

NBER WORKING PAPER SERIES

DISCRETE ADJUSTMENT TO A CHANGING ENVIRONMENT:  
EXPERIMENTAL EVIDENCE

Mel Win Khaw  
Luminita Stevens  
Michael Woodford

Working Paper 22978  
<http://www.nber.org/papers/w22978>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
December 2016

We would like to thank Randy Gallistel for sharing the code for his experiment and for discussion of his work; Fernando Alvarez, Marty Eichenbaum, Cosmin Ilut, Filip Matejka, Alex Wolman, and conference and seminar participants at the Barcelona GSE Summer Forum, the NBER Summer Institute, the University of Maryland, and the JME-SNB-SCG conference for useful comments on earlier versions of this work; and the National Science Foundation for research support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Mel Win Khaw, Luminita Stevens, and Michael Woodford. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Discrete Adjustment to a Changing Environment: Experimental Evidence  
Mel Win Khaw, Luminita Stevens, and Michael Woodford  
NBER Working Paper No. 22978  
December 2016  
JEL No. D84,E03

### **ABSTRACT**

We conduct a laboratory experiment to shed light on the cognitive limitations that may affect the way decision makers respond to changes in their economic environment. The subjects solve a tracking problem: they estimate the probability of a binary event, which changes stochastically. The subjects observe draws and indicate their draw-by-draw estimate. Our subjects depart from the optimal Bayesian benchmark in systematic ways, but these deviations are not simply the result of some boundedly rational, but deterministic rule. Rather, there is a random element in the subjects' response to any given history of evidence. Moreover, subjects adjust their forecast in discrete jumps rather than after each new ring draw, even though there are no explicit adjustment costs. They adjust by both large and small amounts, contrary to the predictions of a simple  $S_s$  model of optimal adjustment subject to a fixed cost. Finally, subjects prefer to report "round number" probabilities, even though that requires exerting additional effort. Each of these regularities resembles the behavior of firms setting prices for their products. We develop a model of inattentive adjustment and compare its quantitative fit with alternative models of stochastic discrete adjustment.

Mel Win Khaw  
Department of Economics  
Columbia University  
420 W. 118th Street  
New York, NY 10027  
mwk2126@columbia.edu

Michael Woodford  
Department of Economics  
Columbia University  
420 W. 118th Street  
New York, NY 10027  
and NBER  
mw2230@columbia.edu

Luminita Stevens  
Department of Economics  
University of Maryland  
7343 Preinkert Drive  
College Park, MD 20742  
stevens@econ.umd.edu

# 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 decisionmakers.

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 decisionmaker’s objective as known, to the extent that we assume that our subjects care only about maximizing their monetary payment from the experiment,<sup>1</sup> and 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 decisionmaker 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.

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. It is explained that the underlying probability will shift from time to time, but also that 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

---

<sup>1</sup>In the experiment reported here, the subjects are paid for their performance, and we take care to explain the strategy that would maximize the expected monetary reward.

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 announced probability estimate), with the payoff from a given action depending on the current value of a continuous state variable, 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 calculate the optimal Bayesian 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.<sup>2</sup>

We further demonstrate that our data are in a number of important respects consistent with the predictions of a particular quantitative model of inattentive choice. 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 model with

---

<sup>2</sup>Our 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 the slider 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 slider will be set when subjects are free to set it anywhere.

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 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. We modify their experiment in a number of respects, because of the different focus of our study, as discussed further below. 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 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. After each ring draw, the subject’s cumulative score is updated and displayed at the top of the screen. The subject may then adjust his or her forecast or leave it unchanged, before drawing the next ring. This process is repeated until the session ends, after 1,000 ring

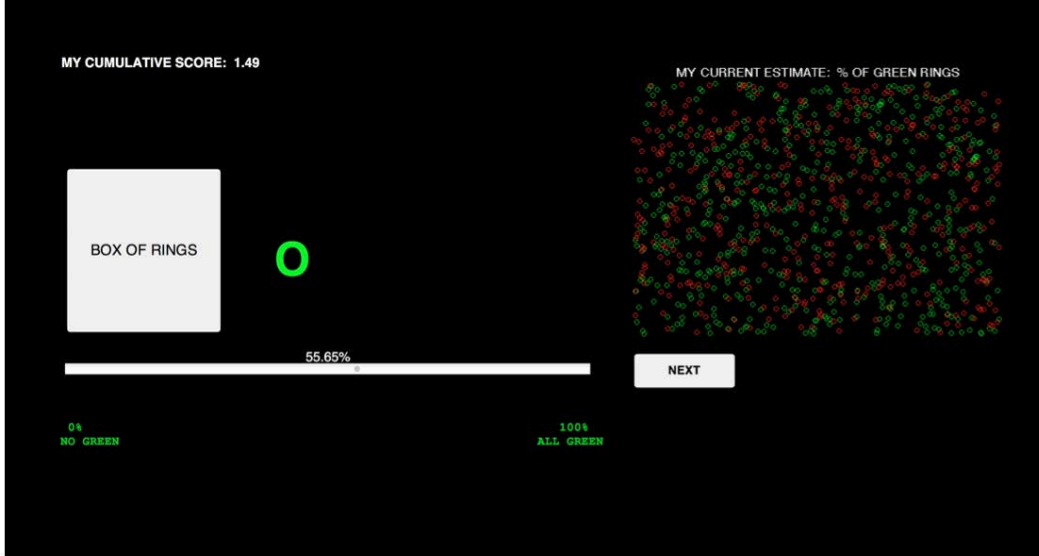


Figure 1: Experiment screenshot. The box of rings contains an unknown number of green and red rings. Subjects use the mouse to adjust the slider to their current estimate of the probability that the next draw from the box is a green ring. Their estimate is displayed numerically above the slider and visually 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.

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

$$r(\hat{p}_t; s_t) = 1 - (s_t - \hat{p}_t)^2. \quad (1)$$

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  $\hat{p}_t = p_t$ .<sup>3</sup> The monetary reward is only a function of the

<sup>3</sup>We 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

history of the subject’s behavior and the sequence of draws thus far. It reveals no information about the number of green rings in the box, either past or present. 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 proportion 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 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.

As noted above, this design is based on the experiment reported by Gallistel *et al.* (2014). We have modified their experimental design in a number of respects, however, to make it more suitable for our concerns. First, we add a monetary reward that varies with performance (as discussed above), and display the subject’s cumulative score in real time, so that they can see how their strategy affects their reward. The reward function allows us to have more control

---

for us to explain the task to subjects, and to make sure that any departures from the rational benchmark do not result from misunderstanding of the optimal strategy.

over the objective of our subjects, rather than having to make assumptions about what they might be maximizing. In addition, we simplify the data-generating process (making each new choice of the underlying probability  $p_t$  independent of the previous history), not only in order to simplify the Bayesian inference problem (analyzed in Appendix A), but more importantly so that it would be simple for us to fully explain the data-generating process to subjects. Indeed, another of the important changes in our procedure is to make a point of explaining all features of the setup, data generating process, and reward function to the subjects.

One of our most important changes is to modify the functioning of the slider, so that subjects can only move it continuously, by dragging it with their mouse. The slider in the experiment of Gallistel *et al.*, 2014, would move in discrete increments of exactly 0.1 to the left or right if the subject clicked on the slider bar to the left or right of the current slider position. Because of our interest in subjects' spontaneous tendency toward discrete adjustment even when nothing about their decision problem would make discrete adjustment desirable, it is important that movement in large discrete jumps *not* be more convenient than continuous adjustment of the slider. (It is for this same reason that we want subjects to report their probability estimate using a slider, rather than typing a number; the latter kind of reporting would make round numbers more convenient, and so provide a possible motive for discrete adjustment.) We also eliminate the feature of the Gallistel *et al.* experiment that asked subjects to press a button at those points in time when they believed that the underlying probability had changed. While eliciting more aspects of subjects' beliefs might seem desirable, this question was posed in a way that required subjects to respond only at discrete points in time, rather than continuously adjusting an estimate, and we feared that this could lead them to also adjust the slider only at discrete points in time, for a reason that would be inapplicable in the case of actual firm decisions such as price-setting.

While we retain the feature of the Gallistel *et al.* display that shows a visual representation of the subject's current probability estimate, we add a numerical readout of the estimate above the slider, allowing each subject three different ways of seeing what estimate they are choosing: the location of the slider relative to the endpoints (marked as 0% and 100%), the visual display of a box of red and green rings with the relative proportions implied by the



estimate, and the numerical readout of the estimate (accurate to four digits). We think it is important to be sure that subjects can understand the probability that they are choosing without necessarily understanding or thinking in terms of decimals; other experiments have found that probabilities are difficult to interpret and to communicate without some visual representation. The box showing them their current estimate is intended to help the subjects to visualize exactly what they are forecasting. On the other hand, some subjects may want to know what probability they are setting with more precision than allowed by the visual representations alone, and for these reasons, we add the numerical readout.<sup>4</sup>

Our setup is designed to minimize adjustment costs and to eliminate reasons for habit preferences, so that there would be no reason for subjects not to constantly adjust their forecasts in the absence of cognitive limitations. Additionally, we fully control the information that subjects have at each point in time, allowing us to test to what extent subjects are making use of 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 decisionmakers' 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 take care to explain all features of the setup to the subjects, they are provided with all of the information needed to form a correct forecast (in accordance with the RE hypothesis), and we can compare their performance to that of an ideal Bayesian observer.

An important question that arises in any 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 very general class of *tracking problems*. Formally, it is analogous to the problem of a pricing manager who sets the price of a product

---

<sup>4</sup>This modification of the experimental design made it possible for us to observe a cognitive “round number” bias, discussed below, that did not appear in the data collected by Gallistel *et al.* (2014), perhaps because their subjects were unable to observe their own probability estimates with the precision with which our subjects could.

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 that problem that also apply to our experimental task: it features a continuous decision variable, it generates profits that depend on other state variables (demand, costs) that fluctuate continuously over time, and finally, the manager making the decision receives a steady stream of information about product demand, competitive pressures, cost conditions, the macroeconomy, and so on, and incorporates these factors in the price of the product.

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.

### 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 decisionmaker who makes full use of all the information available.<sup>5</sup> The figure illustrates key features of the data that we document more systematically below. Both the subject and the benchmark Bayesian decisionmaker track the hidden probability quite well, and they detect changes in this probability relatively quickly. However, one major distinction 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 decisionmaker. 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.

---

<sup>5</sup>The Appendix presents the Bayesian inference problem.



