Sticky Belief Adjustment: A Double Hurdle Model and Experimental Evidence*

Timo Henckel
Australian National University & CAMA

Gordon D. Menzies
University of Technology & CAMA

Peter Moffatt
CBESS & School of Economics, University of East Anglia

Daniel J. Zizzo
BENC, CAMA & Newcastle University Business School

This Version: 3 April 2017

Abstract

Given a lack of perfect knowledge about the future, agents need to form expectations about variables affecting their decisions. We present an experiment where subjects sequentially receive signals about the true state of the world and need to form beliefs about which one is true, with payoffs related to reported beliefs. We control for risk aversion using the Offerman et al. (2009) technique. Against the baseline of Bayesian updating, we test for belief adjustment under-reaction and over-reaction and model the decision making process of the agent as a double hurdle model where agents first decide whether to adjust their beliefs and, if so, then decide by how much. We find evidence for sticky belief adjustment. This is due to a combination of: random belief adjustment; state-dependent belief adjustment, with many subjects requiring considerable evidence to change their beliefs; and Quasi-Bayesian belief adjustment, with insufficient belief adjustment when a belief change does occur.

Keywords: belief revision; double hurdle model; expectations; overreaction; under-reaction.

JEL classification codes: C34; C91; D03; D84 ; E03.

* Daniel John Zizzo (corresponding author), Newcastle University Business School, Newcastle University, Newcastle upon Tyne, Tyne and Wear NE1 4SE, United Kingdom. E-mail address: daniel.zizzo@ncl.ac.uk. Phone: +44-191-20858544.

1 Introduction

Agents form and update their beliefs when they receive new information. The assumptions about how they do this are fundamental to a plethora of theoretical and empirical models, both in macro- and microeconomics. In the presence of rational expectations, new information leads to smooth and continuous belief updating according to Bayes' rule. In reality agents often systematically misunderstand basic statistics, and complexity and inattention may contribute to deviations from Bayesian predictions (Rabin, 2013). Such violations of rational expectations have been studied in static settings, where all the information is presented at once to subjects who discount priors (Tversky and Kahneman, 1982; Camerer, 1987; El-Gamal and Grether, 1995). In contrast, we study a dynamic setting in which new information arrives sequentially. We conduct a simple dynamic experiment and present a double hurdle econometric model to test the explanatory power of three different types of belief adjustment: Bayesian belief adjustment, a simple version of Quasi Bayesian belief adjustment, as well as random and state-dependent belief adjustment. We control for risk aversion using the Offerman et al.'s (2009) technique. We also consider how increased task complexity or scope for inattention affects our results.

Within microeconomic research, Quasi Bayesian (QB) belief adjustment has been the preferred route to think about bounded-rational belief adjustment. Rabin (2013) distinguishes between warped Bayesian models which encapsulate a false model of how signals are generated, for example ignoring the law of large numbers (Benjamin et al., 2015); and information-misreading Bayesian models that misinterpret signals as supporting agents'hypotheses, thus giving rise to confirmation bias (Rabin and Schrag, 1999), and therefore lead to underweighting of information. While various anomalies have been considered within this framework, one simple way of modeling QB adjustment is that the agent adjusts beliefs continuously in response to new information – in the sense that it takes place whenever there is new information – but this adjustment is either too big or too small (Massey and Wu, 2005; Ambuehl and Li, 2014).

There is a large body of macroeconomic research looking at time-dependent versus state-dependent price adjustment, with mixed empirical findings (e.g., Costain and Nakov, 2011; Aucremann and Dhyne, 2005; Stahl, 2005; Dias et al., 2007; Klenow and Kryvtsov, 2008; Midrigan, 2010). Sticky belief adjustment can be seen as a possible microfoundation of sticky price adjustment, for example as a result of inattention and observation costs (Alvarez et al., 2016), information costs (Abel et al., 2013), cognitive costs (Magnani et al., 2016) and the consultation of experts by inattentive agents (Carroll, 2003). State-dependence in beliefs implies a dependence of belief adjustment on the economic state, which in turn may depend on new information flowing in. As discussed later, a way of conceptualizing state-dependent beliefs is in the inferential expectations (IE) model of Menzies and Zizzo (2009): that is, agents hold a belief until enough evidence has accumulated for a statistical test of a given test size to result significant, at which point beliefs switch. Agents can be conceived of as drawing a test size α from a distribution, and this distribution provides a nuanced account of the extent of sticky belief adjustment. Time-dependence in beliefs is often viewed stochastically (as for example in Caballero. 1989) and therefore yields random belief adjustment, following some underlying data generating process. In both state-dependent and time-dependent macroeconomic models, beliefs are normally seen as adjusting partially to new information; furthermore, differently from Quasi-Bayesian models, in IE belief adjustment is not continuous but discrete, occurring at state-dependent or time-dependent time intervals.

One source of deviation from Bayesian updating could be task complexity (Caplin et al., 2011). Charness and Levin (2005) suggest that subjects are able to calculate Bayes' rule when the math is simple, but have difficulty calculating it when the math becomes complicated in which case they apply a different heuristic other than Bayes' rule when making decisions. Within consumer markets, complexity has often been blamed for suboptimality of consumer choices (e.g., Joskow, 2008; Ofgem, 2011; Independent Commission on Banking, 2011); the evidence from consumer experiments is less clear but consistent with at least some effect of complexity on consumer choice (Kalayci and Potters, 2011; Sitzia and Zizzo, 2011; Sitzia et al., 2015).

In the spirit of Sims (2003), inattention to the task could also lead to greater deviation from Bayesian updating (Alvarez et al., 2016; Magnani et al., 2016; Carroll, 2003). Our inattention manipulation consists of an alternative task being available and is closest to Corgnet et al. (2014), who find an effect on team effort in a work experiment, and Sitzia and Zizzo (2015), who find an effect on consumption choices.

We are not aware of research on complexity and inattention that has identified their effect on belief updating with sequential information flow. In brief, our results are as follows. Subjects choose to change their beliefs about half of the time, which is consistent with random belief adjustment, but they also consider the amount of evidence available, which is consistent with state-dependent belief adjustment. Indeed, we estimate that almost half of our subjects have various degrees of sticky belief adjustment. When subjects do change beliefs, they do so by around 40 per cent, which is consistent with our version of Quasi-Bayesian belief adjustment. Our results are perhaps surprisingly robust to either the task complexity or inattention manipulation, but there is consistent evidence that task complexity reduces the extent of a belief update.

Section 2 presents our experimental design and treatments, section 3 our expectation models, Sections 4 and 5 our results. Section 5 provides a discussion and concludes.

2 Experimental Design and Treatments

Our experiment was fully computerized and run in the experimental laboratory of a British university with subjects who were separated by partitions. The experiment was divided in two parts, labelled the practice part and the *main part.* Experimental instructions were provided at the beginning of each part for the tasks in that part (see the online appendix for a copy of the instructions). A questionnaire was administered to ensure understanding after each batch of instructions.

Main part of the experiment. After the practice part described below, in the main part of the experiment subjects played 7 stages, each with 8 rounds. At the beginning of each stage the computer randomly chose an urn out of two (Urn 1 or Urn 2), with Urn 1 being selected at a known probability of 0.6. Each urn represents a different state of the world. While this prior probability was known and it was known that the urn would remain the same throughout the stage, the chosen urn was not known to subjects. It was known that Urn 1 had seven white balls and three orange balls, and Urn 2 had three white balls and seven orange balls. At the beginning of each of the 8 rounds (round $= t$), there was a draw from the chosen urn (with replacement) and subjects were told the color of the drawn ball. These were therefore signals that could be used by subjects to update their beliefs. It was made clear to the subjects that the probability an urn was chosen in each of the seven stages was entirely independent of the choices of urns in previous stages.

Once they saw the draw for the round, subjects were asked to make a probability guess between 0% and 100%, on how likely it was that the chosen urn was Urn 1. The corresponding variable for analysis is their probability guess expressed as a proportion, denoted g. Once a round was completed, the following round got started with a new ball draw, up to the end of the 8^{th} round.

Payment for the main part of the experiment was based on the guess made in a randomly chosen stage and round picked at the end of the experiment. A standard quadratic scoring rule (e.g., Davis and Holt, 1993) was used in relation to this round to penalize incorrect answers. The payoff for each agent was equal to 18 pounds minus a quadratic penalty defined over the difference between a randomly chosen guess g and the correct probability (either 0 or 100%) that Urn 1 had been chosen.

Practice part of the experiment. The practice part was similar to the main part but simpler and therefore genuinely useful as practice. It was modelled after Offerman et al. (2009) to enable us to infer people's risk attitude, as detailed in section 3.

It consisted of 10 rounds. Each round was essentially the same as the guessing task in the main part, but with subjects being told the prior probability of Urn 1 being chosen (in sequence, 0.05, 0.1, 0.15, 0.2, 0.25, 0.75, 0.8, 0.85, 0.9, 0.95) and not receiving any further information – that is, no ball draws were forthcoming and the subjects were to conceive of each round as a potential new urn draw. Payment for the practice part of the experiment was based on the guess made in a randomly chosen round picked at the end of the experiment. A quadratic scoring rule was applied as in the main part, but this time with a top prize of 3 pounds rather than 18 pounds.[1]

[1] This ensured similar marginal incentives for each round in the practice part (3 pounds prize picked up from 1 out of 10 rounds) and the main part (18 pounds prize picked up from 1 out of 56 rounds).

Experimental treatments. There were three treatments. The practice parts were identical across treatments, and the main part of the Baseline treatment was as described.

In the main part (only) of the Complex treatment, the information on the ball drawn from the chosen urn at the beginning of each round was presented in a complex way. Specifically, it was presented as a statement about whether the sum of three numbers (of three digits each) is true or false. If true (e.g., 731 + 443 + 927= 2101), this meant that a white ball draw was drawn. If false (e.g., 731 + 443 + 927= 2121), this meant that an orange ball draw was drawn.

