Building Central Bank Credibility: The Role of Forecast Performance*

Michael McMahon† Ryan Rholes‡

This Draft: November 2022

Abstract

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

Keywords: Expectation formation, Forecasting, Monetary Policy Communication

JEL Codes: E52, E58

*McMahon gratefully acknowledges financial support from the European Research Council (Consolidator Grant Agreement 819131). We are grateful to seminar participants at the University of Oxford, the Economic Science Association's (ESA) World Meeting 2022 at MIT, and the ESA European Meeting 2022 at the University of Bologna . We also thank Luba Petersen and Johannes Abeler for helpful comments. The views presented in the paper are those of the authors alone and do not represent the official views of anyone else. Any errors remain ours alone.

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

‡University of Oxford. Email: ryan.rholes@economics.ox.ac.uk

1 Introduction

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

Central bank communication arguably serves as central banks' primary tool of expectations management which typically involves high-frequency and lower-frequency communication. Such communication policies are widespread; most central banks now devote considerable resources to communicating about the evolution of important macroeconomic variables via projections, public engagement, and policy reports. Both types of communication can achieve inflation control by influencing how agents form their inflation expectations, as described above. Hence, open-mouth operations are now an indispensable component of monetary policy.

Does communication influence expectations? In workhorse monetary models, rational agents appropriately incorporate information from the central bank and other sources to form the best possible expectations. In a sense, these models assume the central bank is always fully credible, which limits our ability to study the determinants and dynamics of credibility or the role of communication when banks aren't fully credible. In practice, central banks worry about their credibility, which is necessary for the transmission of communication policy (Blinder 2000).

In this project, we introduce a novel experimental framework that we use to provide causal evidence on the relationship between forecast performance and forecast credibility.

We focus on inflation forecast credibility for at least a few reasons. Foremost, central bankers care deeply about credibility, viewing it as a necessary component of effective communication (Blinder (2000)). Further, price stability is a mandate for many central banks and our workhorse models tell us our expectations about future prices can impact current prices. Thus, we focus on what policy makers seem to believe is necessary component of for effectively achieving price stability, which is a mandate for many central banks. This relationship is important because inflation forecasts are a key input into the policy decision in an inflation targeting framework (Svensson 1997).

While there are many dimensions to credibility, we focus on forecast credibility. Given the dictionary definition of credibility – "the ability to have one's statements accepted as factual or one's professed motives accepted as the true ones" (Blinder 2000) – forecast credibility involves the central bank's forecast being used as a valuable signal of the likely evolution of future inflation. We measure this as the weight that agents place on the central bank's inflation forecast when updating their estimate of inflation.

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

In this paper, we develop an experimental framework to study the determinants and dynamics of central bank credibility. Specifically, we study how perceptions of a central bank's forecast credibility depend on its historical forecast performance. We focus on the banks historical average forecast performance, how credibility responds discrete changes in performance, and whether banks can use high-frequency communication to augment credibility (i.e. talk its way out of a low credibility position).

Macroeconomists have typically used the learning-to-forecast (LTF) framework to study expectations experimentally, which at its core consists of 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)).

We relate most closely 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.

Participants in our experiment act as atomistic inflation forecasters who provide two sets of one-period-ahead point and range forecasts of inflation in three independent decision periods (e.g. 'Initial Forecasts' and 'Updated Forecasts'). We begin each decision period by providing subjects 12 quarters of historical economic data consisting of the central bank's inflation forecasts alongside actual inflation. Subjects provide Initial Forecasts (priors) for quarter 13 based on this historical data. We then reveal the central bank's quarter 13 inflation forecast and allow subjects to update their own density projection (i.e. Updated Forecasts or posterior). This approach allows us to measure forecast credibility, as defined above, by comparing their Updated Forecasts and their Initial Forecasts. Differences in economic histories, which we simulate to match moments real-world data, constitute within-subject treatment variation in our experiment.[1]

Our design departs from the typical LTF framework in two key ways. First, inflation does not evolve endogenously in our setting.[2] Second we use an individual-choice setting to eliminate strategic uncertainty. These design choices yield an environment where participants necessarily view inflation as exogenous and do not view expectation formation as a coordination game, which we think better aligns our participants' experiences with those of participants in household surveys like the University of Michigan's Survey of

[1] 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 as the way to calibrate the size of the forecast errors.

