Building Central Bank Credibility: The Role of Forecast Performance*

Michael McMahon† Ryan Rholes‡

First Draft: May 2020
This Draft: July 2022

Abstract

Does a central bank influence inflation expectations through its publicised forecast? Does such influence depend on how accurate the central bank's forecasts have been? Given the importance of anchoring inflation expectations to inflation targeting monetary frameworks, and given the central role of forecast in such frameworks, understanding the answers to these questions is important. We show, using an incentivised individual-choice experiment, that forecast performance seems to matter a lot. In particular, the weight that agents attach to the central bank forecast is strongly related to the forecast performance in the last 4 quarters. We also show that central bank communication can play a role in mitigating, though not fully, the effect of poor recent forecast performance.

Keywords: Expectation formation, Forecasting, Monetary Policy Communication

JEL Codes: E52, E58

*McMahon gratefully acknowledges financial support from the European Research Council (Consolidator Grant Agreement 819131). We are grateful to seminar participants at the University of Oxford, and the MIT Conference, as well as Johannes Abeler for helpful comments. The views presented in the paper are those of the authors alone and do not represent the official views of anyone else. Any errors remain ours alone.

†University of Oxford, CEPR, CfM (LSE). Email: michael.mcmahon@economics.ox.ac.uk
‡University of Oxford. Email: ryan.rholes@economics.ox.ac.uk

1 Introduction

Monetary policy frameworks now, largely, involve the management of expectations (Woodford 2005, King et al. 2008). For instance, expectations management is a key tenet of the widely adopted inflation targeting monetary policy framework. This follows from New Keynesian models which have underpinned most recent theoretical research on the effects of monetary policy (Clarida et al. 1999, Woodford 2003, Galí 2008, forexample). This is because, in these frameworks, inflation expectations become a vital determinant of inflation.

Central bank communication has emerged as a key part of the toolkit to manage expectations. Some would argue communication is *the* key part of that toolkit. This communication involves both high frequency communication and lower frequency communication, but both, according to the channel above, can play a key role in inflation control by influencing how agents form their expectations. Hence, open mouth operations are now an indispensable component of monetary policy.

Does communication actually work to influence expectations? In workhorse monetary models, agents are rational and appropriately incorporate central bank, and other, information to form the best possible expectations. In practice, central banks worry about their credibility, which is necessary for the transmission of communication policy Blinder (2000).

However, little is known in practice about the determinants, dynamics, or role of central bank credibility. We can imagine a world where successful achievement of the inflation target builds credibility and gives the central bank greater ability to control inflation by better anchoring inflation expectations. But alongside this virtuous cycle may sit a vicious cycle; lower credibility could impinge upon the ability of the central bank to manage inflation which then makes credibility-reducing inflation fluctuations more likely.

In this paper, we develop an experimental framework to study the determinants and dynamics of central bank credibility, and their relationship with expectations management. This relationship is important because inflation forecasts are a key input into the policy decision in an inflation targeting framework (Svensson 1997).

Participants in our experiment act as inflation forecasters tasked with providing two sets of one-period-ahead point and potentially asymmetric range forecasts of inflation (Initial Forecasts and Updated Forecasts) in 3 sequential decision periods. We begin decision periods by exposing subjects to 12 quarters of economic history consisting of realized inflation alongside corresponding central bank inflation forecasts. Subjects provide Initial Forecasts (priors) after the revelation of the economic history. We then show them the central bank's corresponding inflation forecast and allow them to update their own inflation forecast (forming their posterior estimate, i.e. Updated Forecasts). We measure credibility by the weight that agents place on the central bank's inflation forecast when updating their estimate of inflation.

Differences in economic histories constitute within-subject treatment variation in our experiment. We refer to these histories as: *Early*, *Late*, and *Consistent*. In *Early*, the central bank commits significant forecast errors in the first third of the forecasting history, moderate errors in the second third, and minimal errors in the last third. We reverse this pattern for *Late*. For *Consistent*, the central bank exhibits a consistent average annual forecast performance. Finally, *Late* exactly reverses the absolute error structure of *Early*. We simulated economic histories using the New Keynesian model described in Walsh (2017) linearized around the zero-inflation steady state. We employ pre-selected shock sequences that preserve the structure of real-world absolute forecasting errors from the United Kingdom and Bank of England (BoE) for the three-year period beginning in the first quarter of 2010 and ending in the final quarter of 2012. We then shift the resulting inflation data so that is centered around 2%.

This exercise produces economic histories that should lead to no differences in updating for rational, Bayseian-updating subjects. However, we find subjects exhibit a strong recency bias when forming beliefs about the central bank's forecast credibility. This recency bias leads subjects to place the least weight on the central bank's forecast when updating their own expectations in the *Late* history, more in *Consistent* than *Late*, and the most weight in *Early*.

This recency bias also extends to how subjects' beliefs about central bank credibility influence higher moments of subject-level density forecasts of inflation. Subject-level uncertainty responds most strongly to the central bank's forecast in *Early* and c*Consistent*. Additionally, the higher level of central bank credibility in *Early* and *Consistent* lead subject's to reduce the asymmetry of their forecasts while in *Late* , where point forecasts shift significantly less than in other treatments, subjects increase the skewness of their forecasts toward the central bank's forecast.

2 Literature Review

2.1 Experiments

Macroeconomists have typically studied expectations in the laboratory using the learning-to-forecast (LTF) framework, wherein s experimental economies evolve endogenously according to the incentivized expectations of participants. In LTF experiments, subjects forecast a (some) macroeconomic aggregate(s) that are aggregated and fed into an underlying data generating process, which causes aggregates to evolve and the experiment to progress to a new decision period. Central bank credibility – often a secondary consideration – is usually measured in these experiments by projecting expectations onto the central bank's forecast (if present) and other conditioning information (i.e. an inflation target if communicated, interest rates, shocks, etc.) to determine the extent to which central bank forecasts factor into subjects' forecasts.

Researchers have used this framework to study extensively the design and efficacy of central bank communication the design of central bank communication (Kryvtsov and Petersen (2021); Arifovic and Petersen (2017); Cornand and M'baye (2018); (**?**)), expectation formation and equilibria selection (Adam (2007); Bao et al. (2012)), and how various monetary policy rules and targets affect expectation formation (Ahrens et al.

(2019); Pfajfar and Žakelj (2014); Pfajfar and Žakelj (2018); Assenza et al. (2013); Hommes et al. (2019); Hommes et al. (2019); Cornand and M'baye (2018)).

3 Experimental Design

Participants in our individual-choice experiment act as atomistic inflation forecasters tasked with providing two sets of one-period-ahead inflation forecasts (Initial Forecasts and Updated Forecasts)in three independent decision periods. Each set of forecasts comprises an incentivized point forecast of inflation coupled with an incentivized measure of corresponding forecast uncertainty.

We attempt to better align the experience of participants in our experiment with that of real-world households by departing meaningfully from the typical LTF framework in at least three ways.

First, our subjects form inflation forecasts that do not influence the evolution of state variables as is typical in the LTF literature (i.e. subjects are atomistic).[1]. This design choice to break the endogenous link between expectations and aggregate inflation likely better aligns participants in our experiment with the outlook of participants in typically household surveys like the University of Michigan's Survey of Consumers, the New York Federal Reserve's Survey of Consumer Expectations, or the European Central Bank's Consumer Expectations Survey.

