of this as reflecting decisions made on the basis of imperfect information, it is not imperfect information that results from the structure of the physical environment, as in the model of Lucas (1972); instead, behavior seems to make use of less precise information about the changing environment than has been presented to the subjects, reflecting some form of inattention, limited memory, or other limit in information processing capacity.

Our experiment allows us to reject many familiar hypotheses about adjustment to a changing environment, at least as an explanation of the behavior of our subjects. These include, most obviously, the benchmark of rational expectations (or perfect Bayesian inference from the available data). Yet our data also clearly reject any model that implies continuous adjustment each time a subject has a reason to adjust their beliefs, such as the “sticky information” models of Mankiw and Reis (2002) or Reis (2006), the “noisy information” model of Woodford (2003), or the rational inattention model of Mackowiak and Wiederholt (2009).¹⁴ We can also reject time-dependent models of discrete adjustment, such as the Taylor (1980) model of staggered adjustment, since our data show that adjustments often occur very soon after a previous adjustment of the slider (and there is also no evidence of a preferred time length between adjustments).

Our experimental data on the timing of discrete adjustments are inconsistent with the assumption of the Calvo (1983) model of price adjustment, according to which an adjustment is equally likely to occur over any time interval, for we show that it is possible to define measures of the degree of inappropriateness of the existing slider position which have some ability to predict the timing of slider adjustments. Yet the

¹⁴The rational inattention model in which the subject must choose the slider position in each period would generate discreteness in the slider positions themselves (as shown by Matejka (2015) and Jung et al. (2015)), because our state variable is drawn from a distribution with bounded support. However, it would not imply discreteness in the timing of adjustments, of the kind that we observe here: subjects would be predicted to alternate with high frequency among the different values of a discrete set of slider positions.

kind of state-dependence indicated by our data is not the kind assumed in a simple “ S_s ” model. In the case of none of the gap measures that we consider do we find evidence of sharp thresholds such that adjustment occurs if and only if a threshold is crossed. Moreover, under an “ S_s ” model, if adjustments do not occur constantly, then one should never observe small adjustments, and one should observe few really large adjustments. The distribution of adjustment sizes that we observe is inconsistent with an “ S_s ” model on both counts.

The kind of models that are instead consistent with at least certain broad features of our data are models of discrete adjustment in which the timing of adjustment is stochastic, though the probability of adjustment is a continuous, increasing function of some measure of the inappropriateness of the current slider setting; such models include those proposed by Caballero and Engel (1993, 1999), Dotsey, King and Wolman (1999), Woodford (2009), and Costain and Nakov (2014). Among the several models of this type that we fit to the timing of slider adjustments in our data, we find that the best-fitting model (under a BIC criterion) is a generalization of the rational inattention model of Woodford (2009). In this type of model, the randomness of decision whether to adjust the slider at a given point in time is interpreted as resulting from inattention, or more generally from imprecision in the subject’s subjective awareness of the precise situation that has been revealed by the sequence of ring draws observed to that point (due, perhaps, to a memory limitation).¹⁵ This interpretation of the random timing of the slider adjustments is especially appealing in that it makes it natural to expect that the position of the slider conditional on adjustment will also be random, unlike a model where discrete adjustment is motivated by (possibly random) fixed costs but decision-makers are assumed to be perfectly aware of the state at all times. Our data are much better fit by a model in which the slider location decision is also assumed to be stochastic.

An interesting departure of the experimental data from our model of inattentive

¹⁵Magnani, Gorry and Oprea (2016) document a failure to precisely optimize in a related task, which can similarly be attributed to inattention, though there is no need to rely on memory in order to make an optimal decision in their experiment.

adjustment is the positive serial correlation in the subjects' errors, conditional on adjustment. As we have already shown, subjects do not perfectly close the gap between their forecast and the optimal forecast upon adjustment. We further find that the post-adjustment gap is serially correlated. Conversely, our model predicts zero autocorrelation: the remaining gap is attributed to random noise in the signal received at the time of the adjustment decision, assumed to be independent of all previous signals and actions. To a significant extent, the serial correlation in our experimental data is driven by serial correlation in the errors associated with slider adjustments that are nearby in time. For example, among adjustments that occur less than five ring draws apart, the first autocorrelation coefficient is 64.3%, whereas for adjustments that occur more than 20 ring draws apart, the first autocorrelation coefficient is 21.5%. A possible interpretation of this finding is that rather than the signal upon which the decision is based being a function of the complete history of ring draws (limited only by the cost of a more precise signal about that history), it is based on an imprecise memory of this history, with errors that propagate over time. We leave for future work an investigation of more complex models of inattentive adjustment of this kind.

The similarity of the pattern of adjustment of the slider in our experiment to retailers' adjustments of the prices that they charge raises the possibility that the failure of such prices to more perfectly track the retailer's currently optimal (profit-maximizing) price should be similarly attributed to rational inattention of the kind that we model. Such a conclusion would have important consequences for monetary economics, as discussed in Woodford (2009) and Stevens (2015). We believe that experimental studies such as this one can play an important role in advancing our understanding, not simply of the general importance of inattention in adjustment dynamics, but of the specific models of inattentive decision-making that are most consistent with what we know about human capabilities — and by doing so, can contribute to the construction of more empirically realistic macroeconomic models.

References

- Albers, Wulf & Glsela Albers (1983), “On the prominence structure of the decimal system,” *Advances in Psychology* 16: 271–287.
- Alvarez, Fernando E., Francesco Lippi & Luigi Paciello (2011), “Optimal price setting with observation and menu costs,” *The Quarterly Journal of Economics* 126(4): 1909–1960.
- Binder, Carola C. (2017), “Measuring uncertainty based on rounding: New method and application to inflation expectations,” *Journal of Monetary Economics* .
- Burnham, Kenneth P. & David R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer, New York, 2d ed.
- Caballero, Ricardo J. & Eduardo M.R.A. Engel (1993), “Microeconomic adjustment hazards and aggregate dynamics,” *The Quarterly Journal of Economics* 108(2): 359–383.
- Caballero, Ricardo J. & Eduardo M.R.A. Engel (1999), “Explaining investment dynamics in U.S. manufacturing: A dynamic (S,s) approach,” *Econometrica* 67(4): 783–826.
- Caballero, Ricardo J. & Eduardo M.R.A. Engel (2007), “Price stickiness in Ss models: New interpretations of old results,” *Journal of Monetary Economics* 54(S): 100–121.
- Calvo, Guillermo (1983), “Staggered prices in a utility-maximizing framework,” *Journal of Monetary Economics* 12: 383–398.
- Cheremukhin, Anton, Anna Popova & Antonella Tutino (2011), “Experimental evidence on rational inattention,” *Federal Reserve Bank of Dallas Working Paper* 1112.
- Costain, James & Anton Nakov (2014), “Logit price dynamics,” *European Central Bank Working Paper* 1693.
- Dotsey, Michael & Robert G. King (2005), “Implications of state-dependent pricing for dynamic macroeconomic models,” *Journal of Monetary Economics* 52(S): 213–242.
- Dotsey, Michael, Robert G. King & Alexander L. Wolman (1999), “State-dependent pricing and the general equilibrium dynamics of money and output,” *Quarterly Journal of Economics* 114: 655–690.
- Eichenbaum, Martin, Nir Jaimovich & Sergio Rebelo (2011), “Reference prices, costs, and nominal rigidities,” *The American Economic Review* 101(1): 234–262.

- Gallistel, Charles R., Monika Krishan, Ye Liu, Reilly Miller & Peter E. Latham (2014a), “The perception of probability,” *Psychological Review* 121(1): 96–123.
- Gallistel, Charles R., Monika Krishan, Ye Liu, Reilly Miller & Peter E. Latham (2014b), “The Perception of Probability,” *Psychological Review* 121(1): 96–123.
- Jung, Junehyuk, Jeong-Ho Kim, Filip Matejka & Christopher A. Sims (2015), “Discrete actions in information-constrained decision problems,” Tech. rep.
- Kaplan, Greg & Guido Menzio (2015), “The morphology of price dispersion,” *International Economic Review* 56(4): 1165–1206.
- Khaw, Mel Win, Luminita Stevens & Michael Woodford (2016), “Empirical models of inattentive choice,” *unpublished* .
- Levy, Daniel, Dongwon Lee, Haipeng Chen, Robert J. Kauffman & Mark Bergen (2011), “Price points and price rigidity,” *Review of Economics and Statistics* 93(4): 1417–1431.
- Lucas, Jr., Robert E. (1972), “Expectations and the neutrality of money,” *Journal of Economic Theory* 4: 103–124.
- Maćkowiak, Bartosz A. & Mirko Wiederholt (2009), “Optimal sticky prices under rational inattention,” *The American Economic Review* 99(3): 769–803.
- Magnani, Jacopo, Aspen Gorry & Ryan Oprea (2016), “Time and state dependence in an Ss decision experiment,” *American Economic Journal: Macroeconomics* 8(1): 285–310.
- Mankiw, N. Gregory & Ricardo Reis (2002), “Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips Curve,” *The Quarterly Journal of Economics* 117(4): 1295–1328.
- Matějka, Filip (2015), “Rationally inattentive seller: Sales and discrete pricing,” *The Review of Economic Studies* 83(3): 1125–1155.
- McKelvey, Richard D. & Talbot Page (1990), “Public and private information: An experimental study of information pooling,” *Econometrica* 58(6): 1321–1339.
- Muth, John F (1961a), “Rational expectations and the theory of price movements,” *Econometrica: Journal of the Econometric Society* pp. 315–335.
- Muth, John F (1961b), “Rational expectations and the theory of price movements,” *Econometrica: Journal of the Econometric Society* pp. 315–335.
- Peterson, Cameron R. & Lee R. Beach (1967a), “Man as an intuitive statistician,” *Psychological Bulletin* 68: 29–46.