[2] We are not the first to make this design decision. For a previous example, see Burke and Manz (2014), Armantier et al. (2016).

Consumers, the New York Federal Reserve's Survey of Consumer Expectations, or the European Central Bank's Consumer Expectations Survey.

Several key findings emerge. We consider how historical forecast performance influences the central bank's forecast credibility. To do this, we create a series of economic histories that control for the pattern of historical forecasts errors while scaling their magnitude. We find that participants behave qualitatively like Bayesians. Central bank forecast credibility does depend significantly on historical performance. However, the link between performance and credibility is not as sharp as predicted by theory, with participants updating toward the central bank's forecast too much following poor performance and too little following stellar forecast performance.

Second, we explore how the timing of forecast errors influences the central bank's forecast credibility. To do this, we create a set of three economic histories than hold constant the bank's historical performance while varying the pattern of forecast errors. In theory, a Bayesian agent should find the central bank equivalently credible across these histories. However, we find that our participants exhibit considerable recency bias. Subjects tend to overweight signals from a central bank whose forecast performance was well above average in recent quarters, and severely underweight signals whenever the central bank's performance was well below average in recent quarters.

Additionally, we test whether a central bank can use high-frequency communication to bolster credibility. 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 policy errors, and report performing better or worse than peer forecasting institutions. First, we find that high-frequency communication that reinforces forecasts without providing additional information can 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 high-frequency communication. Instead reporting that the bank out performed peer institutions bolsters credibility. We see that high-frequency communication can recover forecast credibility whenever forecast errors are less bad than peer forecasters.

One might question whether our results are applicable to real-world markets or are instead artifacts of our stylistic setting. We make some 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 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.

2 Central bank signals and forecast updating

We use a Bayesian framework (similar to Morris and Shin (2002), for example) to emphasize how central bank signals optimally influence a Bayesian participant's decision to update her inflation forecasts. We use this framework to elucidate how we will measure updating in our results section and guide our hypotheses.

Suppose participant i forms a prior belief about inflation given by:

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

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

The central bank provides a potentially biased signal:

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

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 return to the effect of bias later (very little changes in the analysis), 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 subjects initial range forecasts as incentivized measures 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 percieve 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:

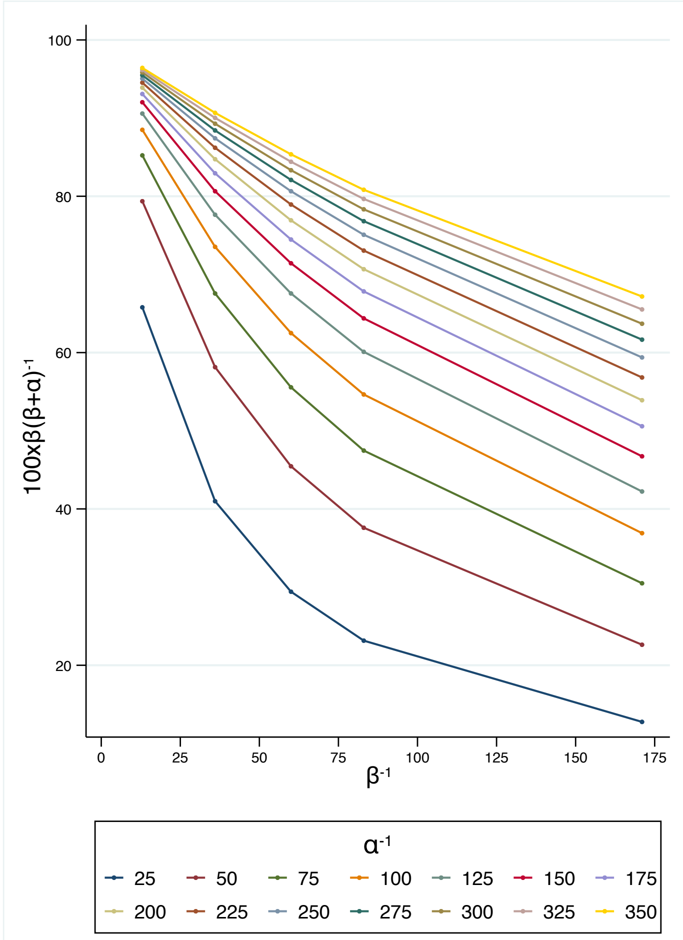

Figure 1: This figure shows the optimal level of updating in percentage terms (y-axis) prescribed by Equation (7) for different levels of a central bank precision (x-axis). Each line denotes a different level of participant forecast uncertainty ranging from 25 basis points (bottom line) to 350 basis points (top line) in increments of 25 basis points.

1. For any given precision of the central bank signal, as the precision of the prior increases, $\alpha \downarrow$, 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, $\beta \downarrow$, 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.

2.1 The Role of Bias

We now return to the issue of a bias in the central bank forecast given by $\gamma \neq 0$. In theory, we typically model no systematic component to the central bank's forecast error. However, even if this were true, we may have $\gamma \neq 0$ 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). To account for this, we can rewrite Equation (2) as:[3]

$$\pi_{cb} - \gamma = \pi + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \frac{1}{\beta}\right). \tag{6}$$

Equation (6) 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 (7), 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 (α_i^{-1}, β^{-1}).

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

$$u^*_{\gamma,i} \equiv \frac{\mathbb{E}(\pi|\pi_{cb}) - \bar{\pi}_i}{(\pi_{cb} - \gamma - \bar{\pi}_i)} \tag{7}$$

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 $\gamma \neq 0$, $\frac{u^*_{\gamma,i} - u^*_i}{u^*_i} = \frac{\gamma}{(\pi_{cb} - \gamma - \bar{\pi}_i)}$.

3 Experimental Design

Participants in our individual-choice experiment act as atomistic inflation forecasters tasked with providing two sets of one-period-ahead inflation forecasts (Initial Forecasts and Updated Forecasts) in three independent decision periods. Each set of forecasts comprises an incentivized point forecast of inflation coupled with an incentivized measure of forecast uncertainty. Treating participants' initial density forecasts as priors on the distribution of inflation, we can proxy the central bank's forecast credibility by measuring how subjects weight the central bank's forecast whenever providing their Updated Forecasts (i.e. their posterior inflation inflation distribution).

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 institutions, their preferences for obtaining economic information, and their familiarity with prevailing economic conditions. We then provided subjects on-screen instructions that explained the inflation forecasting task, the information available to aide in their forecasting task, how to interact with the available information, how to interact with our software, and how we incentivized their forecasts. These instructions remained available to subjects throughout the experiment via a toggle button on all screens.

[3] 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.

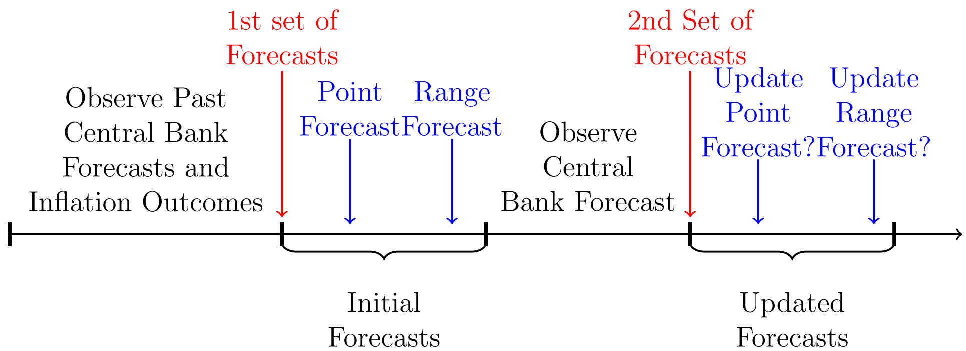

Figure 2: Experimental Timeline: A single decision period

Following the instructions, subjects completed a comprehension quiz. The comprehension quiz consisted of five questions designed to test subjects' understanding of our experimental instructions. Subjects had to answer all five questions correctly to proceed. Our software ended the experiment early for subject who submitted the quiz more than twice with at least one wrong answer.[4] Subjects who successfully completed the quiz proceeded to the forecasting task.