Second, we use an individual-choice setting, which removes strategic uncertainty from the forecasting problem. We do this because we believe it unlikely that the average individual household considers expectation formation to be a coordination problem.

Further, subjects in our experiment have only information regarding past inflation dynamics and central bank forecasting performance to guide expectations. Elicitation of the initial inflation forecast captures how past inflation dynamics influence inflation expectations. The degree to which subjects update their expectations following revelation of the central bank's inflation forecast then cleanly isolates central bank forecast credibility as a function of historical forecast performance.

3.1 Implementation

Participants began the experiment by completing a short survey that measured their level of economics knowledge, their level of understanding of and trust for various public institutions, their preferences for obtaining economic information, and their familiarity with prevailing economic conditions. We then provided subjects on-screen instructions that explained the inflation forecasting task, the information available to aide in their forecasting task, how to interact with the available information, how to interact with our

[1] We can easily make this environment endogenous by feeding subjects' expectations back into the data generating process, described in Section 3.3, that we use to produce our historical economic information

software, and how we incentivized their forecasts. These instructions remained available to subjects throughout the experiment via a toggle button on all screens.

Following the instructions, subjects completed a comprehension quiz. The comprehension quiz consisted of five questions designed to test subjects' understanding of our experimental instructions. Subjects had to answer all five questions correctly to proceed. Our software ended the experiment early for subject who submitted the quiz more than twice with at least one wrong answer.[2] Subjects who successfully completed the quiz proceeded to the forecasting task, which consisted of three independent decision periods. We paid a participant based on her forecasting performance in one of the three independent decision periods she faced.

Following the decision periods, we informed subjects for which decision period they would received payment and of earnings. Participants ended the experiment with a non-compulsory survey-of-decisions.

We programmed our experiment in oTree (Chen, Shonger, and Wickens; 2016). We conducted our experiment online via Prolific, restricting our subject sample to experienced Prolific users from the United States.

3.2 Decision Periods

We began each decision period by providing a participant with a 12-quarter economic history consisting of realized inflation alongside corresponding central bank inflation forecasts. We revealed historical observations sequentially with a one-second lag between observations so that participants carefully considered the full economic history before forming Initial Forecasts. We displayed this historical data graphically and numerically and all information, once revealed, remained available for the duration of that decision period.

After our software revealed the full economic history for a decision period, participants provided a point forecast of one-period-ahead inflation (i.e. $\mathbb{E}_{i,12}\pi_{13}$) in percentage terms with up to two decimal accuracy. We incentivized point forecasts according to Equation (1), which follows previous LTF literature (Rholes and Petersen 2021, Mokhtarzadeh and Petersen 2021, Petersen and Rholes 2022)

$$F_{i,t} = 2^{-|\mathbb{E}_{i,t-1}\{\pi_t\}-\pi_t|}. \tag{1}$$

Note that a perfect forecast yields $F_{i,t} = 1$ and that this forecasting score is reduced by $\frac{1}{2}$ each time the forecast error increases by one percentage point.

Participants could submit point casts two ways. First, they could create a point forecast by clicking on the interactive chart used to display historical economic information. They could subsequently alter this forecast by dragging and dropping this point anywhere inside the forecast region of the graph. Alternatively, participants could type forecasts

[2] We provide examples of the instructions and comprehension quiz in APPENDIX.

directly into an available input field. Participants faced no time pressure and could visualize as many forecasts as they desired before submitting the initial point forecast. Once a subject submits the initial point forecast, our software updates to reflect this value graphically and numerically.

Participants next submit a measure of forecast uncertainty corresponding to their initial point forecast. To start, our experimental software randomly generated upper and lower uncertainty bounds that bracketed the participant's initial point forecast. The area between these two bounds appeared to participants as a shaded region, denoting a visual representation of the participant's forecast uncertainty. Participants could then changed the uncertainty bounds to reflect their true forecast uncertainty. They could do this by dragging and dropping the two bounds independently, dragging and dropping both bounds simultaneously, or by typing numbers directly into corresponding input fields. Our software prevented subjects from inputting values for the upper bound that were below the point forecast and vice versa for lower-bound values. Our software also prevented subjects from visualizing upper and lower bounds that violated these same bounding conditions.

We incentivize range forecasts using the scoring rule given in Equation (2), which follows Pfajfar and Žakelj (2016), Rholes and Petersen (2021), Petersen and Rholes (2022)

$$U_{i,t}(r_{i,t}) = \begin{cases} 0 & \pi_{i,13} \notin [\underline{u_{i,t}}, \overline{u_{i,t}}] \\ \alpha \left(\frac{1}{r_{i,t}}\right) & \pi_{i,13} \in [\underline{u_{i,t}}, \overline{u_{i,t}}]. \end{cases} \tag{2}$$

Here α is a scalar we can adjust to scale average earnings, $\underline{u_{i,t}}$ is the lower-bound of a participant's forecast uncertainty, $\overline{u_{i,t}}$ the upper-bound of a participant's forecast uncertainty, and $r_{i,t} = \|\overline{u_{i,t}} - \underline{u_{i,t}}\|$ is the magnitude of a participant's forecast uncertainty.

This scoring rule is quite intuitive. A participant earns nothing for her uncertainty measure if realized inflation values fall outside her uncertainty bounds. If realized inflation does fall within a participant's uncertainty bounds, then she earns a payoff that subjects' payoff that is decreasing in the magnitude of her uncertainty.

After collecting a participant's Initial Forecasts (initial point forecast plus corresponding uncertainty), we revealed the central bank's quarter-13 inflation forecast and allowed the participant to update her point forecast of inflation and her corresponding forecast uncertainty. We provided participants with numerical and graphical information about their initial point forecast of inflation and their corresponding forecast uncertainty. We emphasized to participants in our instructions and with an on-screen reminder that they were not obligated to update either measure. If they chose to update, they could update any or all values of $\mathbb{E}_{i,12}(\pi_{13})$, $\underline{u_{i,t}}$, $\overline{u_{i,t}}$.

After collecting updated forecast values, our software would reveal to participants the actual value of quarter-13 inflation alongside their forecasting performance for that decision period.

We summarize the flow of a single decision period in Figure 1. Subjects completed three decision periods, forming a total of six sets of forecasts. We randomly selected one of

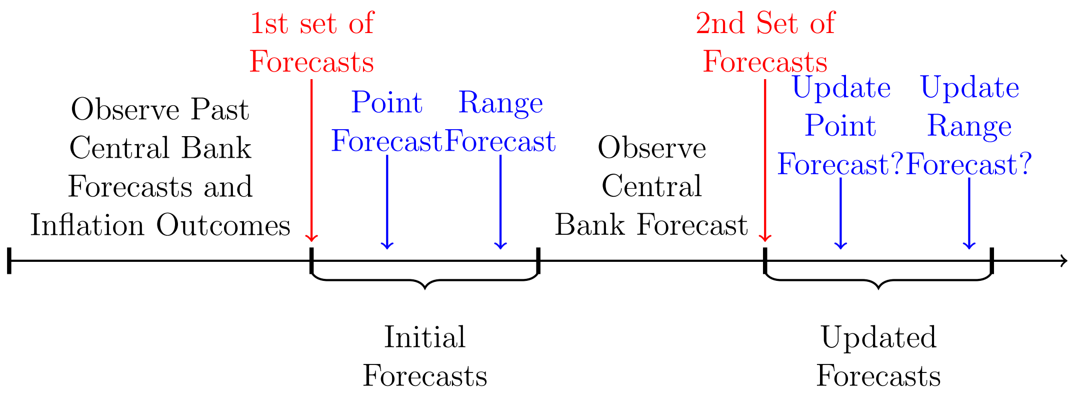


Figure 1: Experimental Timeline: A single decision period

these six sets of forecasts for payment.