- Peterson, Cameron R. & Lee R. Beach (1967b), “Man as an Intuitive Statistician,” *Psychological Bulletin* 68: 29–46.
- Reis, Ricardo (2006), “Inattentive producers,” *The Review of Economic Studies* 73(3): 793–821.
- Ricci, Matthew & Randy Gallistel (2017a), “Accurate step-hold tracking of smoothly varying periodic and aperiodic probability,” *Attention, Perception, & Psychophysics* pp. 1–15.
- Ricci, Matthew & Randy Gallistel (2017b), “Accurate step-hold tracking of smoothly varying periodic and aperiodic probability,” *Attention, Perception, & Psychophysics* pp. 1–15.
- Robinson, Gordon H. (1964), “Continuous estimation of a time-varying probability,” *Ergonomics* 7: 7–21.
- Sheshinski, Eytan & Yoram Weiss (1977), “Inflation and costs of price adjustment,” *Review of Economic Studies* 44: 287–303.
- Sheshinski, Eytan & Yoram Weiss (1983), “Optimum pricing policy under stochastic inflation,” *Review of Economic Studies* 50: 513–529.
- Sims, Christopher A. (2003), “Implications of rational inattention,” *Journal of Monetary Economics* 50(3): 665–690.
- Sims, Christopher A. (2011), “Rational inattention and monetary economics,” in *Handbook of Monetary Economics, volume 3A*, B.M. Friedman & M. Woodford, eds., Elsevier, Amsterdam.
- Stevens, Luminita (2015), “Coarse pricing policies,” *Working paper* .
- Taylor, John B. (1980), “Aggregate dynamics and staggered contracts,” *Journal of Political Economy* 88: 1–23.
- Woodford, Michael (2003), “Imperfect common knowledge and the effects of monetary policy,” in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, J. Stiglitz P. Aghion, R. Frydman & M. Woodford, eds., pp. 25–58, Princeton University Press, Princeton, NJ.
- Woodford, Michael (2009), “Information-constrained state-dependent pricing,” *Journal of Monetary Economics* 56(S): 100–124.

Online Appendix for Discrete Adjustment to a Changing Environment: Experimental Evidence

Mel Win Khaw
Columbia University

Luminita Stevens
University of Maryland

Michael Woodford
Columbia University

August 2017

A Experimental Setup

We modified the experiment of Gallistel et al (2014) along five dimensions, to make it more suitable to the focus of this paper: (1) we added a monetary reward that varies with performance, (2) we simplified the data generating process and the way we elicit responses, to ensure the tractability of the Bayesian inference problem, (3) we added a numerical representation of the subject's estimate, in addition to the visual one, (4) we eliminated the possibility of discrete step sizes of the slider and ensured smooth adjustment of the slider, and (5) we explained all features of the setup and the data generating process to the subjects.

The reward function allows us to have more control over the objective of our subjects, rather than having to make assumptions about what they might be maximizing.

*We would like to thank Fernando Alvarez, Marios Angeletos, Marty Eichenbaum, Eduardo Engel, Benjamin Hebert, Cosmin Ilut, Filip Matějka, Rosemarie Nagel, Erkut Ozbay, Isabel Trevino, Laura Veldkamp, Alex Wolman, and conference and seminar participants at the Barcelona GSE Summer Forum, the NBER Summer Institute, the University of Maryland, the Cowles Summer Conference on Macroeconomics, and the JME-SNB-SCG conference for useful comments, and the National Science Foundation for support of this research.

We incentivized the subjects to do a particular task and we also displayed their cumulative score in real time, so they could see how their strategies affect their reward. We showed the subjects both a visual and a numerical representation of their current estimate for multiple reasons. On the one hand, some subjects may have lower numeracy than others. Additionally, it has been shown that probabilities are notoriously difficult to interpret and communicate without some visual representation. The box showing them their current estimate helps the subjects visualize exactly what they are trying to estimate. On the other hand, subjects may want to know what probability they are setting with more precision. In fact, displaying the numerical representation allows us to uncover evidence of a cognitive “round number” bias.

We conducted the experiment with 11 subjects, who each completed 10 sessions over the course of several days. Each session had 1,000 ring draws and on average lasted 27 minutes. The subjects were undergraduate and Master’s students at Columbia University. At the start of each session, the subjects received both written and verbal instructions and they completed a 15-minute practice session to familiarize themselves with the task.

B Optimal Bayesian Inference

Given a sample of T observations, we wish to determine the posterior distribution over (n, p) , where p is the most recent probability of drawing a 1 (namely a green ring) and n is the number of periods for which the current regime has lasted so far. A model of the data is specified by a probability p of drawing a 1 in the most recent regime and a partition $\pi = \{n_i\}$ of the sample into successive regimes, where n_i is the length of regime i . Let τ_i denote the last observation of regime i . The likelihood of the most recent n observations if the regime has been p over that time is

$$L(n, p) = p^{k_n} (1 - p)^{n - k_n}, \quad (\text{B.1})$$

where k_n is the number of successes in the n most recent observations. Let

$$L(n) \equiv \int L(n, p) f(p) dp. \quad (\text{B.2})$$

and let $L_\tau(n)$ denote the average likelihood computed using the n observations ending with observation τ . The ex-ante joint probability of the model (π, p) being correct and the data being a particular observed sequence is given by

$$\mu(\pi) \left(\prod_{i=1}^{N(\pi)-1} L_{\tau_i}(n_i) \right) f(p) L(n, p),$$

where $N(\pi)$ is the number of regimes under partition π and μ_π is the ex-ante probability of partition π occurring in a sample of length T ,

$$\mu(\pi) = (1 - \delta)^{T - N(\pi) + 1} (\delta)^{N(\pi) - 1}. \quad (\text{B.3})$$

Summing over the set $\Pi(n)$ of all possible partitions for which the final regime is of length n , we define

$$Q(n) \equiv \sum_{\pi \in \Pi(n)} \mu(\pi) \prod_{i=1}^{N(\pi)-1} L_{\tau_i}(n_i). \quad (\text{B.4})$$

The posterior probability of (n, p) is

$$P(n, p) = \frac{Q(n)f(p)L(n, p)}{\sum_{n \geq 1} Q(n)L(n)}. \quad (\text{B.5})$$

The expected value of p sums over all n and integrates over p using the measure $P(n, p)$. The Bayesian estimate for the probability of drawing a 1 on the next observation takes into account the fact that the regime might change on the next draw, which occurs with probability δ , and in which case, the estimate of the probability is 0.5:

$$B = (1 - \delta) \int \sum_{n \geq 1} pP(n, p)dp + \frac{\delta}{2}. \quad (\text{B.6})$$

Figure B.1 plots the distribution of the size of slider adjustments in our data (left panel) and in the Bayesian benchmark. There is considerable variation in the size of the slider adjustments when they occur, whereas according to the Bayesian benchmark, the slider should be adjusted on every trial, and the adjustments should seldom be very large, as shown by the right panel of the figure.

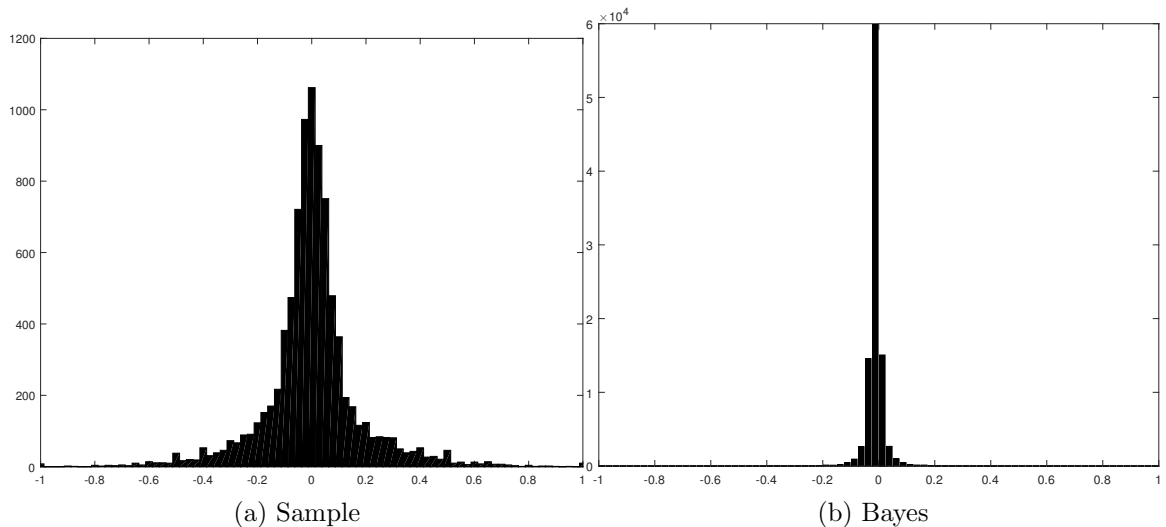


Figure B.1: Distribution of the size of changes in the slider position $\Delta \hat{p}_t$. Left panel: our data, counting only trials on which the slider is moved (all subjects and all sessions). Right panel: prediction under the hypothesis of Bayesian rationality.