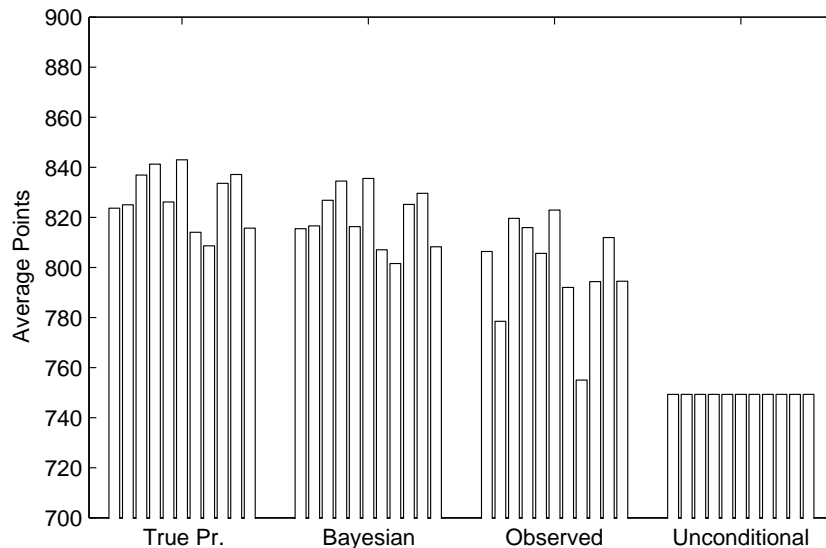


Figure 3: Experiment data for all unique sessions. The scores are broken down by subject. “True p” refers to the score that each subject would have received, given the realized ring draws, had they known the underlying probabilities at all times. “Bayesian” refers to the scores that would have been received 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.

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 the 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).

### 3.2 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

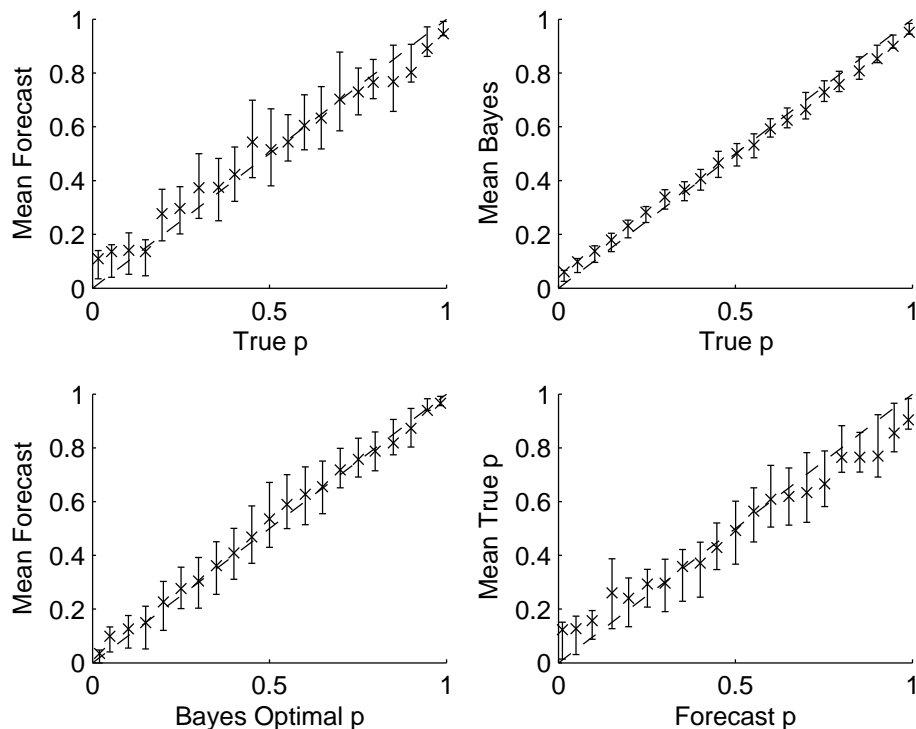


Figure 4: Measures of forecast bias. Upper left:  $E[\hat{p}|p]$  versus  $p$ . Upper right:  $E[p^*|p]$  versus  $p$ . Lower left:  $E[\hat{p}|p^*]$  versus  $p^*$ . Lower right:  $E[p|p^*]$  versus  $p^*$ .

median forecast conditional on the true  $p$ , rather than the mean.<sup>7</sup> 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 4. We pool the data from all unique 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 bin (shown as a vertical interval).<sup>8</sup> As in the corresponding figure in Gallistel *et al.*,<sup>9</sup> the diagonal (i.e., the predicted location of the means under the

<sup>7</sup>Ricci and Gallistel (2016) report a similar finding in a variant of the experiment of Gallistel *et al* (2014) with a different data-generating process.

<sup>8</sup>The 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.

<sup>9</sup>See their Figure 6.

Regression: $y = \alpha + \beta x + \epsilon$				
y	x	$\alpha$	$\beta$	F-stat
$\hat{p}$	$p$	0.085 (0.001)	0.850 (0.002)	3273
$p^*$	$p$	0.052 (0.001)	0.893 (0.001)	5674
$p$	$\hat{p}$	0.087 (0.001)	0.803 (0.002)	6112
$\hat{p}$	$p^*$	0.034 (0.001)	0.955 (0.002)	669

TABLE 1: Regression tests of forecast bias, with alternative choices of the variables  $y$  and  $x$ , corresponding to the four panels of Figure 4. 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  $F(2, 90907)$ . All null hypotheses have p-values of less than  $10^{-288}$ .

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 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).

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 4, where the trials are now binned

according to 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 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 4. 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 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 Bayesian-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 4, 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). The extent to which forecasts are consistent with this simple hypothesis is discussed further below.

### 3.3 Departures from Bayesian Rationality

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.

#### 3.3.1 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.<sup>10</sup> This stochasticity is at odds with traditional economic models in which optimization rules out randomization, as well as with many models of behavioral heuristics.



Figure 5: Experiment data, subject 1 forecasts, all repeated sessions.

Our observation above that certain statistics can be explained by a hypothesis that subjects’ forecasts are equal to the optimal Bayesian forecast plus random noise has already suggested a model of behavior that incorporates random noise. However, the mere fact that the interquartile ranges shown in the lower left panel of Figure 4 are of non-zero length does not in itself prove that subjects’ behavior is random; their forecasts might be deterministic

<sup>10</sup>Magnani, Gorry and Oprea (2016) reach a similar conclusion in the context of a somewhat related task.



functions of some aspect of the sequence of rings that they have seen that is not perfectly captured by the value of  $p_t^*$  (though highly correlated with it). More convincing evidence of stochasticity is provided by sessions in which subjects experienced precisely the same sequence of ring draws as they had observed on a previous occasion.

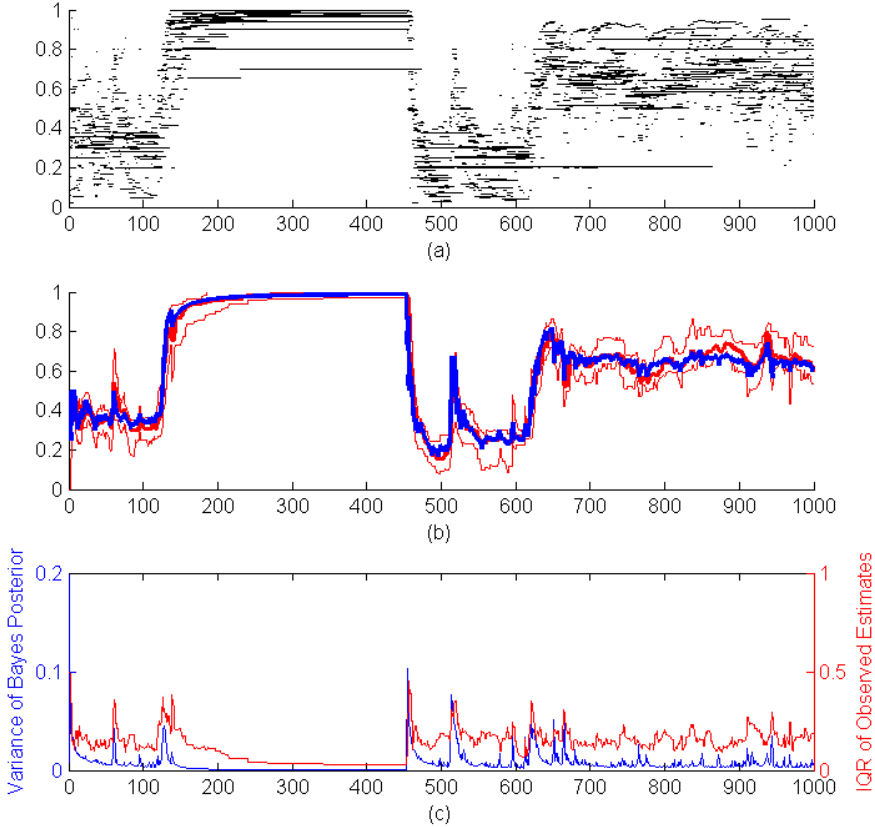


Figure 6: Experiment data, all subjects, all repeated sessions. (a) Individual forecast series, (b) Bayesian posterior mean in blue, median subject forecast and interquartile range in red, (c) Bayesian posterior variance in blue (left axis) and subjective forecasts interquartile range in red (right axis).

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 5 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 6 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.

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.

### 3.3.2 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. This discreteness is common to all of our subjects, albeit to varying degrees.

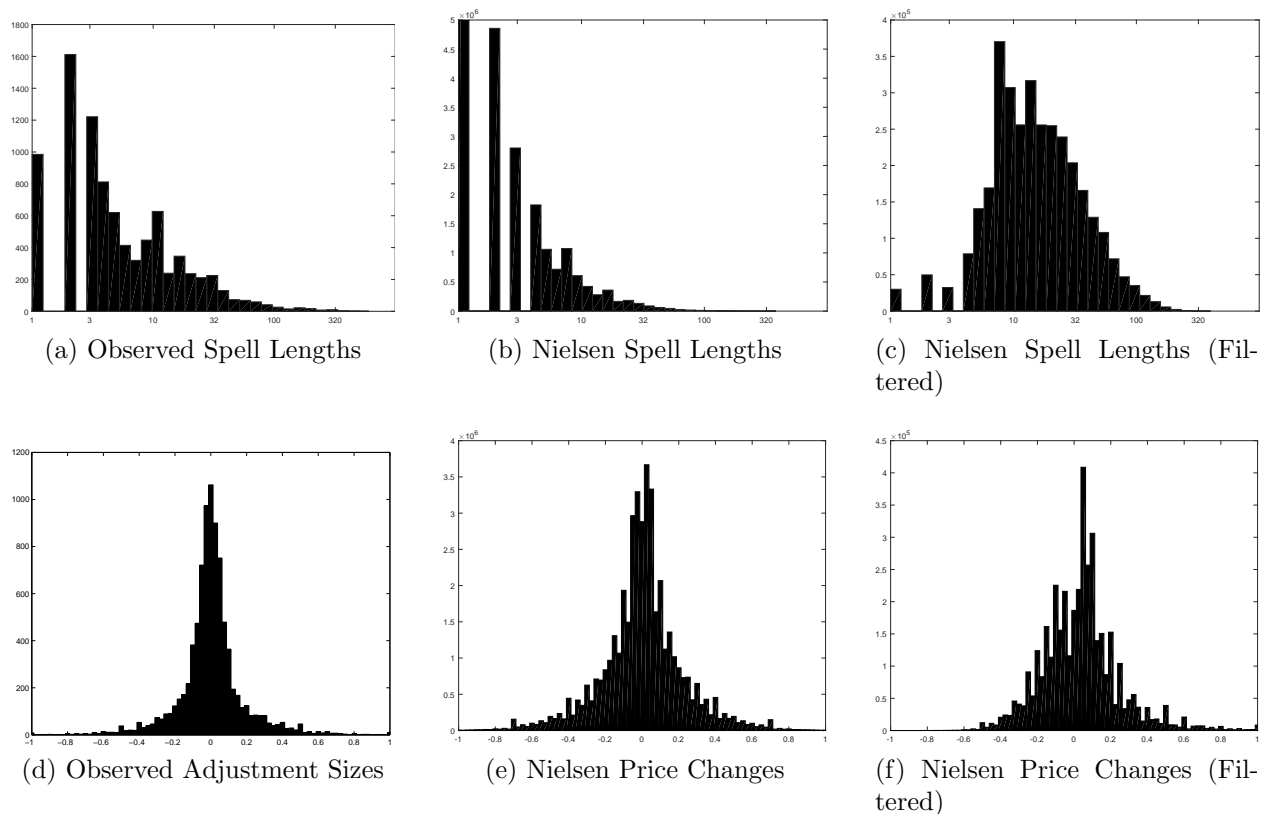


Figure 7: Discreteness. (a) Distribution of the number of ring draws between adjustments (all unique sessions). (b) Distribution of the number of weeks between price changes (AC Nielsen Retail Scanner data; y-axis truncated). (c) The same distribution, but removing transitory price changes. (d) Distribution of the size of changes in the slider position, counting only trials on which the slider is moved (all unique 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.