3.3 Creating the economic histories

In this project, we study the relationship between historical forecast performance and forecast credibility (*ForecastPerformance*), to understand how the timing of forecast errors influences perceived forecast credibility (*Timing*), and whether a central bank can use low-frequency communication to bolster its forecast credibility (*Communication*). Rather than describe all treatments now, we first describe the treatments related to *ForecastPerformance* and detail the related results. Then repeat this organization for *Timing* and *Communication*.

Differences in the economic histories described in Section 3.2 constitute treatment variation in our experiment framework. We rely on a core set of three histories in all three sections of this project, which we will refer to as *Early*, *Late*, and *Consistent*. In *Early*, the central bank commits significant forecast errors in the first third of the forecasting history, moderate errors in the second third, and minimal errors in the last third. We reverse this pattern for *Late*. For *Consistent*, central banks exhibit a consistent average annual forecast performance. Regardless of treatment wave, all participants experiences three economic histories (i.e. completes three independent decision periods) consisting of *Early, Late*, and some version of *Consistent*.

To create these histories, we simulated Equation (3) through Equation (9) (Walsh 2017) using parameters in Table 1 and then shift then center the resulting data around 2%. We assume the central bank in our simulated economy forms rational expectations so that the uncorrelated stochastic components of the per-period shocks (Equation (7), Equation (8), and Equation (9)) drive forecast errors in our simulated data. The central bank's expectation for any per-period shock $\psi_t \in \{g, u, v\}$ is given by $E_t\psi_{t+1} = \rho_{\psi,t}\psi_t$. Note that y_t is the output gap (log-deviation of output from the natural rate), π_t is the quarterly rate of inflation between $t-1$ and t, i_t is the nominal interest rate on funds moving between period t and $t+1$, and r_t is the real interest rate. Finally, $g_t, u_t,$ and v_t are demand, inflation, and monetary policy shocks, respectively.

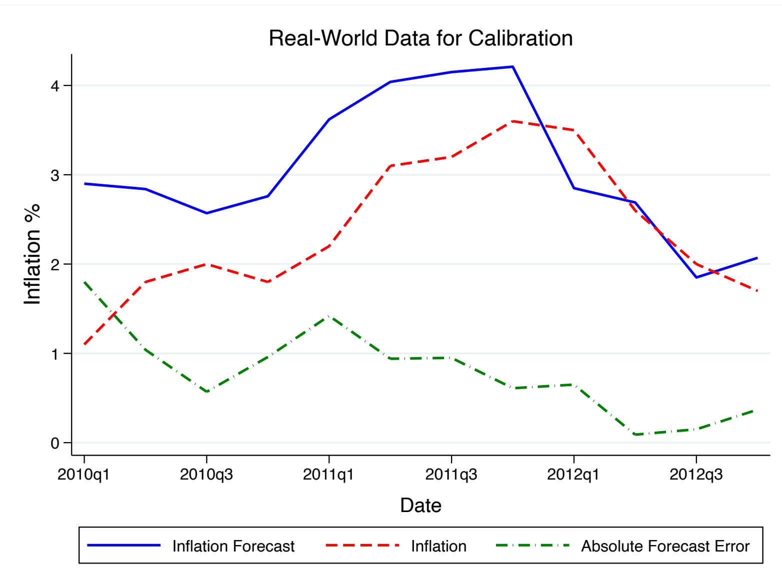


Figure 2: Economic data from the United Kingdom used for calibration of simulated *Early* economic history

$$y_t = E_t y_{t+1} - \sigma^{-1}(i_t - \mathbb{E}_t \pi_{t+1}) + g_t \tag{3}$$
$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa y_t + u_t \tag{4}$$
$$i_t = \phi_x y_t + \phi_\pi \pi_t + v_t \tag{5}$$
$$r_t = i_t - \mathbb{E}_t \pi_{t+1} \tag{6}$$
$$g_{t+1} = \rho_g g_t + \epsilon^g_{t+1} \tag{7}$$
$$u_{t+1} = \rho_u u_t + \epsilon^u_{t+1} \tag{8}$$
$$v_{t+1} = \rho_v v_t + \epsilon^v_{t+1} \tag{9}$$

We base our simulated economic histories on inflation and forecast data from the United Kingdom and Bank of England (BoE) for the three-year period beginning in the first quarter of 2010 and ending in the final quarter of 2012 (see Figure 2). Specifically, we chose shocks for *Early* that qualitatively preserved the observed pattern of absolute forecast errors while exactly matching the sample average absolute forecast error.

Parameter Values

β	$\sigma = \eta$	ω	κ	ρ	ϕ_π	ϕ_y	ρ_g	ρ_u	ρ_v
.99	1	.8	.104	.9	1.5	0	.5	.5	.9

Table 1: Parameter values for simulation exercise

We then reversed this pattern of forecast errors to produce *Late*. Call the central bank's forecast error $\delta^{history}_{\pi,t} = E^{history}_{t-1}(\pi_t) - \pi_t$. Then we solved for the series of shocks that gave $\delta^{Late}_{\pi,12} =^{Early}_{\pi,1}, ..., \delta^{Late}_{\pi,1} = \delta^{Early}_{\pi,12}$. This exactly preserves the absolute average forecast error between *Early* and *Late*. We then added additional noise to *Late* so that subjects wouldn't recognize the economic history as an exact reversal of *Early*. We produce data for all variations of *Consistent* by choosing shock sequences that yield parity between the annual average and sample average absolute forecast errors. Finally, we created inflation forecasts and inflation values for the forecast quarter (i.e. quarter 13) in each economic history using shocks that roughly preserved the average forecast error of the final year of economic history.

Simulating historical economic data offers several benefits. First, this allows us to preserve important features of real-world data while mitigating the chance that participants recognize data patterns that aide them in the forecasting task. Second, this approach allows us to generate forecasting errors and corresponding macroeconomic data by either isolating or blending shocks, which could allow us to cleanly study the relationship between forecasting, credibility, and the source(s) of economic volatility. Finally, simulating data allows precise control over error structures, creating a causal connection between past forecast performance and forecast credibility.

We summarise forecasting performance for our real-world data sample and each of our simulated economic histories in Table 2. We provide more details on our different variations of *Consistent* in Section 4.1.

	Summary of Forecast Performance by History (bps)			
	Year 1	**Year 2**	**Year 3**	**Full Sample**
Calibration Data	110	95	34	80
Early	171	65	13	83
Late	13	65	171	83
Consistent - Great	13	13	13	13
Consistent - Good	36	36	36	36
Consistent - Moderate	60	60	60	60
Consistent - Bad	83	83	83	83
Consistent - Terrible	171	171	171	171

Table 2: Summary of economic histories. Numbers are average absolute forecast error in basis points.

4 Results

This section reports results for *Forecast Performance, Timing, and Communication.* For each question we provide additional details regarding treatments and experimental design, state our hypotheses, and then detail our results.

