Building Central Bank Credibility: The Role of Forecast Performance*

Michael McMahon† Ryan Rholes‡

This Draft: November 2023
Click here for latest draft

Abstract

This paper examines how central banks' forecast accuracy impacts the influence publicized forecasts have on inflation expectations. We find, using an incentivized experiment, that forecast performance matters. Though there is evidence of over-precision bias overall, the main, and novel, finding is the presence of recency bias when subjects evaluate forecast accuracy. This bias, which applies to both short-term and medium-term forecasts, is especially strong after poor forecasting performance. In a New Keynesian model, such biases lead to endogenous forecast credibility which increases the persistence of inflation. Importantly, narrative communication can partly mitigate the detrimental effect of recent poor forecasting.

Keywords: Expectation formation, Forecasting, Central Bank Communication

JEL Codes: E52, E58

*We are grateful to seminar participants at the University of Oxford, at the 2022 conference on Monetary Policy: Heterogeneity, Communication and Subjective Inflation Expectations, the Economic Science Association's (ESA) World Meeting 2022 at MIT, the ESA European Meeting 2022 at the University of Bologna, the 2023 meeting of the Southern Europe Experimental Team, the 2023 SUERF-Banca d'Italia-ECB-EIB conference at the Bank of Italy, the 60th SUERF Anniversary 50th OeNB Annual Economic Conference at The Oesterreichische Nationalbank, the 2023 Meeting of the Central Bank Research Association, and at the Workshop on Challenges For Monetary Policy in Times of High Inflation hosted at Leibniz University Hannover. We also thank Luba Petersen and Johannes Abeler for their helpful comments. McMahon gratefully acknowledges financial support from the European Research Council (Consolidator Grant Agreement 819131). The study received IRB approval from the University of Oxford. AEA Trial Registry RCT ID: AEARCTR-0009026. 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

A central tenet of monetary economics, reflected in the widely-adopted inflation targeting framework, is that central banks must control inflation expectations (Woodford 2005, King et al. 2008, Candia et al. 2020). This is because policymakers view inflation expectations as a key determinant of contemporaneous inflation (Clarida et al. 1999, Woodford 2003, Galí 2008, for example). Essentially, inflation-targeting central banks engage in inflation-forecast targeting and use communication as a means for influencing inflation expectations. For instance, the majority of central banks invest considerable resources into developing and publishing numerical inflation forecasts and contextualizing those forecasts with reports, statements, and speeches. Open-mouth operations have become an integral component of monetary policy, though the extent of their use remains debated Coibion et al. (2020).

This paper uses an experimental framework to study to what extent communication influences inflation expectations. We build on a growing literature (Haldane and McMahon 2018, Coibion et al. 2022, for instance). Our unique focus in this paper is on the role of central bank forecast performance in affecting the influence of the central bank in private expectations. In workhorse monetary models populated by fully-informed rational agents, communicating about inflation is irrelevant since the central bank holds neither an informational nor a processing advantage. Under these assumptions, both the central bank and the agent form coincident, optimal inflation expectations. In reality, the central bank may hold advantages along either or both dimensions and should therefore communicate – via numerical forecasts and contextualization of those forecasts – to improve the inflation expectations of agents. However, the crux of the central bank's problem is whether this communication effectively controls inflation expectations.[1] Theoretical work typically assumes, implicitly or explicitly, that the central bank is fully credible so that communication policy is fully effective. In practice, central banks worry about establishing and safeguarding credibility, which is necessary for the transmission of communication policy (Blinder 2000).

In a world where the central bank is highly credible, it can control and anchor inflation expectations, leaving it free to pursue short-run stabilization policies that promote economic stability and reinforce credibility. But what if the central bank's inflation forecast credibility is low? Instead of a virtuous cycle, there may exist 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 consider this question using an experimental environment to assess the extent to which agents are affected by the forecast performance of the central bank precisely in terms of how the central bank's forecast signals lead agents to update their expectations of inflation. Our interest is in whether a central bank's forecast credibility is endogenously related to its historical forecast performance.[2] And if so, how?

[1] Eusepi and Preston (2010) consider the role of communicating less-than-full information about the central bank's policy function, which they argue is akin to partially-credible communication. Such communication in their framework is enough to prevent self-fulfilling expectations and restore macroeconomic stability under optimal policy.

[2] By inflation forecast credibility, we mean the degree to which a central bank can influence the inflation forecasts of economic agents via a publicized numerical inflation forecast.

Macroeconomists who study expectations experimentally typically use the learning-to-forecast (LTF) framework, which comprises experimental economies that evolve endogenously according to the incentivized expectations of participants. Researchers have used this framework to study the design and efficacy of central bank communication (Kryvtsov and Petersen (2021); Arifovic and Petersen (2017); Cornand and M'baye (2018); Rholes and Petersen (2021); Petersen and Rholes (2022)), 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)).

Our experiment, in contrast, employs an individual-choice setting where inflation evolves exogenously so that our participants are atomistic and face no strategic uncertainty. Because of this, our participants do not interpret inflation forecasting as a coordination game, allowing us to more cleanly isolate the causal relationship between features of a central bank's forecasting history and participants' perceptions of the bank's forecast credibility. This experimental model aligns with 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.

Participants in our experiment provide two sets of one-period-ahead point and range forecasts of inflation in each of three independent decision periods (e.g. 'Initial Forecasts' and 'Updated Forecasts'). We begin each decision period by revealing the three most recent years of a central bank's inflation forecasts alongside actual inflation. Subjects provide Initial Forecasts (priors) for the next period based on this historical data. We then reveal the central bank's own inflation forecast and allow subjects to update their own density projection (i.e. Updated Forecasts or posterior). Thus, we collect incentivized measures of each participant's initial outlook on inflation, perceived initial inflation forecast precision, and their updated outlook on inflation. Additionally, we control both the central bank's historical forecast precision and its signal. Using these measures, we can precisely quantify the degree to which a participant incorporates the central bank's signal into their inflation outlook using a Bayesian signal processing framework and relate this causally to historical economic information we reveal at the start of each decision period.[3]

This design relates most closely to Armantier et al. (2016), who use a Bayesian framework to study how inflation expectations respond to historical price information or professional forecasts in an information provision experiment embedded into the Michigan survey. A key difference between our work and theirs is that our design allows us full control over extraneous features of our experimental environment. Because of this, we can introduce precise treatment variation to cleanly isolate features of the central bank's forecasting history to study their causal relationship to forecast credibility.

Using this framework, we study how several key features of the central bank's forecasting history relate to its forecast credibility. First, we consider how the central bank's historical forecast precision influences its forecast credibility (*Forecast Performance*). To

[3]We use the forecast performance of the Bank of England (BoE) for the three-year period beginning in the first quarter of 2010 and ending in the final quarter of 2012 to calibrate the magnitude of forecast errors.

do this, we create a series of economic histories that control for the pattern of historical forecast errors while scaling the magnitude of the historical average absolute forecast error. We find that participants behave qualitatively like Bayesians, in that the central bank's forecast credibility depends significantly on historical forecast precision. However, the link between precision and credibility is not as sharp as theory predicts, with participants overvaluing low precision and undervaluing high precision. This happens because subjects fail to correctly incorporate their own forecast precision when deciding how to judge the central bank's signal.

Second, we hold the central bank's historical forecast precision constant and ask how variation in the timing of historical inflation forecast errors influences the central bank's forecast credibility (*Timing*). We do this using a set of three core economic histories wherein a perfectly rational Bayesian agent should find the central bank equivalently credible unless the timing of forecast errors distorts the value of less temporally proximate historical information. Participants exhibit considerable recency bias, with subjects relying strongly on the most recent historical information when forming a perception of central bank forecast credibility.

Importantly, we find that this recency bias is considerably stronger if the recent forecast performance was poor. This is true despite the fact that the historical forecast precision changes at the same speed and by the same amount in our economic histories. This suggests an important asymmetry in how recency bias interacts with forecast performance. Bad forecast performance seems to hold more value to participants, regardless of when it occurs. Because of this, subjects exhibit less recency bias when the central bank's forecast precision improves than when it deteriorates. This has important implications for the dynamics of credibility; it is much easier for a bank to lose credibility than to rebuild it. Further, the speed with which credibility evaporates suggests that unanticipated shocks leading to poor short-term forecast performance can undermine the efficacy of inflation communication precisely when the monetary authority most needs it as a policy tool.

We explore the robustness of these timing-related results in two ways. First, we show that these results also hold whenever participants forecast long-term average inflation rather than short-term inflation. Second, we introduce a set of treatments wherein we reverse the direction of forecast errors to show that our results hold regardless of whether the bank under- or over-forecasts inflation.

Finally, we test whether providing contextual information can bolster credibility for a central bank that finds itself in a position of low forecast credibility (*Contextual Communication*). That is, can a central bank talk its way out of a low-credibility position? To do this, we create a series of text-based communication interventions that contain a forward-looking component, rationalize forecast mistakes as the result of unforeseeable exogenous shocks or endogenous policy errors, and report performing better or worse than peer forecasting institutions. Communication that reinforces the central bank's numerical inflation forecast without providing additional information can significantly improve credibility. Layering on additional information can further increase credibility but the effect is more nuanced. Reporting that the bank under-performed relative to peer institutions reduces credibility sufficiently to eliminate any gains from contextual communication. Reporting that the bank outperformed peer institutions bolsters credibility. There is little-to-no difference in credibility arising from the source of forecast errors.

One might question whether our results are applicable to real-world markets or are instead artifacts of our stylistic setting. We attempt to assuage these concerns using a high-frequency, event-study framework to determine whether markets in the United Kingdom respond more strongly to Bank of England (BoE) communication whenever the BoE's recent forecast performance is strong. We show this is true for UK guilt's at several maturities on the short-end of the UK's yield curve and that the effect increases as we expand temporally our backward-looking forecast performance measure. Interestingly, we find the effect eventually stabilizes with respect to the temporal span of this forecast performance measure (i.e. performing well for the last $t+1$ quarters rather than t does not change the strength with which markets respond to central bank communication), which aligns with our finding of recency bias.

Finally, we embed these findings into an otherwise standard three-equation New Keynesian model to show that accounting for endogenous credibility and recency bias can lead to significant persistence in inflation dynamics.

The rest of the paper is organized as follows: Section 2 introduces a simple Bayesian updating framework. Section 3 provides details of our experimental design, Section 4 contains our *Forecast Performance* results, and Section 5 our *Timing* results. Section 6 provides some corroborating evidence using observational data and explores the theoretical implications of our *Forecast Performance* and *Timing* results. Finally, Section 7 explores the role that communication can play in attenuating the effect of poor forecast performance, and Section 8 concludes.

2 Central bank signals and forecast updating

In this section, we introduce a simple Bayesian framework (similar to Morris and Shin (2002), for example) to emphasize how central bank signals should influence a Bayesian participant's decision to update her inflation forecasts. This framework, which guides our experimental design, illuminates how an individual should react to inflation forecasts contingent upon his or her outlook on the central bank's forecast credibility. Using this framework, we can further elucidate when and how features of the central bank's forecasting history ought to influence their perceptions of the central bank's forecast credibility. Crucially, this framework also provides a precise measure of forecast credibility in the context of our experiment.

We begin a decision period by revealing to participant i the central bank's forecast of inflation alongside realized inflation for the twelve most recent quarters. Based on this economic history, participant i forms a belief about the value of inflation for the following quarter along with a belief about her own forecast precision:

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

where $\bar{\pi}_i$ is i's initial point forecast and α is a measure of i's forecast precision, which we collect in our experiment via an incentivized range forecast of inflation.

After the participant forms her initial forecasts, the central bank provides the participant with its own forecast of inflation for the following quarter:

$$\pi_{cb} = \pi + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim \mathcal{N}\left(\gamma, \frac{1}{\beta}\right). \tag{2}$$

where β is related to the precision of the central bank forecast, which i can infer from the 12-quarter economic history, and γ represents a possible systematic bias in the central bank's inflation forecast. We show in Section A2 that our results are qualitatively robust to assuming, but for now, assume that the central bank's forecast errors are unbiased as given by the case of $\gamma = 0$. The optimal Bayesian inflation forecast is a precision-weighted, linear combination of the prior, $\bar{\pi}_i$, and the central bank's signal, π_{cb}:

$$\mathbb{E}(\pi|\pi_{cb}) = \frac{\alpha\bar{\pi}_i + \beta\pi_{cb}}{\alpha + \beta} \tag{3}$$

The optimal update, therefore, is:

$$\mathbb{E}(\pi|\pi_{cb}) - \bar{\pi}_i = \frac{\beta}{\alpha + \beta}(\pi_{cb} - \bar{\pi}_i) \tag{4}$$

Rewriting this in terms of an optimal update rate, we define:

$$u_i^* \equiv \frac{\mathbb{E}(\pi|\pi_{cb}) - \bar{\pi}_i}{(\pi_{cb} - \bar{\pi}_i)} \tag{5}$$

Under Bayesian optimal updating, $u_i^* = \frac{\beta}{\alpha+\beta}$. If $\beta \to \infty$, $\alpha \to 0$, or both, the agent updates fully toward the central bank signal and this would give rise to $u_i^* = 1 = 100\%$. In our experiment, we use w participant's initial range forecast as an incentivized measure of α_i^{-1}. This means the more uncertain the participant, the smaller is α_i and, according to Equation (3), the more credibly they perceive the central bank for a given β^{-1}

Figure 1 plots this optimal update rate (in percentage terms, $100 \times u_i^*$) for different levels of β and α. There are three main implications:

1. For any given precision of the central bank signal, as the precision of the prior increases, $\alpha \uparrow$, the agent updates less when they receive the central bank signal.

2. For a given prior precision, as the precision of the central bank signal decreases, $\beta \uparrow$, the agent updates less when they receive the central bank signal.

3. The marginal effect of decreasing precision of the central bank signal is larger when the individual's prior is *more* precise.

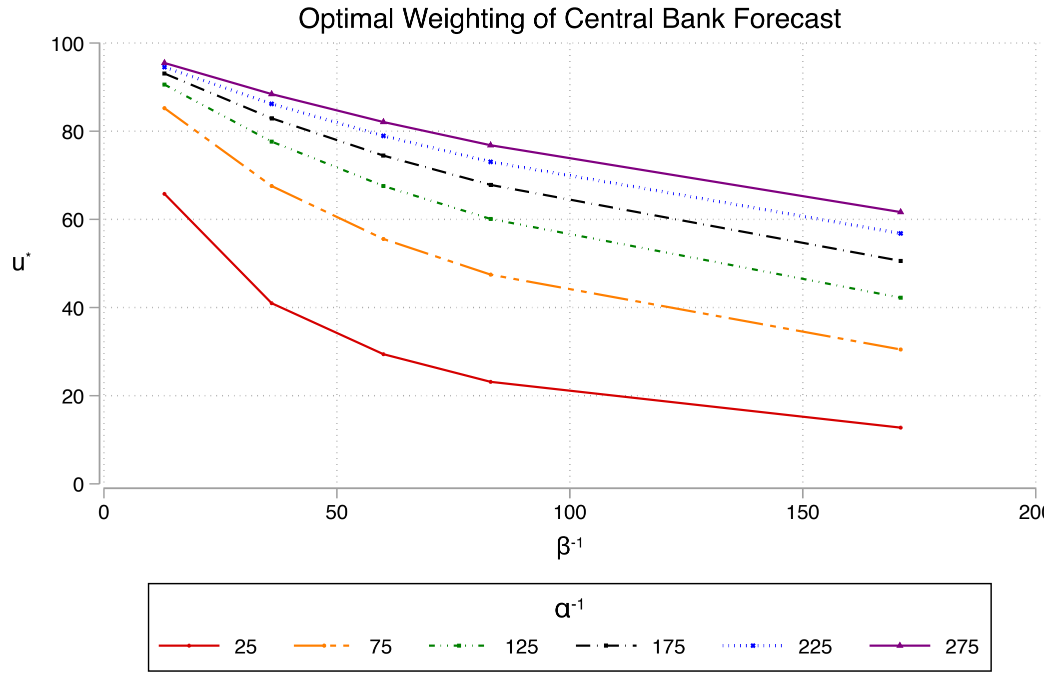


Figure 1: This figure shows the optimal level of updating in percentage terms (y-axis) prescribed by Equation (5) for different levels of a central bank precision (x-axis). Each line denotes a different level of participant forecast uncertainty ranging from 25 basis points (bottom line) to 275 basis points (top line) in increments of 50 basis points.

3 Experimental Design

Our goal in designing this experiment is to isolate the causal relationship between various features of the central bank's forecasting history and its ability to influence inflation expectations. Our interest is in how the historical economic information we reveal to participants influences their perceptions of the central bank's forecast credibility (β^{-1}) by observing how they incorporate (or not) the central bank's forecast into their own inflation outlook (i.e. by observing u^*). As the experimenters, we directly control β^{-1} and π_{cb}, which we reveal to subjects via economic histories and by announcing the central bank's forecast. This means we need to collect from participants incentivized measures of $\bar{(}\pi), \alpha^{-1}$, and $\mathbb{E}(\pi|\pi_{cb})$. Given these values, we obtain an incentivized measure of u^*.

To do this, we introduce an individual-choice experiment where participants 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 (described in Section 3.2). Each set of forecasts comprises an incentivized point forecast of inflation coupled with an incentivized measure of forecast uncertainty. We next provide a detailed description of our experimental model.

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 in various public

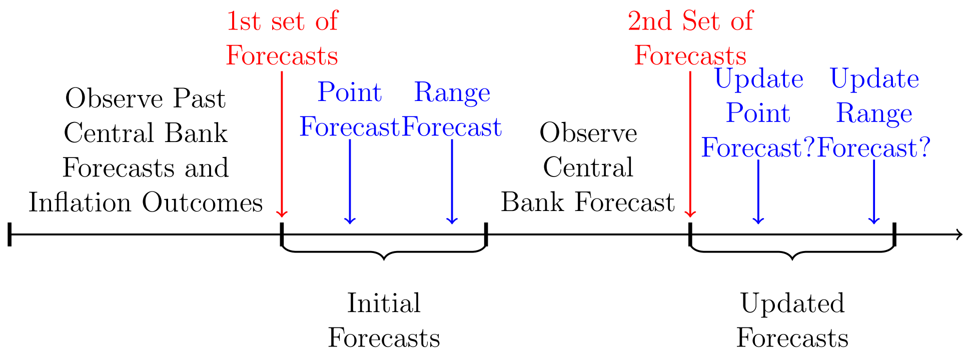


Figure 2: Experimental Timeline: A single decision period

Note: This figure XXXX.

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 when forming forecasts, how to interact with the available information, how to interact with our 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 comprising 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 subjects who submitted the quiz more than twice with at least one wrong answer. Subjects who successfully completed the quiz proceeded to the forecasting task.[4]

In the forecasting task, subjects complete three separate decision periods. Each decision period requires subjects to make an Initial Forecast and an Updated Forecast. This means that our experiment yields a total of six sets of forecasts, with each set consisting of both Point and Range forecasts. Subjects are told that their bonus payment would be based on their performance in one of these randomly selected sets of forecasts.

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

We programmed our experiment in oTree (Chen et al. 2016). We conducted our experiment online and recruited participants via Prolific, restricting our subject sample to experienced Prolific users from the United States.

3.2 Decision Periods

Figure 2 presents the experimental timeline within a decision period. 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

[4]We provide questions from the economic literacy quiz in Section A5.2, our experimental instructions in Section A5, and questions from our comprehension quiz in Section A5.2.

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 two-decimal precision. We incentivized point forecasts according to Equation (6), which follows the previous LTF literature (Rholes and Petersen 2021, Mokhtarzadeh and Petersen 2021, Petersen and Rholes 2022):

$$F_{i,t} = 2^{-|\mathbb{E}_{i,t-1}\{\pi_t\}-\pi_t|}. \tag{6}$$

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 forecasts 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 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 change 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 (7), which follows Pfajfar and Žakelj (2016), Rholes and Petersen (2021), Petersen and Rholes (2022).

$$U_{i,t}(r_{i,t}) = \begin{cases} 0 & \pi_{i,13} \notin [\underline{u_{i,t}}, \overline{u_{i,t}}] \\ \phi\left(\frac{1}{r_{i,t}}\right) & \pi_{i,13} \in [\underline{u_{i,t}}, \overline{u_{i,t}}]. \end{cases} \tag{7}$$

Here ϕ is a scalar we can adjust to scale average earnings, where average earnings are strictly increasing in ϕ. We set $\phi = 1$ for out experiment. $\underline{u_{i,t}}$ 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 (i.e. $\mathbb{E}^{CB}_{i,12}\pi_{13}$) 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 (π_{13}) alongside their forecasting performance for that decision period.

After participants have completed their three decision periods and provided their six sets of forecasts, the participant is informed which of the six forecasts has been selected as the basis for the bonus payment.

3.3 Creating the economic histories

Differences in the economic histories shown to subjects constitute treatment variation in our experimental framework. To create these histories, we simulate the simple 3-equation New Keynesian model in Walsh (2017) linearized around a zero-inflation steady-state, described by Equation (8) through Equation (14). 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.

$$y_t = E_t y_{t+1} - \sigma^{-1}(i_t - \mathbb{E}_t\pi_{t+1}) + g_t \tag{8}$$
$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \kappa y_t + u_t \tag{9}$$
$$i_t = \phi_x y_t + \phi_\pi \pi_t + v_t \tag{10}$$
$$r_t = i_t - \mathbb{E}_t\pi_{t+1} \tag{11}$$
$$g_{t+1} = \rho_g g_t + \epsilon^g_{t+1} \tag{12}$$
$$u_{t+1} = \rho_u u_t + \epsilon^u_{t+1} \tag{13}$$
$$v_{t+1} = \rho_v v_t + \epsilon^v_{t+1} \tag{14}$$

We assume the central bank in our simulated economy forms rational expectations so that the uncorrelated stochastic components of period-specific shocks (Equations (12), (13), and (14)) 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 $E_t\psi_{t+1} = \rho_{\psi,t}\psi_t$. We calibrate

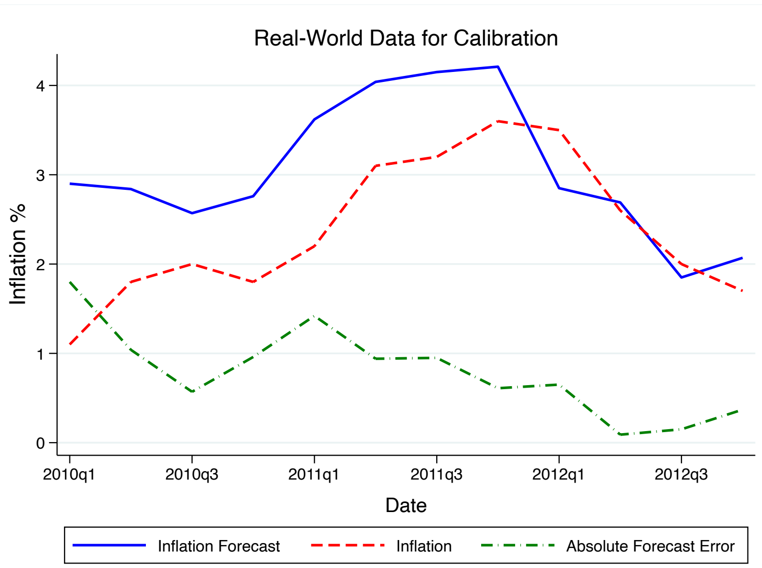


Figure 3: Data used for calibrating economic histories

Note: This figure depicts inflation from the United Kindom (red dashed line) alongside the Bank of England's lagged one-year-ahead inflation forecast (solid blue line) and the corresponding forecast error (green dashed-dotted line) .

this model using parameters in Table 1 and the inflation gap is then converted to inflation data by assuming a target rate of 2%. We created inflation forecasts and inflation values for the forecast quarter ($\mathbb{E}^{CB}_{i,12}\pi_{13}$ and π_{13}) in each economic history using shocks that roughly preserved the average forecast error of the final year of economic history.