As shown in Figure 7, 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.<sup>11</sup> At the same time, the distribution of

<sup>11</sup>They are also both features of subjects' behavior in the similar experiment of Gallistel *et al.*, (2014). Ricci and Gallistel (2016) find similar behavior in a variant experiment in which the underlying probability drifts continuously, rather than jumping on infrequent discrete occasions. 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.

adjustment sizes is also inconsistent with the bimodal distribution that would be predicted by a simple “ $S_s$  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 resemble instead the distributions of product-level price changes in micro price data. For comparison, the right panels of Figure 7 show the distribution of the size of price changes and of the periods between price changes for products in the AC Nielsen Retail 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).

### 3.3.3 Round-Number Bias

Our subjects also exhibit a bias towards selecting “round numbers” when choosing where to position the slider. As shown in Figure 8, 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 continuously<sup>12</sup>). There is also no sense in which the experiment makes it either easier or more profitable to choose certain slider positions over others. Nevertheless, our subjects appear to exert considerable effort to set the slider at or near these preferred numbers. We conclude

---

<sup>12</sup>The 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  $10^{-4}$ )

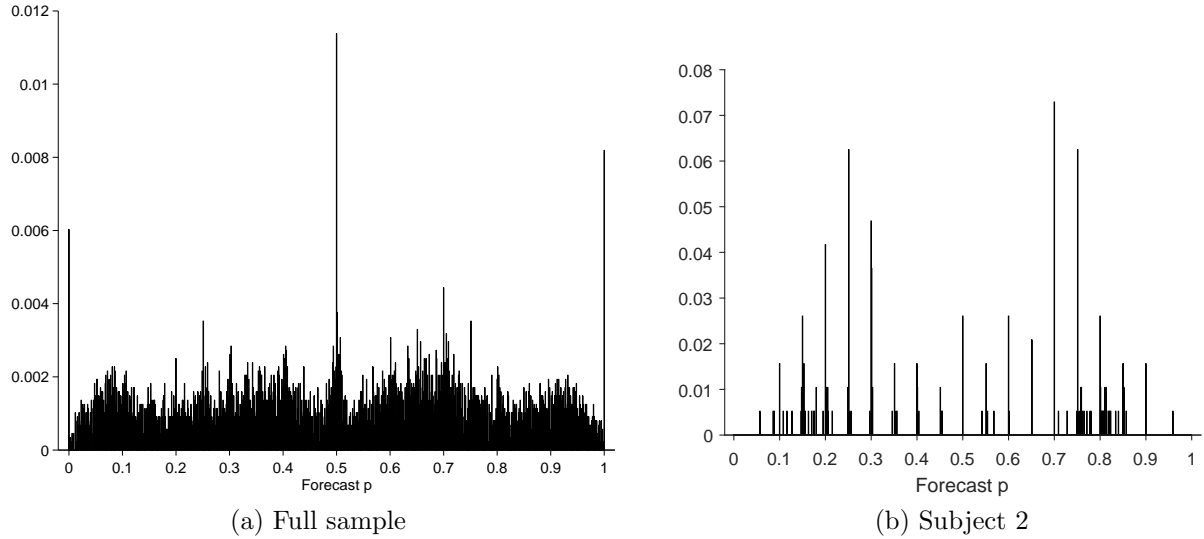


Figure 8: Experiment data. Distribution of the subjects' slider position choices for (a) all unique sessions and (b) the unique sessions for subject 2.

that these round decimal numbers offer some cognitive advantage that compensates for the effort needed to select them.

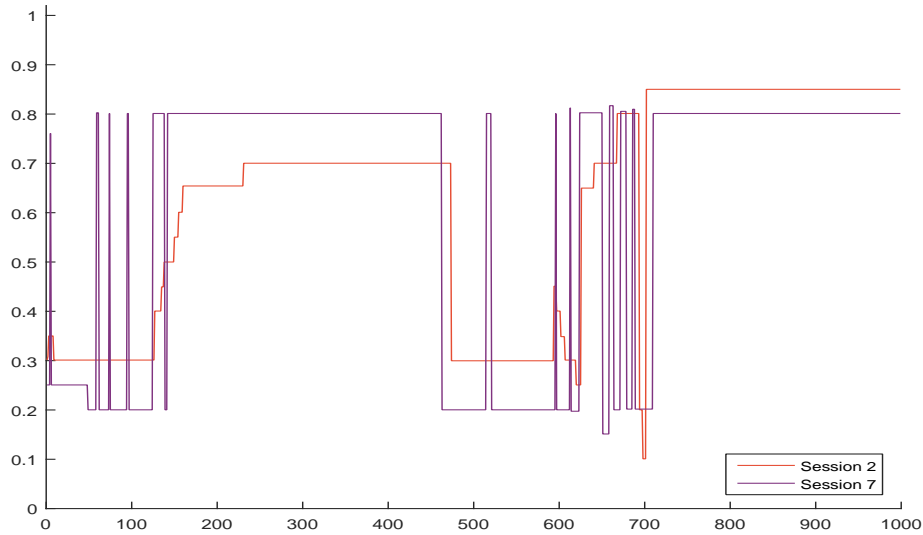


Figure 9: Experiment data, Subject 2 forecasts for repeated sessions.

This hypothesis is reinforced by the patterns in Figure 9, which shows the two repeated sessions for subject 2: during the first session, the subject repeatedly moves the slider to 0.2 and 0.8, while during the second session, the subject largely alternates between 0.3 and 0.7. This pattern suggests that there is some cognitive advantage to picking a small number of slider values to consider.

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. Such a simple rule of thumb corresponds to a myopic form of rational choice in which the decisionmaker cares *only* about the expected payoff on the current trial when deciding whether to adjust and where to adjust. If the decisionmaker must pay a fixed cost to adjust, then it would be optimal to plan to set  $\hat{p}_t$  equal to  $p_t^*$  conditional on adjusting, and to adjust if and only if the expected payoff reduction from non-adjustment exceeds the adjustment cost. The payoff reduction is proportional to the square of the Bayesian gap, thus myopic rational choice with a fixed cost implies adjustment of the slider if and only if the Bayesian gap has a large enough absolute value.

More generally, we might hypothesize that adjustment should occur if and only if the gap

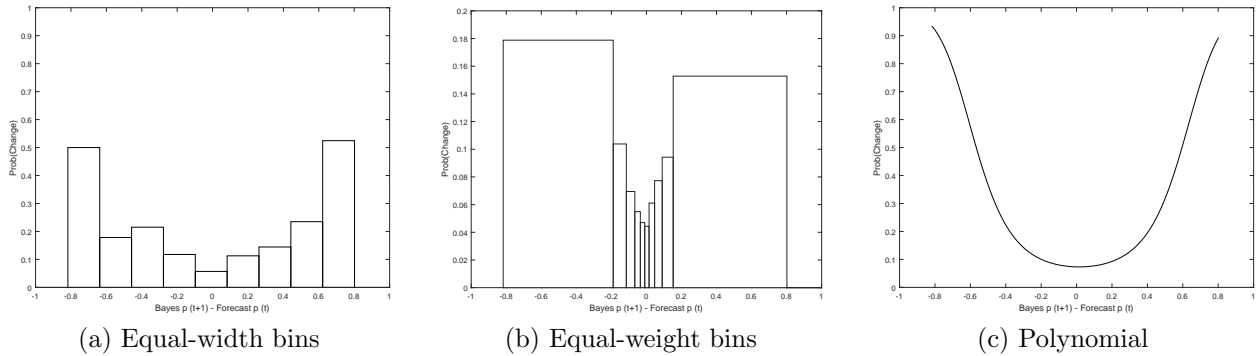


Figure 10: Hazard functions for the slider adjustment probability as a function of the Bayesian gap. (a) Nonparametric empirical hazard plotting the fraction of trials on which adjustment occurs, for each range of values of the gap, with equal-width bins. (b) Similar plot, but with boundaries chosen so that each bin contains an equal number of observations. (c) The best-fitting logistic-polynomial hazard function. (Coefficients given in Table 2.)

is either above some upper threshold or below some lower threshold, as in an “ $S_s$  model.” Even more generally, we might suppose, as in the “generalized  $S_s$  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  $\Lambda(\Delta_t)$ , non-increasing for  $\Delta < 0$  and non-decreasing for  $\Delta > 0$ . In the case that this “hazard function”  $\Lambda(\Delta)$  is symmetric (i.e.,  $\Lambda(-\Delta) = \Lambda(\Delta)$  for all  $\Delta$ ), a generalized  $S_s$  model could also be interpreted as resulting from myopic optimization, under the assumption of a cost of adjustment that is a random draw from a continuous distribution, as in the random-menu-cost models of Caballero and Engel (1999) and Dotsey, King and Wolman (1999).

Figure 10 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. In panel (a), the bins are chosen to be of equal width. This leads to a possibly misleading sense of the degree to which the hazard function is well-estimated in the case of more extreme Bayesian gaps, since the more extreme bins contain many fewer observations than do the central bins; panel (b) instead sorts the data into 10 bins each containing the same number of observations, with the boundaries of each bar indicating the boundaries of the corresponding bin. Both versions of the figure are consistent with the existence of a non-uniform hazard function, which is furthermore decreasing for  $\Delta < 0$  and increasing for  $\Delta > 0$ , as assumed by Caballero and Engel. However, our data do not appear

to support the further assumption, often made in the empirical work of Caballero and Engel, that the hazard falls to zero at least when  $\Delta = 0$ , if not over some larger “zone of inaction.”