4.1 Forecast Performance

Our primary question in *ForecastingPerformance* is how a central bank's recent forecast performance influences its perceived credibility as a forecaster. To answer this question, we study how variation in the sample-average absolute forecast errors in otherwise identical economic histories (the five different version of *Consistent* defined in Table 2) effects the willingness of participants to incorporate the central bank's inflation expectation into their updated point forecast. We will refer to these histories as *Great, Good, Moderate, etc.* throughout this section for ease of exposition.

We first generate a version of *Bad* so that the annual and sample average absolute forecast errors match the sample average absolute forecast errors of *Late* and *Early*. Next, leaving inflation unchanged, we amplify or moderate the central bank's forecast errors to create the other versions of consistent listed in Table 2. We chose average absolute forecast errors in *Great* (*Terrible*) to exactly match the average absolute forecast error in the final year of *Early* (*Late*). Finally, we chose absolute error values for *Good* and *Moderate* so that they partitioned the performance difference between *Great* and *Bad*.

Treatment Summary				
	History 1	**History 2**	**History 3**	**Sample Size**
T1a	*Early*	*Late*	*Great*	46
T1b	*Late*	*Early*	*Great*	44
T2a	*Early*	*Late*	*Good*	44
T2b	*Late*	*Early*	*Good*	46
T3a	*Early*	*Late*	*Moderate*	33
T3b	*Late*	*Early*	*Moderate*	44
T4a	*Early*	*Late*	*Bad*	97
T4b	*Late*	*Early*	*Bad*	76
T5a	*Early*	*Late*	*Terrible*	46
T5b	*Late*	*Early*	*Terrible*	50

Table 3: Treatment Summary

4.1.1 Hypotheses - ***ForecastPerformance***

We do two things in this build on Morris and Shin (2002) to first discuss the three measures of updating we will consider throughout our results section and to establish hypotheses about how historical forecast performance influences a central bank's perceived forecast credibility.

Measures of Updating

Suppose participant i has a prior belief about inflation given by

$$\pi_i \sim \mathcal{N}\left(\bar{\pi}_i, \frac{1}{\alpha_i}\right), \tag{10}$$

where $\bar{\pi}_i$ is i's initial point forecast and α is a measure of i's forecast precision that we proxy using i's initial forecast uncertainty.

The central bank introduces a new signal, which may be biased

$$\pi_{cb} = \pi + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}\left(\gamma, \frac{1}{\beta}\right). \tag{11}$$

Here, γ represents a possible systematic bias in the central bank's inflation forecast. If $\gamma = 0$ then the central bank's forecast errors exhibit no auto-correlation and are unbiased. If $\gamma < 0$ then the central bank systematically under-forecasts inflation and if $\gamma > 0$ then the central bank systematically over-forecasts inflation. β is related to the precision of the central bank forecast, which i can infer from the 12-quarter economic history. We can rewrite Equation (11) so that

$$\pi_{cb} = (\pi + \gamma) + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \frac{1}{\beta}\right). \tag{12}$$

For intuition, suppose that $\gamma < 0$ so that the central bank systematically under forecasts inflation. Then Equation (12) this equation says that the true signal from the central bank is equal to the actual value of inflation π minus its bias γ plus some idiosyncratic error term ϵ. This means $\pi_{cb} - \gamma = \pi + \epsilon, \quad \epsilon$ so that:

$$\mathbb{E}(\pi|\pi_{cb}) = \frac{\alpha\bar{\pi}_i + \beta(\pi_{cb} - \gamma)}{\alpha + \beta} \tag{13}$$

Regardless of γ's value, Equation (13) says that a Bayesian participant's updated point forecast will be a precision-weighted, linear combination of her initial point forecast $\bar{\pi}_i$ and the central bank's signal $\pi_{cb} - \gamma$. We plot the optimal level of updating (percentage terms, $100 \times \left(\frac{\beta}{\alpha+\beta}\right)$) for a Bayesian agent for different levels of β and α in Figure 3.

Why does this matter and how does this inform the way we estimate updating in our data?

First, consider the update measure

$$u_{1,i} = |\pi_1 - \pi_{cb}| - |\pi_2 - \pi_{cb}|. \tag{14}$$

This produces a simple measure of updating in terms of basis points, bounded above by $|\pi_1 - \pi_{cb}|$, where positive values correspond to updating toward the central bank's signal and negative values correspond to updating away from the signal. An issue with this measure is that we could see systematic differences in $|\pi_1 - \pi_{cb}|$ across histories so that even if $\pi_2 = \pi_{cb}$ in all instances, we still observe differences in updating across histories. To account for this, we normalize the measure and covert it to percentage terms

$$u_{2,i} = \left(\frac{|\pi_1 - \pi_{cb}| - |\pi_2 - \pi_{cb}|}{|\pi_1 - \pi_{cb}|}\right) \times 100 \tag{15}$$

so that $\pi_2 = \pi_{cb} \implies u_{2,i} = 1$. Now, systematic differences in $|\pi_1 - \pi_{cb}|$ by history are irrelevant – we have a measure we can cleanly compare across histories. Implicit in this though is that $\gamma = 0$ from the perspective of i. If instead i is Bayesian and infers that $\gamma \neq 0$ and we don't account for it in our update estimate, we would estimate:

$$u_{2,i} = \frac{|\pi_1 - \pi_{cb}| - |\pi_2 - \pi_{cb}|}{|\pi_1 - \pi_{cb}|} u_{2,i} = 1 - \frac{|\gamma|}{|\pi_1 - \pi_{cb}|} \tag{16}$$

since $\pi_2 = \pi_{cb} - \gamma \implies \frac{|\pi_2 - \pi_{cb}|}{|\pi_1 - \pi_{cb}|} = \frac{|(\pi_{cb} - \gamma) - \pi_{cb}|}{|\pi_1 - \pi_{cb}|}$. This means we would underestimate updating. To fix this, we can consider a third update measure:

$$u_{3,i} = \left(\frac{|\pi_1 - (\pi_{cb} - \gamma)| - |\pi_2 - (\pi_{cb} - \gamma)|}{|\pi_1 - (\pi_{cb} - \gamma)|} \right) \times 100 \tag{17}$$

where now we would estimate $u_{3,i} = 100\%$ after accounting for (or assuming the Bayesian subject accounts for) the CB's bias γ. Estimating $u_{3,i} = 100\%$ would imply either that $\beta \to \infty$, $\alpha \to 0$, or both, which is akin to placing full weight on the central bank's signal in Equation (13).

Equation (13) also allows us form clear hypotheses about the relationship between historical forecast performance and the central bank's forecast credibility, as measured by u_1, u_2, and u_3. Using the inverse of a histories sample-average absolute forecasting error as proxy for precision, we have the following:

Hypothesis 1. *A central bank's forecast credibility is decreasing in its historical average absolute forecast error.*

Equation (13) also clarifies that a bank's forecast credibility doesn't depend simply on its own performance. Instead, it depends on the bank's performance relative to a participant's belief about her own forecasting credibility

Hypothesis 2. *For a given economic history, a central bank's forecast credibility is increasing in a participant i's own forecast uncertainty.*

This hypothesis says that participants who exhibit higher levels of forecast uncertainty in the Initial Forecast will update more. Another way to think about this is that forecast errors should be least costly among the subset of participants who exhibit the most forecast uncertainty.

4.1.2 Results

We start by considering updating in each of our five *ForecastPerformance* treatments alongside corresponding measures of the Bayesian optimal level of updating. We show this in Figure 4. Panel (a) depicts box-and-whisker plots (Tukey et al. 1977) of individual-level updates and Bayesian optimal updates. Panel (b) shows mean deviations from the Bayesian optimal update (blue dots) surrounded by 90% confidence intervals.