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 received 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 participants carefully considered the full economic history before forming Initial Forecasts. We displayed this historical data graphically and numerically and all information, once revealed, remained available for the duration of that decision period.

After our software revealed the full economic history for a decision period, participants provided a point forecast of one-period-ahead inflation (i.e. $\mathbb{E}_{i,12}\pi_{13}$) in percentage terms with up to two decimal accuracy. We incentivized point forecasts according to

[4] We provide examples of the instructions and comprehension quiz in Appendix Section A1.1.

Equation (8), which follows the previous LTF literature (Rholes and Petersen 2021, Mokhtarzadeh and Petersen 2021, Petersen and Rholes 2022):

$$F_{i,t} = 2^{-|\mathbb{E}_{i,t-1}\{\pi_t\}-\pi_t|}. \tag{8}$$

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 (9), which follows Pfajfar and Žakelj (2016), Rholes and Petersen (2021), Petersen and Rholes (2022).

$$U_{i,t}(r_{i,t}) = \begin{cases} 0 & \pi_{i,13} \notin [\underline{u_{i,t}}, \overline{u_{i,t}}] \\ \phi\left(\frac{1}{r_{i,t}}\right) & \pi_{i,13} \in [\underline{u_{i,t}}, \overline{u_{i,t}}]. \end{cases} \tag{9}$$

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 simulated a standard 3-equation monetary model taken from Walsh (2017) and given by Equation (10) through Equation (16). 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.

$$y_t = E_t y_{t+1} - \sigma^{-1}(i_t - \mathbb{E}_t\pi_{t+1}) + g_t \tag{10}$$
$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \kappa y_t + u_t \tag{11}$$
$$i_t = \phi_x y_t + \phi_\pi \pi_t + v_t \tag{12}$$
$$r_t = i_t - \mathbb{E}_t\pi_{t+1} \tag{13}$$
$$g_{t+1} = \rho_g g_t + \epsilon^g_{t+1} \tag{14}$$
$$u_{t+1} = \rho_u u_t + \epsilon^u_{t+1} \tag{15}$$
$$v_{t+1} = \rho_v v_t + \epsilon^v_{t+1} \tag{16}$$

We assume the central bank in our simulated economy forms rational expectations so that the uncorrelated stochastic components of the per-period shocks (Equation (14), Equation (15), and Equation (16)) 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 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	ρ	ϕ_π	ϕ_y	ρ_g	ρ_u	ρ_v
.99	1	.8	.104	.9	1.5	0	.5	.5	.9

Table 1: Parameter values for simulation exercise

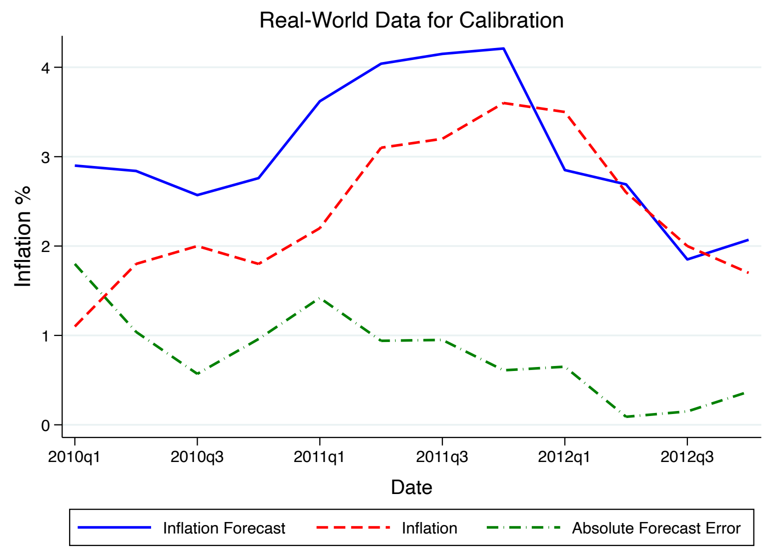

Figure 3: Economic data from the United Kingdom used for calibration of the economic histories

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 $\delta_{\pi,t}^{history} = E_{t-1}^{history}(\pi_t) - \pi_t$.