These plots suggest that 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. If we assume that the adjustment decision  $\alpha_t$  (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

$$LL = \sum_t \{\alpha_t \log \Lambda(\Delta_t) + (1 - \alpha_t) \log(1 - \Lambda(\Delta_t))\}. \quad (2)$$

We consider parametric families of possible hazard functions  $\Lambda_k(\Delta; \theta)$ , where  $\theta$  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  $\theta$ ) so as to minimize the Bayes Information Criterion (BIC)<sup>13</sup> with  $N$  observations,

$$BIC \equiv -2LL + k \log N. \quad (3)$$

Suppose that the log odds of adjustment,  $\lambda(\Delta) \equiv \log \frac{\Lambda(\Delta)}{1-\Lambda(\Delta)}$ , is a polynomial function of  $\Delta$ , with  $k - 1 \geq 0$  the order of the polynomial,<sup>14</sup>

$$\lambda_k(\Delta; \theta) = \sum_{j=0}^{k-1} \theta_j \Delta^j. \quad (4)$$

---

<sup>13</sup>See, for example, Burnham and Anderson (2002, chap. 6) for discussion of this approach to model selection.

<sup>14</sup>The choice of this particular nested sequence of parametric families is motivated by the fact that the more specific theory based on attention costs that we consider below corresponds to a particular family of models of this kind, though with a different measure of the relevant “gap.” A further appealing feature of the assumption that  $\lambda(\Delta)$  is a polynomial function — and hence that  $\Lambda(\Delta)$  is a logistic function of a polynomial function of  $\Delta$  — rather than simply assuming that  $\Lambda(\Delta)$  is a polynomial function, is that the former assumption, unlike the latter, implies a hazard between 0 and 1 for all values of the gap, without any need to truncate the function implied by a given parameter vector. Moreover, this specification implies that the hazard will be greater than zero and less than 1 for all values of  $\Delta$ , so that any sequence of observations will have a finite log likelihood, regardless of the parameter vector.



TABLE 2: Best fitting models

<i>Gap-Based Models</i>								
Model	k	$\theta_0$	$\theta_1$	$\theta_2$				BIC
Constant hazard	1	-2.33	–	–				54,286
Polynomial hazard	3	-2.54	-0.22	7.52				53,112.4
Symmetric poly.	2	-2.53	–	7.58				53,111.5
Model	k	$\epsilon$	$\underline{\Delta}$	$\bar{\Delta}$				BIC
<i>Ss</i> with errors	3	0.088	-0.64	0.77				54,291
Symmetric <i>Ss</i>	2	0.088	-0.88	0.88				54,298
<i>Optimizing Models</i>								
Model	k	$\theta_0$	$\theta_1$	$\epsilon_1$	$\epsilon_2$	$\bar{\Delta}$		BIC
Constant hazard	1	-2.33	–	–	–	–		54,286
Polynomial hazard	2	-2.50	0.92	–	–	–		53,247
<i>Ss</i> with errors	2	–	–	0.088	0.912	3.96		54,247
Step function	3	–	–	0.082	0.20	0.58		53,652
<i>Inattention Model</i>								
Model	k	$\theta_0$	$\theta_1$	$\psi_1$	$\tilde{\Lambda}$			BIC
Rational inattention	2	-2.07	0.77	1.30	0.112			52,780

Top panel: Polynomial 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: Model assuming that adjustment economizes on information costs.

In the case of our data (pooling the data from the 91 unique sessions<sup>15</sup>), the BIC is minimized when  $\lambda(\Delta)$  is a quadratic function, with coefficients  $\{\theta_j\}$  given in the top panel of Table 2. The implied hazard function  $\Lambda(\Delta)$  is shown in the last panel of Figure 10. 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. Specifically, for two model families  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , the relative posterior probability that the true model

<sup>15</sup>The main text presents results for the unique sessions only, while the Appendix presents corresponding results using the data from all 110 sessions, including the repeated sessions. The results are fairly similar with regard to all of the issues discussed below; we prefer to emphasize the results using only the unique sessions, which represent 91 independent samples from the data-generating process, and do not over-sample any part of the state space.

belongs to one of these families rather than the other, after observing the data, is given by

$$\log \frac{P_{post}(\mathcal{M}_1)}{P_{post}(\mathcal{M}_2)} = \log \frac{P_{prior}(\mathcal{M}_1)}{P_{prior}(\mathcal{M}_2)} - \frac{1}{2} [BIC(\mathcal{M}_1) - BIC(\mathcal{M}_2)]. \quad (5)$$

If we assume a constant hazard function (as in the Calvo (1983) model of adjustment), then as indicated in Table 2, the minimum BIC statistic is higher by 1,174 natural log points. Thus the data increases the relative posterior probability that the quadratic log odds model is correct by a factor greater than  $10^{254}$ , relative to the hypothesis of no dependence of the adjustment decision on the Bayesian gap.

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, and the BIC statistic is 0.9 log points lower than in the unconstrained case. Allowing a non-zero linear term increases the log-likelihood of the data, but not enough to outweigh the penalty for the additional free parameter.

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  $Ss$  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  $Ss$  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 the decision is to keep the slider unchanged, the slider will nonetheless be adjusted with probability  $\epsilon$ , for some  $0 < \epsilon < 1/2$ , and when the decision to adjust is made, the slider will only be adjusted with probability  $1 - \epsilon$ . But even allowing for errors of this kind, the minimum value of the BIC statistic remains 1,178 log points higher under a model with sharp thresholds than under the quadratic log odds model (reducing the relative posterior probability by  $10^{255}$ , relative to best fitting quadratic hazard function). 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) assume, as in the classic “ $S_s$ ” 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 (a) of Figure 11 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 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.

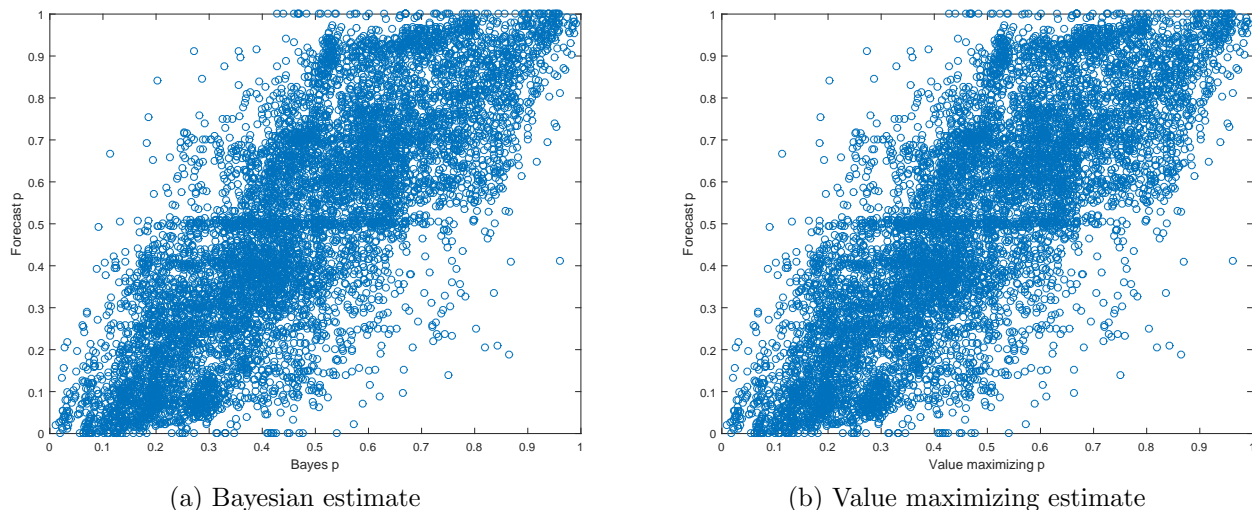


Figure 11: The distribution of slider positions that are chosen, conditional on adjustment. (a) Scatter plot of the slider position [vertical axis] against the Bayesian optimal setting; (b) scatter plot of the slider position [vertical axis] against the one that would maximize the continuation value function.

## 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,<sup>16</sup> or a cognitive cost of having to decide where to move the slider.

Consider first the joint hypothesis that there is a constant fixed cost of moving the slider at all times, and that the slider will be adjusted to its optimal location in the case that the fixed cost is born, as in classic “menu cost” models of price adjustment such as the model of Sheshinski and Weiss (1977, 1983). Then the slider should be adjusted on trial  $t$  if and only if

$$V_t(\hat{p}_{t-1}) < \max_p V_t(p) - \kappa, \tag{6}$$

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  $p$  at the end of trial  $t$ , and  $\kappa > 0$  is 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  $\pi_t$  at that time (i.e., the posterior taking into account the rings drawn on trial  $t$  and earlier); the argument  $\pi_t$  has been suppressed to simplify the notation. Although the pattern of behavior documented in the previous section rules out models with a fixed cost of adjustment, we can still determine how well such a model can predict the decision of whether to adjust the slider.

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  $G(\kappa)$ .<sup>17</sup> If we again assume that the slider will be moved to its optimal location conditional on adjustment, then the probability of adjustment on trial  $t$  should be given by a non-decreasing hazard

---

<sup>16</sup>We 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.

<sup>17</sup>Caballero and Engel (1999) propose a similar model of optimization with random fixed costs as a micro-foundation for the type of “generalized  $S_s$  model” proposed in Caballero and Engel (1993).

function

$$\Lambda(\Delta_t) = G(\Delta_t), \tag{7}$$

where the “value gap”  $\Delta_t$  is defined as

$$\Delta_t \equiv \max_p V_t(p) - V_t(\hat{p}_{t-1}). \tag{8}$$

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).<sup>18</sup>

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  $\pi_t$  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.

In the numerical results presented below, we examine how well the adjustment decision can be explained by variation in the value gap, when we use a particular measure of the value based on the best-fitting value gap model. The contour lines associated with equal magnitudes for the value gap are plotted in Figure 12. 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

---

<sup>18</sup>The model of Calvo (1983) also assumes that the cost of adjustment is effectively infinite, except at those times when a zero adjustment cost is drawn; but one can combine the assumption of a positive probability of drawing a zero adjustment cost with an assumption that the adjustment cost is finite at all other times, so that the hazard approaches 1 for all large enough values of the “value gap,” contrary to what is assumed in the Calvo model.