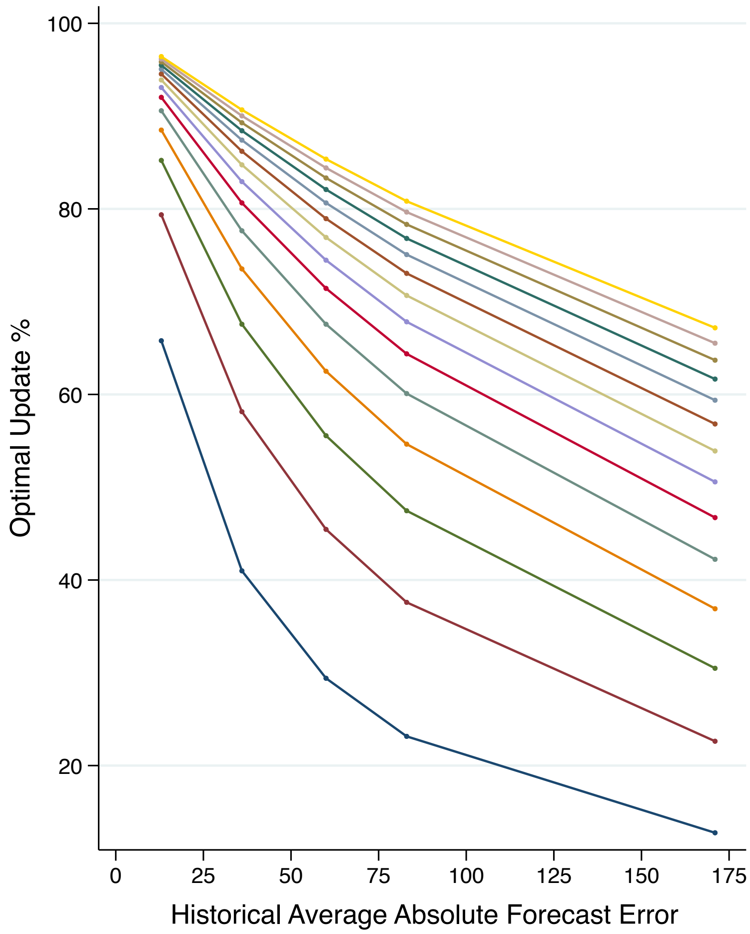
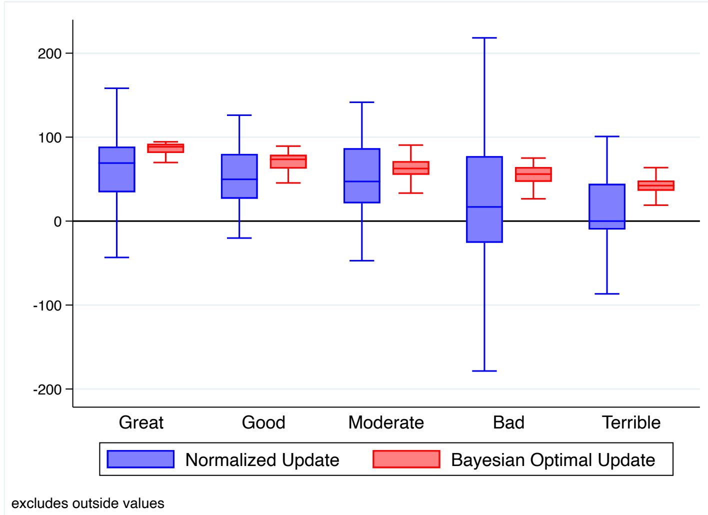
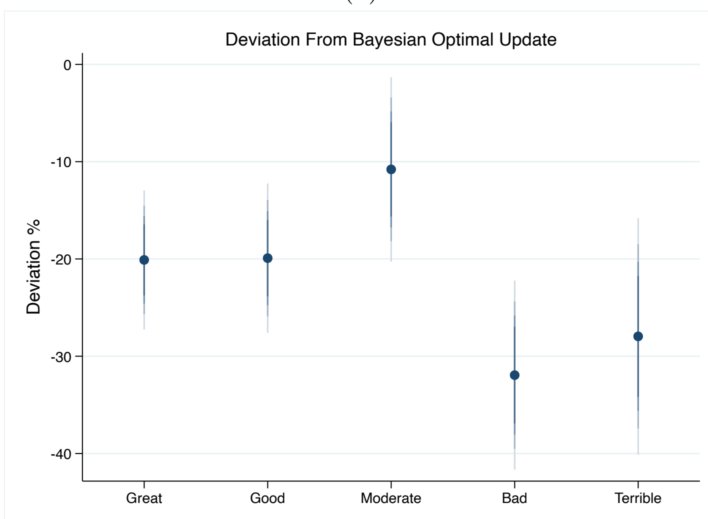


Figure 3: This figure shows the optimal level of updating in percentage terms (y-axis) prescribed by Equation (13) for different levels of a central bank precision (x-axis). Each line denotes a different level participant forecast uncertainty ranging from 25 basis points (top-most line) to 350 basis points (bottom-most line) in increments of 25 basis points.



(a)



(b)

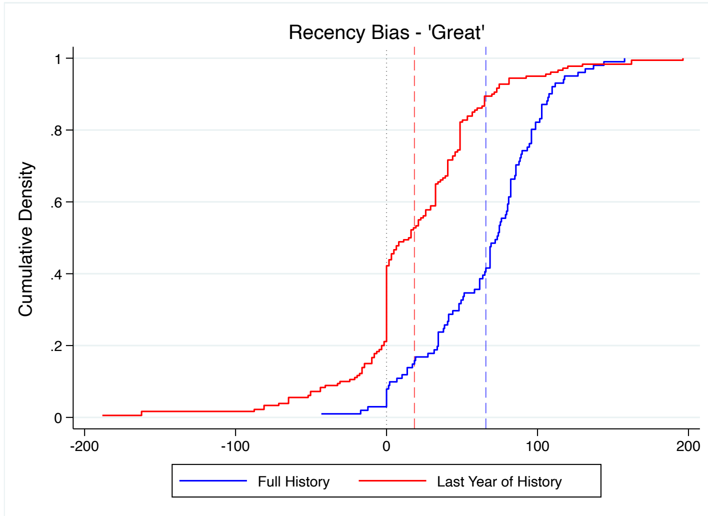
Figure 4

4.2 Timing

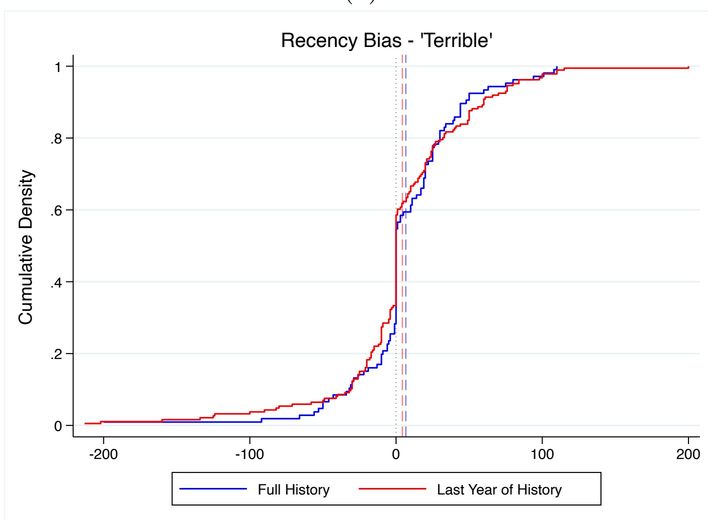
We also study whether the timing of forecast errors influences how participants weight information. To do this, we expose participants to each of use *Early, Late,* and *Bad* (described in Table 2) where the historical average forecast performance is identical but the timing of forecast errors throughout the economic histories vary. If subjects equally weight historical information when deciding on a central bank's forecast credibility, then a participant forms a perception of the bank's precision, β according to

$$\beta^{-1} = \frac{\sum_{j=1}^{j=12} \left|\mathbb{E}_{j-1}^{CB}\left(\pi_j\right) - \pi_j\right|}{12} \tag{18}$$

and $\beta_i = \beta_j \; \forall \; i, j \in \{Early,\ Late,\ Consisitent\}$. If instead the timing of errors matters, then$\beta_i \neq \beta_j$ for some set of histories i, j. Suppose instead that participants think of the