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

Table 1: Parameter values for simulation exercise

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 3). To calibrate our model, we choose model shocks that qualitatively preserved the observed pattern of central bank forecast errors $\delta^{history}_{\pi,t} = E^{history}_{t-1}(\pi_t) - \pi_t$.

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.

During this period, we see that the BoE initially made relatively large forecast errors (in 2010 the annual average absolute forecast error was 110bps), but gradually improved such that the forecast errors in 2012 were around one-third as large (34bps). This motivates our core set of three histories which we refer to as *Early*, *Late*, and *Consistent*.

- For *Consistent*, the central bank exhibits a consistent average annual forecast performance. The key characteristic of *Consistent* is that each of the annual (4-quarter) average absolute forecast errors is the same as the full sample average absolute forecast error.
  - To isolate the causal relationship between historical forecast precision and forecast credibility (i.e. *ForecastPerformance*), we produce five versions of *Consistent* that preserve to time profile of forecast errors but vary the central bank's historical forecast precision. Precision in these alternative versions of *Consistent* vary from *Consistent-Great* performance, through *Consistent-Good*, *Consistent-Moderate*, *Consistent-Bad* and down to *Consistent-Terrible*.
  - We first generate a version of *Consistent-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 *Consistent-Great* (*Consistent-Terrible*) to exactly match the average absolute forecast error in the final year of *Consistent-Early* (*Consistent-Late*). Finally, we chose absolute error values for *Consistent-Good* and *Consistent-Moderate* so that they partitioned the performance difference between *Consistent-Great* and *Consistent-Bad*.
- 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.
- In *Late*, we reverse the pattern of forecast errors observed in *Early*, exactly preserving the absolute average forecast error between *Early* and *Late*. This means that both the magnitude and speed by which historical forecast precision changes are identical across these two histories. All that varies is whether the central bank has recently experienced a spate of poor or great forecast performance.

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 Table 3.

All participants completed three independent decision periods consisting of *Early*, *Late*, and some version of *Consistent*. In *Forecast Performance*, subjects first see some ordering of *Early* and *Late*, and then see one of the five possible versions of *Consistent*.[5] In (*Timing*), a participant experienced some ordering of *Early*, *Late*, and *Consistent-Bad*. Subjects in this wave of treatments. In *Contextual Communication*, participants see *Early* and *Consistent-Bad*. In their third decision period, these participants see *Late*, where we augment the central bank's forecast with additional written communication. In the following sections we cover the *Forecast Performance*, *Timing*, and *Contextual*

[5] We show these in Figure A-16.

	Numerical Summary of Economic Histories (bps)				
	Year 1	**Year 2**	**Year 3**	**Full Sample**	$\gamma_{HistAvg}$
Calibration Data	110	95	34	80	
			Forecast Performance		
Consistent - Great	13	13	13	13	06
Consistent - Good	36	36	36	36	10
Consistent - Moderate	60	60	60	60	06
Consistent - Bad	83	83	83	83	02
Consistent - Terrible	171	171	171	171	-06
			Timing & Contextual Communication		
Consistent - Bad	83	83	83	83	02
Early	171	65	13	83	-51
Late	13	65	171	83	-52

Table 2: This table provides a numerical summary of our economic histories. Numbers are average absolute forecast errors expressed in basis points. The column labeled $\gamma_{HistAvg}$ lists of historical average forecast errors by economic history, which we explore in Section A2.

Communication treatment waves and, for each, provide additional details regarding treatments and experimental design, state our hypotheses, and detail our results.

4 *Forecast Performance*

In *Forecast Performance* we study how a central bank's historical forecast precision changes its perceived forecast credibility, which we measure as the willingness of participants to incorporate the central bank's inflation forecast into their own updated point forecast. As discussed in Section 3, we answer this question using alternative versions of our *Consistent* history vary only in the central bank's historical average absolute forecast errors (i.e. β^{-1}). Recall that we refer to these histories as *Consistent-Great*, *Consistent-Good*, *Consistent-Moderate*, *Consistent-Bad* and *Consistent-Terrible*. For brevity, we drop the *Consistent* prefix from treatment names for the remainder of this section.

More precisely, we randomize participants in this wave of treatment into one of ten different possible treatments, described by the rows of Table 3. Results are based on a between-subjects comparison of the average perceived forecast credibility across *Great* through *Terrible*. Note that sample sizes are relatively consistent across treatments, with the exception of *Bad* for which we draw on strictly comparable treatments that arose in the *Timing* waves.

4.1 Hypothesis 1

Equation (5) provides a clear hypothesis about the relationship between historical forecast performance and the central bank's forecast credibility, as measured by u^*. Using

	Treatment Summary: *Forecast Performance*			
	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: Forecast Performance

Note: This table describes our *Forecast Performance* treatments and provides samples sizes for each. Each subject participated in a single treatment, denoted by rows *T1a* through *T5b*. Within this treatment, a subject experienced three histories, the order of which is outlined by the columns denoted *History 1* through *History 3*. *History 3* always presented some version of our *Consistent* forecasting history, which we detail in Table 2.

the inverse of a history's sample-average absolute forecasting error as a proxy for precision, we have the following:

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

To test this hypothesis, we average over individual-level estimates of perceived forecast credibility, u_i^*, to produce estimated average treatment effects (blue dots), which we compare to the treatment-average Bayesian optimal benchmarks (red triangles) in Figure 4. Here, the treatment average optimal response is given as $u_T^* = \frac{1}{N_T} \sum_{n \in N_T} \dfrac{\beta_T}{\alpha_n + \beta_T}$, where β_T denotes historical forecast precision in the relevant treatment. Finally, we note that we Winsorize the top and bottom 5% of our data to mitigate the impact of outliers on our results. Unless noted otherwise, we do this for all results in our text.[6] Additionally, Table 4 presents results from a series of OLS regressions that capture all pairwise treatment-level comparisons of forecast credibility across *Forecast Performance* treatments.

Estimates indicate that the central bank's forecast credibility is negatively correlated with its historical forecast precision, β^{-1}, which constitutes support for hypothesis one – subjects respond to decreases in historical forecast precision like Bayesians. However, the empirical relationship we observe between forecast credibility and forecast precision is flatter than predicted by theory. In fact, estimates in Table 4 suggest there is no significant loss in credibility if a central bank's forecast deteriorates from *Good* to *Terrible*. In some sense, this is good news because the forecast credibility cost of large errors may not be quite as high as the theory predicts. However, we observe much stronger

[6] We provide a sensitivity analysis of these cut points in Section A3.

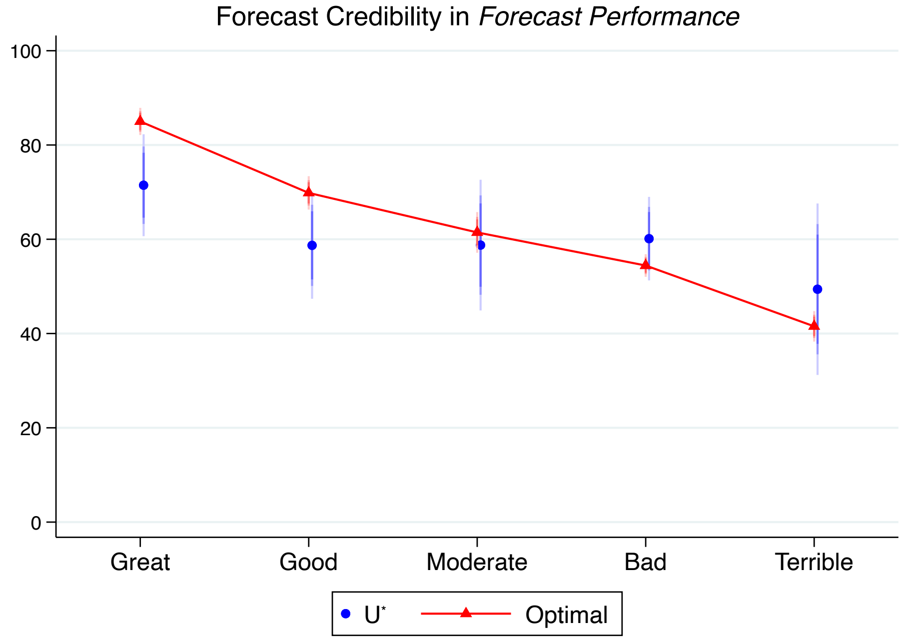


Figure 4: *Forecast Performance* treatment estimates.
Note: This figure presents estimates of central bank forecast credibility in *Forecast Performance* treatments. Blue and red shaded bands surrounding point estimates depict 99% (lightest), 95%, and 90% confidence (darkest) intervals.

evidence that subjects underuse signals from highly-precise central banks.[7] To some extent, a central bank may not fully reap the reward, in terms of forecast credibility, of high forecast precision.

We explore the robustness of these results via regression analysis wherein we project individual-level perceptions of central bank forecast credibility, u_i^*, onto a set of indicator variables denoting treatment (i.e. *Great*, *Good*, ..., *Terrible*), demographic characteristics, a subject's own uncertainty regarding future inflation, and controls for economic literacy. We depict the results of this exercise in Table 5. Estimates in column (1) of Table 5 correspond directly to the unconditional mean estimates of u^* depicted in Figure 4, column (2) additionally controls for forecast uncertainty, (3) layers in controls for demographics, and (4) includes controls based on our survey of economic literacy and for how much a participant trusts and understands the central bank. The main point of this table is that our baseline result – that subjects qualitatively behave like Bayesians – is robust.

This leaves us with an obvious question – why are subjects exhibiting a muted response to forecast precision relative to theoretical predictions? We try to answer this question when addressing our second *Forecast Performance* hypothesis.

[7] We show in Table A-1 that deviations from the equal-weighting Bayesian benchmark are statistically significant ($p < .01$) for *Great* and *Good* but not for *Bad* or *Terrible*.

Table 4: Comparing Credibility Across Treatments

	(1) *Great*	(2) *Good*	(3) *Moderate*	(4) Bad	(5) *Terrible*
Great	-				
	-				
Good	-12.79**	-			
	(6.020)	-			
Moderate	-12.76*	0.0250	-		
	(6.854)	(6.952)	-		
Bad	-11.45**	1.336	1.311	-	
	(5.399)	(5.523)	(6.421)	-	
Terrible	-23.84***	-11.06	-11.08	-12.39	-
	(8.123)	(8.206)	(8.836)	(7.762)	-
Control	71.16***	58.37***	58.39***	59.70***	47.31***
	(4.177)	(4.336)	(5.434)	(3.421)	(6.967)
N	528	528	528	528	528

Note: This table presents the results of a series of OLS regressions capturing the pairwise differences in forecast credibility across *Forecast Performance* treatments. Dashes along the diagonal indicate the omitted treatment in each regression. We report the mean and standard error of u^* for treatment listed in a column header each in the row labeled 'Control'. For an example of how to interpret this table, column (1) indicates that forecast credibility is significantly lower in *Moderate* than *Great* ($-12.76,\ p < .1$) but column (2) indicates that forecast credibility in *Moderate* is not statistically different than in *Good* ($.025,\ p > .1$). * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5: Regression Table for *Forecast Performance*

	(1) u^*	(2) u^*	(3) u^*	(4) u^*
Great	71.16***	71.13***	77.89***	69.74***
	(4.177)	(5.088)	(9.014)	(17.20)
Good	58.37***	58.34***	62.95***	55.18***
	(4.336)	(4.847)	(8.327)	(16.97)
Moderate	58.39***	58.36***	65.13***	58.35***
	(5.434)	(6.114)	(9.404)	(17.57)
Bad	59.70***	59.67***	65.03***	59.16***
	(3.421)	(4.288)	(7.297)	(16.05)
Terrible	47.31***	47.28***	53.49***	46.61***
	(6.967)	(7.665)	(11.06)	(17.62)
Uncertainty		0.000275	0.00126	0.00414
		(0.0241)	(0.0245)	(0.0260)
Demographics			✓	✓
Survey Responses				✓
N	528	528	520	520

Note: This table presents OLS regressions. Robust standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

4.2 Hypothesis 2

Equation (5) also elucidates that a central bank's forecast credibility doesn't depend entirely on things it can control. Instead, forecast credibility depends both on the central bank's forecast precision and a participant's belief about her own forecasting credibility. This leads us to our second hypothesis:

Hypothesis 2. *For a given economic history, the central bank's forecast credibility increases in a participant's forecast uncertainty.*

Intuitively, this hypothesis says that a participant who exhibits more forecast uncertainty in the Initial Forecast will update more toward the central bank's forecast. Put differently, participants who are highly uncertain of their own forecasts should be more forgiving of historical forecast errors than subjects who are more certain about their forecast.

We turn again to Table 5 and consider the coefficient estimates in the ForecastUncertainty row, which estimate the relationship, on average, between forecast uncertainty α^{-1} and perceived forecast credibility u^*. Note that, regardless of specification, participants exhibit no significant response to forecast uncertainty when updating their inflation forecasts. This aligns with Figure 5, which also clearly shows that there is no relationship between forecast uncertainty and perceived central bank credibility. This result violates the logic of Equation (5) and suggests that, for a given forecast history, people would respond to new signals from the central bank in the same way regardless of their own uncertainty about future economic conditions.

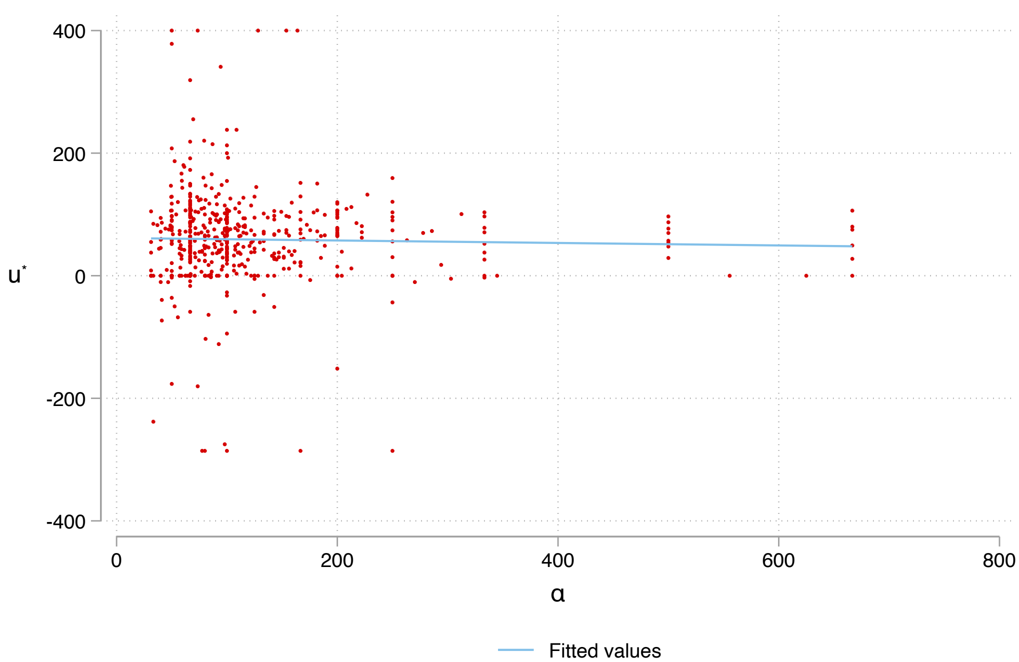


Figure 5: Individual-level forecast uncertainty and forecast update weight.
Note: This figure presents a scatter plot of individual-level forecast uncertainty and the forecast update weight as a measure of perceived forecast credibility of the central bank.

This is surprising. Intuitively, a signal that conveys some clarifying information ought to be more valuable in instances of higher confusion, which is what Equation (5) says

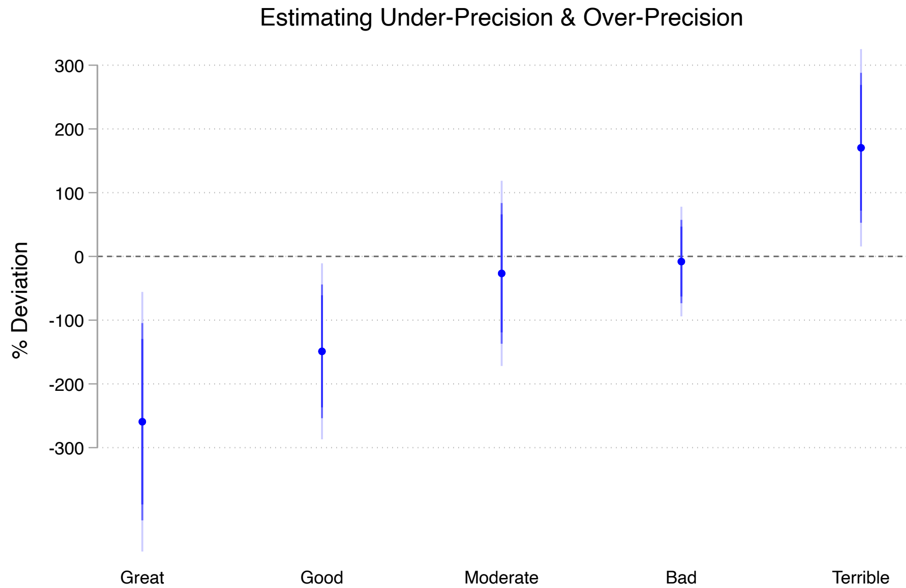


Figure 6: Implied undervaluation (< 0) or overvaluation (> 0) of precision.
Note: This figure presents the percentage by which the average participants undervalues (< 0) or overvalues (> 0) her own precision when incorporating the central bank's forecast into his or her updated forecast. We calculate this measure as $\frac{1}{N}\sum_{j=1}^{j=N}\frac{\alpha_{j,implied}-\alpha_j}{\alpha_{j,implied}}$. Point estimates (blue circles) are surrounded by 99%, 95%, and 90% confidence intervals (blue shading).

– an uncertain agent should more highly value new signals that help her better predict the evolution of important aggregates than a 'certain' agent who thinks she has a good grasp on how those aggregates will evolve.

Suppose subjects correctly infer β. We can quantify the extent to which participants' incorrect perceptions of their own uncertainty distort updating away from the Bayesian optimal benchmark. Recall that $u_i^* = \frac{\beta_T}{\alpha_i+\beta_T}$, where β_T reflects the central bank's true forecast precision for a given treatment. Suppose a participant under-weights the central bank's forecast relative to the Bayesian benchmark when updating her point forecast of inflation. Because we assume that the participant correctly perceives β^{-1}, this implies that α_i is too large in her updating function. Put differently, the forecast uncertainty , α_i^{-1}, implied by her updated forecast is too small. This would yield $\frac{u_i^*}{\beta_T(1-u_i^*)} - \alpha_i^{-1} < 0$ where we treat $\frac{u_i^*}{\beta_T(1-u_i^*)}$ as the implied forecast uncertainty. Intuitively, this says that the participant's implied uncertainty is smaller than the incentivized measure of forecast uncertainty she provided in her initial forecast.

We show the results of this exercise in Figure 6, which suggest that at least some of the sub-optimal behavior we observe in Figure 4 is driven by participants incorrectly accounting for their own forecast uncertainty when forming a perception of the central bank's forecast credibility.[8]

Our finding relates to the broad literature on overprecision, which is an idiosyncratic bias that leads Bayesian agents to treat private information as overly precise (Moore and Healy 2008, Moore and Schatz 2017). This is akin to underreacting to own forecast

[8]Note that results in the two graphs do not perfectly align since Figure 6 necessarily omits participants for whom $u_i^* = 0$, which is not true for results in Figure 4.

uncertainty in our experiment, which is what we observe in our treatment where the central bank's historical forecast performance is best.

5 *Timing*

A possible implication of our *Forecast Performance* results is that the central bank is afforded some leniency when it makes worse forecast errors. This is important from a policy perspective because it informs policymakers about the efficacy of central bank signals following inflation misses. However, these results may depend, quite critically, on the consistency of forecast errors throughout the *Consistent* economic histories.

Thus, we next assess the extent to which the timing of forecast errors influences participants' perceptions of central bank forecast credibility using our *Timing* treatments. These treatments comprise all possible permutations of our three core economic histories, *Early*, *Late*, and *Consistent - Bad*. Recall that these economic histories feature identical historical forecast precision but allow for variation in the pattern of historical forecast errors. By comparing our measure of forecast credibility across histories, we learn whether and how the time profile of forecast errors impacts the central banks ability to influence inflation expectations via forecasting.

5.1 Hypotheses 3 & 4

Given that the presented sample history is only 12-quarters long, we might expect subjects to use the full history to estimate the central bank's precision. If subjects equally weight all historical information when doing this, they estimate

$$\beta^{-1} = \frac{\sum_{j=1}^{j=12} \left|\mathbb{E}_{j-1}^{CB}\left(\pi_j\right) - \pi_j\right|}{12}. \tag{15}$$

If Equation (15) is correct, then the average level of perceived forecast credibility across *Early, Late*, and *Consistent* ought to be identical, assuming average forecast uncertainty is constant across histories. This is because participants would discern, on average, an identical level of historical forecast precision across these histories. We summarise this into the following hypothesis:

Hypothesis 3. *Subjects weigh observed histories equally such that the timing of forecast errors does not lead to a difference in the average level of forecast credibility across economic histories in Timing.*

Though averaging across all three available years of historical performance seems like the natural thing to do (we would fail to reject the null of Hypothesis 3), the results of the previous section suggest that the participants place different weights on very large and very small errors. This does not necessarily translate into timing effects however; if they underweight (overweight) large (small) errors, but the timing does not matter in and of itself, then we might expect that the effects net out over *Early* and *Late* such that $u^*_{Early} = u^*_{Late} = u^*_{Consistent}$.

There is, however, literature that suggests people exhibit time-dependency in economic decision-making in related contexts. Malmendier and Nagel (2016) show that people born at different times s and $s+j,\ j>0$, can weight information at $t>s+j$ differently due to differences in life experiences. Thakral and Tô (2021) show that expectations-based reference points adjust dynamically and exhibit recency bias. If this holds in the context of forecast credibility, then we might observe significant differences in forecast credibility across these economic histories.

For instance, if the economic agent views the central bank's forecast credibility as ever-changing and accounts for this by more heavily weighting recent performance, then they might calculate β as:

$$\beta^{-1} = \lambda \sum_{j=0}^{j=11} (1-\lambda)^j |\mathbb{E}_{t-2-j}^{CB}(\pi_{t-1-j}) - \pi_{t-1-j}| \tag{16}$$

where the weighting function exhibits exponential decay in time. Figure 7 depicts the implied weighting functions from Equation (16) for different values of λ. This is akin to constant-gain learning models of expectation formation common in the learning literature (Evans et al. 2001).[9]This gives us our second *Timing* hypothesis:

Hypothesis 4. *Subjects exhibit recency bias when forming a perception of central bank credibility.*

5.2 Timing Treatments and Results

We test these hypotheses using a within-subject design that exposes each participant to some ordering of *Early, Late,* and *Consistent - Bad*. Because we use a within-subjects design, we implement a full factorial design to nullify concerns about order and learning effects as potential confounds. This yields the *Timing* treatments described in Table 6. Note that, because we only use *Consistent-Bad* in this wave, we will refer to *Consistent - Bad* as *Consistent* throughout the remainder of the *Timing* section.

Table 6: Treatment Summary: *Timing*

Treatment Summary: Timing				
	History 1	**History 2**	**History 3**	**Sample Size**
T6	*Early*	*Late*	*Consistent*	97
T7	*Early*	*Consistent*	*Late*	94
T8	*Late*	*Early*	*Consistent*	76
T9	*Late*	*Consistent*	*Early*	88
T10	*Consistent*	*Late*	*Early*	91
T11	*Consistent*	*Early*	*Late*	79

Note: This table summarizes our *Timing* treatments. Note that T6 and T9 are the same as T4a and T4b in Table 3.

[9]In that context, economists typically motivate these models as a way for an agent to account for structural change in whatever macroeconomic time series an agent is forecasting.