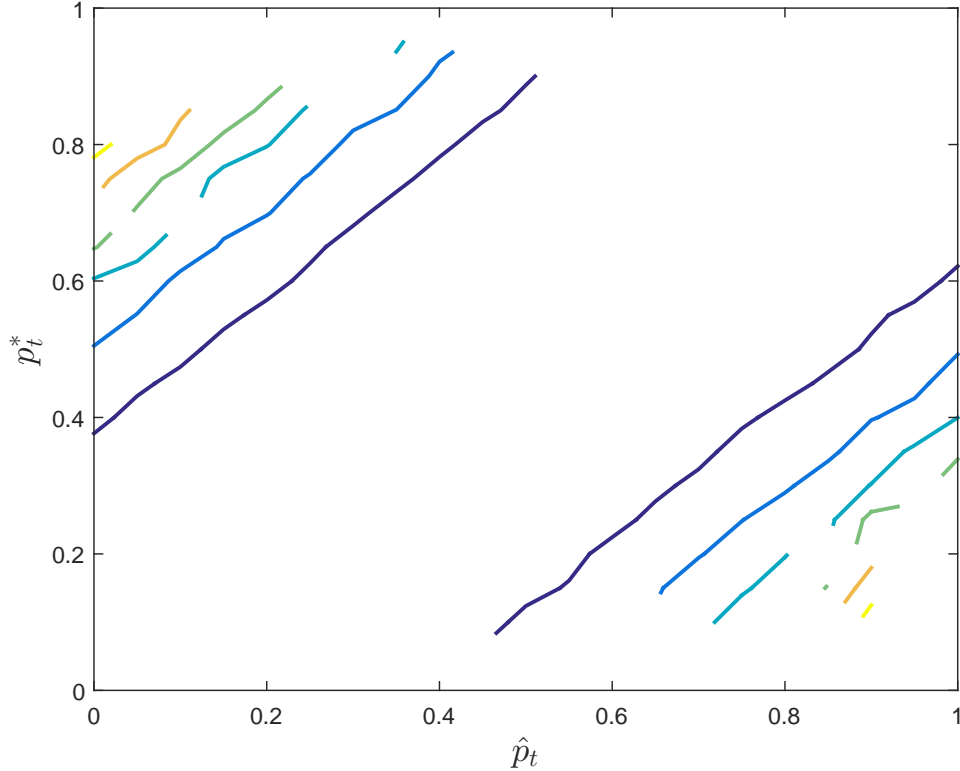


Figure 12: A contour plot of the average value gap  $\Delta_t$  associated with alternative pairs of values for the slider position  $\hat{p}_t$  [the horizontal axis] and the Bayesian posterior mean  $p_t^*$  [the vertical axis], when the continuation value function  $V_t(p)$  is numerically approximated on the basis of an empirical model of the subsequent adjustment dynamics.

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 13 plots the fraction of trials in which the slider is adjusted, grouping the trials in bins with an equal number of observations, but now using the value gap, rather than the Bayesian gap, as the basis for classification.<sup>19</sup> The probability of adjustment is monotonically increasing in the value gap. Moreover, adjustment only occurs with positive probability if the value gap is positive. On the other hand, there does not appear to be a sharp threshold of the kind required by an optimizing  $Ss$  model of adjustment. Nor is there any indication

<sup>19</sup>The value gap is measured, in the horizontal axis of this figure, in terms of the number of additional points (net of any increase in expected subsequent adjustment costs) that the subject can expect to obtain, on average, by adjusting the slider, assuming that it is adjusted to the optimal position.

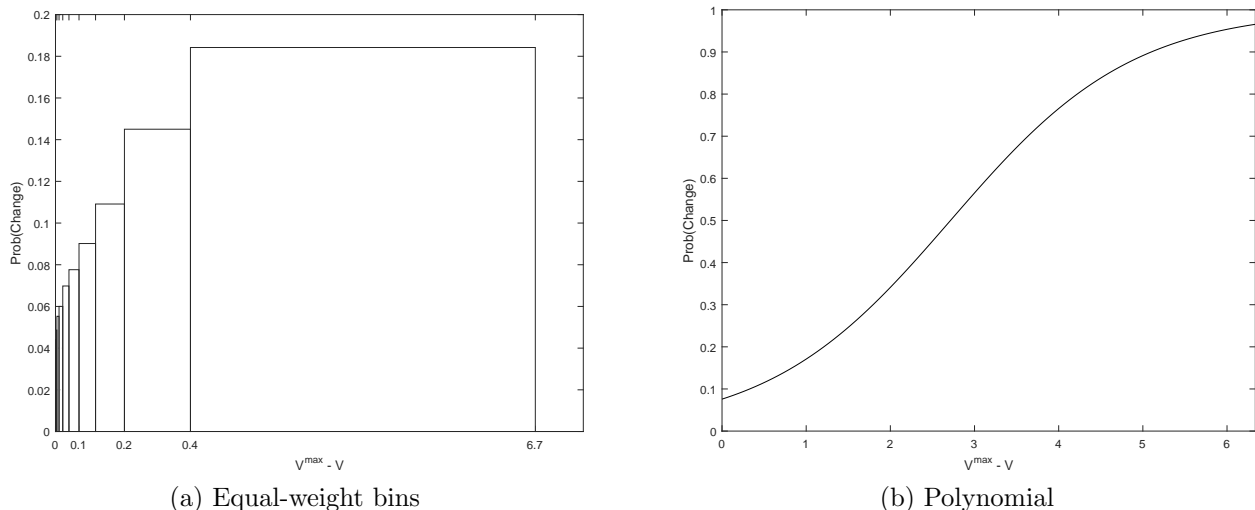


Figure 13: Adjustment probability as a function of the value gap, computed under an assumption that if the slider is adjusted, it will be moved to the currently optimal position. (a) Nonparametric empirical hazard, with boundaries chosen so that each bin contains an equal number of observations. (b) The best-fitting logistic-polynomial hazard function as a function of the value gap. (Coefficients given in Table 2.)

that the adjustment hazard falls to zero as the value gap approaches zero, as the random menu cost model would require in the absence of an atom at zero in the distribution of fixed costs.

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 2. The implied best-fitting value-based hazard function is shown in Panel (b) of Figure 13. 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; the fixed cost can be unboundedly large, but it 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 BIC statistic of 53,247 points, which is 1,039 natural log points lower than that of the best-fitting constant hazard. This implies a relative

posterior probability that is larger by a factor of  $10^{225}$ . We also reject the hypothesis of a constant fixed cost (the optimizing  $S_s$  model). Even if we assume that it is not possible to reduce the probability of adjustment below some  $\epsilon_1 > 0$  or to raise it above  $1 - \epsilon_1$  (with  $\epsilon_1$  a free parameter that can be chosen to fit the data), the lowest BIC statistic obtainable with such a model remains 999 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  $\bar{\Delta}$ , 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 405 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  $V_t(p)$ . This is a standard assumption in state-dependent pricing models (and many models of inattentive adjustment as well, such as the models of Calvo, 1983, or Woodford, 2009). However, the assumption that upon adjustment the chosen slider position is a deterministic function of the value maximizing position is not consistent with our evidence. Panel (b) of Figure 11 shows a scatter plot of the values of  $\hat{p}_t$  and  $\arg \max_p V_t(p)$ . This plot shows a large dispersion in the actual



slider positions that are chosen, conditioning on the currently optimal slider position — especially when the currently optimal position is in the middle of the interval of possible positions. 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 numerical error alone. It appears that the decision where to move the slider when it is adjusted involves a significant degree 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 decisionmakers 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).<sup>20</sup> The theory generalizes the models of inattentive adjustment by Woodford (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  $\mu_t$  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

---

<sup>20</sup>See 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.

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).<sup>21</sup> Under the hypothesis of rational inattention, the subject's decision rule maximizes an objective of the form

$$\mathbb{E} \left\{ \sum_{t=1}^T \left[ r(\hat{p}_t; s_t) - \psi_1 I_1 - \psi_2 \Lambda_t I_2 - \Lambda_t c^{adj} - (1 - \Lambda_t) c^{non} - \Lambda_t \int c(p) d\mu_t(p) \right] \right\}, \quad (9)$$

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 information 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  $\mathbb{E}\{\cdot\}$  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

$$\log \frac{\Lambda_t}{1 - \Lambda_t} = \lambda(\Delta_t) \equiv \log \frac{\tilde{\Lambda}}{1 - \tilde{\Lambda}} + \psi_1^{-1} \Delta_t \quad (10)$$

---

<sup>21</sup>Note 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.

and an optimal measure over new slider positions

$$\mu_t(p) = \frac{\exp\{\psi_2^{-1}V_t(p)\}\tilde{\mu}(p)}{\int \exp\{\psi_2^{-1}V_t(p')\}d\tilde{\mu}(p')}. \quad (11)$$

Here  $V_t(p)$  is again the continuation value function, and the  $\Delta_t$  in (10) is now the “RI value gap” defined as<sup>22</sup>

$$\Delta_t \equiv \int V_t(p)d\mu_t(p) - V_t(\hat{p}_{t-1}). \quad (12)$$

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 (10) implies that the probability of adjusting is again given by a continuous 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  $\Delta_t$ , as in the case of the best-fitting random-fixed-cost model discussed in section 4.2, though here the definition of  $\Delta_t$  is different. Unlike the models considered above, (11) 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  $\psi_2 \rightarrow 0$ .

We consider the parameterization of this model that best fits our experimental data. We wish to determine (i) the values for the information-cost parameter  $\psi_1$  and the reference adjustment probability  $\tilde{\Lambda}$  that best fit our data on the *timing* of slider adjustments, and (ii) the values for the information-cost parameter  $\psi_2$  and the reference measure  $\tilde{\mu}$  over new slider positions that best fit our data on the *positions* to which the slider is moved when it is adjusted. In the first case, we are choosing parameters for prediction (10) for the adjustment

---

<sup>22</sup>The appendix describes the approximation of the value function and the associated value gap in the inattentive adjustment model.

hazard, and in the second we are choosing parameters for prediction (11) for the distribution from which the new slider position is drawn. These problems are not independent, however, since the expected value of adjusting depends on the choice of a distribution of new slider positions, and likewise, this latter choice depends on the probability of adjustment in future periods. Hence, we estimate these parameters jointly.

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.<sup>23</sup> Within this family, the best-fitting reference distribution corresponds to  $\gamma = 0.47$ . The implied distribution  $\tilde{\mu}(p)$  is shown in the left panel of Figure 14; note that that this distribution is significantly flatter than the empirical distribution of slider location choices, also shown in the figure.

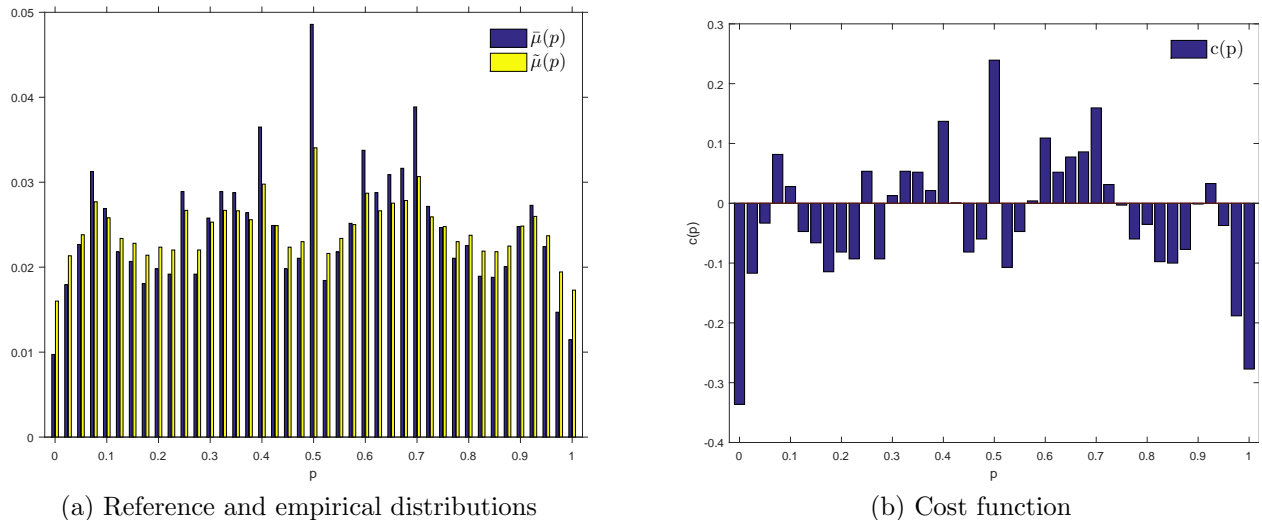


Figure 14: Intrinsic costs. (a) The model-implied reference distribution  $\tilde{\mu}$  of slider positions and the empirical distribution  $\mu$ , conditional on adjustment. (b) The model-implied intrinsic relative cost function,  $c(p)$  associated with different slider positions.