In the main part (only) of the Inattention treatment, subjects were given a non-incentivized alternative counting task which they could do instead of working on the probability. The counting exercise was a standard one from the real effort experimental literature (see Abeler et al., 2011, for an example) and consisted in counting the number of 1s in matrices of 0s and 1s. Subjects were told that they could do this exercise for as little or as long as they liked within 60 seconds for each round, and that we were not asking them in any way to engage in this exercise at all unless they wanted to.[2]

3 Model Variables, Risk and Expectation Process

3.1 Model Variables and Risk Attitude Correction

Table 1 lays out the main model variables and the interrelationships between them, when the event is described in terms of the chosen urn (row 1) and when it is described in terms of the probability of a white ball being drawn (row 2). The two descriptions are equivalent since the subject's subjective guess of the probability that Urn 1 was chosen generates an implied subjective probability that a white ball is drawn.[3] In our modelling, we sometimes use the former probability—that Urn 1 was chosen—and it will be useful to transform this probability guess using the inverse cumulative Normal distribution, so that the support has the same dimensionality as a classic z-statistic. Alternatively, we sometimes describe agents' guesses in terms of the probability that a white ball is drawn.

Time is measured by t, the draw (round) number for the ball draws in each stage. We define the value of t for which subjects last moved their guess (viz. updated their beliefs) to be m (for 'last Move'). Thus, for any sequence of ball draws at time t, the

[2] They were also told that, if they did not make a guess in the guessing task within 60 seconds, they would automatically keep the guess from the previous round and move to the next round (or to the next stage). The length of 60 seconds was chosen based on piloting, in such a way that this would not be a binding constraint if subjects focused on the guessing task.

[3] For example, at the start of the experiment, before any ball is drawn, subjects know that the chance that Urn 1 was drawn is 0.6. It therefore follows that the chance of a white ball draw the very first time is 0.7*0.6+0.3*(1-0.6) = 0.54.

time that has elapsed since the last change in the guess is always $t - m$.

Table 1: Model Variables

Event	Estimator of Event probability	Estimator Symbol	P(Event) Subject guesses	Transformed guess	Strength of Evidence z against earlier chosen $P(\text{Event})_k$
U Urn 1	Bayes rule	P_t	g_t: optimal guess (observed) g^*_t: inferred guess	$r^*_t = \Phi^{-1}(g^*_t)$	$z_{p,t} = \Phi^{-1}(P_t) - \Phi^{-1}(P_m)$
W white	Prop. white balls out of t	Pw_t	not guessed	not guessed	$z_{pw,t} = 2\sqrt{t}[Pw_t - \{0.7P_m + 0.3(1-P_m)\}]$

Starting with the top row, the event U refers to Urn 1 being drawn (the urn with seven white balls and three orange ones). The theoretical estimator for the probability that Urn 1 was drawn is provided by Bayes rule, which we denote by P_t after t ball draws. Subjects do not use Bayes rule when they are guessing the probability of U, though some guesses are closer to it than others.

As derived in Offerman et al. (2009), the elicited guess g_t in the fourth column is the result of maximizing utility based on a Constant Relative Risk Aversion (CRRA) utility function, V:

$$V = \frac{\text{Payoff}^{1-\theta} - 1}{1-\theta},$$

and a true guess g_t^* in $E[V(g_t)] = g_t^* V\left(1 - (1-g_t)^2\right) + [1-g_t^*]V(1-g_t^2)$. Expected utility is maximized with respect to g_t and yields the following relationship between g_t^* and g_t:

$$\ln\left(\frac{g_t^*(1-g_t)}{g_t(1-g_t^*)}\right) = \theta \ln\left(\frac{g_t(2-g_t)}{(1+g_t)(1-g_t)}\right). \tag{1}$$

In the practice part rounds the prior probabilities given to the subjects (by way of reminder, 0.05, 0.1, 0.15, 0.2, 0.25, 0.75, 0.8, 0.85, 0.9, 0.95 for 10 separate rounds) are in fact the correct probabilities P_t. We see no reason to not credit subjects with realizing this, and they possess no other information anyway. Offerman et al. (2009) then interpret the deviations of g_t from g_t^* as being due to the subjects' risk preferences, and so do we. We equate the ten provided priors to g_t^* and then use the ten datapoints (g_t^*, g_t) for each subject to estimate θ in a version of (1) appended with a regression error.[4] Armed with a subject-specific value of θ from the practice part, all the observable g_t values in the main experiment can be transformed to a set

[4] If an agent ever declared zero or unity in this preliminary stage, a regression version of (2) cannot be run, and so we set $\theta = -1$ in those cases.

of inferred g_t^*. This transformation is accomplished by exponentiating both sides of (1), and solving for g_t^*. By taking the inverse cumulative Normal function, Φ^{-1}, of g_t^* we move it outside the $[0,1]$ interval and give it the same dimensionality as a test statistic, namely $(-\infty,\infty)$. The variable in the penultimate column, $r_t^* = \Phi^{-1}(g_t^*)$, thus becomes the basis for all subsequent analysis.

In the final column of the first row, we provide a measure $z_{P,t}$ of the strength of evidence against the probability guess at the time of the last change, where the subscript P reminds us that this measure is based on Bayes rule and where we transform both probabilities by the inverse cumulative Normal function. Agents change their guesses from time to time, and z_P tells us if the value of P at the last change, denoted P_m, seems mistaken in the light of subsequent evidence.

This z_P measure is objective rather than based on the guesses of the subjects. It describes how an agent who always calculates the correct Bayesian probability would perceive a period of inertia.[5] Thus, it must stand for an 'as if' assumption for the subjects' beliefs—after all, if they knew P_t they would use it rather than g_t^*—but it is important for our modelling that we can derive the mathematical properties of z_P as this will allow us to derive some key results about sticky belief adjustment.

Turning to the second row, the event W refers to a white ball being drawn, which agents do not explicitly guess. The theoretical estimate for this probability is the proportion of white balls, which we denote by Pw_t after t ball draws. Because subjects are asked to guess the probability of the chosen urn being Urn 1, it is possible to form an implicit guess of Pw_t based on the last guess of the Bayesian probability P, namely P_m. This guess is the probability of a white ball, $0.7P_m + 0.3(1-P_m)$, and is based on the assumption that Urn 1 has a 70% proportion of white balls while Urn 2 has a 30% proportion.

The measure $z_{Pw,t}$ is based on the standard test statistic for a proportion, using the maximal value of the variance of the sampling distribution (namely $(\frac{1}{2})^2$):

$$z_{Pw,t} = \frac{Pw_t - (0.7P_m + 0.3(1-P_m))}{(0.5^2/t)^{1/2}}. \tag{2}$$

When we analyze the inferential expectations of each agent we can equate $z_{P,t}$ and $z_{Pw,t}$ to each other which, together with our subsequent estimates, enables us to recover the entire distribution of the test size as a key component in our description of sticky belief adjustment. For computational ease, all our econometric estimation uses $z_{P,t}$. Appendix 1 shows that the two measures $z_{P,t}$ and $z_{Pw,t}$ are close numerically

[5] Since our decision problem is a double hurdle one where the first decision is whether to adjust beliefs or not, it has to be the case that the greater the evidence against the null hypothesis, the greater the financial cost from sticking to the null belief when the alternative is true, and therefore the more likely the agent chooses to switch belief. In our experimental set up, there is therefore a one-to-one correspondence between our inferential expectations (IE) formalism and state-dependent cost based belief adjustment. Thus, we can use the IE formalism as a way of appropriately representing not just IE but also other state-dependent belief adjustment models based on comparing the costs of adjusting (whether cognitive, inattention, etc.) with the costs of not adjusting.

and theoretically. In the subsequent analysis we will indicate which measure is being used.

3.2 Expectation Processes

Using the notation around Table 1, we define three processes of expectation formation that will be relevant for our double hurdle model in section 4.

Rational Expectations

The rational expectations (RE) solution predicts straightforward Bayesian updating. The (conditional) probability that the subject is being asked to guess is the rational expectation (RE) which is given by P_t. Calling $P_{initial}$ the initial prior probability and noting that the number of white balls is tPw_t we can write down P_t in a number of ways:

$$\begin{aligned} P_t &= \frac{(0.7)^{tPw_t}(0.3)^{t-tPw_t}P_{initial}}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}P_{initial}+(0.3)^{tPw_t}(0.7)^{t-tPw_t}(1-P_{initial})} \\ &= \frac{1}{1+\frac{(0.7)^{tPw_t}(0.3)^{t-tPw_t}(1-P_{initial})}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}P_{initial}}} \\ &= \left(\frac{(0.7)^{tPw_t}(0.3)^{t-tPw_t}}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}P_{initial}+(0.7)^{tPw_t}(0.3)^{t-tPw_t}(1-P_{initial})}\right)P_{initial}. \end{aligned} \tag{3}$$

The second line is a useful simplification (which we use in appendix 1) whereas the bracketed fraction in the third line is the probability of obtaining the tPw_t white balls when Urn 1 is drawn versus the total probability of obtaining this number of white balls.

Quasi-Bayesian Updating

In our version of Quasi-Bayesian updating (QB), agents use Bayesian updating as each new draw is received, but they incorrectly weight the bracketed probability fraction:

$$P_t^{QB} = \left(\frac{(0.7)^{tPw_t}(0.3)^{t-tPw_t}}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}P_{initial}+(0.7)^{tPw_t}(0.3)^{t-tPw_t}(1-P_{initial})}\right)^{\beta}P_{initial} \tag{4}$$

The parameter β may be thought of as the QB parameter, and if $\beta = 1$, we are back to rational expectations. Agents in this framework can exhibit: $\beta > 1$ where they overuse information and under-weight priors, $0 \leq \beta < 1$ where they underuse information and over-weight priors, or even $\beta < 0$ where they respond the wrong way to information—raising the conditional probability when they should be lowering it, and vice versa.

Agents' attitude towards the extent of belief change in the light of evidence can be summarized by the distribution f_β across subjects. If f_β has most probability mass between 0 and 1, most agents only partially adjust, and subjects converge to full adjustment to the extent that the probability mass in f_β is towards unity.