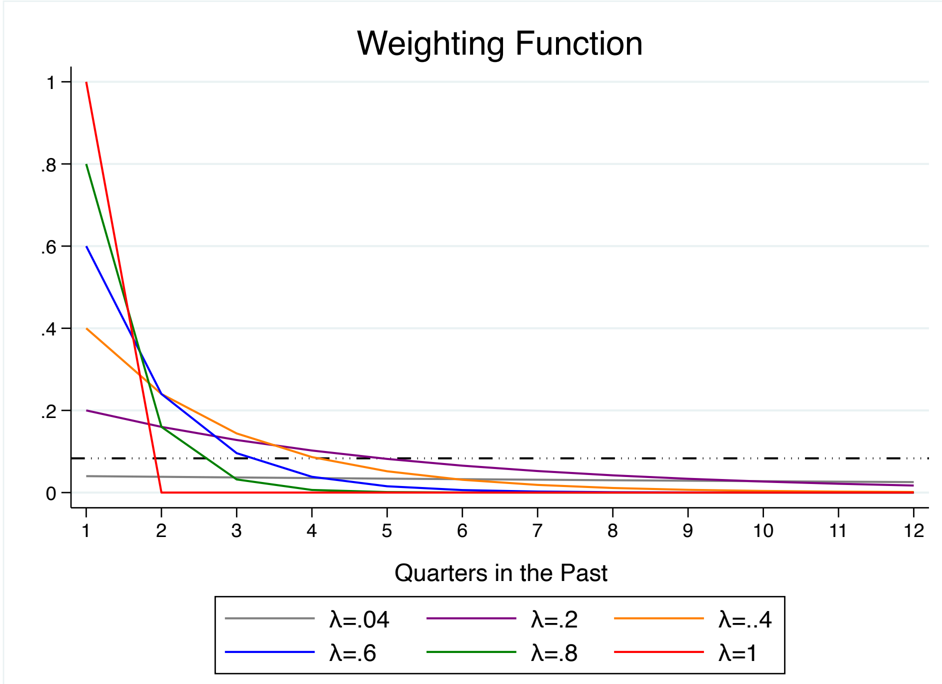


Figure 7: Weighting functions (Equation (16)) with different values of λ

Note: This figure depicts hypothetical weighting functions for alternative values of λ in Equation (16). Note that $\lambda = 1$ means a subject cares only about the most recent historical observation. As λ decreases, subjects begin to more equally consider the central bank's full forecasting history. The black horizontal line is the benchmark case, described by Equation (15), where a participant equally weights all twelve quarters of historical forecast information in our experiment.

Figure 8 reports measures of perceived forecast credibility by economic history (blue circles) alongside the Bayesian optimal level of updating (assuming equal weighting of all historical information, red triangles) and the deviation from this Bayesian benchmark (green diamonds). Shaded bands around each market denote 99%, 95%, and 90% confidence intervals.

Table 7: Results of t-tests and Descriptive Statistics

	Early		*Consistent*		*Late*	
	Mean	SE	Mean	SE	Mean	SE
Early	64.375	3.523				
Consistent	$p \approx 0.163$		58.615	2.142		
Late	$p < .001$		$p < .001$		11.494	2.468

Note: This table provides both the mean and standard error (SE) of u^* for each of the three histories used in our *Timing* treatments along the diagonal. Additionally, the table provides p-values resulting from a series of two-sample, two-sided t-tests, where a p-value at the intersection of two variables compares u^* for those two histories. For example, forecast credibility (u^*) is not significantly different in *Early* and *Consistent* ($p \approx 0.163$).

Participants clearly exhibit recency bias when forming perceptions of central bank forecast credibility in *Early* and *Late*. This is evidenced by the highly-significant deviations from the Bayesian optimal benchmark in both economic histories, where deviations are positive in *Early* and negative in *Late*. This matches with subjects placing more emphasis on more recent information in both economic histories. Taken together, this indicates

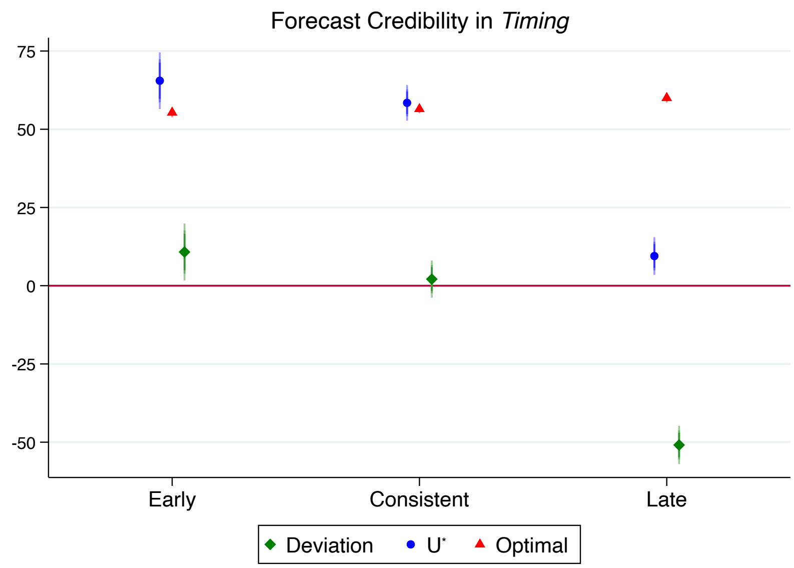


Figure 8: Estimated forecast credibility by economic history in *Timing*
Note: This figure depicts average treatment effects (blue circles) surrounded by confidence intervals ranging from 90 (darkest, thickest lines) to 99 percent (lightest, thinnest lines) from our *Timing* treatments using data Windsorized at the 5% level. Red triangles reflect the Bayesian benchmark weight, while green triangles show participants' average deviations from this benchmark, surrounded by confidence intervals ranging from 90 to 99 percent.

that participants respond strongly to the timing of forecast errors when forming a perception of the central bank's forecast credibility. Thus, we reject the null of Hypothesis 3 and instead, find support for Hypothesis 4.

However, we also observe that the magnitude of the deviation from the Bayesian benchmark is about five times larger in *Late* than it is in *Early*. Recall that we design *Early* and *Late* so that forecasting performance changes at the same speed and by the same magnitude across these two economic histories. Thus, the only difference between them is whether the central bank's most recent forecast performance is better or worse than its historical average forecast performance. This sizable difference in deviations away from the Bayesian benchmark, which is based on equally weighting all available information, suggests an asymmetry in recency bias across the two histories. If so, then it isn't just the *change* in forecast performance that drives recency bias. Instead, it could be that poor forecast performance is more salient for subjects whenever forming a perception of forecast credibility.

We also consider a within-subject measure of perceived forecast credibility, which we report in Figure 9. These measures present forecast credibility in *Early* and *Late* relative to *Consistent*. Specifically, for $X \in \{Early,\ Late\}$:

$$u^*_{X,within} = \frac{1}{N_X} \sum_{n \in N} \left(u^*_{n,X} - u^*_{n,C} \right) \tag{17}$$

This approach acts as a sort of participant-level fixed effect, assuming idiosyncratic biases are invariant to forecast history (for example, a participant's risk preferences are equivalent across the three economic histories). We observe the same pattern of relative updating between *Early* and *Late* in our within-subject measure of perceived forecast

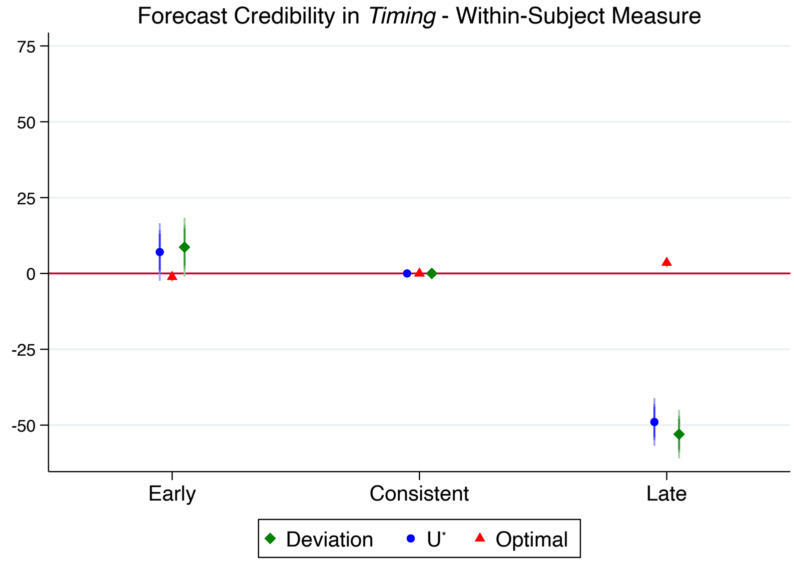


Figure 9: Within-Subject estimated forecast credibility by economic history in *Timing*
Note: This figure presents the average within-subject treatment effects on forecast credibility from the *Timing* treatments, using the 'Consistent' history as a control for idiosyncratic biases. Blue circles illustrate the average treatment effects, bounded by 90 to 99 percent confidence intervals using data Windsorized at the 5% level. Red triangles reflect the Bayesian benchmark weight, while green triangles show participants' average deviations from this benchmark, surrounded by confidence intervals ranging from 90 to 99 percent.

credibility as we do in Figure 8.

As we did in *ForecastPerformance*, we further explore *Timing* results via regression analysis to understand how perceptions of central bank forecast credibility relate to individual characteristics, to forecast uncertainty, and to our measures of economic literacy. Our approach here is identical to that in *Forecast Performance*. We show these results in Table 8, where (1) corresponds directly to the treatment effects depicted in Figure 8, (2) introduces forecast uncertainty as a control, (3) includes demographic controls, and (4) includes controls for economic literacy and how much participants trust and understand the central bank. We note that our baseline estimates change very little in terms of magnitude or statistical significance when we introduce controls.

5.3 Exploring the Extent of Recency Bias

Our *Timing* results show that participants exhibit recency bias, more heavily weighting more recent information when forming a perception of the central bank's forecast precision. Further, we observe a large difference in the magnitude of deviations from the Bayesian benchmark between *Early* and *Late*. This leads us to ask: how strong is this recency bias that we observe and how is it different across economic histories? To answer this, we estimate λ in Equation (16) for these two economic histories.[10]

[10]Combining $\beta^{-1} = (\alpha \times u^*)^{-1}(1 - u^*)$ with Equation (16), gives:

$$\lambda \sum_{j=0}^{j=11} (1-\lambda)^j |\mathbb{E}^{CB}_{t-2-j}(\pi_{t-1-j}) - \pi_{t-1-j}| - (\alpha \times u^*)^{-1}(1-u^*) = 0.$$

Table 8: Regression Table for *Timing*

	(1)	(2)	(3)	(4)
	u^*	u^*	u^*	u^*
Early	64.38***	56.77***	50.69***	62.34***
	(3.525)	(4.332)	(7.023)	(14.09)
Consistent	58.61***	50.97***	44.40***	56.05***
	(2.143)	(3.515)	(6.637)	(13.73)
Late	11.49***	2.192	-3.639	8.175
	(2.470)	(4.058)	(7.008)	(13.85)
Uncertainty		6.363**	5.908**	5.288**
		(2.523)	(2.547)	(2.515)
Demographics			✓	✓
Survey Responses				✓
N	1548	1548	1518	1518
Clusters	516	516	506	506

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

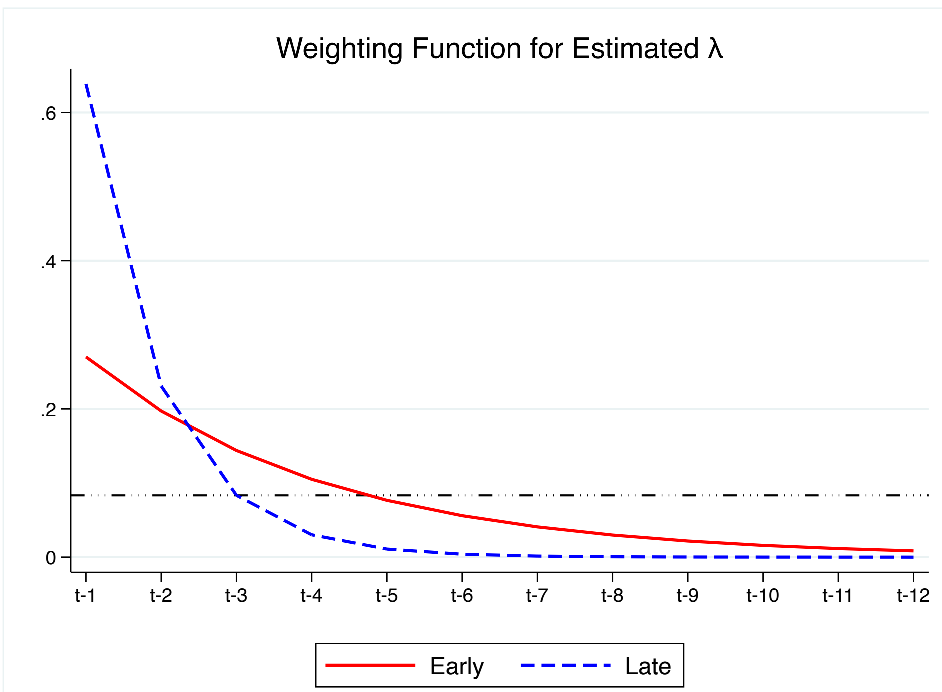


Figure 10: Estimated weighting functions

Note: This figure depicts weighting functions using values of λ that result from estimating λ in Equation (16) using our experimental data. This figure indicates that individuals exhibit significant recency bias and, on average, about threefold more recency bias in *Late* than in *Early*.

We provide results of this estimation exercise in Table 9. Note that participants exhibit recency bias in both *Early* and *Late*.[11] However, the degree of recency bias exhibited in

which we solve for λ via numerical approximation.

[11] We do not show estimates for *Consistent* since, by design, estimating β^{-1} using the last year of

Late is considerably higher than in *Early*. The estimated value of λ for *Late* ($\lambda_{Late} \approx .622$) implies that the average participant in *Timing* bases approximately 97.95% of her perception of central bank forecast precision on the most recent four quarters of economic data. By contrast, this same number for the average participant in *Early* ($\lambda_{Early} \approx .245$) is about 67.5%. This difference in λ_{Late} and λ_{Early} is highly significant ($p < .001$, t-test).

Using these estimates values of λ, we depict estimated weighting functions for *Early* (solid red line) and *Late* (blue dashed line) alongside an equal-weighting benchmark (black dashed line) in Figure 10.

So far, we have shown that the time profile of forecast errors can significantly change how participants use historical information when forming a perception of central bank forecast credibility. Further, we find a stark asymmetry in the extent of recency bias depending upon whether the central bank's most recent performance is significantly better or worse than its historical average performance. Finally, we find that the magnitude of the average deviation from our equal-weighting Bayesian benchmark is about five-fold larger in *Late* than in *Early*.

Our experimental findings in Timing underscore a crucial point: the central bank cannot afford to rest on its laurels in the face of unanticipated shocks that precipitate a sharp decrease in forecast performance, even if confined to short-run forecast performance. Further, such changes in forecast performance undermine the bank's forecast credibility precisely when the bank's need to control and guide inflation expectations is highest. This is so regardless of the bank's long-standing track record of sound predictions; public attention will inevitably gravitate towards the recent subpar performance, thus overshadowing prior successes. Additionally, our results reveal that while a rebound in forecast performance following a negative shock to forecast precision can aid in restoring credibility, the pace of recovery is invariably slower than the rate at which credibility initially eroded. It is paramount that central banks guard against and react swiftly to negative forecasting episodes to prevent credibility loss because the cost of rebuilding credibility may significantly outweigh the cost required to maintain it.

Table 9: Estimated Values of λ

Economic History	Estimated λ
λ_{Early}	0.245
	(0.0170)
λ_{Late}	0.622
	(0.0198)

Note: This table provides numerical measures of recency bias, obtained by estimating λ in Equation (16) using our experimental data. Standard errors in parentheses.

5.4 Dynamics of Perceived Credibility

Our experimental design also allows us to study the dynamics of perceived forecast credibility by comparing episodes where the central bank's historical forecast precision

information should be identical to estimating it over the last two years, all years, etc.

over the full three-year economic history exactly matches the forecast precision of the most recent year of economic history. We can do this for very bad forecast precision and very good forecast precision. We use the fact that the central bank's historical forecast precision in *Terrible* from *Forecast Performance* is identical to the central bank's forecast precision in the final year of *Late* from *Timing*. Similarly, we can also use the fact that the central bank's historical forecast precision in *Great* is identical to the bank's historical forecast precision in the final year of *Early*.

Our experimental design enables us to further study the dynamics of perceived forecast credibility by contrasting scenarios in which the central bank's historical forecast precision matches the forecast precision of the most recent year. This method applies to both terrible and great forecast precision scenarios. By leveraging the fact that the central bank's historical forecast precision from *Terrible* aligns with its forecast precision in the concluding year of the *Late*, we can do this for a low-credibility scenario. Similarly, we exploit the fact that the central bank's historical forecast precision from *Great* matches forecast precision in the final year *Early* to study a high-credibility scenario.

We first consider our low-credibility scenario. We depict the cumulative distribution functions of u^* in *Terrible* and *Late* in Figure 11, which provides means and their corresponding standard errors in both histories. There are two things to note. First, mean forecast credibility is higher in *Terrible* than in *Late* even though the central bank's historical forecast precision is higher in *Late* than *Terrible*. Second, cross-sectional disagreement regarding u^* is much higher in *Terrible* than *Late*. Together, this suggests that witnessing the decline in forecast precision in *Late* leads subjects to significantly discount the central bank's signal. Interestingly, the consistency of forecast errors in *Terrible* leads to much more disagreement about how to use the central bank's signal, with many subjects placing significant weight on the central bank's forecast when forming a posterior belief about inflation. This confusion matches with results in *Forecast Performance*, where we see that subjects overvalue the central bank's inflation forecasts in both *Bad* and *Terrible*.

Turning first to our low-credibility scenario, Figure 11 visualizes the kernel density functions of u^* in *Terrible* (blue line) and *Late* (red dashed line), along with their means and corresponding standard errors. We need to underscore two crucial observations here.

First, we highlight the counter-intuitive finding that *Terrible* yields a higher average forecast credibility than *Late*, despite historical forecast precision being much higher in *Late*. Second, we observe that *Terrible* induces a higher level of disagreement or variation in u^* than *Late*. These observations combined reveal that the visible decrease in forecast precision in *Late* leads participants to significantly undervalue the central bank's inflation forecast *systematically* whereas the consistent pattern of forecast errors in *Terrible* yields greater disagreement on how to use the central bank's signal. A considerable number of participants lean heavily on the central bank's forecast when constructing their posterior beliefs about inflation even though the central bank's historical precision is quite low. This aligns with our main result in *Forecast Precision*, where participants tend to overvalue the central bank's inflation forecasts in both *Bad* and *Terrible*.

We next consider our high-credibility scenario. Figure 12 visualizes the kernel density functions of u^* in *Great* (blue line) and *Early* (red dashed line), along with their means and corresponding standard errors. In some sense, the results here are the opposite of

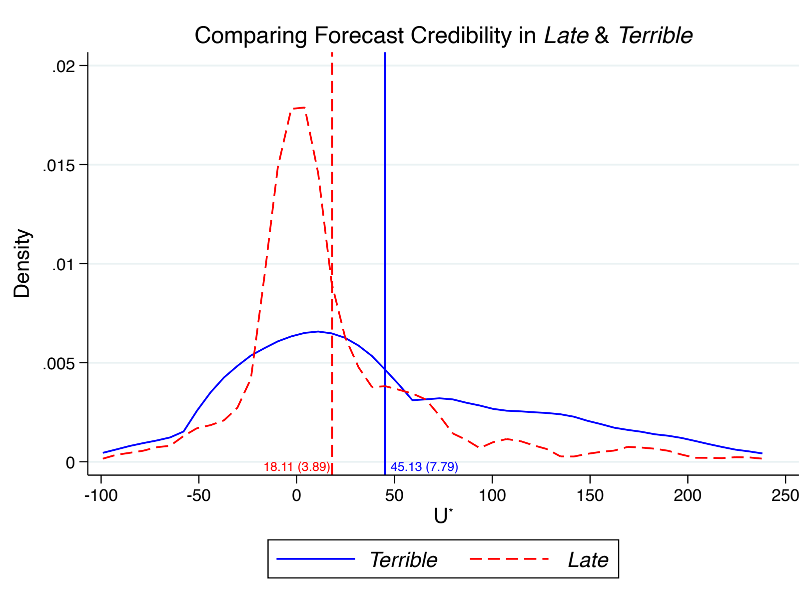


Figure 11: *Consistent-Terrible* vs. *Late*
Note: This figure depicts kernel density estimates of forecast credibility, u^*, from *Late* (red, dashed line) and *Terrible* (blue, solid line).

those in our low-credibility scenario. Consistently high forecast precision in *Great* leads to most participants viewing the central bank as highly credible. By contrast, observing that the central bank exhibited poor forecast precision in its recent history leads many subjects to view the bank as less credible. However, the shift from low to high forecast precision in *Early* induces the same sort of disagreement about how to use the central bank's signal observed in *Terrible*.

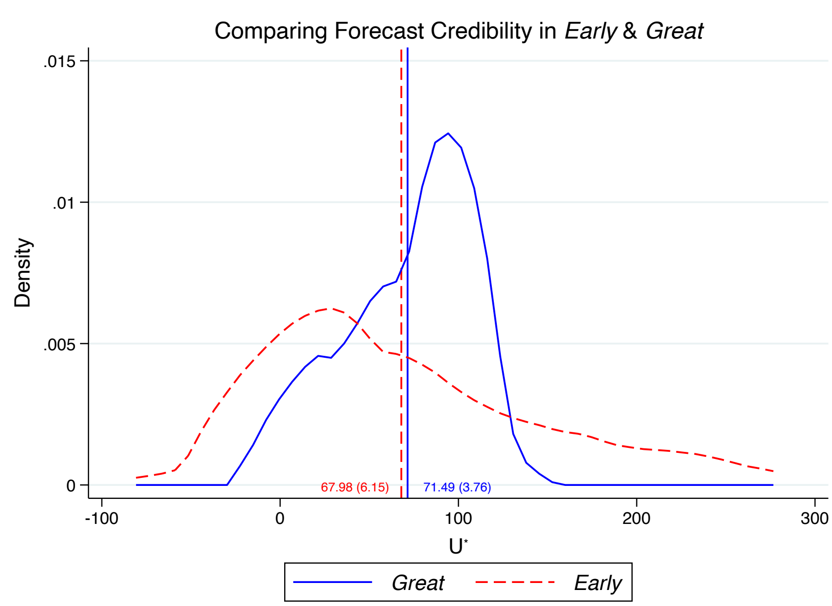


Figure 12: *Consistent-Great* vs. *Early*
Note: This figure depicts kernel density estimates of forecast credibility, u^*, from *Early* (red, dashed line) and *Great* (blue, solid line).

The overarching message from these findings is that both the consistency and the recentness of the central bank's forecast precision significantly influence participants' perceptions of the bank's forecast credibility. Further, these findings suggest that consistency, even in error, may instill a sense of forecast credibility among participants, which would explain why so many subjects in *Terrible* place considerable emphasis on the central bank's signal when forming updated inflation expectations. However, the shift from low to high forecast precision in "Early", akin to the "Terrible" scenario, introduced disagreements among the participants on how to interpret the bank's signal. In essence, these findings illuminate the nuanced role that consistency, recent performance, and changes in precision play in shaping the credibility of the central bank's forecast. This indicates that both past performance and recent changes matter significantly when it comes to trust in forecasts, and highlights the challenge for central banks in maintaining their credibility.