C Model Selection Method

We quantify the performance of different models by computing how much the likelihood of our data is increased by allowing the adjustment hazard Λ to depend on the model-implied gap Δ_t between the optimal forecast and the current position. If we assume that the adjustment decision α_t (taking the value 1 if the slider is adjusted, and zero otherwise) is an independent draw on each trial from a Bernoulli distribution with probability $\Lambda(\Delta_t)$ of adjustment, then the log likelihood of the data sequence $\{\alpha_t\}$ is equal to

$$LL = \sum_t \{\alpha_t \log \Lambda(\Delta_t) + (1 - \alpha_t) \log(1 - \Lambda(\Delta_t))\}.$$

We should not, however, be interested in the question of how large LL can be made through choice of an arbitrarily complex function $\Lambda(\Delta)$, as this is likely to allow severe over-fitting of the particular adjustment sequence that happens to have been observed (with sufficiently finely differentiated measures of Δ_t on the separate trials). In order to avoid such over-fitting, we penalize unduly complex hypotheses about the way that the adjustment hazard can depend on the gap, and we choose both the family of models (the value of k) and a specific model within that family (the value of the parameter vector θ) so as to minimize the Bayes Information Criterion (BIC)¹

$$BIC \equiv -2LL + k \log N, \tag{C.1}$$

where N is the number of observations.

Suppose we have two model families represented by the sets of densities

$$\mathcal{M}_j = \left\{ p(\cdot | \theta_j) : \theta_j \in \Theta_j \right\}, \quad j \in \{1, 2\} \tag{C.2}$$

to explain the random vector y . Given n observations $(y_1, \dots, y_n) = Y$, the relative

¹See, for example, Burnham and Anderson (2002, chap. 6) for discussion of this approach to model selection.

posterior probability that the true model belongs to family 1 rather than alternative family 2 is given by

$$\log \frac{P^{post}(\mathcal{M}_1|Y)}{P^{post}(\mathcal{M}_2|Y)} = \log \frac{P^{prior}(\mathcal{M}_1)}{P^{prior}(\mathcal{M}_2)} + B_{12}, \quad (\text{C.3})$$

where the Bayes factor includes a penalty for the relative number of parameters of the best fitting model in family 1,

$$\begin{aligned} B_{12} &= \log \frac{P(Y|\mathcal{M}_1)}{P(Y|\mathcal{M}_2)} - \frac{1}{2}(k_1 - k_2) \log N \\ &= \frac{1}{2} \left(BIC(\mathcal{M}_1) - BIC(\mathcal{M}_2) \right). \end{aligned} \quad (\text{C.4})$$

D Value Functions Approximation

This appendix describes the estimation method used to compute the value function for the random menu cost model and for the inattention model. We estimate an atheoretical statistical model of subjects' behavior after any trial t conditional on the slider setting p chosen on that trial and the posterior π_t at that point, and we use this atheoretical model to compute an implied value function $V_t(p)$.

D.1 Empirical Model

Using our experimental data, we fit an empirical model of the dynamics for the Bayesian posterior mean p_t^* , for the Bayesian gap $p_t^* - p_{t-1}$, and for the probability of adjustment Λ_t :

$$p_{t+1}^* - \frac{1}{2} = \mu \left(p_t^* - \frac{1}{2} \right) + v_t, \quad \mu < 1 \quad (\text{D.1})$$

$$p_{t+1}^* - p_t = \lambda (p_t^* - p_{t-1}) + u_t, \quad \lambda > 0 \quad (\text{D.2})$$

$$\log \left(\frac{\Lambda_t}{1 - \Lambda_t} \right) = \theta_0 + \theta_2 (p_t^* - p_{t-1})^2 \quad (\text{D.3})$$

where v_t and u_t are two i.i.d. mean-zero random variables with variances σ_u^2 and σ_v^2 , respectively.

Let us also approximate the CDF $G(\kappa)$ of fixed costs by a function of the form

$$\log \left(\frac{G(\kappa)}{1 - G(\kappa)} \right) \approx \gamma_0 + \gamma_1 \kappa, \quad \gamma_1 > 0. \quad (\text{D.4})$$

D.2 Random Fixed Cost Model

In the random fixed cost model, the expected value of a slider position p is the expected value of all future expected monetary rewards, net of the expected value of all future adjustment costs that are incurred:

$$V_t(p) = R_t(p) - K_t(p), \quad (\text{D.5})$$

where note that the continuation value also depends on π_t , the distribution of posterior beliefs at the beginning of trial t , which has been suppressed to reduce notation, and where

$$R_t(p) \equiv \mathbb{E} \left[\sum_{s=t}^T r(p_s; s_s) \middle| p_t = p, \pi_t \right], \quad (\text{D.6})$$

$$K_t(p) \equiv \mathbb{E} \left[\sum_{s=t+1}^T \mathbf{1}_s \kappa_s \middle| p_t = p, \pi_t \right], \quad (\text{D.7})$$

where $\mathbf{1}_s$ is an indicator variable equal to 1 if the slider is adjusted on trial s and 0 otherwise, and κ_s is the value of the menu cost drawn on trial s . We can alternatively write R_t neglecting a term that is independent of the position p_t , and thus irrelevant to our calculation of the continuation value:

$$R_t(p) = - \sum_{s=t}^T \mathbb{E} \left[(p_s - p_s^*)^2 \middle| p_t = p, \pi_t \right] + t.i.p. \quad (\text{D.8})$$

For t perceived by the subjects to be far enough away from the terminal trial T , we can approximate the value of R_t by the value of

$$R^\infty(p) = - \sum_{s=t}^{\infty} \mathbb{E} \left[(p_s - p_s^*)^2 \middle| p_t = p, \pi_t \right] + t.i.p. \quad (\text{D.9})$$

We can now use the empirical model described above, of the joint dynamics of the slider position and of the Bayesian posterior mean to compute a numerical estimate of R^∞ for any hypothetical slider position that may be chosen on trial t , given the posterior π_t at that time:

$$\begin{aligned} R^\infty(p_t) &= \left[\frac{2\lambda^2\mu}{1-\lambda^2} + \frac{2\lambda\mu(1-\mu)}{1-\lambda\mu} \right] \left(p_t^* - \frac{1}{2} \right) (p_t - p_t^*) \\ &+ \frac{\lambda^2}{1-\lambda^2} \left[\left(p_t^* - \frac{1}{2} \right)^2 - \left(p_t - \frac{1}{2} \right)^2 \right] - (p_t - p_t^*)^2 + t.i.p. \end{aligned} \quad (\text{D.10})$$

We next consider the numerical estimation of K_t , which we can alternatively write

as

$$K_t(p) = \sum_{s=t+1}^T \mathbb{E} \left[\Gamma(\Lambda_s) \middle| p_t = p, \pi_t \right], \quad (\text{D.11})$$

where Λ_s is the probability of adjustment on trial s , and where, denoting by G the CDF of fixed adjustment costs,

$$\Gamma(\Lambda) \equiv \Lambda E \left[\kappa \middle| G(\kappa) \leq \Lambda \right] \quad (\text{D.12})$$

multiplies the probability of adjustment by the mean adjustment cost conditional on κ being below the threshold that is required for adjustment on a trial with that particular probability of adjustment.

Given an estimate of the function $\Gamma(\Lambda)$, we can use a purely empirical model of the adjustment hazard (more precisely, of the joint dynamics of the slider position, the Bayesian posterior, and the probability of adjustment) to obtain a numerical estimate of the function $K_t(p)$ for each trial. However, the function $\Gamma(\Lambda)$ depends on the distribution of fixed costs G , which in turn can only be inferred using a measure of the continuation value we are trying to estimate. Hence we must jointly estimate the parameters of the distribution of fixed costs *and* determine the best-fitting model of the adjustment decision, using the estimate of the continuation value that is implied by this parameterization of $\Gamma(\Lambda)$.

As above, we can neglect the terms that are independent of the slider position p_t , and, as long as t is not too close to T , we can use the approximation

$$K^\infty(p) = \sum_{s=t+1}^{\infty} \left\{ \mathbb{E} \left[\Gamma(\Delta_s) \middle| p_t = p, \pi_t \right] - \mathbb{E} \left[\Gamma(\Delta_s) \middle| p_t = p_t^*, \pi_t \right] \right\}. \quad (\text{D.13})$$

Using the approximation for the distribution of fixed costs G , we have that

$$\Gamma(\Lambda) \approx \frac{1}{\gamma_1} D(\Lambda | \underline{\Lambda}), \quad (\text{D.14})$$

where

$$D(\Lambda||\bar{\Lambda}) \equiv \Lambda \log \left(\frac{\Lambda}{\bar{\Lambda}} \right) + (1 - \Lambda) \log \left(\frac{1 - \Lambda}{1 - \bar{\Lambda}} \right). \quad (\text{D.15})$$

Using the empirical model of adjustment, we can approximate the relative entropy D by

$$D(\Lambda||\bar{\Lambda}) \approx \frac{\theta_2^2}{2} \frac{\exp(\theta_0)}{[1 + \exp(\theta_0)]^2} (p_{t+1}^* - p_t)^4. \quad (\text{D.16})$$

Then, using the empirical model for the dynamics of the Bayesian gap, and neglecting terms that are independent of p_t , we obtain

$$K^\infty(p) \approx \frac{\theta_2^2}{2\gamma_1} \frac{\exp(\theta_0)}{[1 + \exp(\theta_0)]^2} [p_t^* k(p_{t+1}^*(G) - p) + (1 - p_t^*) k(p_{t+1}^*(R) - p)], \quad (\text{D.17})$$

where $p_{t+1}^*(x)$ is the Bayesian posterior mean p_{t+1}^* in the case that the ring draw on trial t is x , and where

$$k(p_{t+1}^* - p_t) = \frac{1}{1 - \lambda^4} (p_{t+1}^* - p_t)^4 + \frac{6\lambda^2 \sigma_u^2}{(1 - \lambda^2)(1 - \lambda^4)} (p_{t+1}^* - p_t)^2. \quad (\text{D.18})$$

The parameters that must be evaluated in order to compute V_t for each trial are

1. The parameters $(\lambda, \mu, \sigma_u^2)$ of the empirical laws of motion (D.1)-(D.2).
2. The parameters (θ_0, θ_1) of the empirical hazard function in (D.3).
3. The parameter γ_1 of the estimated distribution of fixed costs in (D.4).