Inferential Expectations

In this version of state dependent belief adjustment, agents form a belief and do not depart from that belief until the weight of evidence against the belief is sufficiently strong, as measured by the result of the hypothesis test applied to the test statistic $z_{Pw,t}$ in the bottom right hand corner of Table 1.

Under inferential expectations (IE), each agent starts with a belief about the probability of U (that is, $P_0 = 0.6$) and an implied probability of a white ball ($Pw_0 = 0.54$) and conducts a test that the latter is true after drawing a test size from their own distribution of α, namely $f_i(\alpha)$. They are assumed to draw this every round during the experiment. The p-value of the test is derived using the measure in the bottom right hand corner of Table 1 as the test statistic. We assume for simplicity that $z_{Pw,t}$ is distributed as a standard Normal.

Agents' attitude towards changing beliefs in the light of evidence can be summarized by the distribution $f_i(\alpha)$. If $f_i(\alpha)$ has most probability mass near zero, agent i is sluggish to adjust. Probability mass in $f_i(\alpha)$ near unity implies a very strong willingness to use evidence, and probability mass at unity implies time dependent updating. That is, if the probability mass at unity in $f_i(\alpha)$ is, say, 0.3, it implies that there is a thirty per cent chance that agent i will update regardless of what the evidence says.

This is because the decision rule in a hypothesis test is to reject H_0, the status quo, if the p-value $\leq \alpha$. A value for α of unity implies the status quo will be rejected, which is the same as updating in this context, for any p-value whatsoever.

Relationship between Expectations Benchmarks

When agent i rejects H_0 within the IE framework we assume she updates her probability guess. This agent can either be fully Bayesian (when she does update) and adopt RE, or she can adopt QB and only update partly (or over-react, if $\beta_i > 1$). In the QB case where $0 < \beta_i < 1$, she moves by a fraction β_i of the distance she should move.

Since each agent has a full distribution of α, namely $f_i(\alpha)$, we need a representative α_i to summarize the extent of sticky belief adjustment for agent i and to relate to her β_i. There are a number of possibilities, but a natural choice which permits analytic solutions is the median α_i from their $f_i(\alpha)$. For the purposes of our empirical analysis a fully rational (Bayesian) agent is one who has $\alpha_i = \beta_i = 1$, whereas any other sort of agent does not have RE.[6]

[6] As we shall see, there is a minor technical qualification to this last sentence because our estimated

We now parameterize all three expectation processes in a double hurdle model. We find evidence for all of them in our data, and importantly we find that the IE representation of $f_i(\alpha)$ has non-zero measure at unity. As discussed above, this is the fraction of agents who undertake random belief adjustment.

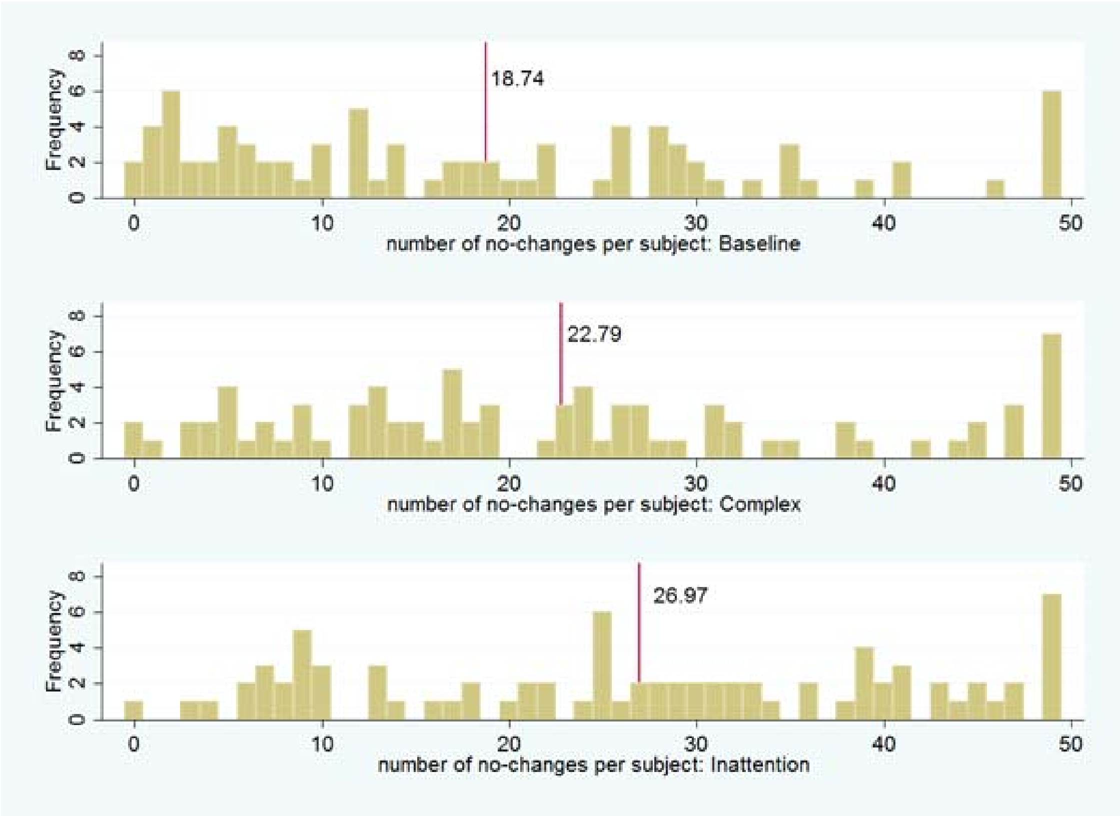
4 Experimental Results

4.1 Preliminary Analysis of "No-change"

The baseline and complex treatments each had 82 subjects, and the inattention treatment had 81 subjects. In this sub-section, we consider the number of times our subjects executed a "no-change", meaning a guessed probability equal to that of the previous period. This is interesting because, given the nature of the information and comparatively small number of draws, incidences of "no-change" are not predicted by either Bayesian or Quasi-Bayesian updating, and so, if such observations are widespread in the data, this is the first piece of evidence that these standard models are unlikely to be enough to describe the data.

The maximum of the number of "no-changes" for each subject is 49: seven opportunities for no change out of eight draws, times the seven stages. The distributions over subjects separately by treatment are shown in Figure 1. The baseline distribution shows a concentration at low values; for both Complex and Inattention, there appears to be a shift in the distribution towards higher values, as one might expect. The means for each treatment are represented by the vertical lines.

$f_i(\alpha)$ densities turn out to be a discrete/continuous mixture.



The mean is higher under C (22.79) than under B (18.74) (Mann-Whitney test gives $p = 0.007$); and higher under I (26.97) than under B ($p < 0.001$).[7] This is expected: complexity and inattention are both expected to increase the tendency to leave guesses unchanged. When C and I are compared, the p-value is 0.06, indicating mild evidence of a difference between the two treatments.

It is clear from the nonparametric evidence in Figure 1, of widespread incidence of "no-changes", that any successful model of our data will have to deal with the phenomenon of whether to adjust, before considering how much to adjust. This in turn can imply that the waiting process is stochastic and is unrelated to the actual information arriving (random belief adjustment) or, that the information that arrives influences the timing (state-dependent belief adjustment). Our double hurdle model enables us to consider both.

[7] All p-values in the paper are two tailed. All bivariate tests use subject level means as the independent observations to avoid the problem of non-independence of within-subject choices.

4.2 A Double Hurdle Model of Belief Adjustment

In this section, we develop a parametric double hurdle model which simultaneously considers the decision to update beliefs and the extent to which beliefs are changed when updates occur. Our econometric task is to model the transformed implied belief $r_t^* = \Phi^{-1}(g_t^*(\theta_i))$, which in turn requires an estimate for risk aversion. We estimate this at the individual level using the technique by Offerman et al. (2009). Appendix 2 contains the subject-level details surrounding the estimation of θ_i and a scatterplot of the resultant guesses g_t^* against g_t.

We will refer to r_{it}^* as subject i's "belief" in period t, as shorthand for 'transformed implied belief'. We will treat r_{it}^* as the focus of the analysis, because r_{it}^* has the same dimensionality as $z_{P,it}$, the test statistic defined in the far right of row 1 of Table 1. That is, both have support $(-\infty, \infty)$. Sometimes r_{it}^* changes between $t-1$ and t; other times, it remains the same. Let Δr_{it}^* be the change in belief of subject i between $t-1$ and t. That is, $\Delta r_{it}^* = r_{it}^* - r_{it-1}^*$.

In the following estimation, justified in Appendix 1, we use changes in the Bayes' probability as an 'as if' proxy for the IE test statistic that we assume agents mentally compute in order to decide whether to update their belief. Since we equate $z_{P,t}$ to $z_{Pw,t}$ we use z_t for both from now on. In round 1 P_m equals the prior 0.6 and the movement of the guess for a given subject is $\Delta r_{it}^* = r_1^* - \Phi^{-1}(0.6)$. That is, both the objective measure of the information change and the subjective guess of the agent are assumed to anchor onto the prior probability that Urn 1 is chosen, 0.6, in the first period.

First Hurdle: The probability that a belief is updated (in either direction) in period t is given by:

$$P(\Delta r_{it}^* \neq 0) = \Phi\left[\delta_i + (\gamma + x_i'\theta_1)|z_{it}|\right], \tag{5}$$

where $\Phi[\cdot]$ is the standard Normal cdf and δ_i represents subject i's idiosyncratic propensity to update beliefs, and therefore models random probabilistic belief adjustment. The probability of an update is assumed to depend (positively) on the absolute value of z_{it}, the test statistic. The expression in round brackets represents the effect of this variable, hence representing the impact of cumulative evidence (in either direction) on the propensity to update. The vector x_i contains treatment (dummy) variables, which are time invariant, and consequently γ represents the impact of cumulative evidence for a baseline subject, which is connected to state-dependent belief adjustment. Finally, the elements of θ_1 tell us how this impact differs by treatment.