5.5 Does the direction of forecast error matter?

So far, we have only tested scenarios where the central bank's historical inflation forecasts were too low. A possible concern then is that our results regarding the relationship between the central bank's historical forecast performance and its ability to influence inflation expectations with new forecasts might vary depending on whether the bank over or underestimates inflation. For example, this could be true if people perceive the costs of inflation as asymmetric. An agent who views high inflation as more costly than low inflation might be less forgiving of under-forecasting inflation than over-forecasting inflation. This is because the former case leads to misspecified expectations that are violated by an even more costly scenario than anticipated whereas the latter to a case where expectations are violated but realized inflation is less costly than anticipated. This would align with Guido et al. (2022), who show that people more harshly punish disappointing violations of expectations than beneficial violations of expectations.

To address this concern, we implement a subset of our original treatments that are identical in every way except that we exactly reverse the direction of forecast errors. We refer to these treatments as our *Timing: Reversed Shock* treatments. We consider our baseline *Timing* results robust to the direction of forecast errors if we are able to qualitative our baseline *Timing* results.

Comparing these results to those from their original counterparts gives us some insight into whether our findings are robust to the direction of forecast errors. We shock results from these treatments in Figure 13.

Similar to baseline results, we find that the time profile of forecast errors matters. Our participants place significant emphasis on more recent information in our *Early* and *Late* treatments, as evidenced by the over-updating in *Early* and under-updating in *Late* relative to the Bayesian benchmark.

5.6 Does this hold for the medium term?

So far, we've considered how perceptions of the central bank's short-term forecast credibility respond to features of its historical short-term forecast performance. Another natural robustness check is whether and how our *Timing* results depend on the forecast horizon. Why? Policymakers are often more concerned with managing longer-term

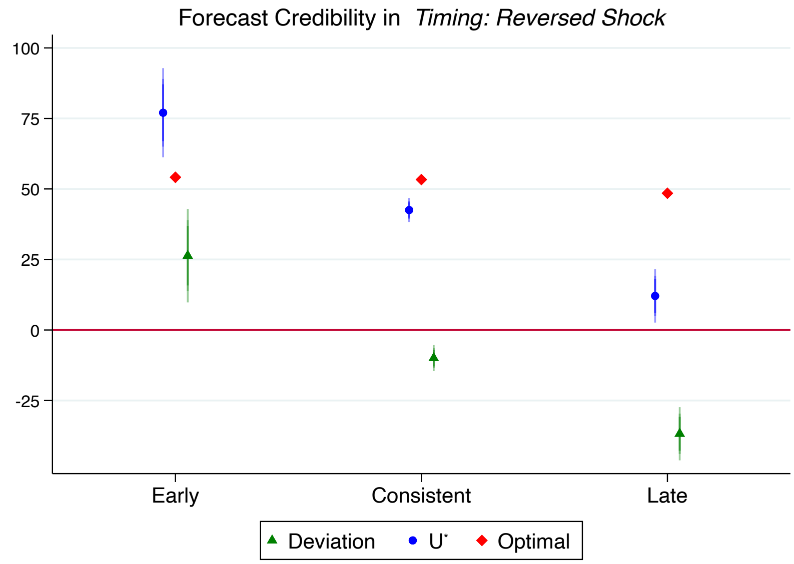


Figure 13: Estimated forecast credibility in *Timing: Reversed Shock* treatments.
Note: This figure depicts average treatment effects (blue circles) surrounded by confidence intervals ranging from 90 (darkest, thickest lines) to 99 percent (lightest, thinnest lines) from our *Timing: Reversed Shock* treatments using data Windsorized at the 5% level. Red triangles reflect the Bayesian benchmark weight, while green triangles show participants' average deviations from this benchmark, surrounded by confidence intervals ranging from 90 to 99 percent.

expectations than short-term expectations, which allows them to pursue short-term stabilization policies in response to transitory shocks. Therefore, we also consider how historical forecast performance impacts the central bank's ability to control longer-term inflation expectations.

To do this, we implement our *Early*, *Consistent*, and *Late* economic histories from our baseline *Timing* treatments but elicit expectations of average inflation over the next three years. We call these our *Timing: Medium-Term* treatments. Additionally, we provide subjects in these treatments with the central bank's outlook on average inflation over the same three-year horizon between the Initial and Updated forecasts. To do this, we average the short-run forecasts of the central bank while allowing our mean-reverting shocks to dissipate over time. We accompany these changes in the forecast and signal horizon with corresponding changes in our instructions and graphical interface. Finally, we included an additional question in our comprehension quiz to ensure participants fully internalized that they were providing medium-term inflation forecasts.[12]

Our interest is in whether the impact of short-term forecast performance on perceived forecast credibility is limited to the short-term or if it affects perceptions of the central bank's longer-term inflation forecast credibility. The idea is similar to Carvalho et al. (2023), which shows that an economic agent's long-term inflation expectations respond endogenously to short-term forecast errors and that the strength of this relationship depends critically on the agent's historical forecast performance. Our litmus is whether or not we replicate the qualitative findings observed in our baseline *Timing* treatments. That is, we consider our primary results – that people exhibit recency bias when forming

[12]We include these instructions, screenshots of the updated graphical interface, and the text of this new question in our appendix.

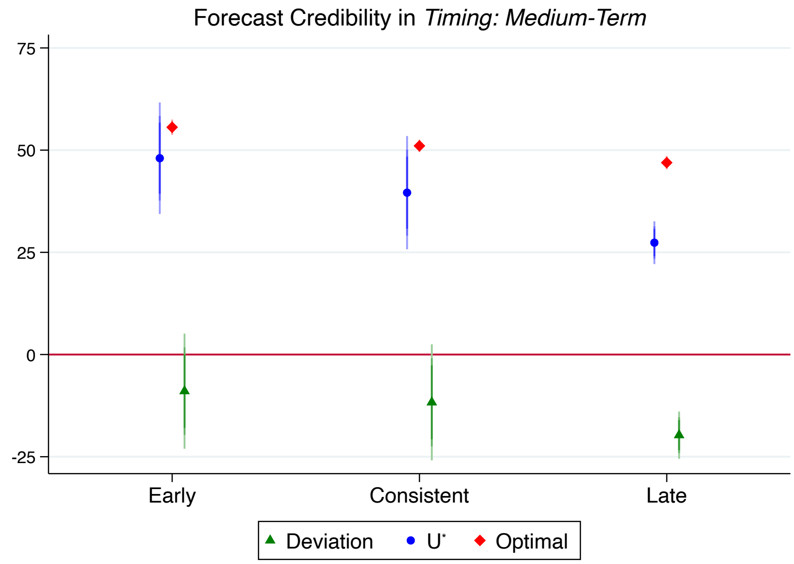


Figure 14: Estimated forecast credibility in *Timing: Medium-Term* treatments.
Note: This figure depicts average treatment effects (blue circles) surrounded by confidence intervals ranging from 90 (darkest, thickest lines) to 99 percent (lightest, thinnest lines) from our *Timing: Medium-Term* treatments using data Windsorized at the 5% level. Red triangles reflect the Bayesian benchmark weight, while green triangles show participants' average deviations from this benchmark, surrounded by confidence intervals ranging from 90 to 99 percent. These estimates omit individuals for whom $u^* = 0$ and their initial forecast matches the central bank's forecast.

a perception of central bank forecast credibility – robust if we qualitatively replicate that result in these treatments.

Note that implementing this litany of changes simultaneously – forecast horizon, signal horizon, instructions, and graphical interface – means that we cannot make causal claims comparing short-term and medium-term *Timing* treatments. Because of this, we are not overly concerned with quantitative differences that may arise across analogous short-term and medium-term treatments.

We show results from these treatments graphically in Figure 14, which depicts a history-contingent pattern of updating that is consistent with our short-term *Timing* treatments.

Before proceeding, we note a peculiarity in our *Early* data from these medium-term treatments. We observe 100 instances where a participant's initial point forecast of inflation perfectly coincides with the central bank's inflation forecast. This is likely because shocks in our *Early* simulation had almost entirely abated during the 12-quarter economic history initially revealed to subjects so that the central bank's inflation forecast for the subsequent three years in this version of *Early* was 2%, which a participant exhibiting satisficing would likely choose over some small deviation from 2%, which is the approximate value of inflation in the last year of *Early*.

Of these 100 instances, there are 75 observations where the participant does not update her point forecast of inflation after viewing the central bank's forecast. We exclude these observations from our main results in Figure 14. However, we could also reasonably assume that these observations represent instances of full central bank credibility. After all, 25% of the subjects whose initial point forecast matched the central bank's projection chose to update their point forecast. If we make this assumption, then find obtain even sharper results, which we depict in Figure 15.

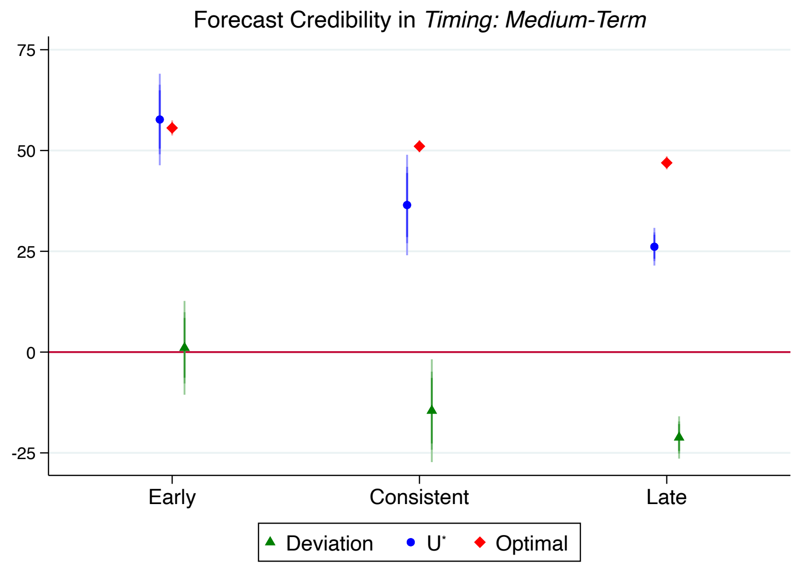


Figure 15: Estimated forecast credibility in *Timing: Medium-Term* treatments.
Note: This figure depicts average treatment effects (blue circles) surrounded by confidence intervals ranging from 90 (darkest, thickest lines) to 99 percent (lightest, thinnest lines) from our *Timing: Medium-Term* treatments using data Windsorized at the 5% level. Red triangles reflect the Bayesian benchmark weight, while green triangles show participants' average deviations from this benchmark, surrounded by confidence intervals ranging from 90 to 99 percent. These estimates assign $u^* = 100\%$ for individuals who do not update and their initial forecast matches the central bank's forecast.

Whether we treat these observations as uninformative or as instances of full credibility is irrelevant from a qualitative perspective; our results in *Timing* extend to longer-term forecast horizons.

6 External Validity and Theoretical Implications

6.1 Evidence of recency bias in real-world data

Are our results an artifact of our experimental setting or do the results from our experiment generalize to a real-world setting? To explore this, we combine a high-frequency identification approach with the Bank of England's (BoE) quarterly Inflation Report (IR) (now called Monetary Policy Report). We answer the question "do markets react more strongly to the BoE's forecast information whenever the BoE's forecast credibility is high?[13]

Our identification strategy involves projecting changes in real yields of different maturities that occur in the 24-hour window surrounding IR releases onto conditioning

[13] For examples of this identification approach, see Cook and Hahn (1989), Kuttner (2001), Gürkaynak et al. (2005), Cochrane and Piazzesi (2002), Nakamura and Steinsson (2018) and many others. Hubert (2015) explores forecast performance and market news for numerous central banks but focusing on the Bank of England is ideal since it releases the forecast information separately to the policy decision with a lag of about a week in our sample.

information and variables capturing the BoE's recent forecasting performance. Based on our experimental results, we would expect that better forecast performance would lead to stronger market reactions to information in the IR. We use data from between Q3 1997, after the Bank gained operational independence, through Q2 2015 (when the provision of information changed). We have 72 observations (i.e. Inflation Reports) in total.

A crucial part of this strategy is categorizing the BoE's forecast performance over time. In our experiment, the central bank forecasts one-period-ahead inflation. In reality, central banks provide forecasts for many periods into the future, which means that forecast credibility in the real world is itself multi-dimensional. We take an approach that tries to be agnostic about the relationship between forecast errors, forecast horizon, and forecast credibility. To do this, we measure the central bank's forecast error for each forecast horizon during each quarter and collapse these horizon-specific measures of forecast performance into a single dimension using factor analysis. The result is a one-dimensional measure of forecast performance that accounts for forecast errors at each forecast horizon during each quarter. Though the BoE has sometimes provided forecasts with as much as three-year horizons, this practice was not consistent during our time sample. Because of this, we focus on the BoE's nowcast and forecasts for the next eight quarters.

Using this factor, we create a set of indicator variables denoting whether or not the BoE's forecast performance has been above its sample average forecast for the previous one, two, three, or four quarters. We record 39, 34, 29 and 26 instances where these indicators take on a positive value, respectively.

Additionally, we require measures of how markets react to information contained in the IR. For this, we use the one-year, three-year, and five-year gilts. More specifically, we measure how yields at each of these three maturities changes during the 24-hour window surrounding the release of the BoE's IR. Our interest is the causal relationship between these measures and our measure of central bank forecast performance.[14] Following from our experimental results, our hypothesis is:

Hypothesis 5. *Yields will respond more strongly to information contained in the IR whenever the BoE's recent forecast credibility, proxied by its forecast performance, is above the sample average level of forecast credibility.*

To test Hypothesis 5, we estimate the following equation:

$$|\Delta y_i| = \alpha_i + \beta_l \mathbb{I}_{l,t} + \sum_{x,j} \psi_{x,j,i} \Delta PC_{x,j,t} + \eta_{1,i} FTSE_{t-1} + \chi_i X_{i,t} + \eta_{2,i} VIX_{t-1} + \epsilon_{i,t} \quad (18)$$

where $\Delta y_{i,t}$ captures changes in the 1,3, and 5yr gilt that occur in the 24-hour window around the IR release, $\Delta PC_{x,j,t}$ is a set of six factors summarizing new information contained in the contemporaneous IR regarding the first three central moments ($x = \{1, 2, 3\}$) of the BoE's outlook on inflation and output ($j = \{\pi, Y\}$), and $\mathbb{I}_{i,t}$ is an

[14] We use a 24-hour window following Hansen et al. (2019), who argue the longer window is necessary because of the volume of information contained in the BoE's IR. This is compared to policy announcements, where market participants can quickly discern and react to information.

indicator capturing whether the BoE's forecast performance has exceeded its sample average for the last $l = \{1, 2, 3\}$ quarters. As controls, $X_{i,t}$ contains controls that account for prevailing economic conditions (unemployment, output, and inflation), $FTSE_{t-1}$ is a daily, market-based measure of economic uncertainty, and VIX_{t-1} captures general economic uncertainty.

That is, we project the asset price news, $|\Delta y_{i,t}|$, $i = \{1, 3, 5\}$, onto a set of controls and our indicator variables indicating if the BoE's performance was above its historical average. To better isolate the relationship between yield changes and forecast credibility, we control for the information contained in the BoE's density forecasts of output and inflation, for the BoE's output forecast performance, for prevailing economic conditions, and for economic uncertainty. We account for the possibility of autocorrelation and heteroskedasticity using Newey-West errors with 3 lags.[15]

Table 10: Regression Table: Recency Bias in Markets

	(1) $\mathbb{I}_{1,t}$	(2) $\mathbb{I}_{2,t}$	(3) $\mathbb{I}_{3,t}$
1yr Gilt	0.036	0.050**	0.064***
	(0.027)	(0.021)	(0.02)
3yr Gilt	0.0307	0.0358*	0.0564****
	(0.025)	(0.020)	(0.012)
5yr Gilt	0.011	0.0199	0.0369***
	(0.023)	(0.019)	(0.013)

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Note: This table depicts estimates of β_1 in Equation (18).

We report $\hat{\beta}_i$ in Table 10, where columns correspond to different durations of above-average forecasting performance, $\mathbb{I}_{l,t}$, and rows to our outcomes of interest, $|\Delta y_{i,t}|$.

Results from this estimation exercise indicate that markets respond more strongly to the information contained in the BoE's inflation report whenever the BoE's forecast credibility, proxied by its forecast accuracy, is above its historical average.

For each gilt maturity we consider, the estimated effect size of above-average forecast performance is strongest whenever this performance has persisted for at least three quarters and decreases monotonically in the duration of above-average performance. Additionally, the credibility premium garnered by above-average forecast performance has less impact as we extend deeper into the term structure.

These results align with our finding of recency bias in *Timing*. Further, the fact that this effect persists for longer-term maturities aligns with our finding that both short- and longer-term expectations respond in a qualitatively similar way to recent forecast performance. Our findings align with Hubert (2015), who shows that central bank forecasts influence the expectations of private agents. However, that work also finds that forecast performance does not drive this result, which is different than what we find. Whereas that work considers whether survey-based forecasts respond to a gap

[15]Lag selection is based on Newey and West (1987) and Greene (2003).

in forecasting performance between private agents and the central bank, we consider if market participants respond differently to central bank communication depending upon the bank's recent forecast performance relative to its historical performance.

6.2 Endogenous Credibility In the New Keynesian Model

Our experimental results suggest some ways to modify common features of workhorse theoretical models. First, the central bank's forecast is not necessarily fully credible. Rather, central bank forecast credibility can evolve endogenously. Second, economic agents do not incorporate all available historical information when forming a perception of the central bank's forecast credibility. Instead, agents' perception of central bank credibility places more weight on temporally-recent information. This is especially true following large shocks that lead to large forecasting errors. This asymmetry in the degree of recency bias means that forecast credibility is harder to build than lose. Third, agents tend to exhibit over-precision, which leads them to under-react to their own forecast errors whenever forming a perception of the central bank's forecast credibility. Below, we incorporate our experimental findings into an otherwise standard New Keynesian model to demonstrate how endogenous forecast credibility coupled with recency bias and an overly-precise economic agent can impact aggregate inflation dynamics.

Before proceeding, we first note that we are not the first to consider the impact of endogenous central bank credibility on aggregate dynamics. Hommes and Lustenhouwer (2019) introduces a model of endogenous central bank credibility, whereby a continuum of heuristic-switching agents form either naive inflation expectations or inflation expectations that align with the central bank's inflation target. Credibility in their model is endogenous in the sense that the proportion of agents anchored onto the central bank's inflation target depends on the bank's historical ability to achieve its target. Goy et al. (2022) extend Hommes and Lustenhouwer (2019) to include N-step learning to study the effectiveness of forward guidance in a world populated by heterogeneous, boundedly-rational agents. Relative to these prior studies, our focus is on the relationship between historical forecast performance and the central bank's ability to influence inflation expectations via the publication of its own subjective outlook on inflation.

6.3 Endogenous Credibility in the New Keynesian Model

We start with a simple three-equation New Keynesian model linearized around a zero-inflation steady state

$$
\begin{aligned}
y_t &= \mathbb{E}_{i,t}\{y_{t+1}\} - \sigma^{-1}(i_t - \mathbb{E}^{\text{HH}}_{t,\text{post}}\{\pi_{t+1}|\mathbb{E}^{\text{CB}}_t \pi_{t+1}\}) + g_t \\
\pi_t &= \beta\mathbb{E}^{\text{HH}}_{t,\text{post}}\{\pi_{t+1}|\mathbb{E}^{\text{CB}}_t \pi_{t+1}\} + \kappa y_t + u_t \\
i_t &= \phi_y y_t + \phi_\pi \pi_t + v_t
\end{aligned}
$$

where

$$
\begin{aligned}
g_t &= \rho_g g_{t-1} + \epsilon_t^g \\
u_t &= \rho_u u_{t-1} + \epsilon_t^u \\
v_t &= \rho_v v_{t-1} + \epsilon_t^v
\end{aligned}
$$

.

We assume that a Bayesian household[16] forms a composite expectation of π, where some proportion ξ of the household's expectation is noisy model consistent

$$\mathbb{E}_{t,\text{RE}}^{\text{HH}}\{\pi_{t+1}\} = \pi^{\text{RE}} + \epsilon_{\text{HH}}, \quad \epsilon_{\text{HH}} \sim \mathcal{N}\left(0, \frac{1}{\tilde{\alpha}}\right).$$

The household holds a belief about her own forecast precision, $\tilde{\alpha} = \tau\alpha,\ \tau \in [1, \infty)$.[17] Note that if $\tau = 1$, then the economic agent is not susceptible to over-precision. Otherwise, $\tau > 1$ implies the agent perceives her forecasting ability as being better than it actually is, which aligns with our experimental finding of over-precision. The remaining $1-\xi$ of the household's belief follows a backward-looking, adaptive heuristic of the form

$$\mathbb{E}_{t,\text{BR}}^{\text{HH}}\{\pi_{t+1}\} = \mathbb{E}_{t-2}\{\pi_{t-1}\} + \eta_\pi(\pi_{t-1} - \mathbb{E}_{t-2}\{\pi_{t-1}\})$$

where $\eta_\pi \in [0, 1]$ determines how strongly the agent reacts to past forecast errors. Note that $\eta = 1 \implies \mathbb{E}_{t,\text{BR}}^{\text{HH}}\{\pi_{t+1}\} = \pi_{t-1}$, which means the household forms expectations naively in a backward-looking fashion. Alternatively, $\eta_x = 0 \implies \mathbb{E}_{t,\text{BR}}^{\text{HH}}\{\pi_{t+1}\} = \mathbb{E}_{t-1}\{\pi_t\}$, which means the household's expectations are completely anchored. For brevity, we refer to these expectations as AD(1). Intuitively, these forecasting heuristics say that the agent forms her initial expectations in each period t by adjusting her most recent forecasts of inflation and output to account for her most recent forecasting errors.

Thus, the household's prior is a linear combination of a noisy rational component and an adaptive component

$$\mathbb{E}_{t,\text{prior}}^{\text{HH}}\{\pi_{t+1}\} = \xi\mathbb{E}_{t,\text{RE}}^{\text{HH}}\{\pi_{t+1}\} + (1-\xi)\mathbb{E}_{t,\text{BR}}^{\text{HH}}\{\pi_{t+1}\}.$$

Notice that $\xi = 1$ implies the household forms noisy rational expectations while $\xi = 0$ means the household is boundedly-rational.

The central bank in this economy publishes model-consistent forecasts of π in each period, which the household can use to update its belief about π before committing to actions:

$$\mathbb{E}_t^{\text{CB}}\{\pi_{t+1}\} = \pi^{\text{RE}} + \epsilon_{\text{CB}}, \quad \epsilon_{\text{CB}} \sim \mathcal{N}\left(0, \frac{1}{\beta}\right).$$

[16] We refer to a single household for ease of exposition. You could also think of a continuum of households where ξ is the proportion of households forming noisy rational expectations and $1-\xi$ the proportion adhering to a backward-looking, adaptive heuristic.

[17] We can think of this arising from the noise term in $\mathbb{E}_{t,\text{RE}}^{\text{HH}}\{\pi_{t+1}\}$.