The parameter γ_1 however is estimated using our numerical estimate of the value gap, $\Delta_t^{value} \equiv \max_p V_t(p) - V_t(\hat{p}_{t-1})$. Hence we must solve a fixed point problem: a conjectured value of γ_1 is used to compute $V_t(p)$ for arbitrary p , and hence the value of Δ_t^{value} on each trial; these values are then used to estimate a value of γ_1 , and this value must turn out to be the same value as was assumed in order to compute Δ_t^{value} .

D.3 Inattention Model

In our model with information costs, the continuation value can be written in the form

$$V_t(p) = R_t(p) - \psi_1 H_{1t}(p) - \psi_2 H_{2t}(p), \quad (\text{D.19})$$

where $R_t(p)$ is defined as in (D.6), and $H_{1t}(p)$ and $H_{2t}(p)$ are the expected cumulative costs (summing both information costs and the intrinsic costs of different actions) of the two decisions on subsequent trials, if the slider is set at p on trial t , and the posterior at that time is π_t :

$$H_{1t}(p) \equiv \mathbb{E} \left[\sum_{s=t+1}^T D(\Lambda_s | \tilde{\Lambda}) \Big| p_t = p, \pi_t \right], \quad (\text{D.20})$$

$$H_{2t}(p) \equiv \mathbb{E} \left[\sum_{s=t+1}^T \Lambda_s D(\mu_s | \tilde{\mu}) \Big| p_t = p, \pi_t \right], \quad (\text{D.21})$$

where Λ_s is the probability of adjustment on trial s , μ_s is the probability distribution over new slider positions if the slider is adjusted on trial s , $\tilde{\Lambda}$ and $\tilde{\mu}$ are the reference measures, and $D(\mu | \lambda) = \int \log \frac{d\mu}{d\lambda} d\mu$ is the Kullback-Leibler divergence between two distributions μ and λ .

As above, we can compute a numerical estimate of the function $R_t(p)$ for each trial, on the basis of an atheoretical empirical model of the joint dynamics of the slider position and the Bayesian posterior on trials subsequent to t . The numerical estimate that we use is the same as in our estimation of the model of optimization subject to a fixed cost of adjustment.

We similarly compute numerical estimates of the functions $H_{1t}(p)$ and $H_{2t}(p)$, using empirical models of the joint dynamics of the Bayesian posterior, the slider position, the adjustment probability Λ_t , and the time-varying measure over possible new slider positions μ_t . We proceed as follows: First, using the experimental data, we estimate a value of Λ_t for each value of the Bayesian gap $p_t^* - p_{t-1}$. From this, we

compute a value for $D(\Lambda_t||\tilde{\Lambda})$, given $\tilde{\Lambda}$. We then fit a relationship of the form

$$D(\Lambda_t||\tilde{\Lambda}) \approx a(p_t^* - p_{t-1})^4 + b(p_t^* - p_{t-1})^2 + \text{const.} \quad (\text{D.22})$$

Using this approximation together with the law of motion for the Bayesian gap, we can then estimate

$$H_{1t}(p) \approx p_t^* h_1(p_{t+1}^*(G) - p) + (1 - p_t^*) h_1(p_{t+1}^*(R) - p) + t.i.p. \quad (\text{D.23})$$

where

$$h_1(p_{t+1}^* - p_t) = \frac{a}{1 - \lambda^4} (p_{t+1}^* - p_t)^4 + \left[\frac{6a\lambda^2\sigma_u^2}{(1 - \lambda^2)(1 - \lambda^4)} - \frac{b}{1 - \lambda^2} \right] (p_{t+1}^* - p_t)^2. \quad (\text{D.24})$$

Similarly, we can obtain an estimate for $H_{2t}(p)$, by first fitting to the data the functional relationships

$$\Lambda_t \approx m(p_t^* - p_{t-1})^2 + n, \quad (\text{D.25})$$

and

$$D(\mu_t||\tilde{\mu}) \approx c \left(p_t^* - \frac{1}{2} \right)^2 + d. \quad (\text{D.26})$$

We obtain

$$H_{2t}(p) \approx p_t^* h_2 \left(p_{t+1}^*(G) - p, p_{t+1}^*(G) - \frac{1}{2} \right) + (1 - p_t^*) h_2 \left(p_{t+1}^*(R) - p, p_{t+1}^*(R) - \frac{1}{2} \right) + t.i.p. \quad (\text{D.27})$$

where

$$h_2 \left(p_{t+1}^* - p_t, p_{t+1}^* - \frac{1}{2} \right) = (mc \quad md \quad 0 \quad nc) (I - N)^{-1} \begin{pmatrix} (p_{t+1}^* - p_t)^2 (p_{t+1}^* - \frac{1}{2})^2 \\ (p_{t+1}^* - p_t)^2 \\ (p_{t+1}^* - p_t) (p_{t+1}^* - \frac{1}{2}) \\ (p_{t+1}^* - \frac{1}{2})^2 \end{pmatrix}, \quad (\text{D.28})$$

where the matrix N is given by

$$N \equiv \begin{pmatrix} \lambda^2 \mu^2 & \lambda^2 \sigma_v^2 & 4\lambda\mu\sigma_{uv} & \mu^2 \sigma_u^2 \\ 0 & \lambda^2 & 0 & 0 \\ 0 & 0 & \lambda\mu & 0 \\ 0 & 0 & 0 & \mu^2 \end{pmatrix}. \quad (\text{D.29})$$

Computing the estimates of $H_1(p)$ and $H_2(p)$ requires values for $\tilde{\Lambda}$ and $\tilde{\mu}$ (in order to estimate empirical models of the dynamics of the quantities $D(\Lambda_t|\tilde{\Lambda})$ and $D(\mu_t|\tilde{\mu})$), while the estimate of $V_t(p)$ also requires values for ψ_1 and ψ_2 . We start with a set of parameters and compute an estimate of the value function. Once we obtain a numerical estimate of the function $V_t(p)$, we can find the values of ψ_2 and $\tilde{\mu}$ that maximize the consistency of observed slider position choices with the model prediction for the adjustment decision. Using this estimated model of the slider position decision to estimate V_t^{adj} , we can then find the values of ψ_1 and $\tilde{\Lambda}$ that maximize the consistency of the observed timing of adjustments of the slider with the prediction for the slider position choice conditional on adjustment. Finally, we check whether the estimated values for ψ_1 , ψ_2 , $\tilde{\Lambda}$, and $\tilde{\mu}$ coincide with the values assumed in computing the value function $V_t(p)$. The numerical estimates reported here represent a solution to this fixed-point problem. In order to compute the best-fitting distribution of slider choices, $\tilde{\mu}$, we assume the parametric form $\tilde{\mu}(p) = A\bar{\mu}(p)^\gamma$, where A is a normalizing constant that ensures that $\tilde{\mu}$ is a probability distribution that integrates to 1. The parameter $\gamma \in [0, 1]$ determines how close the model-implied reference distribution is to the empirical unconditional distribution ($\gamma = 1$, which corresponds to the standard rational inattention model) versus the uninformative, uniform distribution ($\gamma = 0$).

E Overall Performance and Forecast Bias

Our subjects demonstrate the ability to track the varying state, although less well than would be possible given the available information. As described in the previous section, the subjects are rewarded using a function that depends on their mean square error. Figure E.1 compares the scores of the subjects and the Bayesian decisionmaker to two polar benchmarks: the *complete information* case in which the subject knows the hidden probability at all times, and the *no information* case in which the subject is fully rational but does not see any rings and therefore sets the slider to the unconditional estimate of 0.5 on all draws. For each observer type, the figure shows 11 bars, each corresponding to the realized draws seen by each of our 11 subjects across all sessions.

All the subjects outperform the *no information* benchmark, indicating that they are tracking the state at least to some extent. They all underperform the Bayesian decisionmaker: on average, the subjects' scores are 2.5% lower than that of the Bayesian observer. There is considerable heterogeneity in performance. One driver of this heterogeneity is the noise in the realized draws, which determines the difficulty of the forecasting problem. This can be seen in the differences in scores in the full information benchmark. Another driver of the heterogeneity in scores is heterogeneity in behavior across the subjects. Some subjects (such as subject 3) are much closer to the fully rational Bayesian benchmark than others (for example, subject 8).