One econometric problem that arises is the endogeneity of the variable $|z_{it}|$: subjects who are averse to updating tend to generate large values of $|z_{it}|$ while subjects who update regularly do not allow it to grow beyond small values. This is likely to create a severe downward bias in the estimate of the parameter γ in the first hurdle. To address this problem, we use an instrumental variables (IV) estimator which uses a variable in place of $|z_{it}|$, where $|\widehat{z}_{it}|$ comprises the fitted values from a regression of $|z_{it}|$ on a set of suitable instruments. This IV procedure is explained in detail in Appendix 3.

Second hurdle: Conditional on one or more subjects choosing to update beliefs in draw t, the next question relates to how much they do so. This is given by:

$$\Delta r_{it}^{*} = \left(\beta_i + x_i'\theta_2\right) z_{it} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N\left(0, \sigma^2\right). \tag{6}$$

As a reminder, the Quasi-Bayesian belief adjustment parameter β_i represents subject i's idiosyncratic responsiveness to the accumulation of new information: if $\beta_i = 1$, subject i responds fully; if $\beta_i = 0$, subject i does not respond at all. Remember that β_i is not constrained to $[0, 1]$. In particular, a value of β_i greater than one would indicate the plausible phenomenon of over-reaction. Again, treatment variables are included: the elements of the vector θ_2 tell us how responsiveness differs by treatment. There are two idiosyncratic variables, δ_i and β_i. These are assumed to be distributed over the population of subjects as follows:

$$\begin{pmatrix} \delta_i \\ \beta_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \eta_1^2 & \rho\eta_1\eta_2 \\ \rho\eta_1\eta_2 & \eta_2^2 \end{pmatrix}\right]. \tag{7}$$

In total, there are eleven parameters to estimate: μ_1, η_1, μ_2, η_2, ρ, γ, σ, and four treatment effects.

The results are presented in Table 2, for four different models.

Table 2: Results of hurdle model with risk adjustment

	QB	*IE*	*IE-QB* (ρ=0)	*IE-QB*
	(1)	(2)	(3)	(4)
Propensity to update:				
μ_1	$+\infty$	-0.170**(0.054)	-0.073(0.050)	-0.063(0.052)
η_1	0	0.915**(0.040)	0.962**(0.039)	0.992**(0.042)
γ	0	0.555**(0.082)	0.572**(0.078)	0.566**(0.078)
Complex	0	0.011(0.101)	-0.033(0.090)	-0.031(0.091)
Inattention	0	-0.060(0.101)	-0.063(0.097)	-0.064(0.100)
Extent of update:				
μ_2	0.326**(0.031)	1	0.501**(0.038)	0.412**(0.037)
η_2	0.286**(0.018)	0	0.339**(0.022)	0.364**(0.024)
Complex	-0.131**(0.040)	0	-0.188**(0.046)	-0.136**(0.053)
Inattention	-0.132**(0.051)	0	-0.122*(0.050)	-0.029(0.050)
σ	0.524**(0.003)	0.851**(0.007)	0.676**(0.006)	0.675**(0.006)
ρ				0.429**(0.065)
LogL	-10,193.67	-16,497.07	-14,964.52	-14945.18
AIC (=2k-2LogL)	N/A	33,006.14	29,949.04	29,912.36
Wald test (df, p-value)	61436(6,0.000)	1192(5,0.000)	33.34(1,0.000)	N/A
Total observations	245	245	245	245
Observations per subject	56	56	56	56

Note: LogL for QB cannot be compared with that of other columns.

Model 1 estimates the QB benchmarks, in which it is assumed that the "first hurdle" is crossed for every observation—that is updates always occur. Zero updates are treated as zero realizations of the update variable in the second hurdle, and their likelihood contribution is a density instead of a probability. Because of this difference in the way the likelihood function is computed, the log-likelihoods and AICs cannot be used to compare the performance of QB to that of the other models.

Model 2 estimates the IE benchmark, in which the update parameter (β_i) is fixed at 1 for all subjects. Consequently the extra residual variation in updates is reflected in the higher estimate of σ. The parameters in the first hurdle are free.

Model 3 combines IE and QB, but constrains the correlation (ρ) between δ and β to be zero. Model 4 is the same model with ρ unconstrained.

The overall performance of a model is best judged using the AIC; the preferred model is the one with the lowest AIC. Among the models that can be compared, the best model is the most general model 4: IE-QB with ρ unrestricted, whose results are presented in the final column of Table 2.

To confirm the superiority of the general model over the restricted models, we conduct Wald tests of the restrictions implied by the three less general models. We see that, in all three cases, the implied restrictions are strongly rejected, implying that the general model is superior. Note in particular that this establishes superiority of the general model 4 (IE-QB with ρ unrestricted) over the QB model 1 (a comparison that was not possible on the basis of AIC).

We interpret the results from the preferred model as follows. Consider the first hurdle (propensity to update). The intercept parameter in the first hurdle (μ_1) tells us that a typical subject has a predicted probability of $\Phi(-0.063) = 0.47$ of updating in any task, in the absence of any evidence (i.e. when $|z_{it}| = 0$).

Result 1 There is evidence of random belief adjustment. In every period subjects update their beliefs idiosyncratically around half of the time.

The large estimate of η_1 tells us that there is however considerable heterogeneity in this propensity to update (see Figure 2 below), something we will explore further in section 4.3. The parameter γ is estimated to be significantly positive, and this tells us, as expected, that the more cumulative evidence there is, in either direction, the greater the probability of an update.

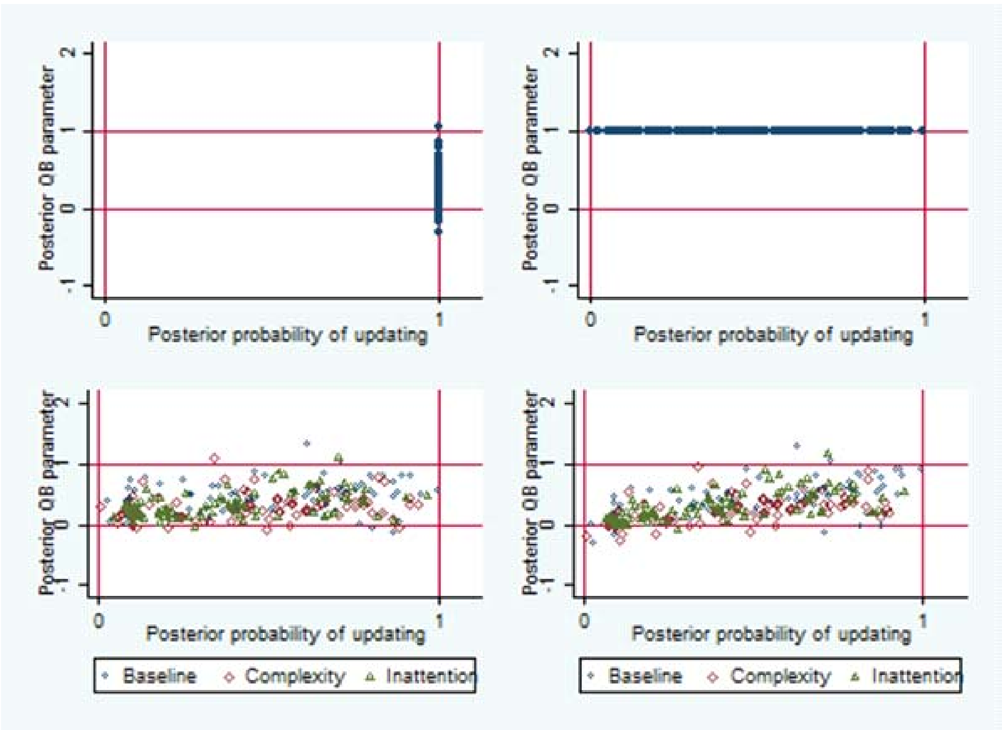
Result 2 There is evidence of state-dependent belief adjustment. Subjects are more likely to adjust if there is more evidence to suggest that an update is appropriate (thus making it costlier not to update).

The two treatments have the expected signs: complexity reduces the impact of information on the propensity to update; inattention also has a negative effect. However, neither of these treatment effects is significant, which suggests the robustness of Results 1 and 2 to different treatment manipulations, i.e. these results do not require additional complexity or potential sources of inattention.

In the second hurdle, the intercept (μ_2) is estimated to be 0.41 in our preferred model 4: when a typical (baseline) subject does update, she updates by a proportion

0.41 of the difference from the Bayes probability. The large estimate of η_2 tells us that there is considerable heterogeneity in this proportion also (see Figure 2 below). However, in all models where the second hurdle is meaningful (models 1, 3 and 4), almost all of the β_i's are between 0 and 1. If we take model 4, only 23 out of 245 subjects have $\beta < 0$, which indicates noise or confused subjects who adjusted in the wrong direction.[8] More interestingly, only 3 out of 245 subjects (~1%) display overreaction to the evidence in model 4, with similar numbers in the other models (1 in model 1 and 4 in model 3).

Result 3 There is evidence of Quasi-Bayesian partial belief adjustment. On average, subjects who adjust do so by around 40%. There is no evidence of prior information under-weighting: virtually none of the subjects overreact to evidence once they decide to adjust.



The estimate of ρ is strongly positive, indicating that subjects who have a higher propensity to update, also tend to update by a higher proportion of the difference from the Bayes probability. This positive correlation is seen in the Model 4 plot in Figure 2.

[8]There were a small number of cases in each of the three treatments (B: 7; C: 11; I: 5 cases).

The treatment effects in the second hurdle are again of the expected sign. The inattention effect is significant at the 5% level in model 1, but, once we model the first hurdle the significance is only at the 10% level, and once we recognize the positive correlation between δ and β, it disappears. In other words, whatever inattention effect there is, it is captured by the fact that subjects who update less update by a lower degree. There is instead evidence of that further complexity in the decision problem reduces the extent of update by almost 15%.