This distribution points to the presence of significant intrinsic gains and costs in the decision of where to position the slider conditional on adjustment, since in the absence of such costs the two distributions would be identical. We can go further and use the estimated reference distribution to infer the intrinsic relative cost  $c(p)$  associated with movement to

<sup>23</sup>Allowing 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).

each possible slider position. The relative cost function is shown in Panel (b) of Figure 14. 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. We find that our subjects prefer extreme slider positions (close to 0 or close to 1) more than would be predicted by a rational inattention model without any intrinsic costs or gains from choosing specific slider positions. In addition, subjects are also more likely to choose certain interior slider positions than others. The associated information cost parameter for the use of more information in choosing the slider position is  $\psi_2 = 0.67$ .

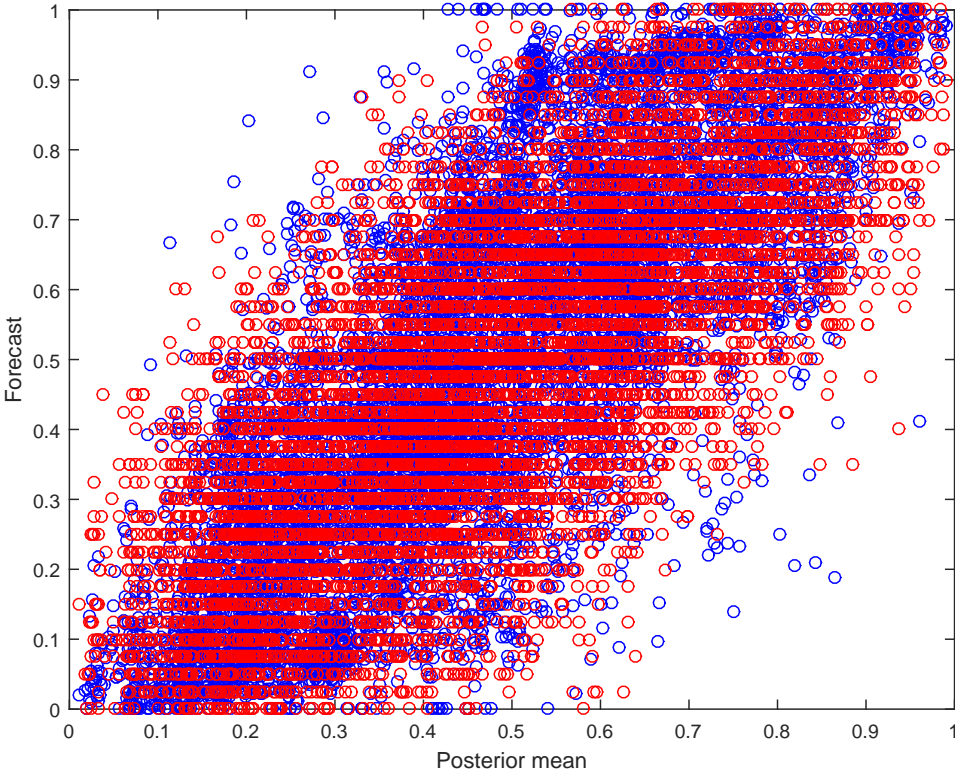


Figure 15: The model-implied joint distribution of slider settings and Bayesian posterior mean, conditional on adjustment. Blue circles: scatter plot of the actual slider positions [vertical axis] against the Bayesian optimal setting. Red circles: scatter plot of the simulated RI slider positions [vertical axis] against the Bayesian optimal setting.

The degree to which this model of the slider location decision matches the experimental data when  $\psi_2 = 0.67$  is shown in Figure 15. Here we show again the scatter plot of slider positions chosen on trials where the slider is adjusted, plotted against that Bayesian posterior mean for the underlying state on that trial (in blue circles, as in panel (a) of Figure 11),

with the corresponding scatter plot from a simulation of the rational inattention model of the overlaid on it (in red circles). 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  $\mu_t$  given by equation (11). We find that the predicted joint distribution of the Bayesian posterior mean and the new slider position is reasonably similar to the one observed in our data.<sup>24</sup>

We turn next to the ability of the RI model to account for the timing of adjustments of the slider. Like the models of the adjustment decision considered in the previous sections, the RI model implies that the adjustment hazard should be a function of the “RI gap” variable. However, the continuation value function is now net of the expected information and intrinsic costs; and the expected value of adjustment is an average of the continuation values associated with different slider positions, rather than solely the maximum continuation value. The value gap implied by the RI model, under our estimated parameter values, is graphed in Figure 16, using the same format as Figure 12.

As a nonparametric illustration of the degree to which this gap variable is able to explain the timing of adjustments, Figure 17 plots the fraction of trials in which the slider is adjusted, grouping the trials in bins with an equal number of observations. We see that the adjustment probability is monotonically increasing with this gap measure, as predicted by the model.

The RI model makes a very specific prediction, that  $\lambda(\Delta)$  should be an increasing linear function of the value gap. When we consider polynomials of arbitrary order, but use the BIC criterion to select the best model family, we find that in fact our data are best fit by a model according to which the log odds are an increasing linear function of the gap, as predicted by the theory. The best fitting parameters and the associated BIC value are shown in the

---

<sup>24</sup>We do not provide a quantitative measure of the goodness of fit, as there is no alternative stochastic model of the slider location decision with which we wish to compare this model. Because of the substantial stochastic variation in the chosen slider position, the rational inattention model is clearly much more successful in accounting for the data than the deterministic hypotheses considered above (that when the slider is adjusted, it is set to the current Bayesian posterior mean, or that it is set to the position that maximizes the continuation value).

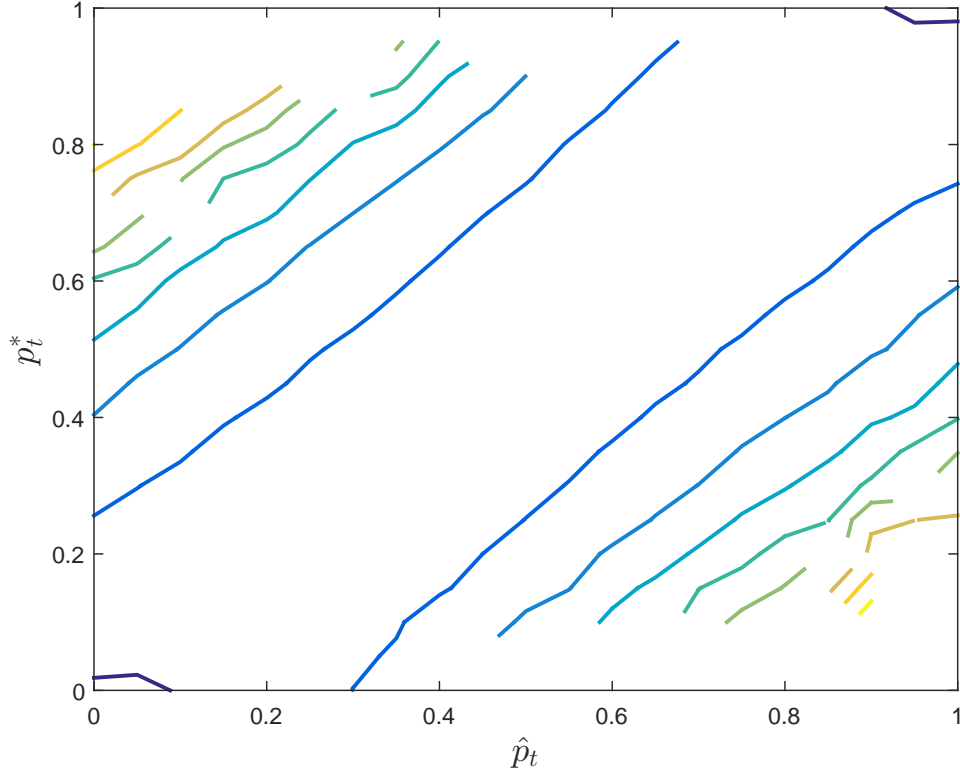


Figure 16: A contour plot of the average RI value gap  $\Delta_t$  associated with alternative pairs of values for the slider position  $\hat{p}_t$  [the horizontal axis] and the Bayesian posterior mean  $p_t^*$  [the vertical axis].

bottom panel of Table 2. The best-fitting parametric hazard function is plotted in the second panel of Figure 17.

Equation (10) allows us to interpret the coefficients  $(\theta_0, \theta_1)$  of the logistic-linear hazard function in terms of the information cost  $\psi_1$  and the reference frequency of adjustment  $\tilde{\Lambda}$  of the rational inattention model; the implied values of these two parameters are also given in Table 2. Interestingly, the information cost  $\psi_1$  for the decision whether to adjust is estimated to be roughly twice as large as the information cost  $\psi_2$  for the decision where to move the slider conditional on adjustment.<sup>25</sup>

The estimated reference probability of adjustment,  $\tilde{\Lambda} = 11.2\%$ , is larger than the empirical probability of adjustment,  $\bar{\Lambda} = 8.9\%$ . This 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

<sup>25</sup>This 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.

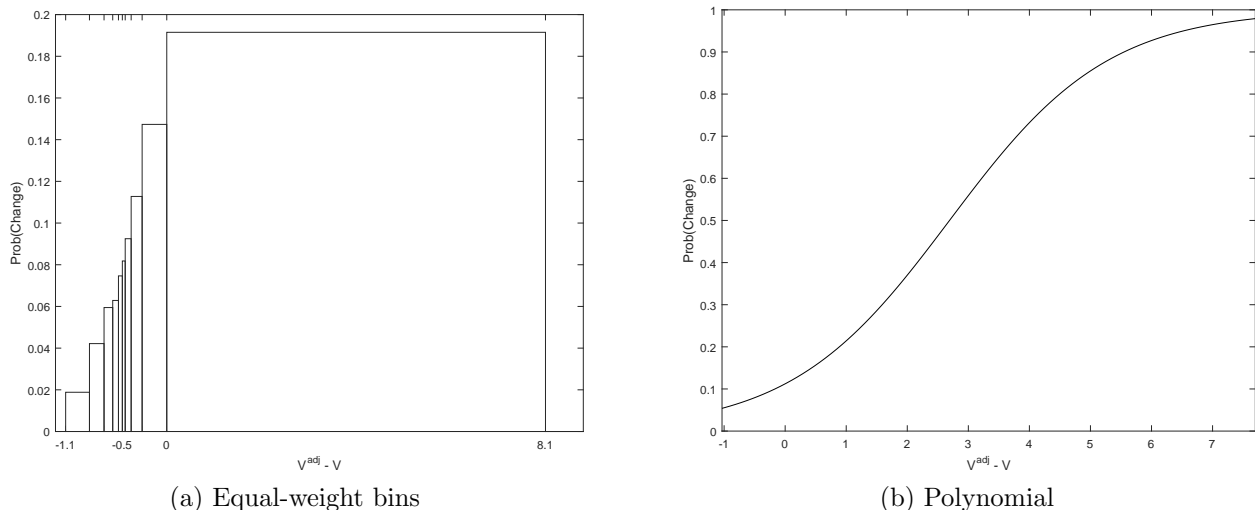


Figure 17: Adjustment probability as a function of the value gap implied by the rational inattention model. (a) Nonparametric empirical hazard with the boundaries chosen so that each bin contains an equal number of observations. (b) The best-fitting logistic-polynomial hazard function as a function of the value gap implied by the rational inattention model. (Coefficients given in Table 2.)

slider. In fact, in our data subjects often adjust when the RI value gap is negative.