E.1 Forecast Bias

A relatively simple way of characterizing the degree to which behavior resembles or differs from the rational (perfect Bayesian) benchmark is to ask to what extent forecasts correctly track the true state on average. One familiar diagnostic considers whether forecasts are *unbiased*, in the sense that $E[\hat{p}_t|p_t]$ is equal to p_t , the true state. Gallistel *et al.* (2014) emphasize the extent to which their subjects are close to perfectly unbiased, though they compare the median forecast conditional on the

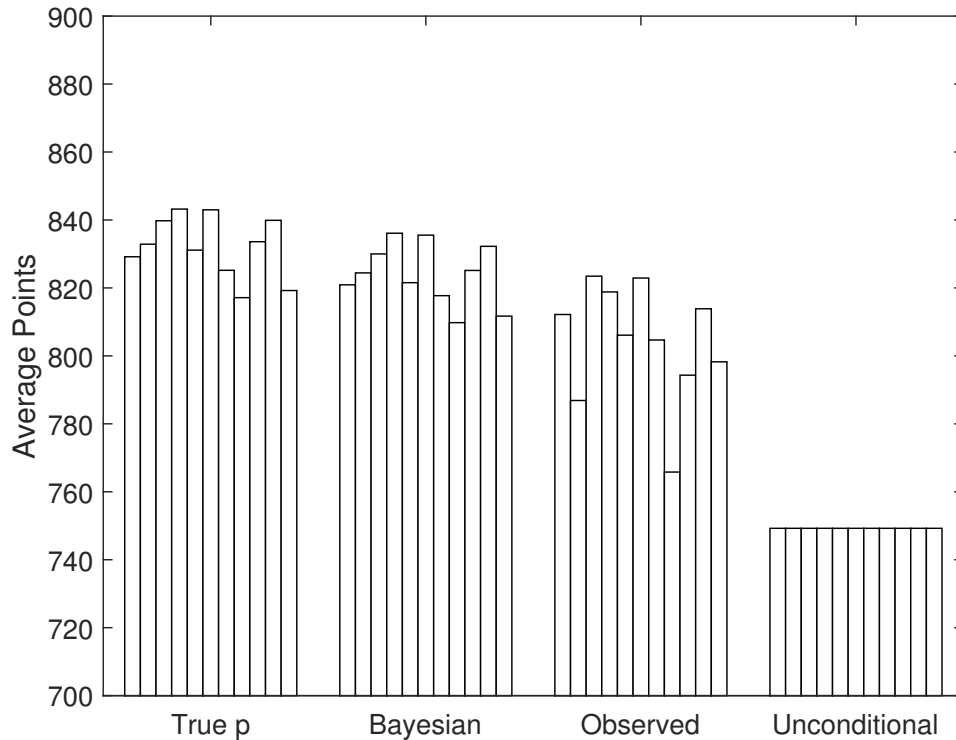


Figure E.1: Experiment data for all subjects and sessions. The scores are broken down by subject. “True p” refers to the score that each subject would have received, given the realized ring draws, had they known the underlying probabilities at all times. “Bayesian” refers to the scores that would have been received by each subject, given the realized ring draws, had they acted like the fully rational optimal Bayesian decisionmaker. “Unconditional” refers to the scores received under a constant forecast equal to the unconditional prior of 0.5.

true p , rather than the mean.² Robinson (1964) had similarly reported no significant forecast bias (using the more conventional definition based on the conditional mean) for any value of p , and this had been confirmed by additional studies summarized in Peterson and Beach (1967).

We examine this question in the case of our data in the upper left panel of Figure E.2. We pool the data from all subjects and sessions, and sort the data into 21 bins centered on multiples of 0.05, according to the true probability p_t on that trial. For each bin, the plot shows the mean slider position \hat{p}_t across those trials (as an x), together with the inter-quartile range of slider positions associated with that

²Ricci and Gallistel (2016) report a similar finding in a variant of the experiment of Gallistel *et al.* (2014) with a different data-generating process.

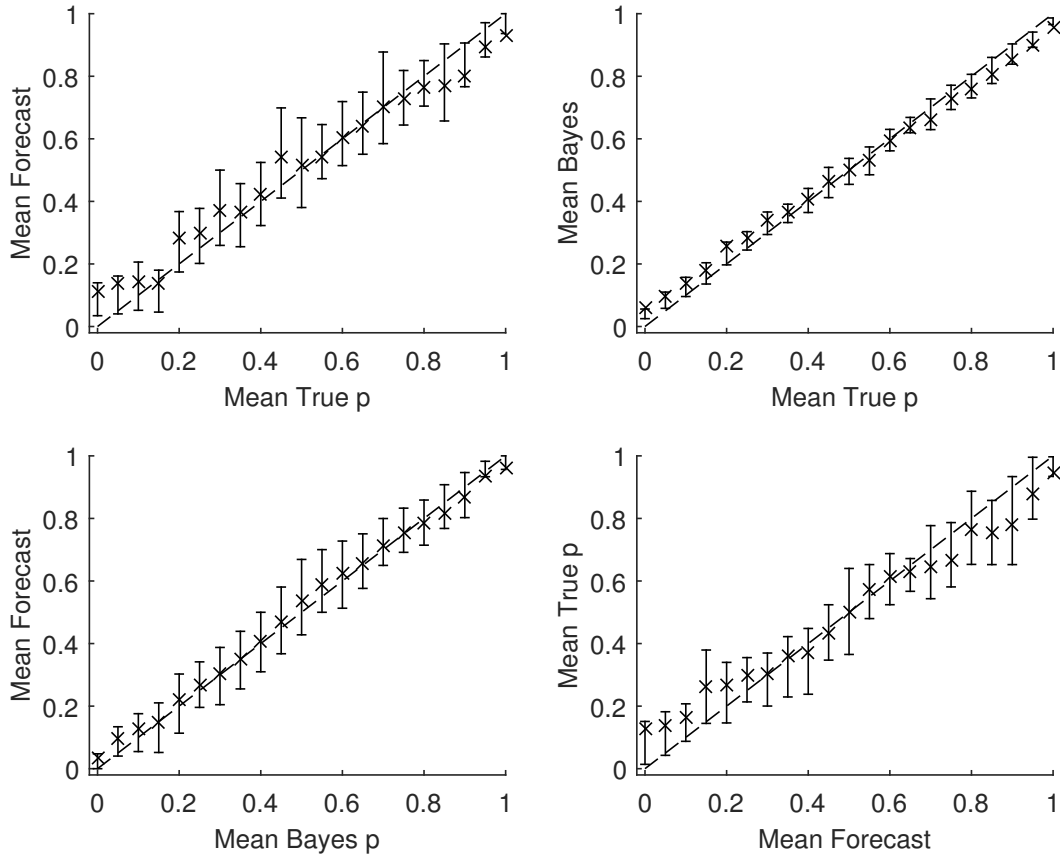


Figure E.2: Measures of forecast bias. Upper left: $E[\hat{p}|p]$ versus p . Upper right: $E[p^*|p]$ versus p . Lower left: $E[\hat{p}|p^*]$ versus p^* . Lower right: $E[p|p^*]$ versus p^* .

bin (shown as a vertical interval).³ As in the corresponding figure in Gallistel *et al.*,⁴ the diagonal (i.e., the predicted location of the means under the hypothesis of unbiasedness) is contained within the IQR in nearly every case, over the entire range of true probabilities. However, it is not quite right to conclude from this that the average slider position, conditional on any true state, is exactly that value. As indicated in Table E.1, a linear regression of p_t on \hat{p}_t , using our pooled data, yields a regression line with a slope slightly less than 1; and the null hypothesis that the regression line is the diagonal can be rejected with a p-value that is indistinguishable from zero (at the precision allowed by Matlab).

³The horizontal coordinate of the x that is plotted for each bin is the mean value of p_t for the trials in that bin. Under the hypothesis that $E[\hat{p}_t|p_t] = p_t$ for all values of p_t , we should also observe that $E[\hat{p}_t|p_t \in b_j] = E[p_t|p_t \in b_j]$, for any bin b_j , i.e., that the X for bin b_j should lie on the dashed diagonal line.

⁴See their Figure 6.

Regression: $y = \alpha + \beta x + \epsilon$				
y	x	α	β	F-stat
\hat{p}	p	0.086 (0.001)	0.849 (0.002)	4056
p^*	p	0.054 (0.001)	0.894 (0.001)	6901
p	\hat{p}	0.089 (0.001)	0.818 (0.002)	6273
\hat{p}	p^*	0.033 (0.001)	0.952 (0.002)	647

TABLE E.1: Regression tests of forecast bias, with alternative choices of the variables y and x , corresponding to the four panels of Figure E.2. Standard errors are shown in parentheses below the regression coefficients. The F statistic in each case is for a test of the null hypothesis that $\alpha = 0$ and $\beta = 1$, and under the null hypothesis should be distributed as $F(2, 109888)$. All null hypotheses have p-values of less than 10^{-279} .

Unbiasedness in this sense is not, however, the only hypothesis of interest, and is not even a prediction of rationality. Even an ideal Bayesian observer would not be able to track the hidden true probability perfectly, and as shown in the upper right panel of the figure, the forecasts of an ideal Bayesian observer would not be “unbiased” in the sense just mentioned: we should not expect $E[p_t^*|p_t]$ to equal p_t . Because of noise in the series of ring draws as an indication of the true probability, the ideal observer would on average over-estimate the probability when p_t is low, and under-estimate it when p_t is high. Bayesian rationality instead would require that (under the prior implied by our data-generating process) $E[p_t|p_t^*] = p_t^*$.

In fact, even if subjects do not produce fully optimal forecasts (due for example to memory limitations), if their forecasts are Bayes-rational *conditional on the information used to make the forecast*, they should satisfy the property that $E[p_t|\hat{p}_t] = \hat{p}_t$ for all slider positions \hat{p}_t . We check for this property in the lower right panel of Figure E.2, where the trials are now binned according the value of \hat{p}_t and the mean value of p_t associated with each bin is indicated by the vertical coordinate of the x. This rationality condition is not grossly rejected; the diagonal falls within the IQR (if only barely) for each of the bins. However, we note a fairly clear pattern in the figure, with the mean true p above the forecast whenever \hat{p} is below 0.3, and below the forecast

whenever \hat{p} is above 0.65. And indeed, a regression of p_t on \hat{p}_t using our pooled data yields a slope coefficient significantly below 1, as shown in Table E.1.