Result 4 Any inattention effect is captured by the positive correlation between δ and β. Complexity does not affect whether subjects decide to update or not, but, if they do, they partially adjust by almost 15% less on average.

4.3 The Empirical Distribution of α and β

To get a better sense of the population heterogeneity in sticky belief adjustment, this subsection maps out the empirical distribution of the IE α_i across subjects and QB β_i parameters across subjects against each other. Estimating f_β, the distribution of β is easy enough to infer from our double hurdle models and we have done so in Figure 2. We next use the first hurdle information to generate $f_i(\alpha)$, the empirical distribution of α_i.

As we flagged earlier each agent has a full distribution of α and so we need a representative α_i to summarize the extent of sticky belief adjustment for agent i, to then relate to their β_i. As will be clear below, the choice that permits analytic solutions is the median α_i from $f_i(\alpha)$.

The econometric equation for the first hurdle is equivalent to the probability of rejecting the null under IE. We omit the dummy variables, which are insignificant. We begin by re-writing the first hurdle, namely (5) without dummies:

$$\Pr(\text{reject } H_0)_{it} = \Phi(\delta_i + \gamma |z_{it}|), \tag{8}$$

where δ_i and γ are estimated parameters and $|z_{it}|$ is the test statistic based on the proportion of white balls.

$$z_{it} = \frac{Pw_{it} - 0.54}{\sqrt{0.5^2/t}}. \tag{9}$$

For any $|z_{it}|$ it is possible to work out an implied p-value and we do so by assuming that (9) is approximately distributed $N(0,1)$. This in turn allows us to work out $f_i(\alpha)$ from the econometric equation for the first hurdle. When $|z_{it}| = 0$, the p-value for a hypothesis test is unity, and so the equation says that a fraction of agents will reject H_0 if the p-value is unity. Since the ciriteria for rejecting H_0 in a hypothesis test is always $\alpha \geq p$ value it implies that there must be a non-zero probability mass on $f_i(\alpha)$ at the value of α *exactly* equal to 1. The distribution of α_i will thus have a discrete 'spike' at unity and be continuous elsewhere. We know what that spike is from equation (8) with $|z_{it}| = 0$, namely $\Phi(\delta_i)$.

The probability of rejecting H_0 depends on the probability that the test size is greater than the p-value, but this is also equal to the econometric equation for the first hurdle.

$$\Pr(\text{reject } H_0)_{it} = \int_{p\text{-value}_{it}}^{1} f_i(\alpha)\, d\alpha_i = 1 - F_i(p\text{-value}_{it}) = \Phi(\delta_i + \gamma |z_{it}|) \tag{10}$$

Upper case F in the last equality is the anti-derivative of the density. We define $F_i(1)$ to be unity since 1 is the upper end of the support of α but we also note that there is a discontinuity such that F jumps from $1 - \Phi(\delta_i)$ to 1 at $\alpha = 1$, as a consequence of the non-zero probability mass on $f_i(\alpha)$ at unity. To solve the equation we use an expression for the p-value of $|z_{it}|$ on a two-sided Normal test.

$$p\text{-value}_{it} = 2(1 - \Phi(|z_{it}|)). \tag{11}$$

We use a 'single parameter' approximation to the cumulative Normal (see Bowling et al. 2009). For our purposes $\sqrt{3}$ is sufficient for the single parameter.

$$\Phi(|z_{it}|) = \frac{1}{1 + \exp\left(-\sqrt{3}\,|z_{it}|\right)}. \tag{12}$$

We can now write down $|z_{it}|$ as a function of the p-value using (12).

$$|z_{it}| = \frac{1}{\sqrt{3}} \ln\left(\frac{2 - p\text{-value}_{it}}{p\text{-value}_{it}}\right). \tag{13}$$

Intuitively, a p-value of zero implies an infinite $|z_{it}|$ and p-value of unity implies $|zit|$ is zero. We can now use the relationship between $F_{\alpha I}(p\text{-value}_{it})$ and our estimated first hurdle to generate $F_i(\alpha)$.

$$\begin{aligned}
1 - F_i(p\text{-value}_{it}) &= \Phi(\delta_i + \gamma |z_{it}|) \\
\therefore F_i(p\text{-value}_{it}) &= 1 - \Phi(\delta_i + \gamma |z_{it}|) \\
&= \Phi\left(\delta_i + \gamma\left\{\frac{1}{\sqrt{3}} \ln\left(\frac{2 - p\text{-value}_{it}}{p\text{-value}_{it}}\right)\right\}\right).
\end{aligned} \tag{14}$$

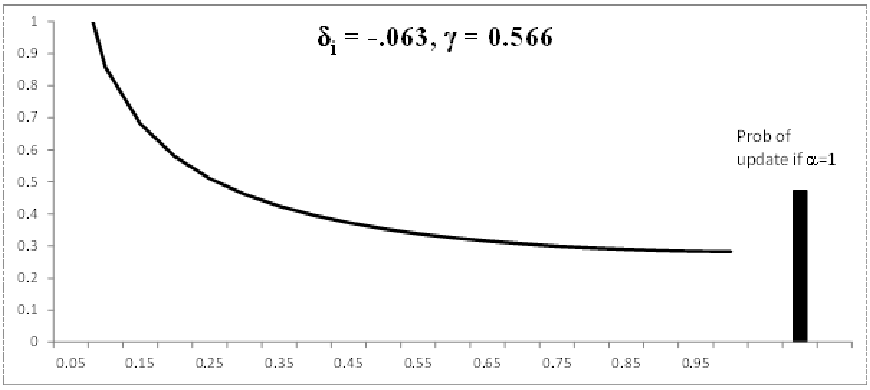
In the above expression the variable 'p-value' is just a place-holder and can be replaced by anything with the same support leaving the meaning of (14) unchanged. Thus, it can be replaced by α giving the cumulative density of α.

$$\begin{aligned}
F_i(\alpha) &= 1 - \Phi\left(\delta_i + \gamma\left\{\frac{1}{\sqrt{3}} \ln\left(\frac{2-\alpha}{\alpha}\right)\right\}\right) \\
&= 1 - \frac{1}{1 + \left[\frac{\alpha}{2-\alpha}\right]^{\gamma} \exp\left(-\sqrt{3}\delta_i\right)}.
\end{aligned} \tag{15}$$

Substitution of $\alpha = 1$ does not give unity, which is what we earlier assumed for the value of $F_i(1)$. However, it does give $1 - \Phi(\delta_i)$, which of course concurs with the econometric equation for the first hurdle when $|z_{it}| = 0$. This discontinuity in F_i is consistent with a discrete probability mass in $f_i(\alpha)$ at unity, as we noted earlier. It now just remains to differentiate F_i to obtain the continuous density $f_i(\alpha)$ for α strictly less than unity. The description of the function at the upper end of the support (unity) is completed with a discrete mass at unity of $\Phi(\delta_i)$.

$$\left.\begin{array}{lr} f_i(\alpha) = \frac{2\gamma \exp\left(-\sqrt{3}\delta_i\right)\alpha^{\gamma-1}}{(2-\alpha)^{1+\gamma}\left\{1\left[\frac{\alpha}{2-\alpha}\right]^{\gamma}\exp\left(-\sqrt{3}\delta_i\right)\right\}^2}, & \alpha < 1 \\ \Pr_i(\alpha = 1) = \frac{1}{1+\exp\left(-\sqrt{3}\delta_i\right)} = \Phi(\delta_i), & \alpha = 1 \end{array}\right\} \qquad (16)$$

Figure 3 illustrates the distribution $f_i(\alpha)$ for $\delta_i = -0.063$ and $\gamma = 0.566$. The former is the mean of δ across subjects, from our estimation. On the right-most of the chart is the probability mass when $\alpha = 1$. As discussed earlier, this corresponds to the proportion of agents who update without any evidence at all ($|z_{it}| = 0$).



Since there are idiosyncratic values of δ_i there will be an equivalent for Figure 3 for every subject varying over δ_i. So we must use a summary statistic for $f_i(\alpha)$, and the one which comes to hand is the median α value, obtained by solving $F_i(\alpha) = 0.5$ in equation (15). In Figure 4, we plot the collection of subject i's (median α, β) duples for models 3 and 4 from Figure 2, where model 4 was our preferred equation. Table 3 lists the percentage of subjects in each (median α_i, β_i) 0.2 bracket.

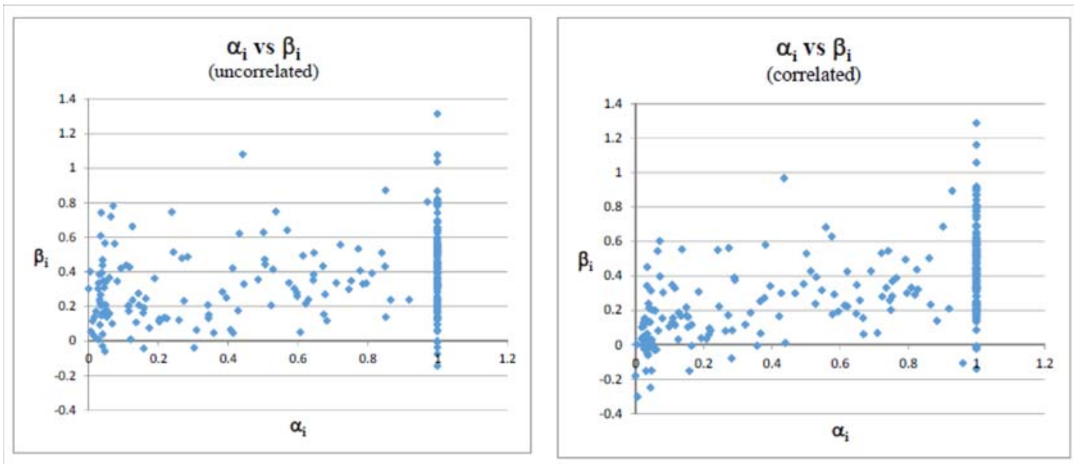


Table 3 – Percentage of subjects in each (α_i, β_i) bracket in model 4

	β_i								
α_i	-0.4 - - 0.2	-0.2 - 0	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	1-1.2	1.2-1.4	Total
0.0-0.2	0.4%	2.4%	11.0%	6.9%	1.2%	0.8%	0.0%	0.0%	0.0%	22.9%
0.2-0.4	0.0%	0.8%	4.5%	2.0%	0.8%	0.8%	0.0%	0.0%	0.0%	9.0%
0.4-0.6	0.0%	0.0%	1.2%	3.3%	1.6%	0.4%	0.0%	0.4%	0.0%	6.9%
0.6-0.8	0.0%	0.0%	0.8%	5.3%	1.6%	0.8%	0.0%	0.0%	0.0%	8.6%
0.8-1.0	0.0%	0.8%	2.9%	17.1%	15.9%	8.6%	6.1%	0.4%	0.8%	52.7%
Total	0.4%	4.1%	20.4%	34.7%	21.2%	11.4%	6.1%	0.8%	0.8%	100.0%

Fully rational agents, whose α_i are always identically equal to unity, are hard to come by since they would require a modelled probability mass of unity at $\alpha = 1$ in the distribution of α_i (16), which in turn would require an infinite δ_i in (5). So, our procedure in Table 3 is to describe agents as rational on the α-dimension (the rows of Table 3) if they have a median in the bottom range (0.8 to 1.0).