The household has some subjective outlook on the central bank's forecast precision, $\frac{1}{\beta}$, approximating its outlook on the central bank's forecast credibility. Further, households may exhibit recency bias when evaluating the central bank's forecast precision. Specifically, we assume the household only uses information from the central bank's N most recent forecasts and cares most about the most recent information. This gives

$$\tilde{\beta}_{\pi,t}^{-1} = \sum_{k=1}^{k=N} \lambda_k \big(\| \mathbb{E}_{CB,t-k}\{\pi_{t-k}\} - \pi_{t-k} \| \big) \tag{19}$$

where $\sum_{k=1}^{k=N} \lambda_k = 1$ and $\lambda_k > \lambda_{k-1} > \lambda_{k-2} > ... > \lambda_{k-N}$. Given this, the household forms an updated expectation (i.e. a posterior) according to:

$$\mathbb{E}^{\text{HH}}_{t,\text{post}}\{\pi_{t+1} | \mathbb{E}^{\text{CB}}_t \pi_{t+1}\} = (1 - \Gamma_{\pi,t}) \Big(\mathbb{E}^{\text{HH}}_{t,\text{prior}}\{\pi_{t+1}\} \Big) + \Gamma_{\pi,t} \Big(\mathbb{E}^{\text{CB}}_t \{\pi_{t+1}\}\} \Big) \tag{20}$$

where $\Gamma_{\pi,t} = \frac{\tilde{\beta}_{\pi,t}}{\tilde{\alpha}_{i,t} + \tilde{\beta}_{\pi,t}}$.[18]

Notice in Equation (20) that $\Gamma_{\pi,t}$ determines how strongly the household's updated inflation expectation depends on the central bank's forecast. Thus, $\Gamma_{\pi,t}$ constitutes a measure of central bank forecast credibility in our model. Further, notice that $\Gamma_{\pi,t}$ depends on the households outlook on the central bank's forecast precision, $\tilde{\beta}^1$ and its own forecast precision, $\tilde{\alpha}^{-1}$. The former serves as the channel through which the central bank's forecasting history influences its forecast credibility and the latter where the household's biased perception of its own forecast precision can further distort credibility from the Bayesian benchmark.

Using this simple model, we can consider what happens when an economy sustains an unanticipated cost-push shock while accounting for endogenous credibility and recency bias. Endogenous credibility can matter for at least two reasons. First, the economic agent will not necessarily view the central bank's inflation projections as fully credible. Second, the agent's perception of the central bank's forecast credibility can change over time depending on the central bank's forecast performance. We attempt to separate these effects by first considering the impact of less-than-full but exogenous credibility. We then layer in endogenous credibility and over-precision.

6.4 One-time, permanent credibility shock

We isolate the role of less-than-perfect credibility by first considering how this economy responds to a 1pp cost-push shock that leads to a permanent shift in the central bank's

[18] Note the timing here assumes the agent cannot observe contemporaneous inflation and instead can only adjust expectations based on its forecast of $t-1$ inflation relative to realized inflation. We demonstrate in Section A4 that our results do not depend on this timing assumption by simulating dynamics assuming the agent adjusts expectations based on its forecast and observation of contemporaneous inflation.

forecast credibility.[19] Further, we assume that the household's perception of its own forecast precision remains fixed so that changes in Γ are driven entirely by changes in $\tilde{\beta}$.

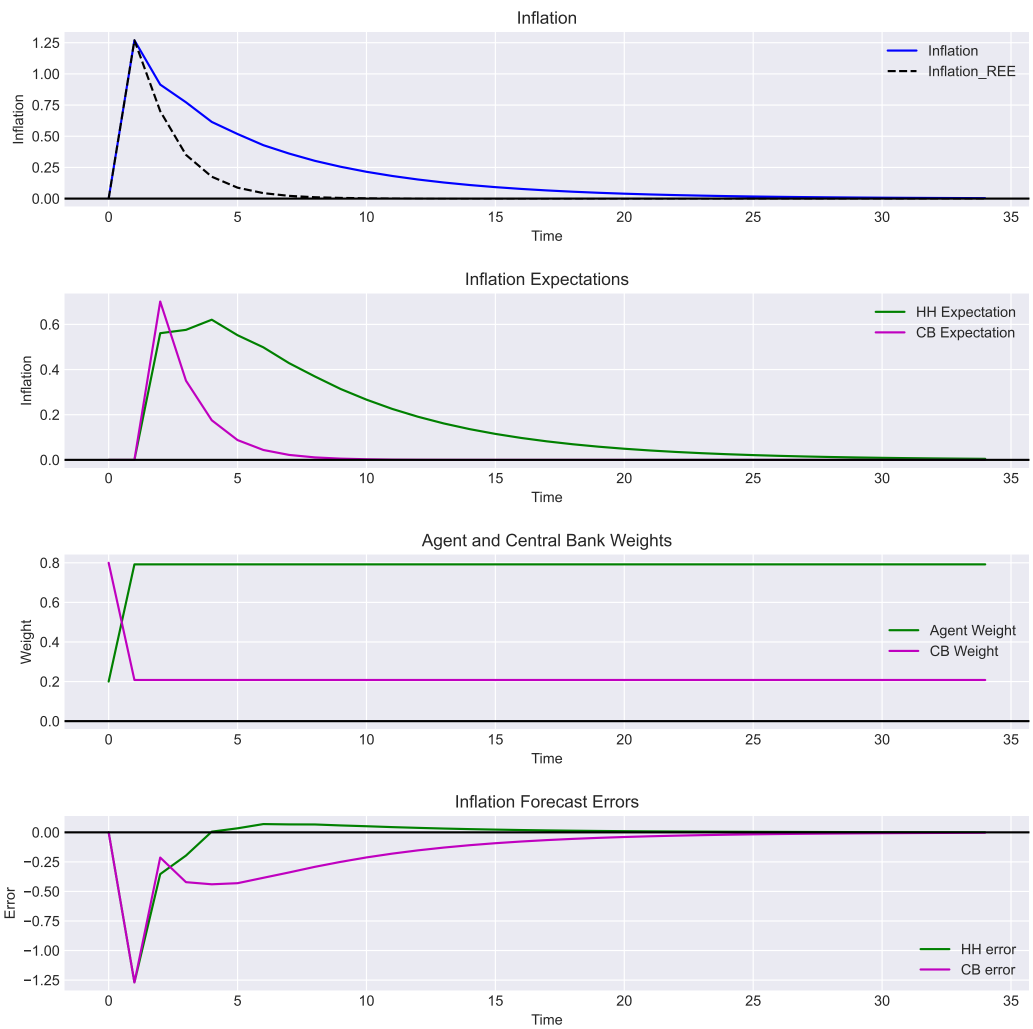


Figure 16: Response to a one-time, permanent credibility shock.

Note: This figure depicts the economic response to a one-time cost-push shock that permanently shifts the central bank's forecast credibility on impact. We have calibrated the shock so that $\Gamma = .8$ before the shock and $\Gamma = .2$ after the shock. Here, $\xi = 0$ and $\eta = .5$. This means the household follows an adaptive learning heuristic in the absence of credible central bank signals.

We show the results of this exercise in Figure 16, where the top panel traces the response of aggregate inflation to this cost-push shock assuming the household forms rational expectations (black dashed line) and when forms expectations according to Equation (20) assuming $\xi = 0$ and $\eta = .5$. Note these parameter values correspond to a household who, in the absence of central bank signals, adheres to an adaptive expectation heuristic. the central bank, on the other hand, forecasts the rational expectations path without

[19]Throughout this section, we calibrate the main model parameters with the same parameter values used in the model to generate our experimental scenarios (described in Table 1).

accounting for the credibility loss that results from the exogenous cost-push shock. The second through fourth panels of Figure 16 depict inflation expectations (second panel), values for $\Gamma_{\pi,t}$ and its complement (third panel), and forecast errors (bottom panel) for the central bank (purple lines) and the household (green lines).

The primary takeaway from this exercise is that a decrease in the central bank's ability to influence the inflation expectations of a boundedly rational agent leads to considerably more persistence in the inflation gap relative to the rational expectations counterfactual. Inflation remains considerably higher for longer, taking more than twice as long to converge to the zero inflation steady-state path. Because the central bank does not incorporate the effects of credibility loss into its inflation forecasts, the central bank's low forecast performance becomes persistent, leading to pronounced forecast errors relative to the household.

Assuming instead $\eta = 1$ (see Figure 17), so that the household is purely backward-looking in the absence of central bank signals, amplifies the dynamics exhibited in Figure 16. Inflation dynamics become hump-shaped, leading to a prolonged period of high inflation following the cost-push shock, with peak inflation occurring in the periods immediately following the shock.

6.5 Endogenous Credibility with fixed self-perceptions

We now consider a slightly different question: how long does it take a fully-credible central bank to recover from an exogenous cost-push shock when changes in Γ result exclusively from changes in a fully-endogenous $\tilde{\beta}$. Intuitively, this means the household updates its perception of the central bank's forecast credibility based on the bank's forecasting history but without ever changing its outlook on its own forecasting ability. To do this, we assume the agent bases her estimates of $\tilde{\beta}$ on the four most recently observed periods where $\lambda_k = [.63, .238, .09032, .03904]$ for $k = \{1, 2, 3, 4\}$, respectively. Note that these are the values produced by Equation (16) assuming λ=.622 and combined account for 99.73% of the total weighting. If $t < 4$, we use the same weighting scheme for all available λ_k and normalize the weights so that they sum to one.

We show the results of this in Figure 18 where we again assume $\xi = 0$ and $\eta = .5$. Interestingly, allowing $\tilde{\beta}$ to adjust endogenously follow the initial credibility loss does little to speed up recovery relative to our exercise in Section 6.4, where $\tilde{\beta}$ remain fixed at some new value following its initial response to the cost-push shock. This is because the initial loss of credibility leads to large and persistent forecast errors, which translate into persistently low levels of $\Gamma_{\pi,t}$.

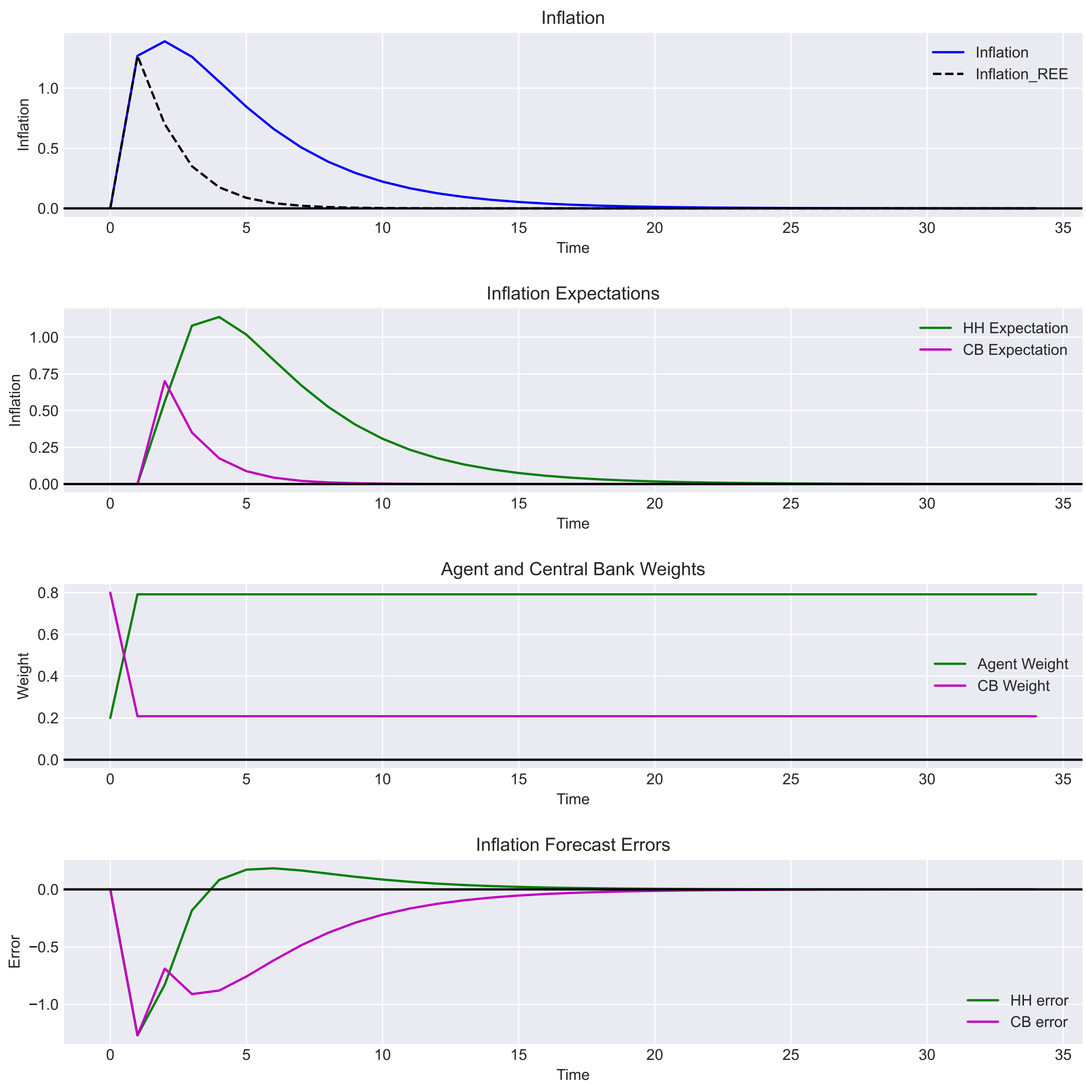


Figure 17: Response to a one-time, permanent credibility shock.
Note: This figure depicts the economic response to a one-time cost-push shock that permanently shifts the central bank's forecast credibility on impact. We have calibrated the shock so that $\Gamma = .8$ before the shock and $\Gamma = .2$ after the shock. Here, $\xi = 0$ and $\eta = 1$. This means the household is naively backward-looking in the absence of central bank signals.

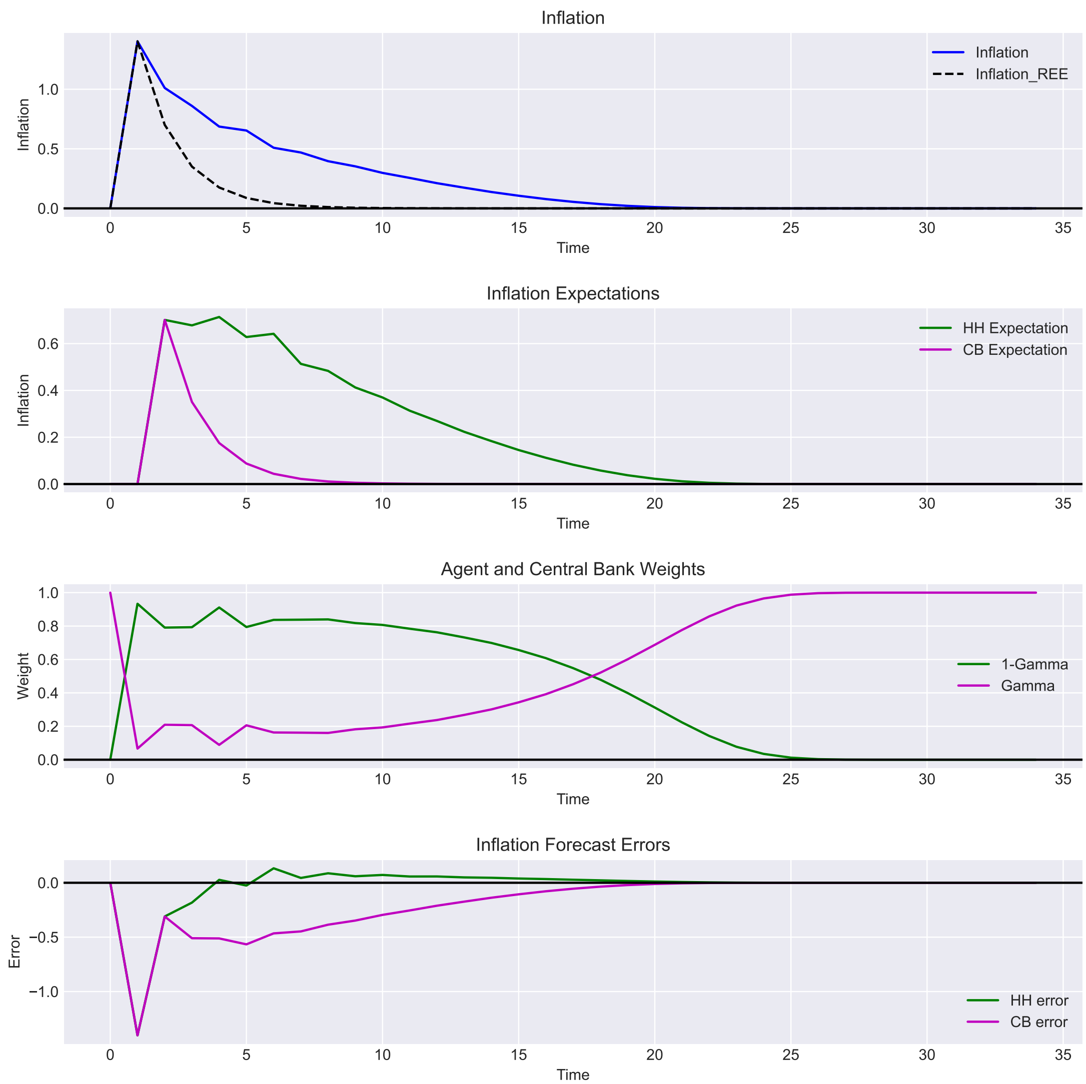


Figure 18: Response to cost-push shock with endogenous credibility with fixed self-perceptions

Note: This figure shows how long it will take a fully credible central bank that naively forecasts the REE path to recover full credibility following a one-time cost-push shock. Here, $\xi = 0$ and $\eta = .5$. In this exercise, $\alpha = 10$ remains fixed but $\tilde{\beta}$ adjusts endogenously. Obviously, the value of α we choose matters a lot for dynamics. Note, $\alpha = 10$ means the household, loosely speaking, thinks its forecasts are accurately within 10 basis points on average.

6.6 Endogenous Credibility

We now allow both $\tilde{\alpha}$, $\tilde{\beta}$ to evolve endogenously by assuming that the estimation scheme for $\tilde{\beta}$ described in Section 6.5 applies both $\tilde{\alpha}$, $\tilde{\beta}$. This means that Γ is fully endogenous since the household updates its perceptions of both its own and the central bank's forecast precision. We show impulse responses to a 1pp cost-push shock under these assumptions in Figure 19.

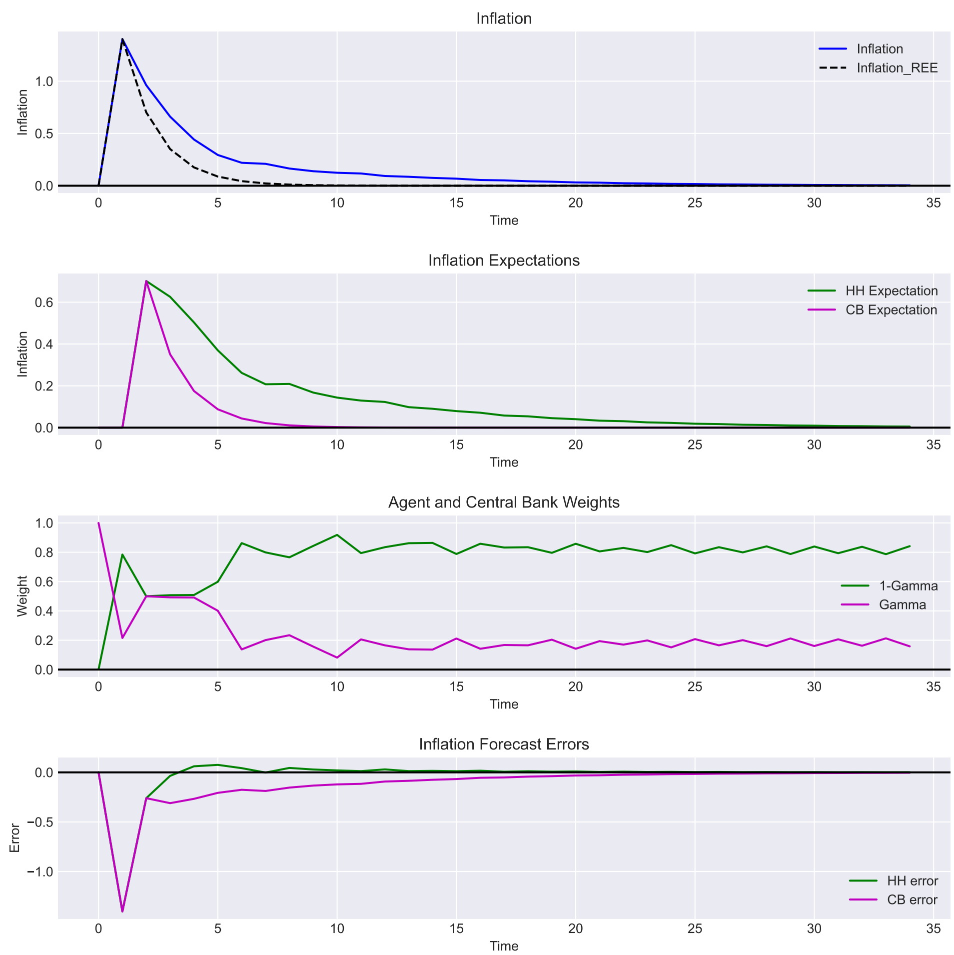


Figure 19: Response to cost-push shock with fully-endogenous credibility
Note: This figure shows how aggregate inflation dynamics, inflation forecast errors, and central bank forecast credibility respond to a 1pp cost-push shock assuming that both $\tilde{\beta}$, $\tilde{\alpha}$ adjust endogenously following the shock. Here, we assume that $\xi = 0$, $\eta = .5$ so that the household follows an adaptive learning heuristic in the absence of credible central bank signals.

The main takeaway from this exercise is that allowing for fully endogenous forecast credibility, where the household is equally recency-biased toward itself and the central bank, again leads to very persistent inflation dynamics relative to an economy populated by a fully informed, rational household. Though this result depends on what we assume about η, ξ, this figure also demonstrates that exogenous shocks can result in a seemingly

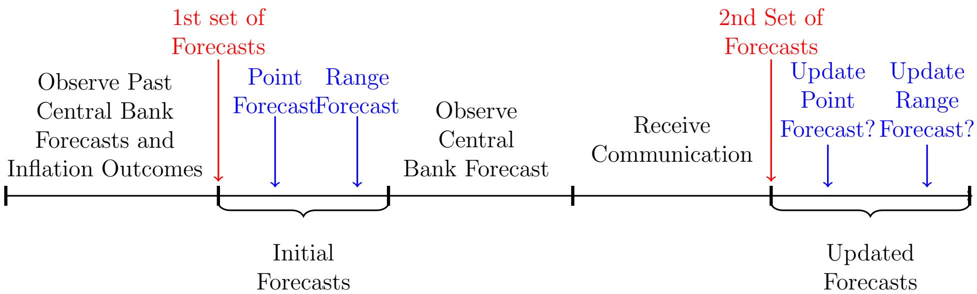


Figure 20: Timing of decision period in *Contextual Communication* treatments
Note: This figure depicts the flow of a single decision period in the *Contextual Communication* treatments of our experiment. The only difference here relative to any other treatments *Contextual Communication* treatments include a written statement from the central bank alongside the central bank's numerical inflation forecast.

permanent decrease in forecast credibility.

7 *Contextual Communication*

In this section, we ask whether contextualizing communication that reinforces the central bank's inflation outlook and rationalizes its forecasting history can influence forecast credibility. From a practical perspective, this question is important because most central banks devote considerable resources to crafting and publishing contextualizing communication, often alongside projections. For example, most central banks now publish a monetary policy report and information contained in these reports influences matters for market behavior (Hansen et al. 2018). Can this sort of communication enhance forecast credibility? If so, to what extent? And which sorts of messaging most effectively allow the central bank to talk its way out of a low-credibility position?

7.1 Contextual Communication Treatments

To answer these questions, we incorporate contextualizing communication into *T7* from our *Timing* treatments. Specifically, alongside the final history (*Late)*, we publish a written statement alongside the central bank's graphical forecast before allowing subjects to update their inflation expectations. We focus on *T7* where participants experience *Late* last, which allows them maximal time to learn the experimental environment before encountering written communication. We summarize the timing of decision periods in *Contextual Communication* in Figure 20.

Written statements in these treatments convey information about the central bank's inflation outlook, whether the source of poor historical forecast performance in *Late* is endogenous or exogenous, and about the central bank's forecast performance relative to peer forecasting institutions. Arguably, introducing written statements alongside projections increases information complexity from the perspective of participants. To address this, we also include a control text that provides generic information about the central bank. Because our control text is uninformative about the central bank's outlook or its forecasting history, we can use this control text as a baseline for our treatments that either reinforce the bank's outlook or rationalize its forecasting history. We provide additional details about each statement in Figure 21 and provide the full text of each statement in Section A5.3.