The pooled data from our subjects is instead more consistent with an alternative hypothesis, namely that subjects’ forecasts are distributed around the rational forecast, and equal to it on average — that is, that $E[\hat{p}_t|p_t^*] = p_t^*$ for all values of p_t^* . This is related to the hypothesis of unbiasedness that Gallistel *et al.* test, but recognizes that we can at best expect subjects’ forecasts to reflect the optimal Bayesian estimate of the state, and not the hidden state itself. It is also the hypothesis of “rational expectations” as originally proposed by Muth (1961), though that term has since come to be associated with the hypothesis of full Bayesian rationality, since the work of Lucas (1972).

This hypothesis is tested in the lower left panel of Figure E.2. We see that when the data are binned according to the value of p_t^* (rather than by the value of p_t , as in the upper left panel), the conditional mean indicated by the x is close to the diagonal for all bins. We can nonetheless reject that the hypothesis holds exactly in our data through a regression test, as shown in Table E.1; but note that in this case the F statistic is less gigantic than for the other null hypotheses considered in the table.

Note that even to the extent that this last hypothesis is accepted as an approximate characterization of our data, this does not mean that our subjects’ forecasts are Bayes-rational; if they (always) were, not only would the x marks be exactly on the diagonal, but the IQR would be an interval of zero length for each bin in the lower left panel of Figure E.2, which is not the case. Subjects’ forecasts can be more accurately characterized as equal to the rational forecast plus random noise (a characterization that also explains the slope less than 1 in the lower right panel).

F Additional Figures

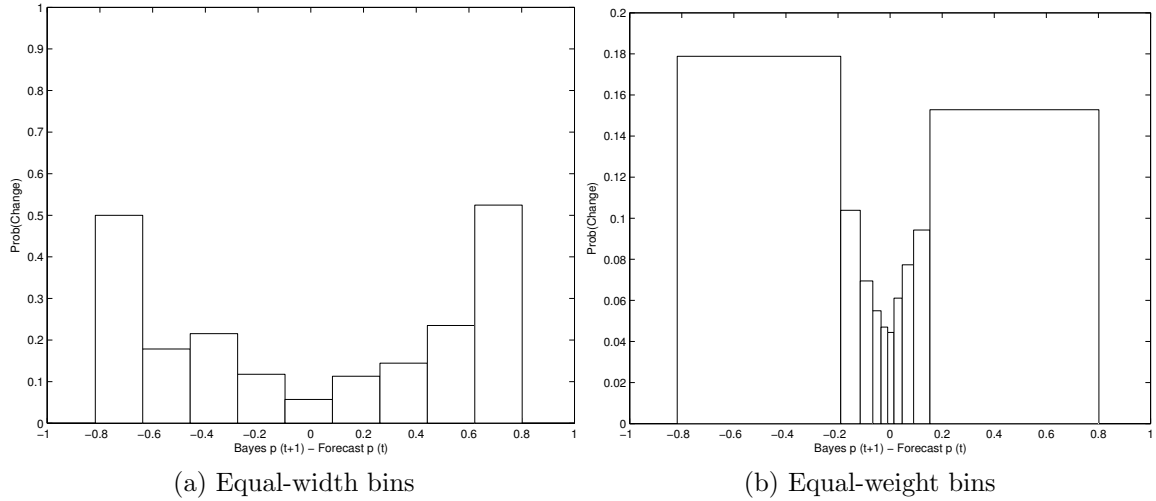


Figure F.1: Hazard functions for the slider adjustment probability as a function of the Bayesian gap. (a) Nonparametric empirical hazard plotting the fraction of trials on which adjustment occurs, for each range of values of the gap, with equal-width bins. (b) Similar plot, but with boundaries chosen so that each bin contains an equal number of observations.

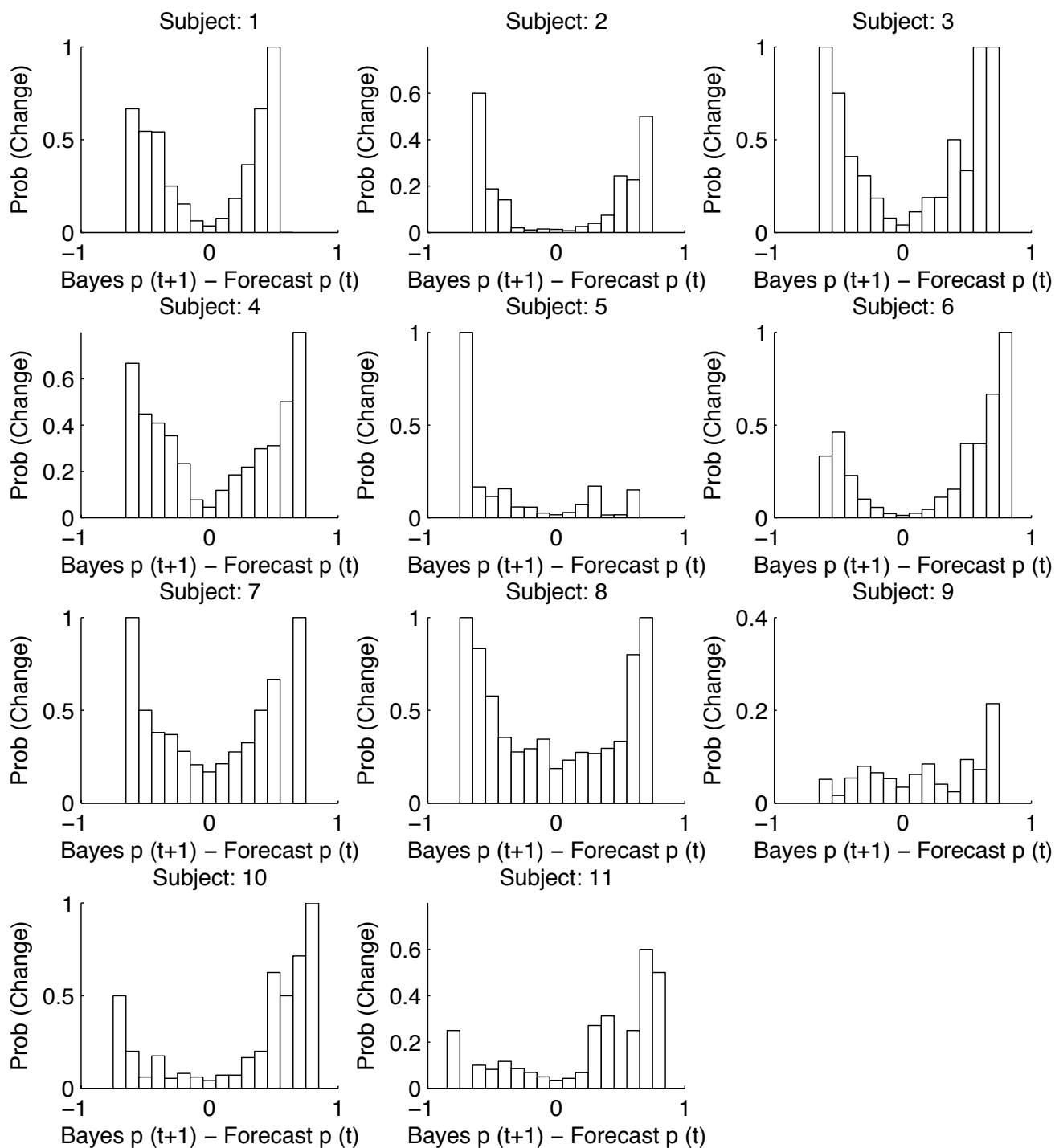


Figure F.2: Hazard functions for the slider adjustment probability as a function of the Bayesian gap for each individual subject.

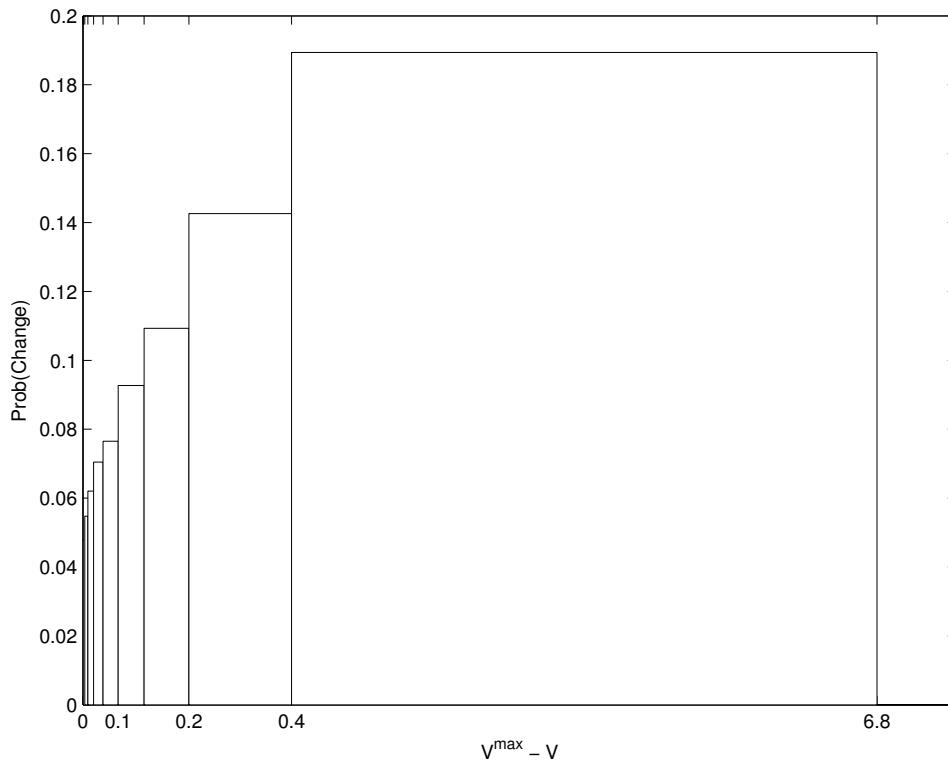


Figure F.3: Adjustment probability as a function of the value gap, computed under an assumption that if the slider is adjusted, it will be moved to the currently optimal position. Nonparametric empirical hazard with the boundaries are chosen so that each bin contains an equal number of observations.