With that in mind, we can now comment on the subjects' use of information. Roughly half of the subjects update regardless of evidence, so the median α's cluster at unity along the bottom axis with over half of them (53 per cent) in the range at or above 0.8. Roughly one quarter of median α's point to strong sticky belief adjustment, with α-values no more than 0.2. The remaining quarter of median α's point to weaker forms of sticky belief adjustment.

Regarding the size of updating, we already know from Result 3 that it is far from complete. In Table 3 just over one third of subjects (35 per cent) only update between

20 and 40 per cent of what they should, and we have already noted from model 4 of Table 2 that the average amount of updating over all subjects is just over 40 per cent (41.2 per cent). Figure 4 and Table 3 show that those agents who are relatively likely to update ($\alpha \to 1$) are likely to accomplish relatively more complete updating than those who do not.

Result 5 Estimated test sizes spread over the whole support $[0, 1]$. There is a positive correlation between the median α_i and the extent of belief adjustment when it occurs.

This positive correlation between α_i and β_i suggest the existence of a small group (not more than 10 per cent) of rational agents. They inhabit the bottom RHS of Table 3, where α_i and β_i both exceed 0.8.

5 Discussion and Conclusion

We present a novel and general double hurdle model to consider how agents update their beliefs in an environment where multiple pieces of new information arrive dynamically and sequentially. Our experiment shows how, in such an environment, there is no evidence of underweighting of prior information. It uses a quadratic scoring rule to incentivize beliefs and employs the Offerman et al. (2009) technique to control for risk attitudes. Our regression models in the paper use risk attitude adjusted beliefs, but our key findings are unchanged if raw elicited beliefs are used instead, though the goodness of fit is lower.

We observe random belief adjustment taking place in aggregate around half of the time, which is consistent with stochastic time-dependent belief adjustment. Deviations from Bayesian updating are systematically in the direction of sticky belief adjustment, with only 1% of the subjects showing belief under-weighting in our sample. The likelihood of a belief change increases as the amount of evidence against the null hypothesis increases. This is consistent with state-dependent models of sticky belief adjustment. Because of the incentive mechanism, in our experimental setting the greater the amount of evidence against the currently held belief, the greater the expected costs of keeping to such belief. Because of this, in our experiment threshold-based rational inattention models of belief adjustment (Sims, 2003) are equivalent to agents holding inferential expectations (Menzies and Zizzo, 2009). That is, they stick to their current belief as the null hypothesis until the evidence cumulating against it passes a threshold determined by the test size α, at which point they switch. We estimate that roughly 23% of agents are belief conservative with $\alpha \leq 0.2$, 53 per cent have $\alpha \approx 1$ in terms of likelihood of changing beliefs, and the rest is somewhere in the middle.

We also find that, when beliefs change, in our preferred regression specification they do so by only around 40% of the amount required by the Bayesian prediction.

We therefore also find support for a partial adjustment version of Quasi-Bayesian modeling. Furthermore, subjects who are less likely to adjust their beliefs are also subjects who adjust their beliefs less when they do adjust them.

There are different ways to model sticky belief adjustment, which gives the literature the appearance of divergence. However, what we found overall is that there is support for stochastic time-dependence (beliefs change around 50% other things being equal) and state-dependent sticky belief adjustment (around one quarter of subjects are strongly belief conservative on this respect) and Quasi-Bayesian partial belief adjustment (when adjustment takes place, it is by around 40%). We also find significant population heterogeneity in the extent of the sticky belief adjustment bias.

Current applications consider each of these channels separately, and this has clearly been a good starting point to identify their effect. Sticky belief adjustment is not a novel idea, and initial experimental evidence for it in settings with evidence presented all at once were discussed as long ago as Phillips and Edwards (1966) and Edwards (1968). A pilot study described in Menzies and Zizzo (2005) found evidence for sticky belief adjustment in an experiment with dynamically provided information, but, apart from the small nature of the study, it neither controlled for risk aversion nor did it test among different forms of sticky belief adjustment. Massey and Wu (2005) contains a related but different experiment with dynamically provided information where the goal of the subjects is to identify whether a regime shift has taken place, but they are allowed to change their mind only once; they identify conditions where, in a decision problem of this kind, their subjects display underweighting or overweighting of priors.

In the context of an experiment in which there is only one piece of information provided at the beginning of trading, Camerer (1987) argues that probability updating anomalies wash away in the light of market discipline. Conversely, again in a setting where information is provided all at once, Menzies and Zizzo (2012) find greater evidence of stickiness in market prices in a Walrasian auction market setting intended to model an exchange rate market, than in the corresponding individual beliefs as revealed by the market choices of traders. While this is obviously an area for future research, there is a range of empirical applications where belief stickiness appears plausible in natural economic environments, including markets. Applications of sticky belief adjustment include, among others, optimal principal agent contracts (Rabin and Schrag, 1999), individual responses to market signals (Sims, 2003), a micro-foundation for the New Keynesian Phillips curve (Mankiw and Reis, 2002), consumer and producer behavior (Reis, 2006a, 2006), and pricing under information costs (Woodford, 2009). Inferential expectations modeling has been applied to explain the uncovered interest rate parity failure (Menzies and Zizzo, 2009, 2012), central bank credibility (Henckel et al., 2011, 2013) and merger decisions by competition regulators (Lyons et al., 2012).To conclude, there are different ways to model sticky belief adjustment. What we found overall is that there is support for stochastic time-dependence (beliefs change around 50% of the time, other things being equal) and state-dependent sticky belief adjustment (almost half of the subjects are belief conservative to some degree) and Quasi-Bayesian partial belief adjustment (when adjustment takes place, it is by around 40%). We also find significant population heterogeneity in the extent of the sticky belief adjustment bias.

References

[1] Abeler, Johannes, Falk, Armin, Goette, Lorenz and David Huffman. 2011. Reference Points and Effort Provision. American Economic Review 101(2), 470-492.

[2] Alvarez, Fernando, Francesco Lippi and Juan Passadore (2016). "Are State and Time dependent models really different?" NBER Macroeconomics Annual 2016, Vol. 31.

[3] Ambuehl, Sandro and Li, Shengwu (2014), Belief Updating and the Demand for Information. SSRN Discussion Paper, June.

[4] Aucremanne, Luc, and Emmanuel Dhyne (2005). "Time-Dependent versus State-Dependent Pricing: A Panel Data Approach to the Determinants of Belgian Consumer Price Changes," ECB Working Paper No. 462, March.

[5] Benjamin, Daniel J., Matthew Rabin, and Collin Raymond (2015). "A Model of Non-Belief in the Law of Large Numbers." Journal of the European Economics Association 14(2), 515-544.

[6] Bowling, S. Khasawnch, M., Kaewkuekool, S., and B. Cho, (2009), A logistic approximation to the cumulative normal distribution, Journal of Industrial Engineering and Management, 2(1), 114-127.

[7] Caballero, R.J. (1989). "Time dependent rules, aggregate stickiness and information externalities." Columbia University Department of Economics Discussion Paper 428.

[8] Camerer, Colin F. (1987). "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence," American Economic Review 77(5), 981-997.

[9] Caplin, Andrew, Mark Dean, and Daniel Martin (2011). "Search and Satisficing," American Economic Review 101(7), 2899-2922.

[10] Carroll, Christopher D. (2003), "Macroeconomic Expectations of Households and Professional Forecasters" Quarterly Journal of Economics, 118, 269-298.

[11] Charness, Gary, and Dan Levin (2005). "When Optimal Choices Feel Wrong: A laboratory Study of Bayesian Updating, Complexity and Affect," American Economic Review 95, 1300-1309.

[12] Corgnet, Brice, Roberto Hernan-Gonzalez and Eric Schniter (2015). "Why real leisure really matters: incentive effects on real effort in the laboratory," Experimental Economics 18, 284-301.

[13] Costain, James, and Anton Nakov (2011). "Price Adjustments in a General Model of State-Dependent Pricing," Journal of Money, Credit and Banking 43(2-3), 385-406.

[14] Dias, D. A., C. Robalo Marques, and J. M. C. Santos Silva (2007). "Time- or state-dependent price setting rules? Evidence from micro data," European Economic Review 51, 1589-1613.

[15] Dias, Daniel A., Carlos Robalo Marques, and Fernando Martins (2012). "Choosing between time and state dependence: Micro evidence on firms' price-reviewing strategies, mimeo, University of Illinois at Urbana-Champaign.

[16] Edwards, W. (1968). "Conservatism in human information processing," in: Kleinmuntz, B. (Ed.). Formal representation of human judgment, New York: Wiley, 17-52.