(a)



(b)

Figure 5

central bank's forecast credibility as ever-changing and account for this by more heavily weighting more recent performance. Then we might instead see something like

$$\beta^{-1} = \lambda \sum_{j=0}^{j=11} (1-\lambda)^j \left| \mathbb{E}_{t-2-j}^{CB} \left(\pi_{t-1-j} \right) - \pi_{t-1-j} \right| \tag{19}$$

where the weighting function exhibits exponential decay in time, as measured by quarters j. This is akin to constant-gain learning models of expectations formation common in learning literature (Evans et al. 2001).

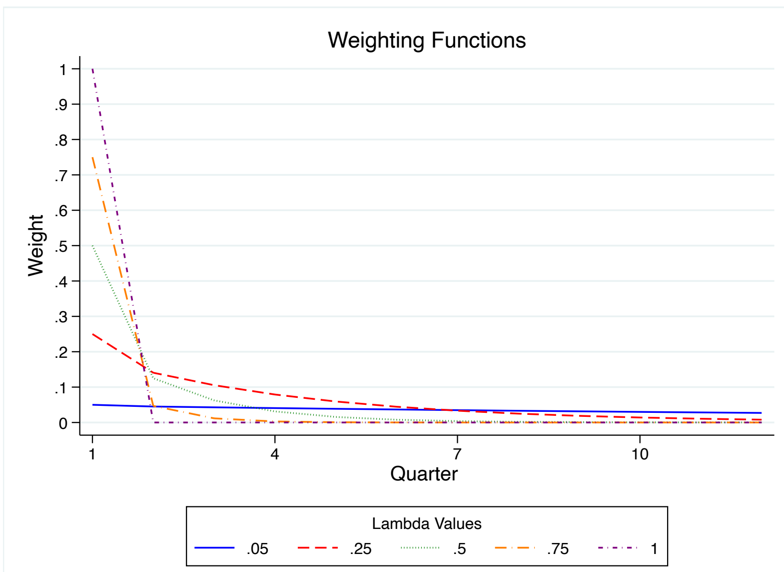


Figure 6: Weighting functions (Equation (19)) with different values of λ

Notes:

- Equation (19) is essentially a model of adaptive expectations with exponentially declining weights. This sort of model is typically motivated as being a response to structural change in whatever macroeconomic time series an agent is forecasting. In our setting, there is no structural change in forecast performance. However, subjects may think there is? So, one question we can ask/answer is whether there is an asymmetry in how perceived structural change maps into lambda whenever that change increases or decreases forecast performance.

- This relates to Malmendier and Nagel (2016), who show that two people born at different times s and $s + j,\ j > 0$, can perceive information at $t > s + j$ different due to different life experiences. Suppose these two agents are now at time t and must forecast $t + 1$ inflation. Suppose each agent thinks inflation is an AR(1) process:

$$\pi_{t+1} = \gamma + \alpha_t \pi_t + \nu_{t+1}.$$

  To forecast $t + 1$ inflation, both agents must estimate γ and α. MN argue the agent born in $s + j$ bases estimates of γ and α on less information than does the agent born at s and so will more heavily weight any new information received in t than will the agent born in s when estimating these parameters. The thing that matters here is age. The thing that matters in our context is time. What makes what we are doing different is that we are considering how agents relate historical forecast performance to perceptions of credibility rather than describing how past realizations of a time series map into future realizations of the time series.

5 Conclusion

Bibliography

Adam, K. (2007). Experimental evidence on the persistence of output and inflation. *The Economic Journal 117*(520), 603–636.

Ahrens, S., J. Lustenhouwer, and M. Tettamanzi (2019). The stabilizing effects of publishing strategic central bank projections. *Macroeconomic Dynamics*, 1–43.

Arifovic, J. and L. Petersen (2017). Stabilizing expectations at the zero lower bound: Experimental evidence. *Journal of Economic Dynamics and Control 82*, 21–43.

Assenza, T., P. Heemeijer, C. H. Hommes, and D. Massaro (2013). Individual expectations and aggregate macro behavior.

Bao, T., C. Hommes, J. Sonnemans, and J. Tuinstra (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control 36*(8), 1101–1120.

Blinder, A. S. (2000). Central-bank credibility: Why do we care? how do we build it? *American economic review 90*(5), 1421–1431.

Clarida, R., J. Galí, and M. Gertler (1999, December). The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature 37*(4), 1661–1707.

Cornand, C. and C. K. M'baye (2018). Does inflation targeting matter? an experimental investigation. *Macroeconomic Dynamics 22*(2), 362–401.

Cornand, C. and C. K. M'baye (2018). Band or point inflation targeting? an experimental approach. *Journal of Economic Interaction and Coordination 13*(2), 283–309.

Evans, G. W., S. Honkapohja, and R. Marimon (2001). Convergence in monetary inflation models with heterogeneous learning rules. *Macroeconomic Dynamics 5*(1), 1–31.

Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton, USA and Oxford, UK: Princeton University Press.

Hommes, C., D. Massaro, and I. Salle (2019). Monetary and fiscal policy design at the zero lower bound: Evidence from the lab. *Economic Inquiry 57*(2), 1120–1140.

Hommes, C., D. Massaro, and M. Weber (2019). Monetary policy under behavioral expectations: Theory and experiment. *European Economic Review 118*, 193–212.

King, R. G., Y. K. Lu, and E. S. Pastén (2008, December). Managing expectations. *Journal of Money, Credit and Banking 40*(8), 1625–1666.

Kryvtsov, O. and L. Petersen (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics 117*, 760–780.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics 131*(1), 53–87.

Mokhtarzadeh, F. and L. Petersen (2021). Coordinating expectations through central bank projections. *Experimental economics 24*(3), 883–918.

Morris, S. and H. S. Shin (2002). Social value of public information. *american economic review 92*(5), 1521–1534.

Petersen, L. and R. Rholes (2022). Macroeconomic expectations, central bank communication, and background uncertainty: a covid-19 laboratory experiment.

Pfajfar, D. and B. Žakelj (2014). Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control 44*, 147–168.

Pfajfar, D. and B. Žakelj (2016). Uncertainty in forecasting inflation and monetary policy design: Evidence from the laboratory. *International Journal of Forecasting 32*(3), 849–864.

Pfajfar, D. and B. Žakelj (2018). Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics 22*(4), 1035–1075.