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 symmetric quadratic polynomial), the BIC of the RI model is lower by 332 log points, implying a relative posterior probability of the RI model greater than  $10^{72}$ .

Hence the rational inattention model augmented with intrinsic costs succeeds in capturing the main features of our data, at least as far as the pooled data, describing the average behavior of our population of subjects, are concerned. It predicts both the timing of adjustments (that these are 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 (this 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 (10) that is predicted for the adjustment hazard function fits our data better than any of the other models considered above. And 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 discreteness 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). 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 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 kinds 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).<sup>26</sup> 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).<sup>27</sup> This interpretation of the random timing of the slider adjustments is especially appealing in

---

<sup>26</sup>As discussed in Khaw, Stevens, and Woodford (2016), this model can also be viewed as a generalization of the logit model of Costain and Nakov (2014).

<sup>27</sup>Magnani, 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.

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 decisionmakers 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.

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 decisionmaking 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

- 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.
- 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 (2014), “The Perception of Probability,” *Psychological Review* 121(1): 96–123.
- 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 Adjustment,” *in progress* .
- 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.
- McKelvey, Richard D. & Talbot Page (1990), “Public and Private Information: An Experimental Study of Information Pooling,” *Econometrica* 58(6): 1321–1339.
- Muth, John F (1961), “Rational expectations and the theory of price movements,” *Econometrica: Journal of the Econometric Society* pp. 315–335.
- Peterson, Cameron R. & Lee R. Beach (1967), “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 (2016), “Accuracy and Structure Detection in the Perception of Probability,” *unpublished* .
- 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,” *unpublished* .
- 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.

# A 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  $\tau_i$  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

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

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

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

and let  $L_\tau(n)$  denote the average likelihood computed using the  $n$  observations ending with observation  $\tau$ . The ex-ante joint probability of the model  $(\pi, p)$  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  $\pi$  and  $\mu_\pi$  is the ex-ante probability of partition  $\pi$  occurring in a sample of length  $T$ ,

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

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

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

The posterior probability of  $(n, p)$  is

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

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  $\delta$ , and in which case, the estimate of the probability is 0.5:

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

Figure A.1 plots the distribution of the size of slider adjustments in our data and in the Bayesian benchmark for the sample of unique sessions (top panels) and for the full sample, including the repeated sessions (bottom panels). In the data, 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.

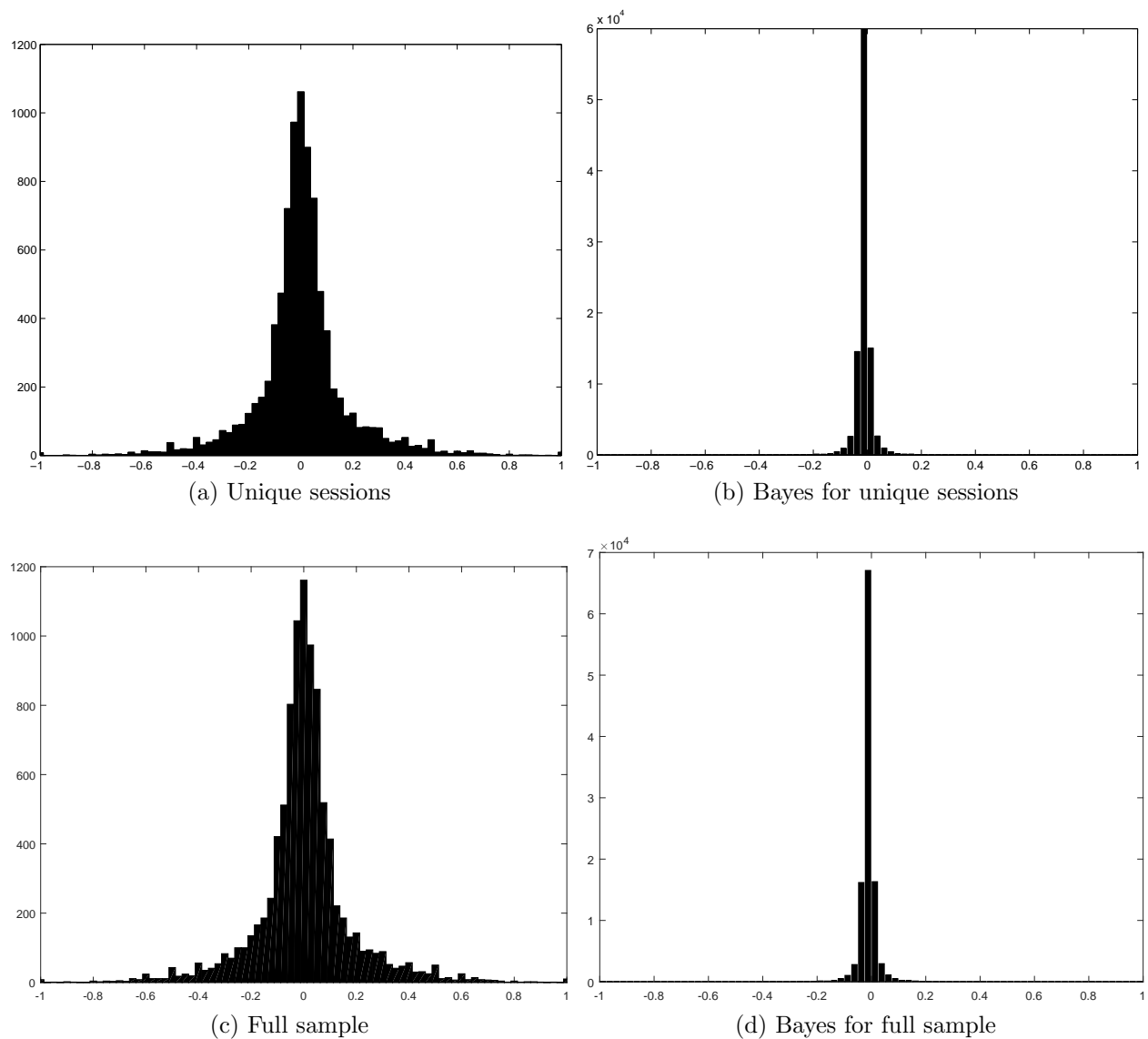


Figure A.1: Distribution of the size of changes in the slider position  $\Delta \hat{p}_t$ . Left panels: our data, counting only trials on which the slider is moved (unique sessions in the top panel and all sessions in the bottom panel). Right panel: prediction under the hypothesis of Bayesian rationality, corresponding to the two samples.

## B Approximation of the Value Function

This appendix describes the estimation methods used to compute the value functions for the random menu cost model and for the inattention model. In each case, 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  $\pi_t$  at that point, and we use this atheoretical model to compute an implied value function  $V_t(p)$ .

### B.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  $\Lambda_t$ :

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

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

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

where  $v_t$  and  $u_t$  are two i.i.d. mean-zero random variables with variances  $\sigma_u^2$  and  $\sigma_v^2$ , respectively.

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

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

### B.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:

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

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

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

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

where  $\mathbf{1}_s$  is an indicator variable equal to 1 if the slider is adjusted on trial  $s$  and 0 otherwise, and  $\kappa_s$  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:

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

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

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

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  $R^\infty$  for any hypothetical slider position that may be chosen on trial  $t$ , given the posterior  $\pi_t$  at that time:

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

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

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

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

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

multiplies the probability of adjustment by the mean adjustment cost conditional on  $\kappa$  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

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

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

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

where

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

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

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

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

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

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

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

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 (B.1)-(B.2).
2. The parameters  $(\theta_0, \theta_1)$  of the empirical hazard function in (B.3).
3. The parameter  $\gamma_1$  of the estimated distribution of fixed costs in (B.4).

The parameter  $\gamma_1$  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  $\gamma_1$  is used to compute  $V_t(p)$  for arbitrary  $p$ , and hence the value of  $\Delta_t^{value}$  on each trial; these values are then used to estimate a value of  $\gamma_1$ , and this value must turn out to be the same value as was assumed in order to compute  $\Delta_t^{value}$ .

### B.3 Inattention Model

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

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

where  $R_t(p)$  is defined as in (B.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  $\pi_t$ :

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

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

where  $\Lambda_s$  is the probability of adjustment on trial  $s$ ,  $\mu_s$  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  $\mu$  and  $\lambda$ .

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  $\Lambda_t$ , and the time-varying measure over possible new slider positions  $\mu_t$ . We proceed as follows: First, using the experimental data, we estimate a value of  $\Lambda_t$  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

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

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

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

where

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

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

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

and

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

We obtain

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

where

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

where the matrix  $N$  is given by

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

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  $\psi_1$  and  $\psi_2$ . 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  $\psi_2$  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  $\psi_1$  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  $\psi_1$ ,  $\psi_2$ ,  $\tilde{\Lambda}$ , 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 ( $\gamma = 0$ ).

## C Full sample results

In this appendix we present results for the full sample (all 110 sessions), including the repeated sessions.

### C.1 Observed Behavior

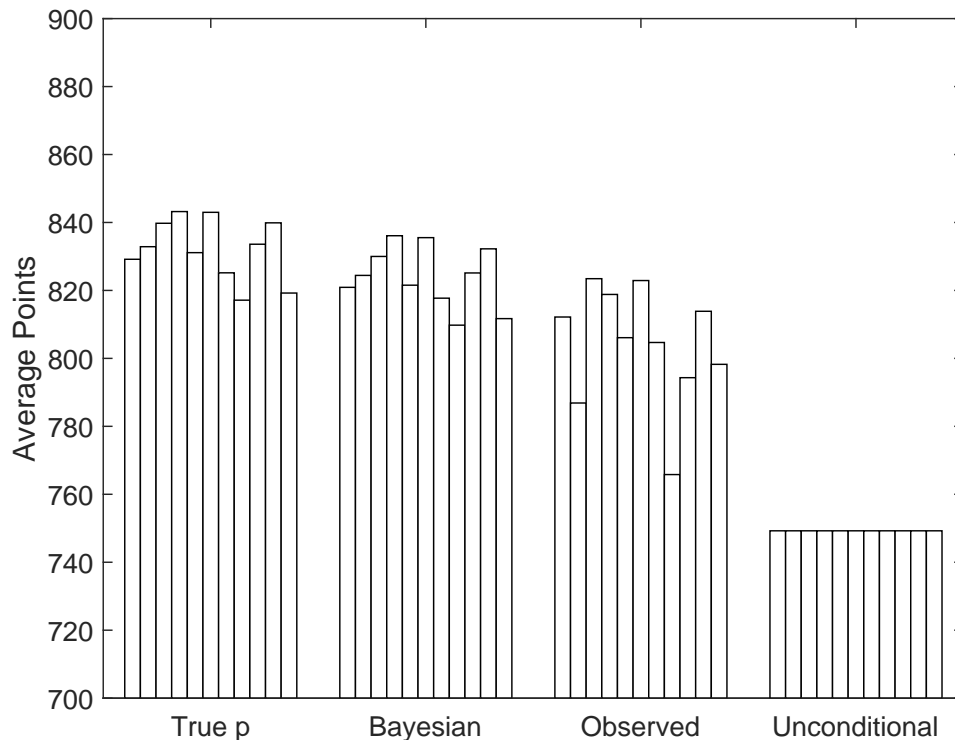


Figure C.1: Experiment data for all subjects and sessions. The scores are broken down by subject. “True p” refers to the score that each subject would have received, given the realized ring draws, had they known the underlying probabilities at all times. “Bayesian” refers to the scores that would have 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.