[17] El-Gamal, Mahmoud A., and Grether David M. "Are People Bayesian? Uncovering Behavioral Strategies." Journal of the American Statistical Association 90, no. 432 (1995): 1137-145.

[18] Grether, David M. (1980). "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," Quarterly Journal of Economics 95(3), 537-57.

[19] Henckel, Timo, Menzies, Gordon D. and Daniel J. Zizzo, "The Great Recession and the Two Dimensions of European Central Bank Credibility". Australian National University Centre for Applied Macroeconomic Analysis Working Paper 55/2013, August 2013.

[20] Henckel, Timo, Menzies, Gordon D., Prokhovnik, Nicholas and Daniel J. Zizzo, "Barro-Gordon Revisited: Reputational Equilibria with Inferential Expectations", Economics Letters, 112(2), August 2011, 144-147.

[21] Independent Commission on Banking, 2011. Final Report Recommendations. London: U.K. Government.

[22] Joskow, Paul L., 2008. Lessons Learned From Electricity Market Liberalization. Energy Journal 29(Special Issue 2), 9-42.

[23] Klenow, Peter J., and Oleksiy Kryvtsov (2008). "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?" Quarterly Journal of Economics 123(3), 863-904.

[24] Lyons, Bruce, Menzies, Gordon D. and Daniel J. Zizzo, "Conflicting Evidence and Decisions by Agency Professionals: An Experimental Test in the Context of Merger Regulation", Theory and Decision, 73(3), September 2012, 465-499.

[25] Magnani, Jacopo, Aspen Gorry, and Ryan Oprea (2016). "Time and State Dependence in an Ss Decision Experiment," American Economic Journal: Macroeconomics 8(1), 285-310.

[26] Mankiw, N. Gregory and Reis, Ricardo, Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. Quarterly Journal of Economics, 117 (2002), pp. 1295–1328

[27] Martins, F. (2005). "The price setting behaviour of Portuguese firms: Evidence from survey data," ECB Working Paper No. 562.

[28] Massey, Cade and George Wu (2005). Detecting regime shifts: The causes of under- and overreaction. Management Science 51(6), 932-947.

[29] Menzies G. and Daniel J. Zizzo (2012), 'Monetary Policy and Inferential Expectations of Exchange Rates', Journal of International Financial Markets, Institutions & Money, 22(2), April, 359 380.

[30] Menzies, Gordon D. and Daniel J. Zizzo (2009), "Inferential Expectations", B.E. Journal of Macroeconomics 9(1) (Advances), Article 42.

[31] Menzies, Gordon D. and Daniel J. Zizzo, "Inferential Expectations", Australian National University Centre for Applied Macroeconomic Analysis Discussion Paper n. 12, June 2005.

[32] Midrigan, Virgiliu (2010). "Is Firm Pricing State or Time Dependent? Evidence from U.S. Manufacturing," Review of Economics and Statistics 92(3), 643-656.

[33] Offerman, Theo, Joep Sonnemans, Gijs Van De Kuilen, and Peter P. Wakker (2009), "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," Review of Economic Studies, 76 (4): 1461-1489.

[34] Ofgem, 2011. Simpler Energy Tariffs. London: U.K. Government.

[35] Phillips, L.D., and W. Edwards (1966). "Conservatism in a simple Probability Inference Task.' Journal of Experimental Psychology 72, 346-354.

[36] Rabin, Matthew (2013). "Incorporating Limited Rationality into Economics," Journal of Economic Literature 51(2), 528-43.

[37] Rabin, Matthew, and Joel L. Schrag (1999). "First Impressions Matter: A Model of Confirmatory Bias," Quarterly Journal of Economics 114(1), 37-82.

[38] Reis, Ricardo (2006a), Inattentive Consumers. Journal of Monetary Economics, 53 (8), 1761-1800.

[39] Reis, Ricardo (2006b), Inattentive Producers. Review of Economic Studies, 73 (3), 793-821.

[40] Sims, Christopher A. 2003. "Implications of Rational Inattention." Journal of Monetary Economics, 50(3): 665–690.

[41] Sitzia, Stefania, Jiwei Zheng and Daniel J. Zizzo, "Inattentive Consumers in Markets for Services", Theory and Decision, 79(2), September 2015, 307-332.

[42] Stahl, Harald (2005). "Time Dependent of State Dependent Price Setting? Micro-Evidence from German Metal-Working Industries," ECB Working Paper No. 534, October.

[43] Tversky, Amos and Daniel Kahneman (1982), "Judgments of and by Representativeness," in D. Kahneman, P. Slovic, and A. Tversky, eds., Judgment under Uncertainty: Heuristics and Biases, Cambridge: Cambridge University Press.

[44] Woodford, Michael (2009), Information-constrained state-dependent pricing. Journal of Monetary Economics, 56(15), S100-S124.

6 Appendices

6.1 Appendix 1: Closeness of Two Strength-of-evidence Measures

The task of this appendix is to explain why the two measures in the final column of Table 1 are numerically similar when simulated in an online appendix 'Two z measures as functions of $t \cdot xls$'. We focus here on the expected values of $z_{P,t}$ and $z_{Pw,t}$ to make a statement of similarity that is not clouded by sampling variability.

We first consider the evolution of the expected value of $z_{Pw,t}$ at the start of the experiment, when $P(U) = 0.6$, or equivalently when the expected probability of a white ball is $P(W) = 0.7\,(0.6) + 0.3\,(1 - 0.6) = 0.54$. We now calculated the expected value of $z_{Pw,t}$ from the second row of the table, when Urn 1 is in fact drawn, so that the true $P(W) = 0.7$.

$$\begin{aligned} E\left(z_{Pw,t}\right) &= 2\sqrt{t}\left[E\left(Pw_t\right) - \left(0.7\,(0.6) + 0.3\,(0.4)\right)\right] \\ &\approx 2\sqrt{t}\left[0.7 - 0.54\right] \\ &= 0.32\sqrt{t} \end{aligned} \tag{19}$$

If Urn 2 is chosen, $E(Pw_t) = 0.3$, and making this substitution gives $E(z_{Pw,t}) = 0.48\sqrt{t}$.

We now compare these to the expected value of $z_{P,t} = \Phi^{-1}(P_t) - \Phi^{-1}(0.6)$. The cumulative Normal Φ is approximated by a 'one parameter equation' logistic approximation (Bowling et al. 2009). We use $\sqrt{3}$ below, but one could use the other standard parameters $\pi\sqrt{3}$ or 1.7 without any substantive difference to the results. Here is the approximation we use together with its inverse.

$$\Phi(x) \approx \frac{1}{1+\exp\left(-\sqrt{3}x\right)} \qquad \Longleftrightarrow \qquad \Phi^{-1}(x) = \frac{1}{\sqrt{3}}\ln\left(\frac{x}{1-x}\right).$$

The components of $z_{P,t}$ at the beginning of the experiment are thus:

$$\Phi^{-1}(P_t) = \frac{1}{\sqrt{3}}\ln\left(\frac{P_t}{1-P_t}\right) \qquad \& \qquad \Phi^{-1}(0.6) = \frac{1}{\sqrt{3}}\ln\left(\frac{3}{2}\right). \tag{20}$$

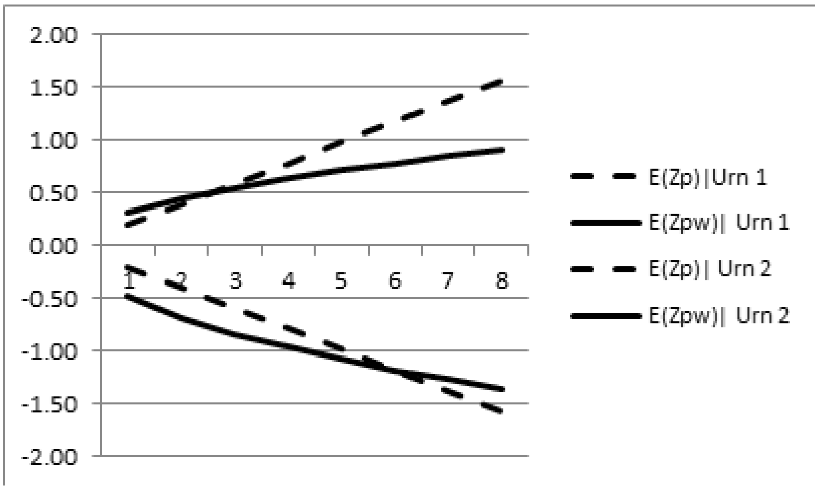
Writing P_t in terms of t, and noting that the number of white balls is tPw_t, we obtain:

$$\begin{aligned}
P_t &= \frac{(0.7)^{tPw_t}(0.3)^{t-tPw_t}(0.6)}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}(0.6)+(0.3)^{tPw_t}(0.7)^{t-tPw_t}(1-0.6)} \\
&= \frac{1}{1+\frac{(0.3)^{tPw_t}(0.7)^{t-tPw_t}}{(0.7)^{tPw_t}(0.3)^{t-tPw_t}}\frac{2}{3}} \qquad \Longrightarrow \qquad \frac{P_t}{1-P_t} = \frac{3}{2}\left(\frac{7}{2}\right)^{2tPw_t-t} \\
z_{P,t} &= \Phi^{-1}(P_t) - \Phi^{-1}(0.6) = \frac{1}{\sqrt{3}}\ln\left(\frac{3}{2}\left(\frac{7}{2}\right)^{2tPw_t-t}\right) - \frac{1}{\sqrt{3}}\ln\left(\frac{3}{2}\right) \\
&= \frac{n}{\sqrt{3}}(2Pw_t-1)\ln\left(\frac{7}{3}\right).
\end{aligned} \tag{20}$$

Finally, the expectation of $2Pw_t - 1$ is 0.4 if Urn 1 is chosen and -0.4 if Urn 2 is chosen.

$$E(z_{P,t}) = \left\{ \begin{array}{c} \frac{0.4t}{\sqrt{3}}\ln\left(\frac{7}{3}\right) \approx 0.2t \text{ for Urn 1} \\ -\frac{0.4t}{\sqrt{3}}\ln\left(\frac{7}{3}\right) \approx -0.2t \text{ for Urn 1} \end{array} \right\} \tag{22}$$

In Figure A1 we show $E(z_{Pw,t})$ and $E(z_{P,t})$ as a function of t when Urn 1 is chosen and when Urn 2 is chosen. $E(z_{Pw,t})$ under the two states of nature (Urn 1 and Urn 2) is solid, and based on (19). $E(z_P)$ is dashed and based on (22). If Urn 1 is chosen, both strength-of-evidence measures can be expected to raise the prior probability of 0.6 upwards as t increases (towards the true value of unity). If Urn 2 is chosen, both strength-of-evidence measures can be expected to draw the prior probability down (towards the true value of zero). As is clear from looking at the numerical simulations, the highly stochastic paths of z differ from those shown in Figure A1, but as one observes a number of simulations are done, the patterns of (19) and (22) shown in Figure A1 can indeed be discerned.