Control We provide a general description of central banking.

Control + Outlook Repeats text from *Control* but also includes a written outlook on inflation that matches the graphical forecast and adds no new information for participants. This allows us to discern whether reinforcing graphical information via text can better convey important economic information.

Exogenous + Relative Performance *Control + Outlook* but includes an additional paragraph explaining that the decline in historical forecast performance resulted from exogenous forces and also says whether the bank performed better or worse than peer forecasting institutions.

Endogenous + Relative Performance As *Exogenous + Relative Performance* except that the central bank explains the decline in historical forecast performance resulted from endogenous forces.

Figure 21: *Contextual Communication* treatments

Note: This table provides general descriptions of the statements provided to subjects in our *Contextual Communication* treatments.

Additionally, we developed our contextual communication treatments so that complexity is roughly identical across the written statements where we measure complexity using the Flesch-Kincaid reading grade level. The Flesch-Kincaid reading grade level is a readability metric that gauges the complexity of a text and estimates the grade level at which an individual can comprehend a piece of writing. The formula takes into account the total words, sentences, and syllables in a text, producing a numerical score that corresponds to a U.S. grade level. We summarize these treatments in Table 11.

Treatment Summary - Communication				
	Name	**Sample Size**	**Flesch-Kincaid**	
			Score	*Reading Level*
T12	*Control*	160	8	10th-12th
T13	*Control + Outlook*	151	8.3	10th-12th
T14	*Exogenous + Better*	131	8.5	10th-12th
T15	*Exogenous + Worse*	152	8.5	10th-12th
T16	*Endogenous + Better*	157	8.4	10th-12th
T17	*Endogenous + Worse*	137	8.4	10th-12th

Table 11: Treatment summary for *Contextual Communication*

Note: The table summarizes our *Contextual Communication* treatments. Subjects completed a single treatment, described by rows *T12* through *T17*. Note that we include Flesch-Kincaid scores (*Score* column) and reading levels (*Reading Level* column) to demonstrate the similarity in the complexity of the written statement included in each of our six *Contextual Communication* treatments.

7.2 Finding on the Effects of Communication

We present estimates of central bank forecast credibility in *Contextual Communication* in Figure 22. For ease of interpretation, we center forecast credibility in *Control* to zero and then normalize estimates of forecast credibility in the remaining treatments to *Control*. Thus, this figure reports estimates of effects relative to *Control*.

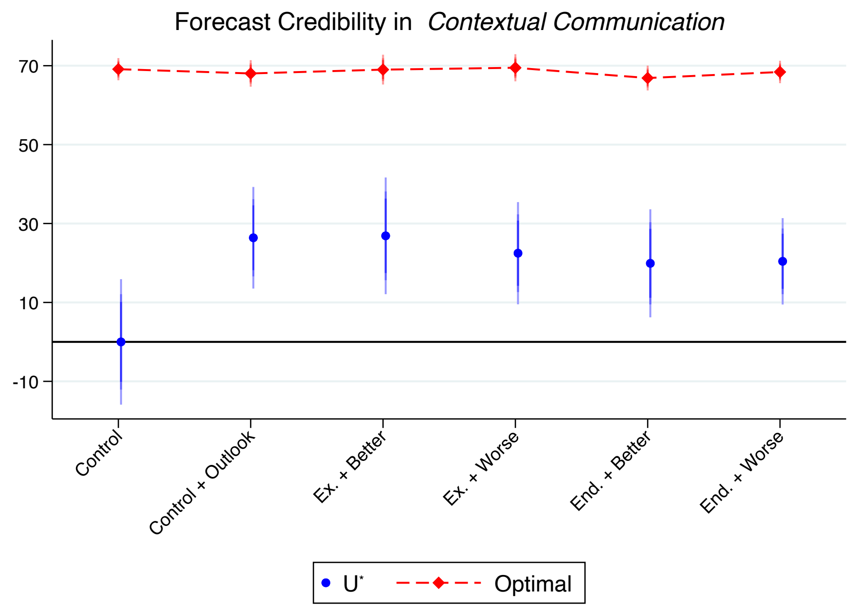


Figure 22: Estimated forecast credibility in *Contextual Communication* treatments.
Note: Forecast credibility (blue circles) in *Contextual Communication* treatments. Red diamonds indicate the optimal weighting of the central bank's forecasting in the updated inflation expectation of the average participant, assuming this participant is a rational Bayesian who equally weights all available historical information. Shading around both types of markers indicate 99%, 95%, and 90% confidence intervals.

Our primary result is that contextualizing communication can significantly increase the central bank's forecast credibility. Interestingly, this is true even with *Control + Outlook*, which reinforces the central bank's inflation outlook without providing any new information to our participants. This could be because participants are better at extracting qualitative or narrative information from text, if the process of reading text yields a better synthesis of information, or if simply seeing the information again but in text form somehow reinforces learning. Additionally, it could be that being seen to attempt to communicate helpfully is beneficial for the central bank's reputation as suggested in Haldane and McMahon (2018).

Taken together, these findings suggest that the design and the delivery of central bank communication is important. The communication can provide important and useful context in the delivery of forecast performance, especially where that performance may not be so strong for a period of time.

Table 12: Regression Table for Contextual Communication

	(1) u^*	(2) u^*	(3) u^*	(4) u^*
Control + Outlook	29.83****	29.82****	30.58****	30.00****
	(8.257)	(8.274)	(8.497)	(8.401)
Ex. + Better	27.84****	27.82***	27.71***	28.58****
	(8.387)	(8.463)	(8.511)	(8.425)
Ex. + Worse	23.43***	23.41***	23.74***	23.95***
	(7.920)	(8.039)	(8.234)	(7.884)
End. + Better	20.87**	20.88***	21.09**	20.64**
	(8.105)	(8.053)	(8.183)	(8.040)
End. + Worse	21.38***	21.38***	21.65***	21.84***
	(7.453)	(7.434)	(7.510)	(7.365)
Uncertainty		0.00162	-0.00167	0.00275
		(0.0351)	(0.0355)	(0.0359)
Demographics			✓	✓
Survey Responses				✓
Control	0.000	-0.222	7.099	-2.913
	(6.132)	(6.741)	(9.012)	(17.81)
N	679	679	674	674

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Note: This table provides estimates of u^* for each of our *Contextual Communication* treatments obtained via OLS. Column 1 provides baseline estimates of forecast credibility, column 2 controls for forecast uncertainty, column 3 layers in demographic controls, and column 4 includes controls from our pre-experiment survey questions.

8 Conclusion

Central bank communication has emerged over the last few decades as a mainstay of central banking because it offers policymakers an effective way to manage expectations. Arguably, the key component of communication is the central bank's economic outlook, which banks often publish as forecasts of key economic variables. Yet, this newly-established tool carries with it new concerns. Primarily, policymakers must now worry about how best to build and safeguard their forecast credibility so that publishing forecasts and communicating about their economic outlook remains potent. Though we know in practice that policymakers care deeply about forecast credibility (Blinder 2000), very little is known in theory about the determinants and dynamics of this credibility. To address this shortcoming, we've used a novel experimental framework to study the causal relationship between features of historical forecast performance and forecast credibility.

We show that the link between historical forecast performance and forecast credibility is not as sharp as theory might predict, which is perhaps due to an inability of people to accurately reflect on their forecast precision when considering new signals from the central bank. Additionally, we show that it isn't just a central bank's historical forecast performance that matters. Instead, our subjects exhibit considerable recency bias when evaluating forecast performance to form a perception of the central bank's forecast credibility. Taken together, this suggests that historical forecast performance can influence a central bank's forecast credibility, but that discrete changes in forecast performance can quickly shift perceived credibility.

An implication of this is that forecast credibility is not static. Instead, credibility is an endogenous component of communication that central banks can both win and lose, a feature conspicuously absent in most theoretical work on the topic. Though this implies that banks can lose their ability to manage expectations via forecasting whenever unexpected economic shocks lead to poor forecast performance, it also implies that banks can rebuild forecast credibility. However, we show that these dynamics of credibility are asymmetric – building credibility is a much slower process than losing it. We also demonstrate that low-frequency communication can bolster a bank's forecast credibility even when it does not convey new information about the bank's economic outlook or about the conditions underlying historical forecast performance.

We embed these ideas into an otherwise standard New Keynesian model in an ad-hoc way to demonstrate how accounting for endogenous credibility can matter for aggregate inflation dynamics. We show that less-than-perfect forecast credibility can generate persistent inflation dynamics in an intuitive way and that this effect is made worse when we include in the model a channel through which over-precision can impact the weight agents place on central bank signals whenever updating their beliefs.

References

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.

Armantier, O., S. Nelson, G. Topa, W. Van der Klaauw, and B. Zafar (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics 98*(3), 503–523.

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.

Candia, B., O. Coibion, and Y. Gorodnichenko (2020). Communication and the beliefs of economic agents. In *Navigating the Decade Ahead: Implications for Monetary Policy* (Aug. 27-28, 2020 ed.)., EConomic Policy Symposium.

Carvalho, C., S. Eusepi, E. Moench, and B. Preston (2023). Anchored inflation expectations. *American Economic Journal: Macroeconomics 15*(1), 1–47.

Chen, D. L., M. Schonger, and C. Wickens (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

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.

Cochrane, J. H. and M. Piazzesi (2002). The fed and interest rates - a high-frequency identification. *American Economic Review 92*(2), 90–95.

Coibion, O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte (2020). Inflation expectations as a policy tool? *Journal of International Economics 124*, 103297.

Coibion, O., Y. Gorodnichenko, and M. Weber (2022). Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy 130*(6), 1537–1584.

Cook, T. and T. Hahn (1989). The effect of changes in the federal funds rate target on market interest rates in the 1970s. *Journal of monetary economics 24*(3), 331–351.

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.

Eusepi, S. and B. Preston (2010). Central bank communication and expectations stabilization. *American Economic Journal: Macroeconomics 2*(3), 235–71.

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.

Goy, G., C. Hommes, and K. Mavromatis (2022). Forward guidance and the role of central bank credibility under heterogeneous beliefs. *Journal of Economic Behavior & Organization 200*, 1240–1274.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Guido, A., A. Martinez-Marquina, and R. Rholes (2022). Reference dependence and the role of information frictions.

Gürkaynak, R. S., B. P. Sack, and E. T. Swanson (2005). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *International Journal of Central Banking*.

Haldane, A. and M. McMahon (2018). Central bank communication and the general public. *AEA Papers and Proceedings 1*(1), Forthcoming.

Hansen, S., M. McMahon, and M. Tong (2018). The Long-Run Information Effect of Central Bank Narrative. CEPR Working Paper.

Hansen, S., M. McMahon, and M. Tong (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics 108*, 185–202.

Hommes, C. and J. Lustenhouwer (2019). Inflation targeting and liquidity traps under endogenous credibility. *Journal of Monetary Economics 107*, 48–62.

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.

Hubert, P. (2015). Do central bank forecasts influence private agents? forecasting performance versus signals. *Journal of Money, Credit and Banking 47*(4), 771–789.

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.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of monetary economics 47*(3), 523–544.

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.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review 115*(2), 502.

Moore, D. A. and D. Schatz (2017). The three faces of overconfidence. *Social and Personality Psychology Compass 11*(8), e12331.

Morris, S. and H. S. Shin (2002). Social value of public information. *American Economic Review 92*(5), 1521–1534.

Nakamura, E. and J. Steinsson (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics 133*(3), 1283–1330.

Newey, W. K. and K. D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777–787.

Petersen, L. and R. Rholes (2022). Macroeconomic expectations, central bank communication, and background uncertainty: A covid-19 laboratory experiment. *Journal of Economic Dynamics and Control*, 104460.

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.

Thakral, N. and L. T. Tô (2021). Daily labor supply and adaptive reference points. *American Economic Review 111*(8), 2417–43.

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.

9 Appendix

A1 Solving the model for cost-push shocks

What happens when the economy sustains an unanticipated cost-push shock? Assuming for the moment that the system depends only on u_t, we can rewrite the model as

$$\begin{aligned} y_t &= \frac{\sigma}{\sigma+\phi_y}\Big(\mathbb{E}_t\{y_{t+1}\} + \frac{1}{\sigma}\big[\mathbb{E}_t\{\pi_{t+1}\} - \phi_\pi \pi_t\}\big]\Big) \\ \pi_t &= \beta \mathbb{E}_t \pi_{t+1} + \kappa y_t + u_t \\ u_{t+1} &= \rho_u u_t + \epsilon^u_{t+1} \end{aligned}$$

Assuming the solution of this linear system is of the form:

$$y_t = \bar{y} + a u_t$$

$$\pi_t = \bar{\pi} + b u_t$$

where $\bar{y} = \bar{\pi} = 0$ in our system, which is linearized around a zero inflation steady-state, implies that the central bank's expectations for inflation and output are of the form:

$$\mathbb{E}_t\{y_{t+1}\} = a\rho_u u_t$$

$$\mathbb{E}_t\{\pi_{t+1}\} = b\rho_u u_t$$

If we assume $\psi = \frac{\sigma}{\sigma+\phi_y}$, then solving for a, b via substitution yields

$$\begin{aligned} a &= \frac{-\psi(\phi_\pi - \rho_u)}{\sigma(1-\beta\rho_g)(1-\psi\rho_g) + \kappa\psi(\phi_\pi - \rho_g)} \\ b &= \frac{\sigma(1-\psi\rho_u)}{\sigma(1-\beta\rho_g)(1-\psi\rho_g) + \kappa\psi(\phi_\pi - \rho_g)} \end{aligned}$$

which pin down analytical forms for the central bank's rational expectations of both output and inflation. Using this, we can consider how incorporating our experimental findings into an otherwise standard New Keynesian model changes inflation dynamics following a one-percentage-point cost-push shock via impulse response analysis.

A2 The role of forecast bias

What if participants believe that the central bank's forecast error is biased? In theory, we typically model no systematic component to the central bank's forecast error. However, participants may perceive $\gamma \neq 0$ because our experimental histories contain only twelve quarters of data based on volatile, real-world time series (we provide details on how we create these histories in Section 3.3). If this is true, then not accounting for this bias can lead to systematically

To account for this, we can rewrite Equation (2) as:[1]

$$\pi_{cb} - \gamma = \pi + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \frac{1}{\beta}\right). \tag{A.21}$$

Equation (A.21) says that once we adjust the central bank signal for its bias, we can apply the same logic as before. Intuitively, suppose that $\gamma < 0$ so that the central bank systematically under forecasts inflation. When the central banks signals its inflation forecast, the true signal from the central bank is adjusted upward and this new, higher, signal is used in the optimal update. That is, in Equation (5), we use $\pi_{cb} - \gamma \; > \; \pi_{cb}$ as the central bank's signal. Note that Figure 1 is unchanged once we make this bias correction since optimal updating scheme depends only on forecast precision $(\alpha_i^{-1},\ \beta^{-1})$.

Of course, our measure of the optimal update rate should also reflect the bias adjustment:

$$u^*_{\gamma,i} \equiv \frac{\mathbb{E}(\pi|\pi_{cb}) - \bar{\pi}_i}{(\pi_{cb} - \gamma - \bar{\pi}_i)} \tag{A.22}$$

Once this adjustment is done correctly, and assuming i updates according to the Bayesian optimal, $u^*_{\gamma,i} = \frac{\beta}{\alpha+\beta}$. Where $\gamma = 0$, $u^*_i = u^*_{\gamma,i}$ but if $\gamma \neq 0$, $\frac{u^*_{\gamma,i} - u^*_i}{u^*_i} = \frac{\gamma}{(\pi_{cb} - \gamma - \bar{\pi}_i)}$.

A2.1 Re-estimating our main results

We now account for the possibility of perceived bias ($\gamma \neq 0$) by adjusting our estimates of central bank forecast credibility according to Equation (A.22), where we assume that participants believe that the central bank's forecast bias is equal to the historical average forecast error so that

$$\gamma_{HistAvg} = \frac{1}{12} \sum_{k=t-1}^{k=t-12} (\mathbb{E}_{cb}\{i_k\} - i_k)$$

. We provide values of $\gamma_{HistAvg}$ for all economic histories in Table 2.

A2.1.1 *Forecast Performance*

Figure A-1 plots average treatment effects assuming our participants observe no systematic component in the central bank's forecast error (blue dots, baseline results) and also assuming that participants account for a systematic error component of the central bank's inflation forecast (purple squares). For these biased estimates, we assume that subjects use the entire forecast history to discern the magnitude and direction of this systematic error component. Adjusting our estimates of forecast credibility to account for forecast bias preserves our results qualitatively and leads to little quantitative change (see also Table A-1).

[1] We replace $\tilde{\epsilon} \sim \mathcal{N}\left(\gamma, \frac{1}{\beta}\right)$ in Equation (2) with $\gamma + \epsilon$ where $\epsilon \sim \mathcal{N}\left(0, \frac{1}{\beta}\right)$. If $\gamma = 0$, $\tilde{\epsilon} \equiv \epsilon$ trivially.

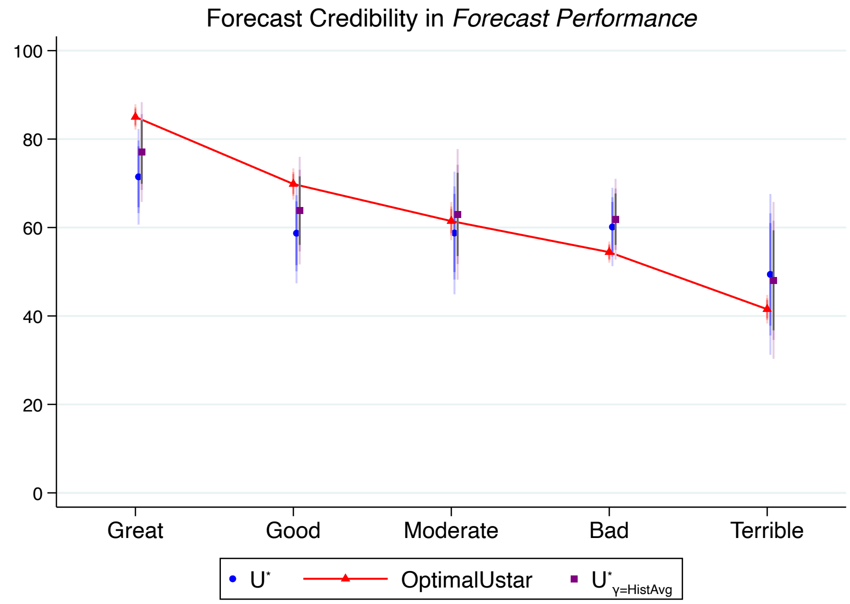


Figure A-1: *Forecast Performance* using $u^*_{\gamma_{HistAvg}}$
Note: This figure presents estimates of central bank forecast credibility in the *Forecast Performance* treatments. It compares scenarios assuming that subjects do not perceive any systematic forecast errors (as in Equation (5), blue circles) with scenarios where subjects do perceive systematic forecast errors (as in Equation (A.22), purple squares). Shaded bands surrounding the point estimates depict 99% (lightest), 95%, and 90% (darkest) confidence intervals. The connected red triangles depict the optimal level of credibility, on average, from the perspective of a rational Bayesian agent.

Bias-adjusting our credibility estimates effectively leads to higher measures of forecast credibility in all but our *Terrible* treatments. This is unsurprising, given that $\gamma_{HistAvg} > 0$ for *Great*,...,*Bad* but not for *Terrible*. This is because adjusting Equation (A.22) for positive values of $\gamma_{HistAvg}$ shrinks the denominator of our estimation equation.

Intuitively, this resembles the assumption that the central bank consistently over-predicts inflation. Rectifying this over-prediction suggests that subjects align more closely with the central bank's signal than if we suppose participants perceive $\gamma = 0$. This is because the true signals participants received were lower than the central bank's published forecast.

Despite this mechanical increase in estimated forecast credibility, our finding that participants exhibit over-precision survives. We note this in Table A-1, which presents the output of a series of regressions capturing unconditional estimates of u^* and $u^*_{\gamma_{HistAvg}}$ (columns one and two) and their corresponding deviations from the equal-weighting Bayesian benchmark adopted throughout this paper (columns 3 and 4). Comparing columns 3 and 4 shows that, qualitatively, deviations from the equal-weighting Bayesian benchmark are robust to bias adjustment. However, the mechanical increase in estimates forecast credibility that results from bias adjustment decrease significance.

A2.1.2 *Timing*

We note in Table 2 that average forecast errors are larger in *Early* and *Late*. This subsection considers our *Timing* results assuming that our participants perceive a sys-

Table A-1: Regression Table for Forecast Performance: Bias Adjusted

	(1) u^*	(2) $u^*_{\gamma HistAvg}$	(3) $u^* - u^*_{optimal}$	(4) $u^*_{\gamma HistAvg} - u^*_{optimal}$
Great	71.16****	76.71****	-12.60***	-9.167**
	(4.177)	(4.273)	(4.186)	(4.326)
Good	58.37****	63.47****	-11.60***	-6.511
	(4.336)	(4.608)	(4.443)	(4.700)
Moderate	58.39****	62.67****	-3.061	1.211
	(5.434)	(5.622)	(5.698)	(5.867)
Bad	59.70****	61.52****	5.273	7.084*
	(3.421)	(3.479)	(3.575)	(3.633)
Terrible	47.31****	46.77****	6.282	5.733
	(6.967)	(6.614)	(7.044)	(6.733)
N	528	528	524	524

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Note: This table presents estimates of forecast credibility for each *Forecast Performance* treatment. Column 1 is based on the assumption that there is no perception of a systematic forecast error (as in Equation (5)), whereas Column 2 is based on the assumption that subjects do perceive a systematic forecast error (as in Equation (A.22)). Columns 3 and 4 provide the corresponding estimates of deviations from optimal forecast credibility.

tematic bias in the central bank's forecasts so that we must adjust u^* to account for the values of $\gamma_{HistAvg}$ for *Early* and *Late* given Table 2. Note that we do not revisit *Consistent* results here since credibility estimates for this history remain essentially unchanged (note in Table 2 that the average forecast error for *Consistent* is only 2 basis points). Our interest is in whether our main finding – that the time profile of historical forecast errors – causes participants to more heavily weigh recent information when forming a perception of the central bank's forecast credibility.

Table A-2: Results of t-tests and Descriptive Statistics

	Early		*Late*	
	Mean	SE	Mean	SE
Early	28.25	1.38		
Late	$p < .001$		15.68	3.13

Comparing forecast credibility in *Early* and *Late* using $u^*_{\gamma HistAvg}$

Note: This table reports the mean and standard error (SE) of $u^*_{\gamma HistAvg}$ for the *Early* and *Late* histories in the *Timing* treatments. It also includes the p-value from a two-sample, two-sided t-test comparing the means between *Early* and *Late* groups.

We present these updated estimates in Figure A-2, which shows that subjects are now estimated to underweight the central bank's inflation forecast for both *Early* and *Late*. This is consistent with the results in Figure 4 where participants tend to underweight very good performance; the net effect of over-weighting recent performance but under-weighting great performance is not, ex-ante, obvious. Nonetheless, the finding that timing matters remains. Further, the qualitative result that participants exhibit more recency bias in *Late* than *Early* also survives. We confirm this in Table A-2, which shows that forecast credibility is almost two-fold larger in *Early* than in *Late*, and this difference in estimate forecast credibility across histories is statistically significant ($p < .001$).

Recency Bias for $\gamma \neq 0$

Table A-3: Estimated Values of λ

	γ_0	$\gamma_{HistAvg}$
Early	0.245	0.275
	(0.0170)	(0.0160)
Late	0.622	0.560
	(0.0198)	(0.0222)

Note: This table presents estimates of recency bias in the *Early* and *Late Timing* treatments. It distinguishes between conditions where we assume subjects perceive no systematic forecast error (γ_0 column) and where they do perceive a systematic forecast error ($\gamma_{HistAvg}$ column). Estimates are obtained via OLS regression, with robust standard errors reported in parentheses.