Rholes, R. and L. Petersen (2021). Should central banks communicate uncertainty in their projections? *Journal of Economic Behavior & Organization 183*, 320–341.

Svensson, L. E. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European Economic Review 41*(6), 1111–1146.

Tukey, J. W. et al. (1977). *Exploratory data analysis*, Volume 2. Reading, MA.

Walsh, C. E. (2017). *Monetary theory and policy*. MIT press.

Woodford, M. (2003). *Interest and Prices*. Princeton University Press.

Woodford, M. (2005). Central bank communication and policy effectiveness. *Proceedings - Economic Policy Symposium - Jackson Hole* (Aug), 399–474.

6 Appendix

6.1 Tables and Figures

6.2 Instructions

Experimental Instructions

You will now proceed to our experiment. If you read these instructions carefully and make appropriate decisions, you may earn a considerable bonus payment in addition to the participation payment. The bonus depends directly on the quality of your decisions.

You can access these instructions throughout the experiment. You may toggle the instructions on and off using the button labelled 'Instructions' below the 'Next' button on any page.

We will quiz you over these instructions on the following page. If you submit the quiz with at least one wrong answer more than three times then we will end the experiment early.

Your Objective in the Experiment

Your job in this experiment **is to forecast inflation**. Inflation is a measure of how prices change over an observed period of time. By 'inflation forecast' we mean your best guess of what inflation will be at a certain point in time. The more accurate your inflation forecasts, the more bonus money you earn!

You will provide two types of inflation forecasts:

- **Point Forecast:** Your 'Point Forecast' of inflation is your best guess of the exact value inflation will be at a certain point in time.
- **Range Forecast:** Your 'Range Forecast' of inflation allows for some uncertainty by letting you provide a range of possible values, defined by upper and lower inflation bounds, that you think will almost certainly contain the actual value of inflation.

Additional Definitions:

- **Central bank:** These national institutions provide banking services for the government, issue currency, and set interest rates to control inflation and maintain economic stability. Examples are the Federal Reserve in the United States and the Bank of England in the United Kingdom. An important part of a central bank's job is to provide economic forecasts to the public. Some examples of things central banks forecast are inflation and unemployment.
- **Forecast error:** A forecast error is the difference between an inflation forecast and inflation at a specific time. Your goal in this experiment is to have the smallest forecast error possible.

- **Quarter:** A quarter is a common unit of time for economic data. One quarter is equal to three months so that each year has four quarters. Central banks usually provide quarterly forecasts.

The experiment:

This experiment consists of **three decision periods**. In each decision period, you will form two sets of inflation forecasts. We call these your Initial Forecasts and your Updated Forecasts. The imagine below shows the flow of a decision period.

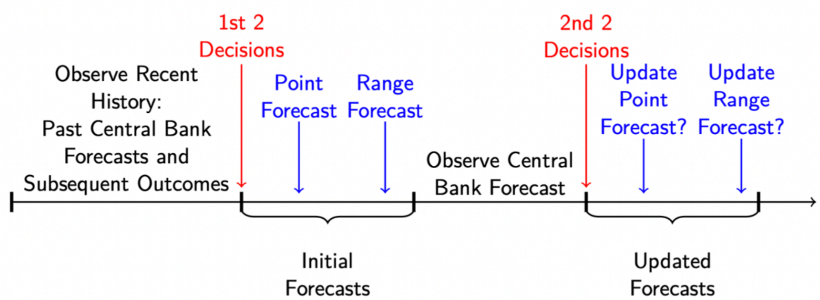


Figure: Experimental Timeline: A single decision period

1. We provide 12 quarters of history (quarters 0 through 11) of inflation (blue line and dots) alongside the central bank's corresponding forecasts for those quarters (black line and dots).
   - Note that the difference between these two dots within the same quarter represents the central bank's forecast error for that quarter.
2. After viewing this history, you will provide your Initial Forecasts:
   - A point forecast (red dot) of inflation for the next quarter (Quarter 12)
   - Your corresponding range forecast of inflation.
3. After forming your Initial Forecasts, we will reveal to you the central bank's inflation forecast (on the next screen).
4. You will then provide your Updated Forecasts:
   - You will again provide a point forecast and a corresponding range forecast of inflation.
   - Your Updated Forecasts can be the same as your Initial Forecasts, use some of the same values, or use completely new values.
   - We provide information about your Initial Forecasts both graphically and numerically whenever you are forming your Updated Forecasts.
5. After providing your Updated Forecasts, we will reveal the actual value of inflation for the forecasted period and inform you of your forecast performance.

6. **You will play through three decision periods with different economic data in each decision period.**

How our software scores your performance:

- *Point forecast:*
  - A perfect forecast earns exactly \$1.
  - The larger your forecast error (above or below), the less you earn.
- *Range forecast:*
  - If inflation does not fall inside your forecast range, you earn nothing for your range forecast.
  - The total range of your forecast is given by the gap between the upper bound of range forecast and the lower bound of range forecast.
  - If actual inflation is inside your forecast range, you score $P = \frac{1}{1+totalrange}$.
  - The larger the range you create the less money you earn for your range forecast.

Suppose that actual inflation turns out to be 2.5%

- If you set your range from 1% to 3% then you would earn $P = \frac{1}{1+2} = \$.33$
- If you set your range from 1% to 5% then you would earn $P = \frac{1}{1+4} = \$.2$
- If you set your range from 3% to 5% then you would earn nothing since actual inflation is not within your range.
- you set your point forecast to 2.5% then you would earn \$1
- If you set your point forecast to 3.5% (or 1.5%) then you would earn \$0.50
- If you set your point forecast to 4.5% (or 0.5%) then you would earn \$0.25

You will get paid for your performance in one set of forecasts (Initial or Updated) in one of the 3 decision periods:

- Our software randomly chooses one of your three decision periods.
- For that decision period, the software chooses randomly either the initial forecasts or the updated forecasts.
- We pay you for this set of inflation forecasts as a bonus payment.

This means you need to take both the Initial Forecasts and the Updated Forecasts equally seriously when making your decisions.

Interacting with the data and inputting your forecasts:

The historical data:

- You may hover your mouse over any dot on the figure to see its exact value, which will appear in the upper left-hand corner of the graph.
- We remind you of your Initial Forecasts graphically (red dot and red shading) and numerically when forming your Updated Forecasts.

Providing your Point Forecast:

- You may submit positive values (prices are going up), negative values (prices are going down), or a value of zero.
- You can input your point forecast of inflation by clicking on the graph in the shaded 'Your Forecast' section and then dragging/dropping the dot that appears there.
- The dot will be red for your Initial Forecast and blue for your Updated Forecast.
- You may also type your forecast into the clearly labelled input text box.

Providing your Range Forecast:

- Our software will randomly generate upper and lower bounds for your range forecast (shaded region surrounding your point forecast).
- You may click on and drag these upper and lower bounds to whatever values you prefer.
- You can also drag the entire forecast range up and down.
- Your forecast range can be as big or small as you prefer.
- You may choose to have more or less range above your point forecast than below, and vice versa.
- Your upper (lower) bound must always be equal to or above (below) your point forecast - the software will prevent impossible range inputs.

Extended Abstract for Central Bank Credibility

Michael McMahon, Ryan Rholes

Monetary policy frameworks now, largely, involve the management of expectations (Woodford 2005, King et al. 2008). For instance, expectations management is a key tenet of the widely adopted inflation targeting monetary policy framework. This follows from New Keynesian models which have underpinned most recent theoretical research on the effects of monetary policy (Clarida et al. 1999, Woodford 2003, Galí 2008, forexample). This is because, in these frameworks, inflation expectations become a vital determinant of inflation.