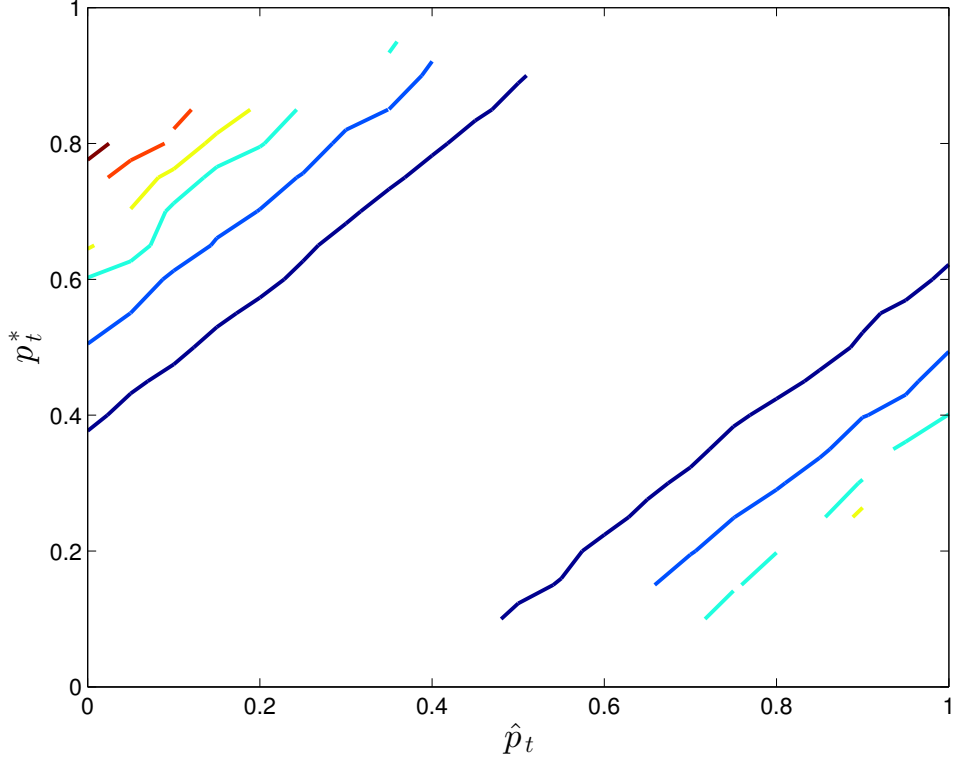


Figure F.4: A contour plot of the average value gap Δ_t associated with alternative pairs of values for the slider position \hat{p}_t [the horizontal axis] and the Bayesian posterior mean p_t^* [the vertical axis], when the continuation value function $V_t(p)$ is numerically approximated on the basis of an empirical model of the subsequent adjustment dynamics. We see that the value gap is related to the Bayesian gap (which corresponds, in the figure, to the distance of points from the diagonal). If the value gap were purely a monotonic function of the absolute value of the Bayesian gap, the contour lines would be the family of parallel straight lines with a slope exactly equal to one. In this case, the predictions of the optimizing model would be identical to those of the gap-based model with a symmetric hazard function. This is not quite true, since in the optimizing model, the value of adjusting also takes into account future expected adjustment costs, in addition to the current gap; nonetheless, the lines are close to those predicted by the Bayesian gap model.

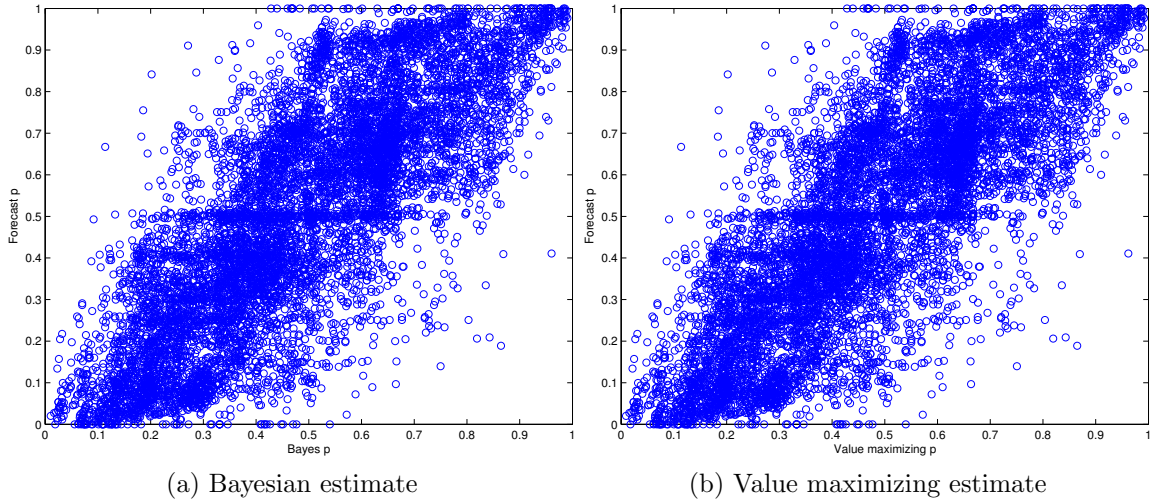


Figure F.5: The distribution of slider positions that are chosen, conditional on adjustment. (a) Scatter plot of the slider position [vertical axis] against the Bayesian optimal setting; (b) scatter plot of the slider position [vertical axis] against the one that would maximize the continuation value function. The relationship between the optimal Bayesian forecast and the position chosen by the subjects upon adjusting the slider is very noisy. Nonetheless, the Bayesian setting has predictive value: running a regression yields a slope that is very close to 1. Likewise, panel (b) also shows a large dispersion in the actual slider positions that are chosen, even conditioning on the currently optimal slider position (at least the one that would be implied by our estimate of the value function) — especially when the currently optimal position is in the middle of the interval of possible positions. This suggests a considerable degree of random error (presumably reflecting inattentiveness) when the slider is moved. Such errors must be taken into account in judging what an optimal decision should be as to whether the slider should be adjusted at all, as they should reduce the expected increase in payoff that can be achieved by adjustment.

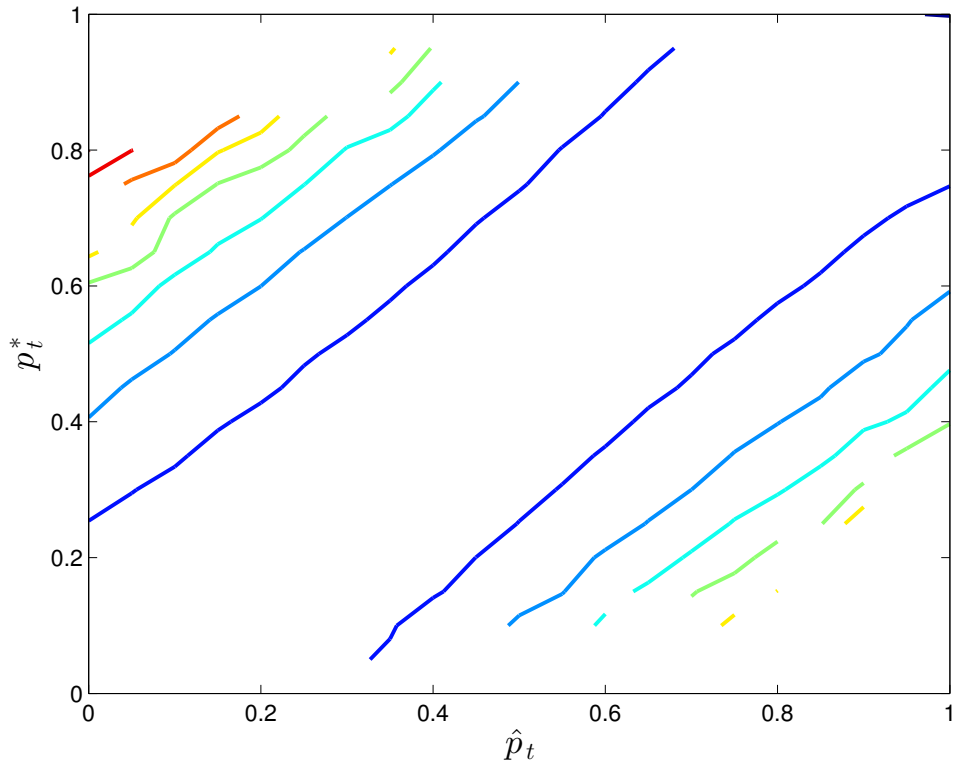


Figure F.6: A contour plot of the average RI value gap Δ_t associated with alternative pairs of values for the slider position \hat{p}_t [the horizontal axis] and the Bayesian posterior mean p_t^* [the vertical axis].