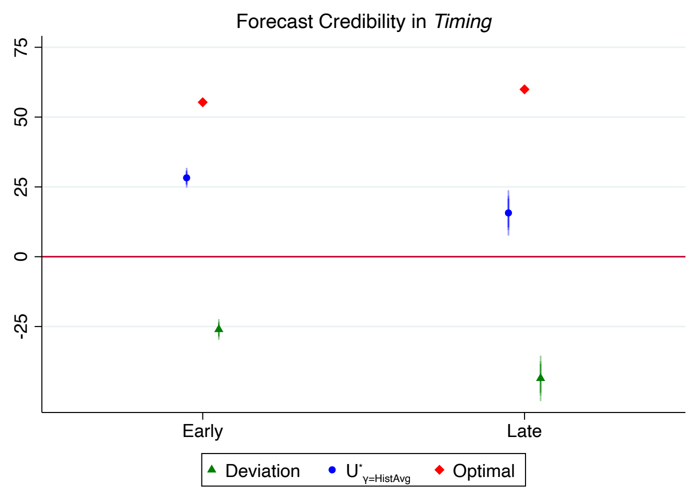


Figure A-2: Perceived forecast credibility in *Timing*: $\gamma_{HistAvg}$
Note: This figure depicts estimates of forecast credibility for the *Early* and *Late Timing* treatments (represented by blue circles), under the assumption that subjects perceive a systematic forecast error, as specified in Equation (A.22). It also illustrates the Bayesian optimal level of updating as a benchmark (red diamonds) and deviations from this benchmark (green triangles). Shaded bands around the point estimates indicate 99% (lightest), 95%, and 90% (darkest) confidence intervals.

Dynamics of Credibility for $\gamma \neq 0$

A2.1.3 Dynamics of forecast credibility

We further explore our bias-adjusted forecast credibility estimates to gain some insight into the dynamics of perceived forecast credibility. To do this, we use the fact that the central bank's historical forecast precision in *Terrible* from *ForecastPerformance* is identical to the central bank's forecast precision in the final year of *Late* from *Timing*. By comparing estimated forecast credibility economic histories, we can learn something

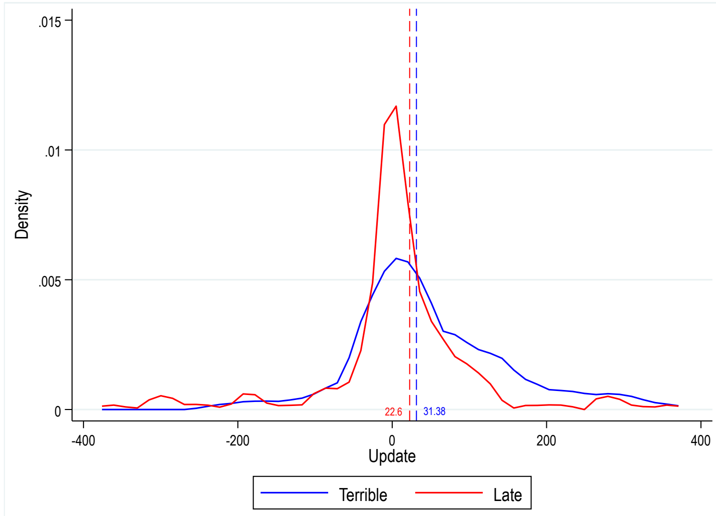


Figure A-3: *Consistent-Terrible* vs. *Late*

Note: This figure depicts kernel density estimates of forecast credibility assuming subjects perceive a systematic forecast error from the central bank,$u^*_{\gamma_{HistAvg}}$ as in Equation (A.22), from *Late* (red, dashed line) and *Terrible* (blue, solid line). The corresponding vertical dashed lines denote the mean of $u^*_{\gamma_{HistAvg}}$.

about how quickly forecast credibility erodes. Similarly, we can also use the fact that the central bank's historical forecast precision in *Great* is identical to the bank's historical forecast precision in the final year of *Early*.

We first compare perceived credibility measures from *Terrible* and *Late*, which we depict as kernel density estimates in Figure A-3. First we note that the mean level of perceived credibility is not statistically different across treatments ($p = .704$, two-sample t-test).[2] Overall, results suggest that *Terrible* forecast precision for a single year leads to perceived forecast credibility that is as low, on average, as if subjects see *Terrible* forecast precision over the entire economic history.

However, this does not hold when comparing *Great* and *Early* in Figure A-4. Instead, we see that the mean level of perceived forecast credibility is significantly higher in *Great* than in *Early* ($p < .001$) and that the distributions are highly significantly different ($p < .01$, Kolmogorov-Smirnov (KS) test). This suggests that seeing *Great* forecast performance over the full sample history leads to significantly higher credibility than seeing it over only the last year.

These results align with our estimated weighting functions. In *Late*, the deterioration of the central bank's forecast performance induces a very strong recency bias. On average, participants in that experiment base the majority of their perception of the central bank's forecast credibility on the very last historical observation. In *Early*, an analogous improvement in forecast precision does not induce the same degree of recency bias. Though participants primarily focus on the final year of forecast performance following both histories, our estimated weighting function from *Early* exhibits a fatter right tail. Intuitively, this suggests that poor forecast performance lingers longer in people's minds when deciding how much faith to place in the central bank's ability to predict inflation accurately.

[2]Results from a Kolmogorov-Smirnov test indicate that the perceived credibility is slightly lower in *Late* than in *Terrible* ($p = .044$). This is driven by the slightly lower mass of positive updates in *Late*.

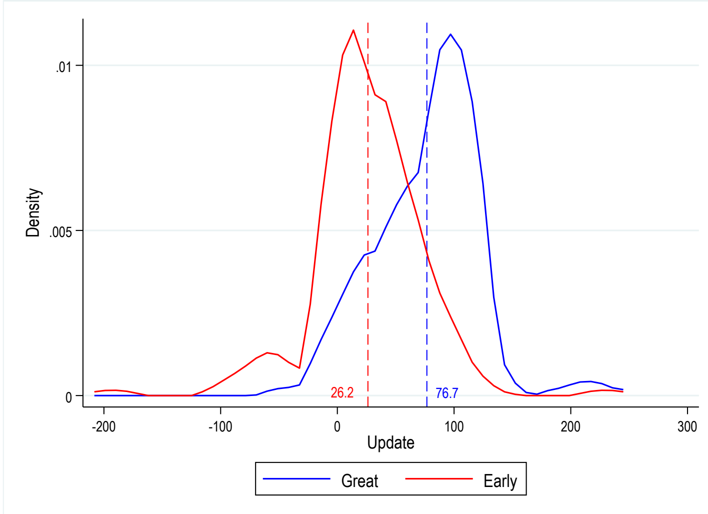


Figure A-4: *Consistent-Great* vs. *Early*

Note: This figure depicts kernel density estimates of forecast credibility assuming subjects perceive a systematic forecast error from the central bank, $u^*_{\gamma_{HistAvg}}$ as in Equation (A.22), from *Early* (red, dashed line) and *Great* (blue, solid line). The corresponding vertical dashed lines denote the mean of $u^*_{\gamma_{HistAvg}}$.

A2.1.4 *Contextual Communication*

This section reconsiders our primary result from *Contextual Communcation* assuming that $\gamma_{HistAvg}$. We depict these results in Figure A-5.

We note two main points. First, communication still increases perceptions of the central bank's forecast credibility. In fact, assuming $\gamma_{HistAvg}$ *strengthens* the estimated credibility gains in most treatments, and so much so in *Exogenous + Better* that the central bank recovers the Bayesian optimal level of forecast credibility via its contextualizing statement.

Second, acknowledging that the central bank's historical average forecast precision was lower than peer forecasting institutions is quite detrimental. This effect is most pronounced in *Exogenous + Worse*, though statistically the effects are similar for *Endogenous + Worse*. Announcing the bank outperformed peer forecasting institutions yields $u^*_{\gamma_{HistAvg}} = 43.85$ while announcing relative under-performance yields $u^*_{\gamma_{HistAvg}} = 10.37$. These differences are highly significant ($p < .001$, two-sided t-test).

A3 Sensitivity Analysis

Our baseline analysis uses data Winsorized at 5th and 95th percentile so that extreme outliers do not drive our results. We chose these cutpoints because they were sufficient to eliminate extreme outliers in all instances for our data so that cutpoints remain consistent throughout. However, we understand that Winsorizing our data introduces a decision point into our analysis. Because of this, we explore in this section the sensitivity of results to the choice of cut points. To do this, we reproduce results from *Forecast Performance* (Figure A-6) and *Timing* (Figure A-7), which together comprise all eight

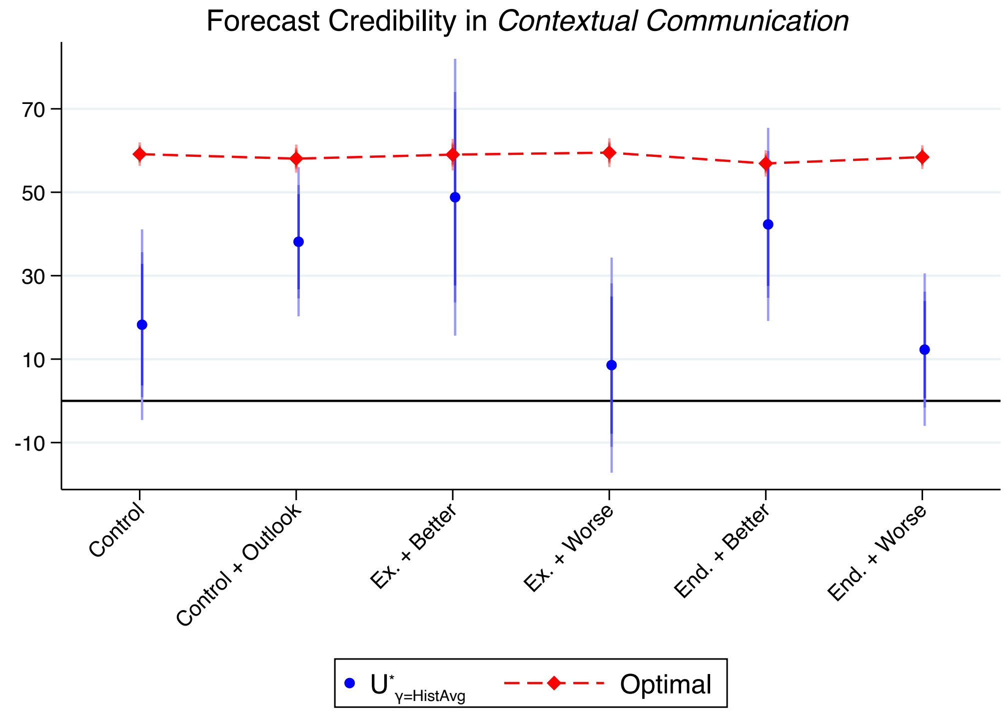


Figure A-5: Estimated forecast credibility in *Contextual Communication* treatments using $u^*_{\gamma_{HistAvg}}$
Note: This figure presents forecast credibility estimates for *Contextual Communication* treatments (blue circles), assuming subjects perceive a systematic forecast error ($\gamma_{HistAvg}$). The red diamonds represent the Bayesian optimal weighting of the central bank's forecast in the updated inflation expectation of a rational Bayesian participant, who equally considers all available historical information. Shaded bands around the markers denote 99% (lightest), 95%, and 90% (darkest) confidence intervals.

histories that we use in this experiment. We show estimates of forecast credibility using cutpoints that retain data relative to our baseline (1st and 99th percentile) and trim additional data relative to the baseline (10th and 90th percentiles).

A3.1 *Forecast Performance*

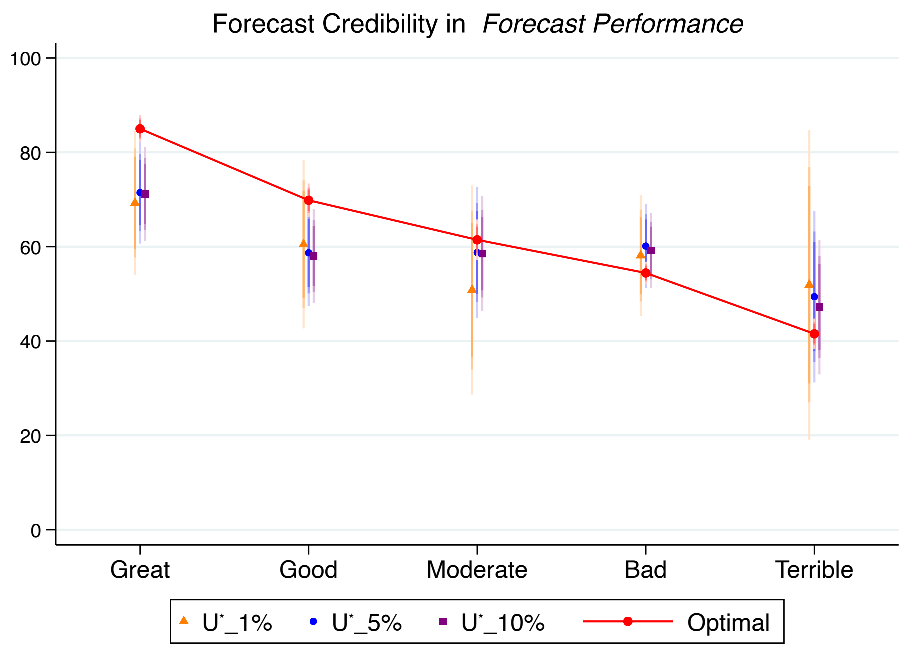


Figure A-6: Sensitivity Analysis for *Forecast Performance* results

Note: This figure provides estimates of forecast credibility, u^*, in *Forecast Performance* using data that is Winsorized at the 1st and 99th percentiles (orange triangles), the 5th and 95th percentiles (blue circles), and the 10th and 90th percentiles (purple squares). Shaded bands around the point estimates indicate 99% (lightest), 95%, and 90% (darkest) confidence intervals.

A3.2 *Timing*

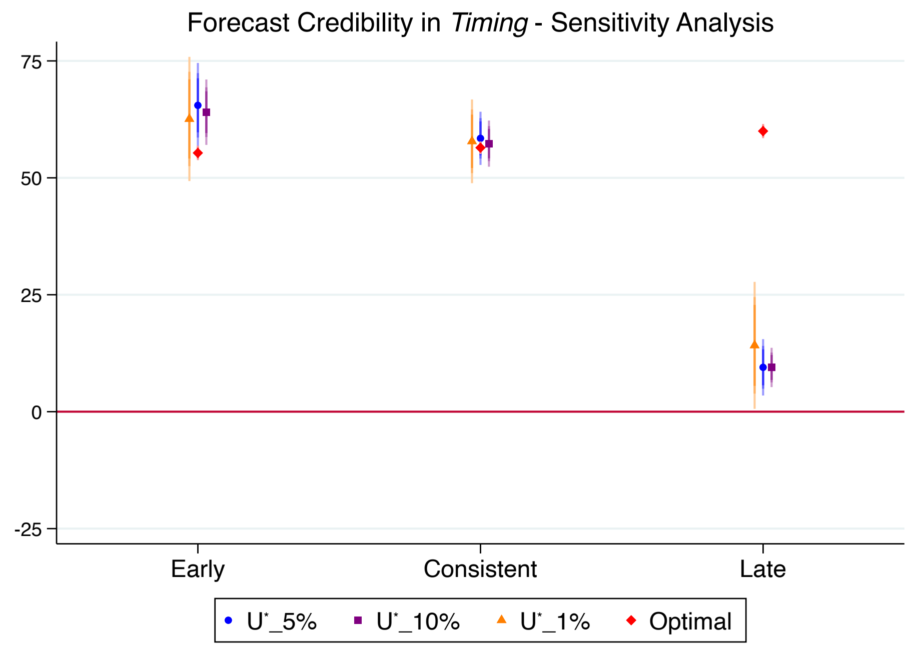


Figure A-7: Sensitivity Analysis for *Timing* results
Note: This figure presents forecast credibility estimates for the *Timing* treatments using Winsorized data. Orange triangles indicate data Winsorized at the 1st and 99th percentiles, blue circles at the 5th and 95th percentiles, and purple squares at the 10th and 90th percentiles. Shaded bands around the point estimates denote confidence intervals of 99% (lightest), 95%, and 90% (darkest).

A4 Alternative Assumptions for Endogenous Credibility

This section explores the robustness of our results Section 6.3. We demonstrate that the non-contemporaneous timing of household expectations formation does drive our result that endogenous credibility leads to persistent inflation dynamics.

A4.1 Alternative timing assumption

In Section 6.3, we assume the boundedly-rational household forms the backward-looking component of its expectation, $\mathbb{E}^{\mathrm{HH}}_{t,\mathrm{BR}}\{\pi_{t+1}\}$, based on inflation observed in the previous period, $t-1$, and its forecast of inflation for that period that it formed in the period before that, $t-2$. In some sense, this assumption captures the idea that a boundedly rational household with imperfect information, imperfect processing, or both does not form its inflation expectation while also realizing inflation in the current period. Instead, the household observes inflation with a lag and forms its expectation based on its observation of lagged inflation and lagged forecast errors.

One might be concerned that assuming this timing scheme introduces inertia into inflation dynamics and that the timing scheme can explain most or all of the persistence in inflation we discuss in Section 6.3. We address this concern by updating the timing scheme so that households form the boundedly rational component of its inflation prior as

$$\mathbb{E}^{\mathrm{HH}}_{t,\mathrm{BR}}\{\pi_{t+1}\} = \mathbb{E}_{t-1}\{\pi_t\} + \eta_\pi(\pi_t - \mathbb{E}_{t-1}\{\pi_t\}) \tag{A.23}$$

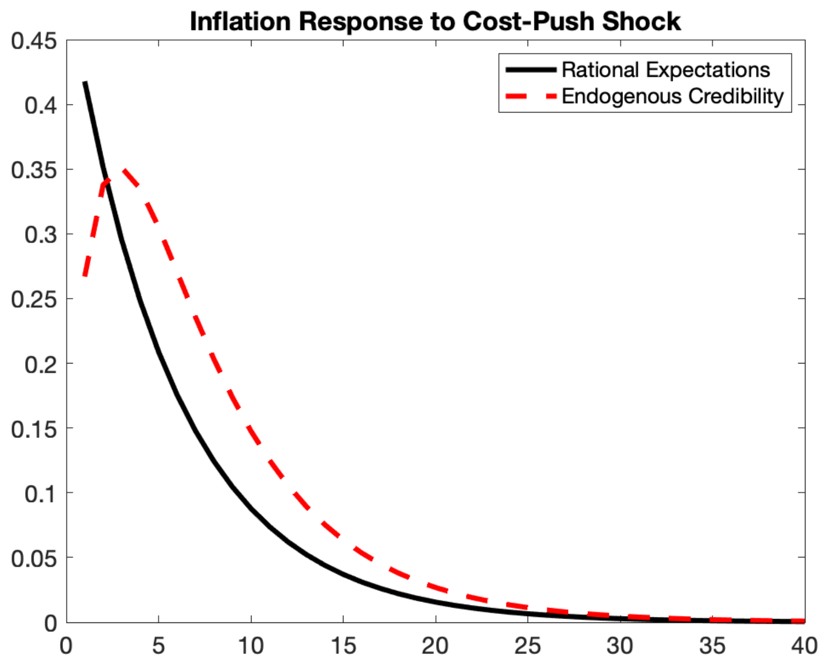


Figure A-8: Response of Inflation to Cost-Push Shock With Endogenous Credibility
Note: This figure depicts the response of inflation to a cost-push shock assuming that central bank forecast credibility is fully endogenous (red dashed line) and assuming a fully-informed and rational agent (solid black line). Here, we assume that the boundedly rational component of the agent's inflation expectation uses a contemporaneous timing scheme, as described by Equation (A.23).

This timing scheme implies that, though boundedly rational, the household is capable of concurrently observing contemporaneous inflation, forming an inflation expectation, and acting on that expectation. We reconsider how inflation responds to a 1pp cost-push shock using this alternative timing. We display the result of this exercise in Figure A-8, which shows inflation dynamics under rational expectations (black solid line) and with endogenous central bank forecast credibility (red dashed line).

Inflation dynamics, though slightly different, are qualitatively very similar to those observed in Section 6.3. Further, our main point remains – endogenous credibility can lead to prolonged bouts of higher inflation following an exogenous supply shock.

A5 Instructions

This section contains our experimental instructions for all treatments.

A5.1 *Contextual Communication*, *ForecastPerformance*, *Timing*, and *ReversedShock* 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 labeled '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.

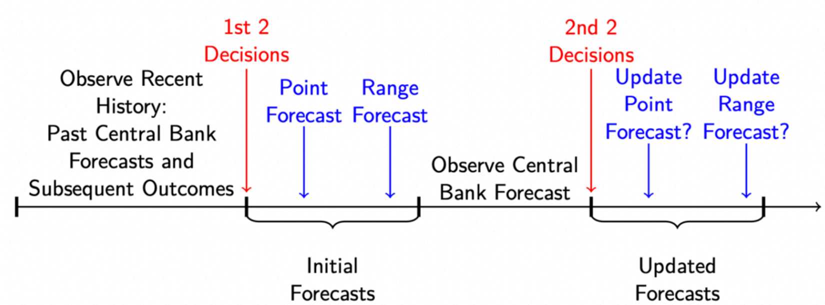


Figure: Experimental Timeline: A single decision period

1. We provide 12 quarters of history of inflation (blue line and dots) (*black line and dots* in *ReversedShock*) alongside the central bank's corresponding forecasts for those quarters (black line and dots) (*blue line and dots* in *ReversedShock* treatments). We also provide a summary of the central bank's historical forecast performance (text, next to the left side of the chart) that includes absolute forecast errors for each year and for the overall historical period.
   - 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 13)
   - 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 and lower bound of your 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 randomly chooses 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.

A5.1.1 *Medium-Term* instructions

Your Objective in the Experiment

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

You will provide two types of average inflation forecasts:

- **Point Forecast:** Your 'Point Forecast' of inflation is your best guess of the exact value inflation will be, on average, over a given time span.
- **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 average 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.
- **Average Inflation:** Average inflation is what you think inflation will be, on average, for a given time span. For example, suppose inflation is 3% in one quarter and then 4% in the next. Average inflation for these two quarters is $\frac{3\%+4\%}{2} = \frac{7}{3}=$ 3.5%. Suppose instead, we want to know of average inflation for a year where inflation was 6% in the first quarter, 1% in the second, 3% in the third, and 2% in the fourth. Average inflation for the year would be $\frac{6\%+1\%+3\%+2\%}{4}$ $\frac{12}{4} = 3\%$. Remember, we are asking you to forecast average inflation.

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.

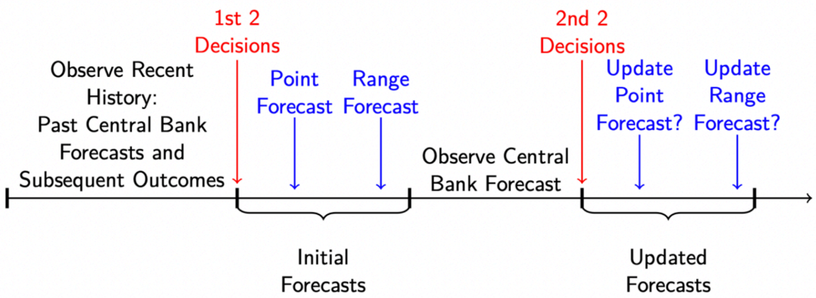


Figure: Experimental Timeline: A single decision period

1. We provide a 12-quarter history of inflation (blue line and dots) alongside the central bank's corresponding one-period-ahead forecasts for those quarters (black line and dots). For example, a one-period-ahead forecast would be if the central bank forecasts inflation for the fourth quarter of a year while in the third quarter of that same year. We also provide a summary of the central bank's historical forecast performance (text, next to left side of chart) that includes absolute forecast errors for each year and for the overall historical period. This historical data is **quarterly**. We provide **12 quarters** worth of historical data, which is equivalent to **three years** of data. We then ask you to forecast average inflation **for the next three years** .
   - Note that the difference between the blue dot and black dot 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 average inflation for the **next three years**.
   - Your corresponding range forecast of average inflation.
3. After forming your Initial Forecasts, we will reveal to you the central bank's forecast of average inflation for the next three years (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 average 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 average inflation for the forecasted time span 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 actual average 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 average 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 average 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 average 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 randomly chooses 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.