While the expected function pairs (one pair for each urn draw) in the figure are not identical, we judge them to be close enough for our approximation in the text.

6.2 Appendix 2: Method for Estimating CRRA Risk Parameter

The log relationship between transformations of g_t and g_t^* in the text has an i.i.d. error added to it and run with 10 observations as an OLS regression. The variable t in (23) refers to the 10 rounds in the practice part, not to the rounds in the main part. The estimated parameter θ_i is subscripted for subjects, because (23) is run for each subject to provide her own θ.

$$\ln\left(\frac{g_t^*\left(1-g_t\right)}{g_t\left(1-g_t^*\right)}\right)=\theta_i\ln\left(\frac{g_t\left(2-g_t\right)}{\left(1+g_t\right)\left(1-g_t\right)}\right)+\eta_t, \qquad t=1,2,\ldots,10. \tag{23}$$

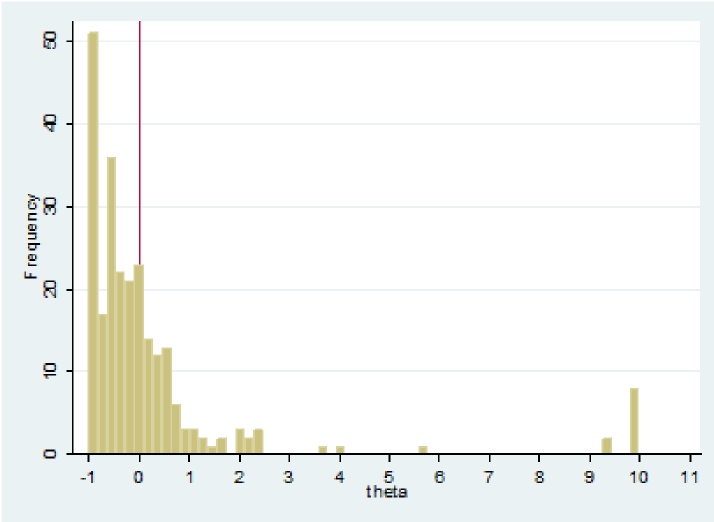
We note the following:

1. The regression has no intercept. If an intercept is included the θ_i estimates are inefficient.

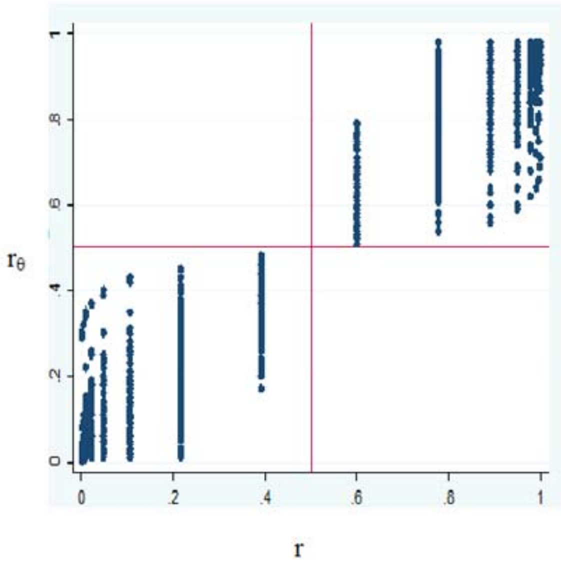
2. For some subjects $g_t = 0.5$ in every period. In this case, the RHS variable is $\ln(1)$ in every period. This means that θ approaches $+\infty$. These estimates need to be re-coded to a high positive number, and we use $+10$.

3. There is a logical requirement that θ cannot be less than -1 in this model. Hence, any estimates less than -1 need to be re-coded to -1.

The above procedure gives rise to the following distribution of θ over the 245 subjects:



Once each subject has her own estimated θ_i, the full set of implied g_t^* values can be generated from the observed guesses g_t in the main experiment. As discussed in the main text, this is accomplished by rewriting (21) without an error, which is identical to equation (1), exponentiating both sides, and then solving for g_t^*. The following figure shows g_t^* against g_t for the full sample. The multiple values of g_t^* for every g_t are due to the different estimated values of θ_i for each subject.



6.3 Appendix 3: IV Estimator

As mentioned in subsection 4.2, there is an endogeneity problem with using the strength of evidence against the previously chosen value as an explanatory variable in the first hurdle, and in this appendix we resolve this problem. (We continue to drop the subscript ‘P’ in what follows, since we have equated the two measures of strength of evidence and refer to both of them as z_{it}). The problem is that the variable is endogenous, because subjects with a low propensity to update are clearly likely to generate large values of $|z_{it}|$ simply by virtue of rarely updating. Hence $|z_{it}|$ always appears to have a perverse negative effect on the propensity to update.

In order to address this problem, we use an IV estimator. We first create a prediction $|\widehat{z}_{it}|$ for use in the second stage of a two-staged least squares estimation. The two instruments for $|z_{it}|$ that we use to create this predictor are the round number (t), and the absolute value of the contribution to $|z_{it}|$ in the *current* round $|\Delta z_{it}|$. This is not to be confused with the difference built into $z_{P,it}$ which spans the current period to period m, namely, $\Phi^{-1}(P_{it}) - \Phi^{-1}(P_{im})$. We note that, because $\Phi^{-1}(P_{im})$ is fixed, a difference operator will eliminate it, leaving Δz_{it} as the change in $\Phi^{-1}(P_{it})$ over the last period.

The stage 1 OLS regression is, therefore:

$$|z_{it}| = \pi_0 + \pi_1 t + \pi_3 |\Delta z_{it}| + \varepsilon_{it}. \tag{24}$$

The results are shown below. Both variables show strong significance in the expected direction, implying that they are good instruments.

	Coefficient	**s.e.**	**t-stat**
constant	0.0954	0.0117	8.18
t	0.0263	0.0018	14.24
$\lvert\Delta z_{it}\rvert$	0.8168	0.0106	77.14
Obs.	13,720		
R-Squared	0.3065		

Having estimated the stage 1 regression we obtain the predicted values, $|\widehat{z}_{it}|$, and use these in place of $|z_{it}|$ in the first hurdle of the main model. To underline the importance of the instruments, we show two plots below. Figure A4 shows the estimated itprobability of updating against $|z_{it}|$, while Figure A5 displays the estimated probability of updating against $|\widehat{z}_{it}|$.

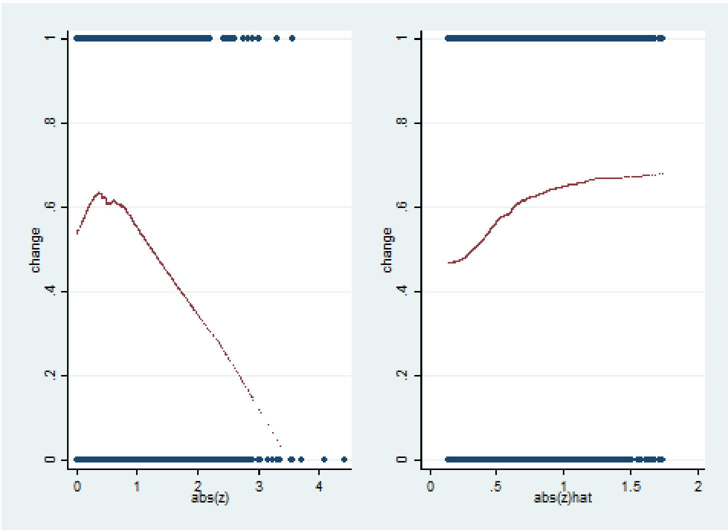


Figure A4 (left panel) makes clear the endogeneity problem identified above: over most of the range of $|z_{it}|$ its effect on the propensity to update is negative. Figure A4 (right panel) is of the same binary variable against $|\widehat{z}_{it}|$ (the prediction from the stage 1 regression). It shows completely the opposite pattern: a monotonically increasing effect of $|\widehat{z}_{it}|$ on the propensity to update.

6.4 Appendix 4: Adjustment costs and beliefs

In Offerman et al. (2008) P is the probability of choosing Urn 1 and r is the guess that maximizes expected utility:

$$E(V) = PV\left(1-(1-r)^2\right) + (1-P)V\left(1-r^2\right).$$

Optimal r, denoted r^*, solves the following expression which is derived from the first order condition:

$$r^* = \frac{P}{P + \frac{(1-P)V'(1-r^{*2})}{V'(1-(1-r^*)^2)}}.$$

If an agent has a standing choice of r, say r', and moving to the optimal r^* costs C (whereas maintaining r' costs nothing), then the agent will move iff:

$$PV(1-(1-r^*)^2-C)+(1-P)V(1-r^{*2}-C) > PV(1-(1-r')^2)+(1-P)V(1-r'^2).$$

Thus the presence of an adjustment cost creates inertia. Note that this is increasing the greater the difference is between r' and r^*, and that this difference will be monotonically greater the greater the evidence against r' and in favour of r^* in an inferential expectations framework. Thus, there is a connection between adjustment costs and the use of inferential expectations.