Central bank communication has emerged as a key part of the toolkit to manage expectations. Some would argue communication is *the* key part of that toolkit. This communication involves both high frequency communication and lower frequency communication, but both, according to the channel above, can play a key role in inflation control by influencing how agents form their expectations. Hence, open mouth operations are now an indispensable component of monetary policy.

Does communication actually work to influence expectations? In workhorse monetary models, agents are rational and appropriately incorporate central bank, and other, information to form the best possible expectations. In practice, central banks worry about their credibility, which is necessary for the transmission of communication policy Blinder (2000).

However, little is known in practice about the determinants, dynamics, or role of central bank credibility. We can imagine a world where successful achievement of the inflation target builds credibility and gives the central bank greater ability to control inflation by better anchoring inflation expectations. But alongside this virtuous cycle may sit a vicious cycle; lower credibility could impinge upon the ability of the central bank to manage inflation which then makes credibility-reducing inflation fluctuations more likely.

In this paper, we develop an experimental framework to study the determinants and dynamics of central bank credibility, and their relationship with expectations management. This relationship is important because inflation forecasts are a key input into the policy decision in an inflation targeting framework (Svensson 1997).

Participants in our experiment act as inflation forecasters tasked with providing two sets of one-period-ahead point and potentially asymmetric range forecasts of inflation (Initial Forecasts and Updated Forecasts) in 3 sequential decision periods. We begin decision periods by exposing subjects to 12 quarters of economic history consisting of realized inflation alongside corresponding central bank inflation forecasts. Subjects provide Initial Forecasts (priors) after the revelation of the economic history. We then show them the central bank's corresponding inflation forecast and allow them to update their own inflation forecast (forming their posterior estimate, i.e. Updated Forecasts). We measure credibility by the weight that agents place on the central bank's inflation forecast when updating their estimate of inflation.

Differences in economic histories constitute within-subject treatment variation in our experiment. We refer to these histories as: *Early*, *Late*, and *Consistent*. In *Early*, the central bank commits significant forecast errors in the first third of the forecasting history, moderate errors in the second third, and minimal errors in the last third. We reverse this pattern for *Late*. For *Consistent*, the central bank exhibits a consistent average annual forecast performance. Finally, *Late* exactly reverses the absolute error structure of *Early*. We simulated economic histories using the New Keynesian model described in Walsh (2017) linearized around the zero-inflation steady state. We employ pre-selected shock sequences that preserve the structure of real-world absolute forecasting errors from the United Kingdom and Bank of England (BoE) for the three-year period beginning in the first quarter of 2010 and ending in the final quarter of 2012. We then shift the resulting inflation data so that is centered around 2%.

This exercise produces economic histories that should lead to no differences in updating for rational, Bayseian-updating subjects. However, we find subjects exhibit a strong recency bias when forming beliefs about the central bank's forecast credibility. This recency bias leads subjects to place the least weight on the central bank's forecast when updating their own expectations in the *Late* history, more in *Consistent* than *Late*, and the most weight in *Early*.

This recency bias also extends to how subjects' beliefs about central bank credibility influence higher moments of subject-level density forecasts of inflation. Subject-level uncertainty responds most strongly to the central bank's forecast in *Early* and c*Consistent*. Additionally, the higher level of central bank credibility in *Early* and *Consistent* lead subject's to reduce the asymmetry of their forecasts while in *Late* , where point forecasts shift significantly less than in other treatments, subjects increase the skewness of their forecasts toward the central bank's forecast.

This project is ongoing with several follow-on treatments planned. First, we will introduce a level difference into the central bank's historical forecast performance to study the relationship between average historical forecast performance and credibility. This produces a series of six treatments identical to those described above but with average absolute forecast errors that are reduced by approximately 23. This allows us to understand how credibility is affected by overall performance, and whether performance affects the nature of the recency bias.

Additionally, we plan to introduce a series of treatments aimed at understanding the specific temporal nature of the recency bias by extending the economic history provided to subjects in such a way that recency in levels (i.e. I only consider the last X periods of the Y-period history) or proportional recency (i.e. I only consider the last Z-percent of period in a Y-Period history) produce predictably different behavior.

Finally, we will introduce a series of treatments to study whether and how the central bank can use textual communication to offset the detriment of recency bias in the *Late* histories. This answers the question as to whether communication can mitigate the credibility reducing effect of poor forecast performance?

Shorter Abstract

Central banks have increasingly relied on communication to manage expectations, which is a key tenet of inflation targeting frameworks. In practice, effective communication relies crucially on central bank credibility. Despite this, little is known about the determinants or dynamics of credibility. To this end, we introduce a novel experimental framework to study credibility in relation to expectations management. We find that subjects acting as inflation forecasters exhibit a strong recency bias when forming beliefs about the central bank's forecast credibility and that this recency bias also impacts subjects' higher-order forecast moments. We introduce additional treatments to understand how level differences in forecast performance impact perceived credibility, the exact temporal nature of the recency bias we observe, and to explore whether contextualization via additional communication can offset this recency bias (i.e. if banks can 'talk their way out of trouble').

6.3 Alternative Words on Effect we are investigating

The basic fact about forecasting and macroeconomic policy is that ex-ante optimal forecasts will still have ex-post errors. When policy makers foresee major economic events, they take action and policy endogenously prevents these big events from arising. So, by definition, the crises that we see or the inflation that happens, are exactly the ones the policymakers didn't see.

But people don't generally understand that. The question that the Queen asked is a reflection of this. So too is the plethora of articles about macroeconomics.

We show that this is true - even though it shouldn't be. Forecasting performance matters a lot to people.

6.4 Notes from Jamboree 2022

Thinking 'out loud'

- Lots of pushback from audience about our framing. In particular, the audience disliked us referring to what we are measuring as 'credibility'. These types of forecasts have no role in the FIRE models typically used to study monetary policy since they neither inform the agent nor serve as a coordination device. Because of this, forecast accuracy doesn't feature in a standard definition of central bank credibility. Instead, people tend to think of credibility as the central bank's ability to achieve an inflation objective (i.e. anchor expectations around a target) or stick to well-articulated policy rules or pursue well-defined policy objectives in a coherent way.
- Nevertheless, central banks care deeply about these sorts of forecasts (dedicate resources to producing them, changed them surrounding CoVID, etc.) and markets

react to these sorts of projections (work with Tatevik for numerical, Michael's JME for textual, etc.). Further, we have suggestive empirical evidence corroborating the idea that markets respond more strongly to central bank information whenever the bank has recently performed 'well' (better than historical average performance).

- And so what are we showing? Well, at minimum, we are showing that people care very much about historical forecast performance when deciding how to make use of contemporary forecasts. And at least one obvious implication of this is that real-world central banks using forecasts as a tool to coordinate and manage expectations should then also care.

	Treatment Summary			
	History 1	**History 2**	**History 3**	**Sample Size**
T1	*Early*	*Late*	*Consistent*	97
T2	*Early*	*Consistent*	*Late*	94
T3	*Late*	*Early*	*Consistent*	80
T4	*Late*	*Consistent*	*Early*	88
T5	*Consistent*	*Early*	*Late*	79
T6	*Consistent*	*Late*	*Early*	91

Table 4: Treatment Summary