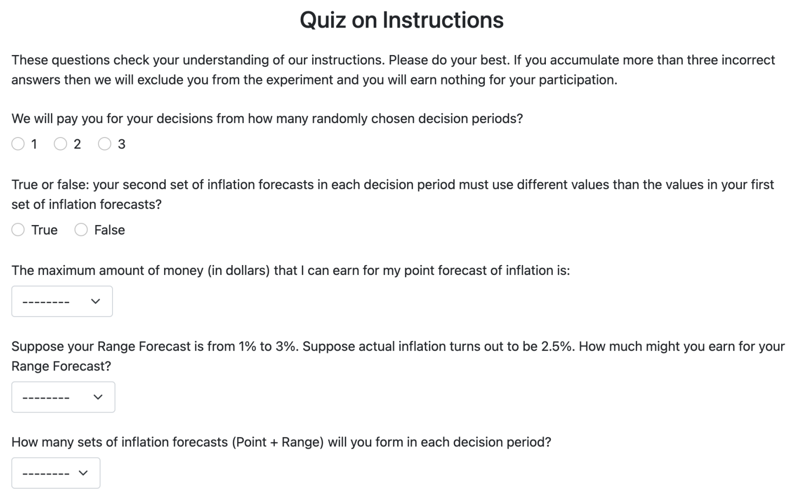
Providing your Point Forecast:

- 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 average 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.

A5.2 Comprehension Quizes, Survey Questions, & Economic Literacy Questions



Quiz on Instructions

These questions check your understanding of our instructions. Please do your best. If you accumulate more than three incorrect answers then we will exclude you from the experiment and you will earn nothing for your participation.

We will pay you for your decisions from how many randomly chosen decision periods?

1 2 3

True or false: your second set of inflation forecasts in each decision period must use different values than the values in your first set of inflation forecasts?

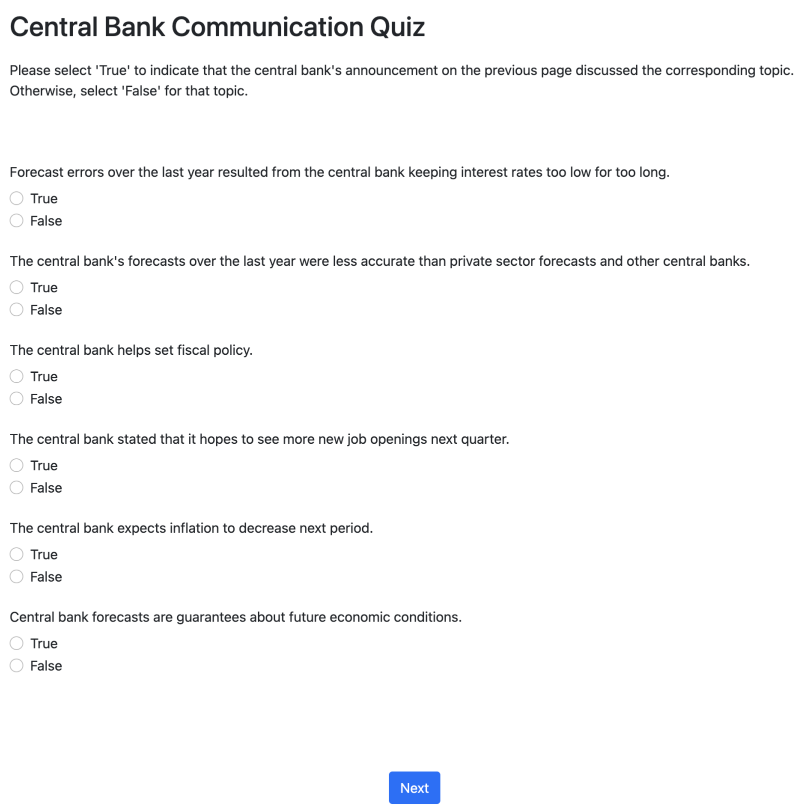
True False

The maximum amount of money (in dollars) that I can earn for my point forecast of inflation is:

Suppose your Range Forecast is from 1% to 3%. Suppose actual inflation turns out to be 2.5%. How much might you earn for your Range Forecast?

How many sets of inflation forecasts (Point + Range) will you form in each decision period?

Figure A-9: This is a screenshot of the comprehension quiz faced by all subjects before beginning our experiment. Subjects who failed the quiz three times were excluded.



Central Bank Communication Quiz

Please select 'True' to indicate that the central bank's announcement on the previous page discussed the corresponding topic. Otherwise, select 'False' for that topic.

Forecast errors over the last year resulted from the central bank keeping interest rates too low for too long.

True

False

The central bank's forecasts over the last year were less accurate than private sector forecasts and other central banks.

True

False

The central bank helps set fiscal policy.

True

False

The central bank stated that it hopes to see more new job openings next quarter.

True

False

The central bank expects inflation to decrease next period.

True

False

Central bank forecasts are guarantees about future economic conditions.

True

False

Next

Figure A-10: This screenshot provides an example of the comprehension quiz we administered to subjects in our low-frequency communication treatments. The quiz came after subjects updated both their point and rage forecasts.

A5.2.1 Survey Questions

Survey - Trust

Please indicate **how much you trust** each institution where 1 means 'not at all' and 5 means 'perfectly':

	1	2	3	4	5
Police	○	○	○	○	○
Legislature	○	○	○	○	○
Judiciary	○	○	○	○	○
Federal Reserve	○	○	○	○	○
Government	○	○	○	○	○

Next

Survey - Information

Please indicate **how much you engage with** with the following forms of communication where 1 means 'not at all' and 5 means 'Engage a lot ':

	1	2	3	4	5
Media reports about the Economy?	○	○	○	○	○
Media reports about the Federal Reserve	○	○	○	○	○
Media reports about the Government	○	○	○	○	○
Communication directly from Federal Reserve	○	○	○	○	○
Communication directly from the Government	○	○	○	○	○

Next

Survey - Understanding

Please indicate **how well you understand the role** of each institution where 1 means 'not at all' and 5 means 'perfectly':

	1	2	3	4	5
Police	○	○	○	○	○
Legislature	○	○	○	○	○
Judiciary	○	○	○	○	○
Federal Reserve	○	○	○	○	○
Government	○	○	○	○	○

Next

Figure A-11: Survey questions asking subjects about their level of trust and understanding for various U.S. institutions, and about preferences for information sources.

A5.2.2 Economics Literacy Questions

1. WagesRecession: What do you think happens to real wages (i.e. purchasing power) during a recession?
    - Fall on average
    - Unchanged on average
    - increase on average
    - Don't know
2. EmploymentRecession: What do you think happens to employment during a recession?
    - Goes up
    - Stays the same
    - Goes down
    - Don't know
3. InflationRecession: On average, what do you think happens to inflation during a recession?
    - It increases
    - It decreases
    - It remains unchanged
    - Don't know
4. BorrowRates: Suppose you need to borrow money. Which condition is best for you?
    - Interest rates are low
    - Interest rates are about average
    - Interest rates are high
    - Interest rates are irrelevant
    - Don't know
5. SavingRates: Suppose you are saving money. Which condition is best for you?
    - Interest rates are low
    - Interest rates are about average
    - Interest rates are high
    - Interest rates are irrelevant
    - Don't know
6. WhoSetsRates: Which U.S. institution sets the interest rate?
    - Congress

- President
- Federal Reserve
- Don't know

7. AvPi: What was the last reported level of inflation in the U.S.?

   - $< -1\%$
   - -1% to 1%
   - 1% to 3%
   - 3% to 5%
   - 5% to 8%
   - $> 8\%0$

8. InflationPreference: How much inflation do you think is good for the economy?

   - $< -1\%$
   - -1% to 1%
   - 1% to 3%
   - 3% to 5%
   - 5% to 8%
   - $> 8\%$

A5.3 Central Bank Messages in *Contextual Communication*

T12 - Control:
The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain and forecasts are never perfect.

T13 - Control+Outlook
The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain and forecasts are never perfect.

Over the last year, our forecasts under-predicted inflation. Our best guess is that inflation will decrease next quarter.

T14 - Exogenous + Better
The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain, and forecasts are never perfect.

Over the last year, our forecasts under-predicted inflation. This is because the pandemic lasted longer than initially expected and caused supply shortages.Our forecasts over this period were **more accurate** than private sector forecasts and other central banks. Our best guess is that inflation will decrease next quarter.

T15 - Exogenous + Worse
The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain and forecasts are never perfect.

Over the last year, our forecasts under-predicted inflation. This is because the pandemic lasted longer than initially expected and caused supply shortages. Our forecasts over this period were **less accurate** than private sector forecasts and other central banks. Our best guess is that inflation will decrease next quarter.

T16 - Endogenous + Better
The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain and forecasts are never perfect.

Over the last year, our forecasts under-predicted inflation. This resulted from interest rates being too low for too long. Our forecasts over this period were **more accurate** than private sector forecasters and other central banks. Our best guess is that inflation will decrease next quarter.

T17 - Endogenous + Worse

The Fed uses interest rate policy to stabilize prices and keep employment high. We base monetary policy on how healthy the economy is now and how healthy we think it will be in the future. We use forecasts to guide our decisions. We do our best when making forecasts but the world is uncertain and forecasts are never perfect.

Over the last year, our forecasts under-predicted inflation. his resulted from interest rates being too low for too long. Our forecasts over this period were **less accurate** than private sector forecasters and other central banks. Our best guess is that inflation will decrease next quarter.

A5.4 Salience of Real-World Inflation

We conducted a series of five treatment waves unrolled sequentially over the course of approximately one year, starting in February 2022 and ending in March 2023. We described the timing of our treatments in Table A-4. During this time, the United States – and many other economies – was experiencing considerable inflation, which constituted the first salient change in price dynamics in more than a decade. Though inducing preferences in an experimental setting ought to insulate results from real-world economic dynamics, Petersen and Rholes (2022) provides some evidence that central bank communication may be susceptible to real-world shocks. In the context of this experiment, one might be concerned about differences in headline inflation throughout our treatment waves, were these differences salient to participants, did it change inflation preferences, and should we be concerned that this matters for our results?

First, we note the last reported value for CPI inflation leading into each treatment wave Table A-4, was at its lowest in our final wave of treatments at approximately 6% and at its highest in our third treatment wave at about 8%. Though the two-percent variation in inflation is not trivial, it does constitute considerably less variation than what participants experienced in the time leading up to our experiment when inflation rose from approximately 2% to 7.5%.

Was this variation salient to participants? We show in Figure A-12 (main treatment waves, and Figure A-13 (robustness treatment waves), that the large majority of subjects were aware of the most recent measure of headline inflation. These same figures also show that preferences for real-world inflation were identical across treatment waves, indicating that the salient difference in actual inflation did not change inflation preferences. In all treatment waves, the majority of participants indicated a desire for low inflation ranging from 1% to 3%, with approximately 90% or more of our participants indicating a preference for inflation between -1% and 3%.

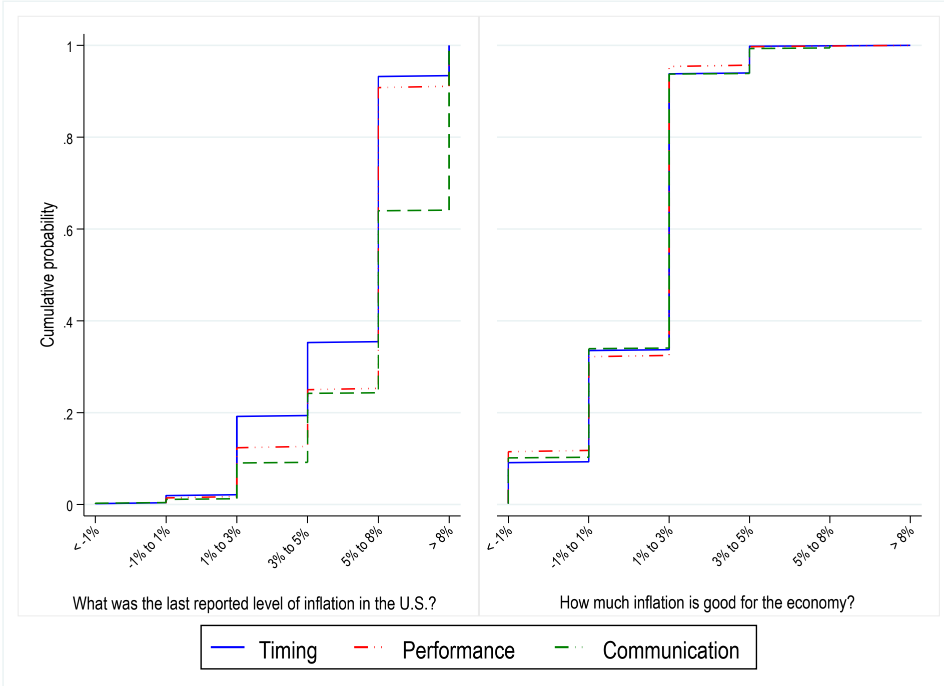


Figure A-12: This figure depicts cumulative distribution functions of participants' inflation preferences alongside perceptions of prevailing inflation for *Timing ForecastPerformance*, and *Contextual Communication* treatments.

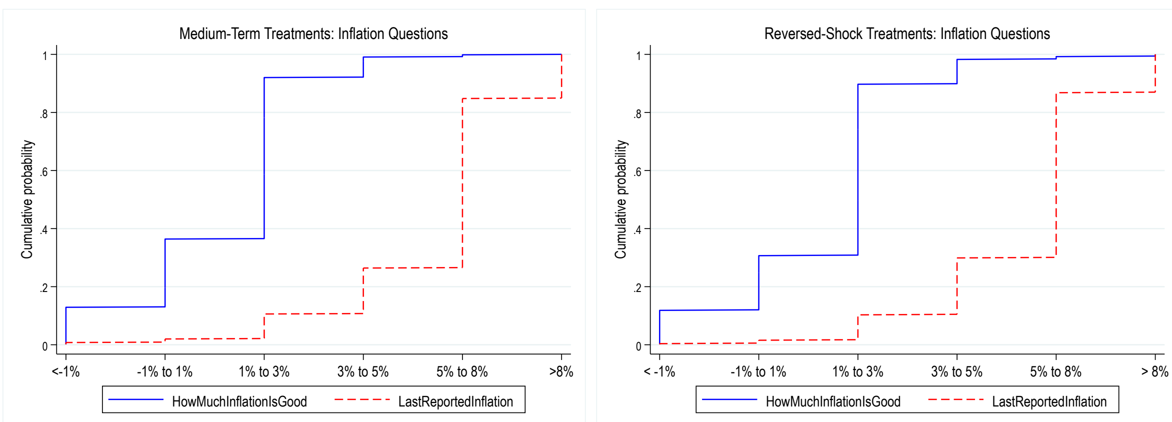


Figure A-13: This figure depicts cumulative distribution functions of participants' inflation preferences alongside perceptions of prevailing inflation for *Medium-Term* treatments.

Treatments	Name	Dates	Last Reported CPI Inflation
T1 - T6	*Timing*	Feb. 25 - Feb. 28, 2022	7.5% in Jan.
T7 - T14	*ForecastPerformance*	Mar. 28 - Mar. 29, 2022	7.9% Feb.
T15 - T20	*Contextual Communication*	May 27 - May 28, 2022	8.2% in Apr.
T21 - T26	*Medium-Term*	Dec. 14 - Dec. 15, 2022	7.13% in Nov.
T27 - T28	*ReversedShock*	Mar. 24 - Mar. 25, 2023	5.98% in Feb. 2023

Table A-4: Actual inflation during each treatment wave.

A6 Tables and Figures

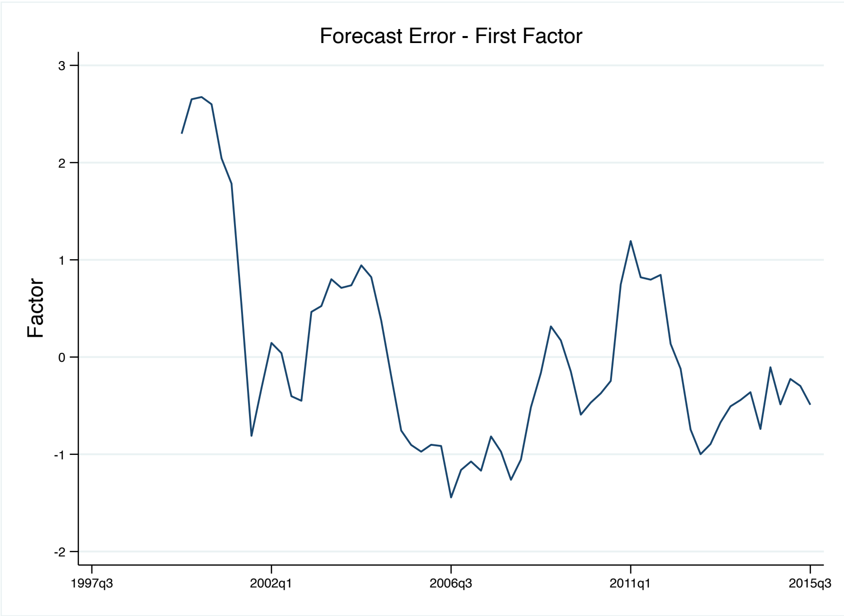


Figure A-14: Bank of England Forecast Performance: Factor Summary Variable
Note: This figure depicts our measure of the Bank of England's forecast error over time. The Bank of England forecasts the current and at least the next eight quarters of inflation in each quarter of our sample. To form our measure of the BoE's composite forecast error, we collapse these nine quarterly forecast errors via principal component factor analysis and use the first principal component that results from this exercise. This represents the linear combination of errors that captures the most common variation across individual forecast errors over time.

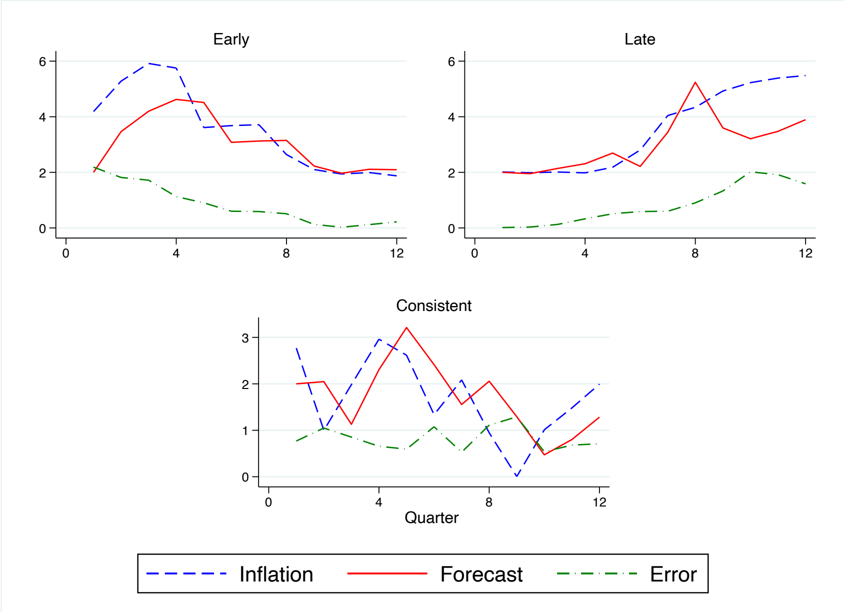


Figure A-15: *Early*, *Late*, and *Consistent-Bad* – Timing Histories
Note: This figure presents three subpanels labeled *Early* (top left), *Late* (top right), and *Consistent* (bottom). Each subpanel shows historical inflation trends (blue dashed lines) alongside inflation forecasts (solid red lines) provided to participants in our study for the respective histories. Green dash-dot lines represent the corresponding absolute forecast errors.

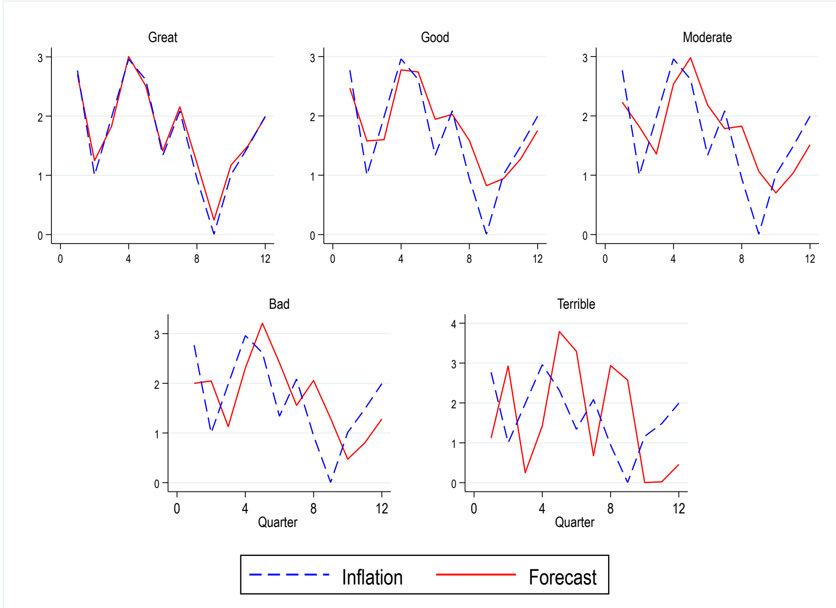


Figure A-16: Alternative versions of *Consistent* used in *Forecast Performance*
Note: *Early*, *Late*, and *Consistent-Bad* – Timing Histories *Note*: This figure presents subpanels labeled *Great* through *Terrible* (bottom right) that show historical inflation trends (blue dashed lines) alongside inflation forecasts (solid red lines) provided to participants in our study for the respective histories.

Summary of Bonus Payments					
Early					
N	**Treatments**	**Initial - Point**	**Initial - Range**	**Updated - Point**	**Updated - Range**
516	*Timing*	.82	.40	.82	.40
348	*ForecastPerformance*	.83	.40	.83	.41
719	*Contextual Communication*	.81	.37	.82	.39
Consistent					
	Timing	.33	.01	.47	.05
	ForecastPerformance	.34	.01	.55	.13
	Contextual Communication	.33	.01	.47	.06
Late					
	Timing	.68	.23	.68	.23
	ForecastPerformance	.67	.22	.66	.22
	Contextual Communication	.69	.23	.70	.25

Table A-5: XXXXX

Note: XXXX

Summary of Bonus Payments - ForecastPerformance					
N	**Treatments**	**Initial - Point**	**Initial - Range**	**Updated - Point**	**Updated - Range**
90	*Great*	.37	.00	.72	.25
90	*Good*	.34	.01	.56	.14
72	*Moderate*	.34	.01	.54	.09
180	*Bad*	.32	.00	.48	.05
96	*Terrible*	.32	.02	.38	.02

Table A-6: XXXXX

Note: XXXX

Summary of Bonus Payments - Communication					
N	**Treatments**	**Initial - Point**	**Initial - Range**	**Updated - Point**	**Updated - Range**
130	*Control*	.70	.23	.69	.25
124	*Control+Outlook*	.71	.21	.75	.25
107	*Exogenous+Good*	.66	.23	.67	.24
121	*Exogenous+Bad*	.69	.24	.69	.22
122	*Endogenous+Good*	.68	.23	.73	.26
115	*Endogenous+Bad*	.67	.22	.69	.25

Table A-7: XXXXX

Note: XXXX