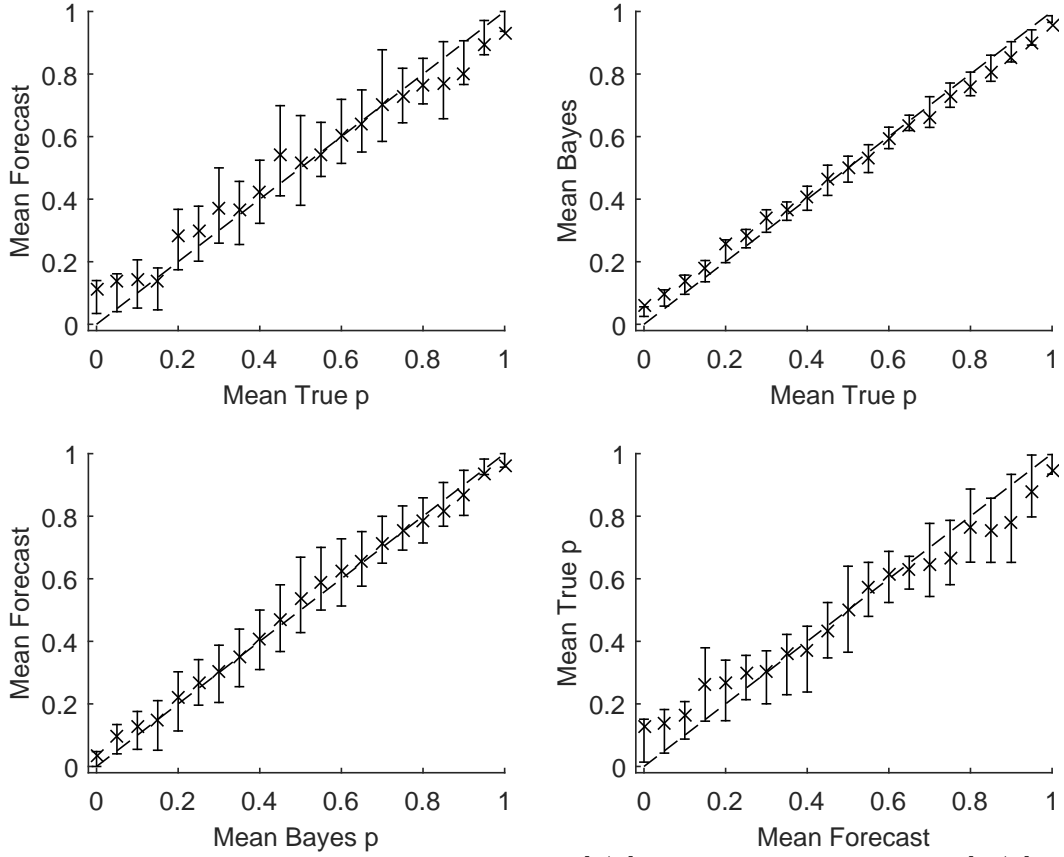


Figure C.2: Measures of forecast bias. Upper left:  $E[\hat{p}|p]$  versus  $p$ . Upper right:  $E[p^*|p]$  versus  $p$ . Lower left:  $E[\hat{p}|p^*]$  versus  $p^*$ . Lower right:  $E[p|p^*]$  versus  $p^*$ .

Regression: $y = \alpha + \beta x + \epsilon$				
$y$	$x$	$\alpha$	$\beta$	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 C.1: Regression tests of forecast bias, with alternative choices of the variables  $y$  and  $x$ . 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  $10^{-279}$ .

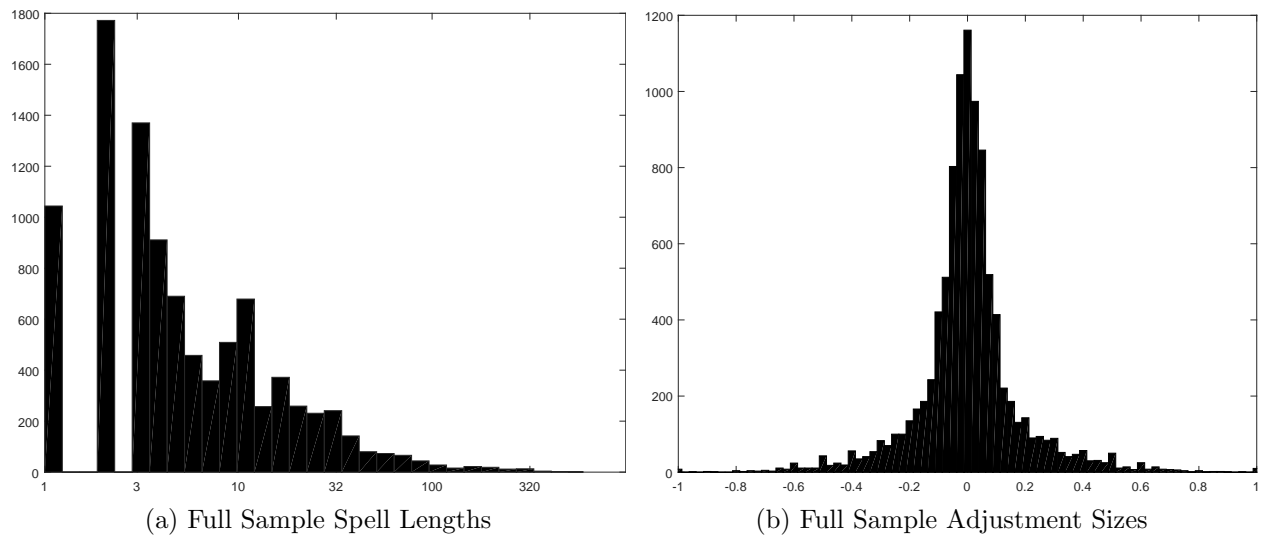


Figure C.3: Discreteness. (a) Distribution of the number of ring draws between adjustments (all sessions). (b) Distribution of the size of changes in the slider position, counting only trials on which the slider is moved (all sessions).

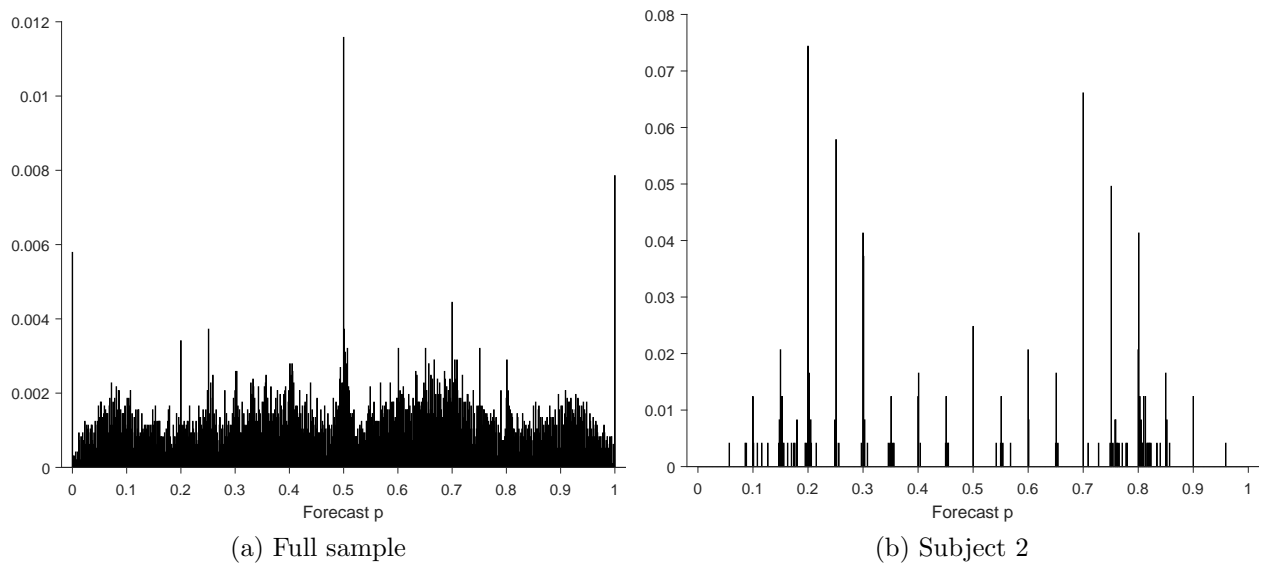


Figure C.4: Experiment data. Distribution of the subjects' slider position choices for (a) the full sample and (b) subject 2.

## C.2 Estimated Models of the Timing of Slider Adjustments

Table C.2 presents the full-sample results corresponding to those reported in Table 2 of the text. The top panel of the table presents the best-fitting Bayesian “gap-based” models of the adjustment decision. The polynomial coefficients are very similar to those obtained using only the unique sessions. The only change is that the full sample now suggest very weak evidence in favor of asymmetry, whereas the unique sessions sample failed to reject the null of a symmetric hazard function. The different magnitudes of the BIC statistics reflect the differences in the sample size. The middle panel of the table presents the best-fitting optimizing models of 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. The bottom panel gives results for the best-fitting rational inattention model, and these are again similar to those obtained using only the non-repeated sessions. Again we find that the rational inattention model offers by far the best fit to our data on the timing of slider adjustments, in terms of the BIC statistic.

TABLE C.2: Best-fitting models for the full sample, including repeated sessions.

<i>Gap-Based Models</i>								
Model	k	$\theta_0$	$\theta_1$	$\theta_2$				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
Model	k	$\epsilon$	$\underline{\Delta}$	$\Delta$				BIC
<i>Ss</i> with errors	3	0.088	-0.64	0.74				65,464
Symmetric <i>Ss</i>	2	0.088	-0.92	0.92				65,485
<i>Optimizing Models</i>								
Model	k	$\theta_0$	$\theta_1$	$\epsilon_1$	$\epsilon_2$	$\bar{\Delta}$	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	0.912	3.87	65,433	
Step function	3	–	–	0.082	0.20	0.60	64,800	
<i>Inattention Model</i>								
Model	k	$\theta_0$	$\theta_1$	$\psi_1$	$\tilde{\Lambda}$	BIC		
Rational inattention	2	-2.07	0.84	1.19	0.112	63,606		