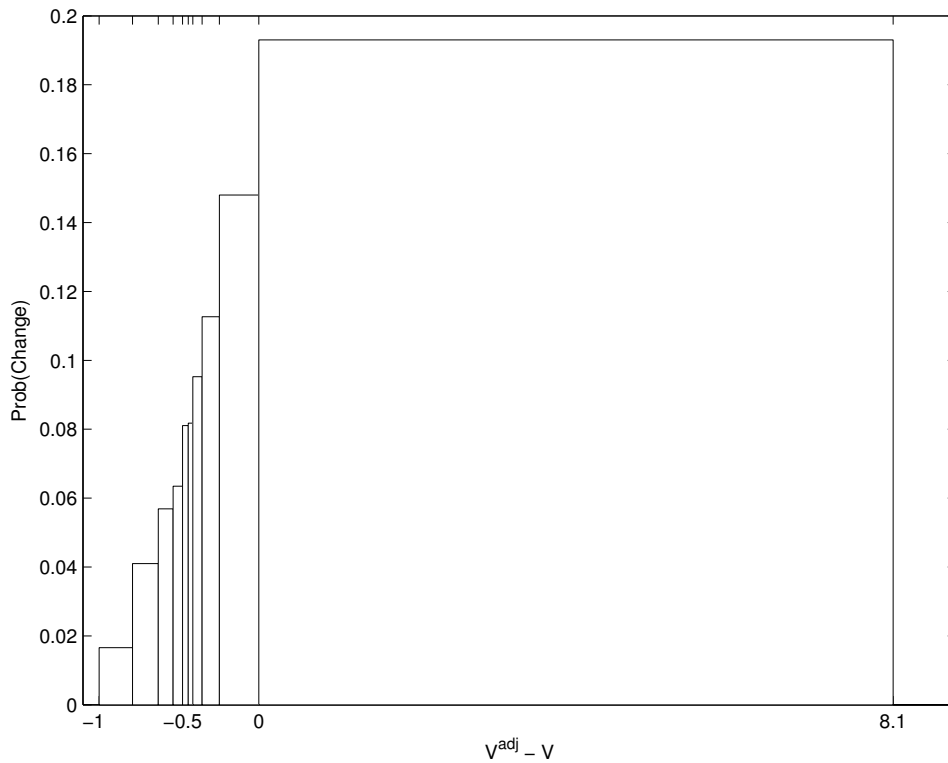


Figure F.7: Adjustment probability as a function of the value gap implied by the rational inattention model. Nonparametric empirical hazard, with the boundaries are chosen so that each bin contains an equal number of observations. We see that the adjustment probability is monotonically increasing with this gap measure, as predicted by the model.

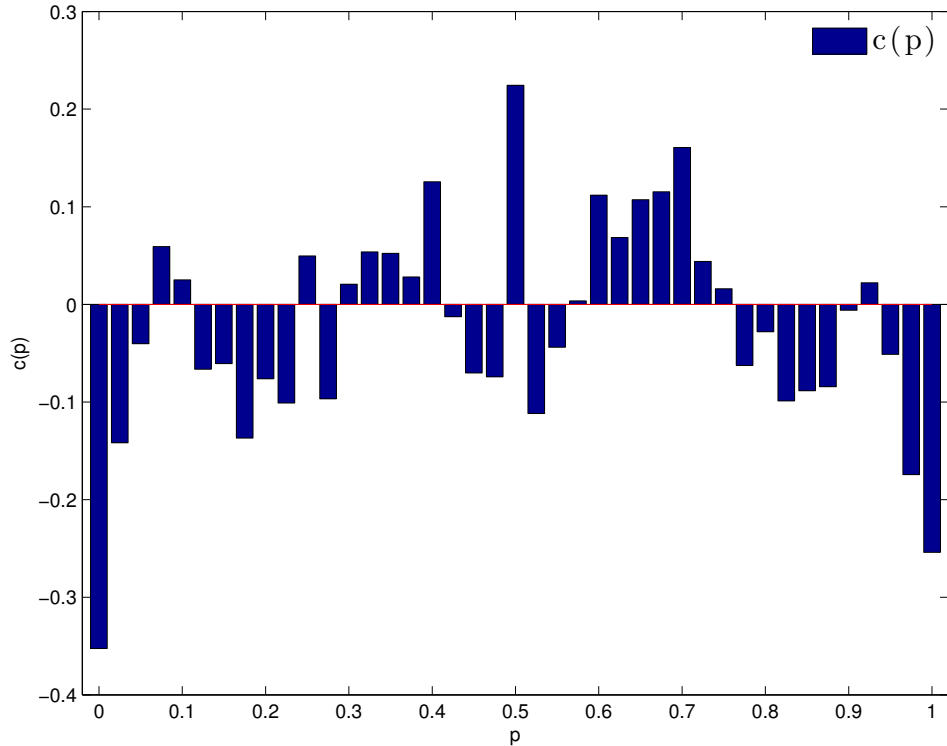


Figure F.8: The inattentive model-implied intrinsic relative cost function, $c(p)$ associated with different slider positions. Negative values of the intrinsic cost function are associated with preferred slider positions (which are chosen by subjects more than would be predicted by the rational inattention model without intrinsic costs), while positive cost values are associated with disliked slider positions. In addition to our subjects preferring extreme values, we also note that subjects are more likely to choose certain interior slider positions than others as well.

G Full sample results

In this appendix we present results for the full sample (all 110 sessions), including the repeated sessions. Figure G.1 shows the fraction of trials on which the slider is adjusted, if the data are sorted into bins according to the value of the Bayesian gap at the time of the decision whether to adjust the position.

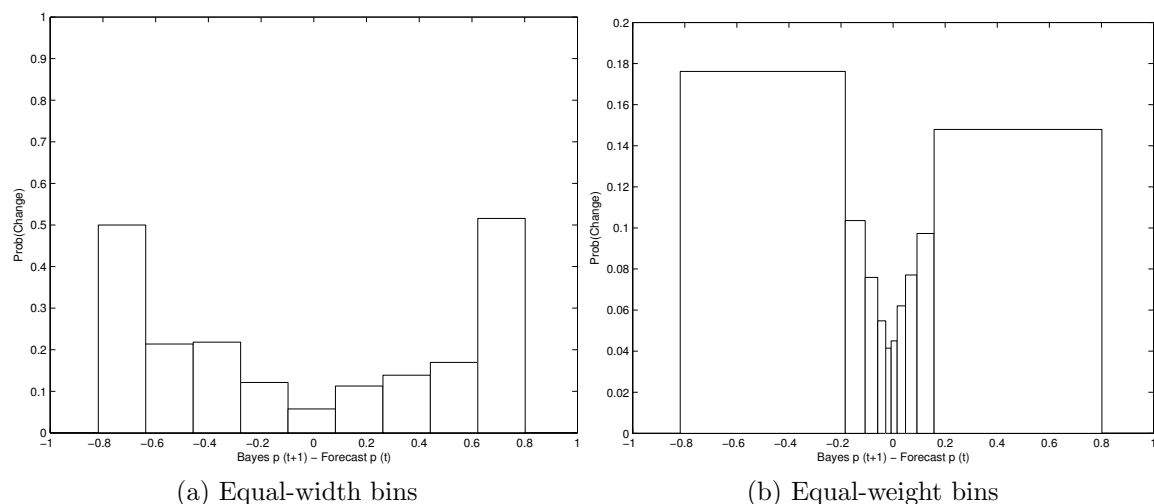


Figure G.1: Nonparametric empirical hazard functions for the slider adjustment probability as a function of the Bayesian gap. (a) Fraction of trials on which adjustment occurs, for each range of values of the gap, with equal-width bins. (b) Similar plot, but with boundaries chosen so that each bin contains an equal number of observations.

The top panel of Table G.1 presents the best fitting Bayesian “gap-based” models of the adjustment for the adjustment decision. The polynomial coefficients are very similar to those obtained only using the unique sessions, and moreover, the ranking of the alternative models, in terms of the BIC statistic, is preserved. The different magnitudes of the BIC statistics reflects the differences in the sample size. The lower panel of the table presents the best fitting “value-based” models of the adjustment for the adjustment decision. Once again, the polynomial coefficients are very similar to those obtained only using the unique sessions, and the ranking of the alternative models, in terms of the BIC statistic, is also preserved.

TABLE G.1: Best fitting models for the full sample, including repeated sessions

Gap Models	k	θ_0	θ_1	θ_2		BIC
Constant hazard	1	-2.34	–	–		65,473
Polynomial hazard	3	-2.53	-0.41	7.37		64,112
Symmetric poly.	2	-2.53	–	7.38		64,145
	k	ϵ	$\underline{\Delta}$	$\bar{\Delta}$		BIC
<i>Ss</i> with errors	3	0.088	-0.641	0.745		65,464
Symmetric <i>Ss</i>	2	0.088	-0.923	0.923		65,485
Value Models	k	θ_0	θ_1	ϵ	$\bar{\Delta}$	BIC
Constant hazard	1	-2.34	–	–	–	65,473
Polynomial hazard	2	-2.50	0.92	–	–	64,287
<i>Ss</i> with errors	2	–	–	0.088	3.87	65,433
	k	θ_0	θ_1	ψ_1	$\tilde{\Lambda}$	BIC
Rational inattention	2	-2.07	0.84	1.19	0.11	63,496