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*:

- In *Early*, as in the observed BoE pattern the central bank commits significant forecast errors in the first third of the forecasting history, moderate errors in the second third, and minimal errors in the last third.

- We reverse this pattern for *Late* but added additional noise to *Late* so that subjects wouldn't recognize the economic history as an exact reversal of *Early*. This exactly preserves the absolute average forecast error between *Early* and *Late*.

- For *Consistent*, central banks exhibit 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. For the different experiments, we produce many variations of *Consistent* by adjusting the shock sequences to achieve higher/lower average absolute forecast errors; the levels 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*.

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.

	Summary of Forecast Performance by History (bps)					
	Year 1	**Year 2**	**Year 3**	**Full Sample**	$\gamma_{HistAvg}$	$\gamma_{LastYear}$
Calibration Data	110	95	34	80		
Consistent - Great	13	13	13	13	06	08
Consistent - Good	36	36	36	36	10	05
Consistent - Moderate	60	60	60	60	06	-07
Consistent - Bad	83	83	83	83	02	-19
Consistent - Terrible	171	171	171	171	-06	-42
Consistent - Bad	83	83	83	83	02	-19
Early	171	65	13	83	-51	12
Late	13	65	171	83	-52	-171

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

Regardless of treatment wave, all participants experiences three economic histories (i.e. completes three independent decision periods) consisting of *Early*, *Late*, and some version of *Consistent*. When we study the relationship between historical forecast performance and forecast credibility, an experimental wave we call *ForecastPerformance*, we vary the consistent history across five histories that differ in the level of forecast error. When we explore the effect of the timing of forecast errors on forecast credibility (*Timing*), we compare the differences between *Early*, *Late* and the comparable *Consistent-Bad*. Looking at whether a central bank can use high-frequency communication to bolster its forecast credibility (*Communication*), participants are shown *Early*, *Consistent-Bad*, and then *Late* but augmented with some additional communication. Rather than describe all treatment waves now, in the following sections we cover the *Forecast Performance*, *Timing*, and *Communication* treatment waves and for each we provide additional details regarding treatments and experimental design, state our hypotheses, and detail our results.

4 Forecast Performance

We start by examining the broad effect of Forecast Performance and ask how a central bank's forecast performance influences its perceived forecast credibility which we measure as the willingness of participants to incorporate the central bank's inflation expectation into their updated point forecast. We label this set of treatments *Forecasting Performance*. As defined in Table 2, we vary the sample-average absolute forecast errors across otherwise identical economic histories and refer to these histories as *Consistent-Great*, *Consistent-Good*, *Consistent-Moderate*, *Consistent-Bad* and *Consistent-Terrible*.

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

Table Table 3 describes the 10 different variations of economic history that participants could have seen – the 5 treatment histories following $\{Early, Late\}$ and $\{Late, Early\}$ respectively. The sample sizes are relatively consistent with around 90 for each of the forecast performance treatments, with the exception of *Bad* for which we also draw on treatments that arose in the *Timing* waves.

4.1 Forecast Performance Hypothesis 1

Equation (7) provides a clear hypothesis about the relationship between historical forecast performance and the central bank's forecast credibility, as measured by u_1, u_2, and u_3. Using the inverse of a 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.*

At the individual-level, we measure updating Equation (7) under various assumptions about the value of γ. We then average over these estimates to produce estimates of treatment effects, which we present in Figure 4 alongside treatment-average Bayesian optimal responses as benchmarks. Here, the treatment average optimal response is given as $u^*_T = \frac{1}{N_T} \sum_{n \in N_T} \frac{\beta_T}{\alpha_n + \beta_T}$.

Figure 4 depicts average treatment effects assuming our participants observe no systematic component in the central bank's forecast error (i.e. $\gamma = 0$, blue diamonds) and also assuming that participants account for a systematic error component of the central bank's inflation forecast (red triangles). For these biased estimates, we assume that subjects use the entire forecast history to discern the magnitude and direction of this systematic error component so that $\gamma_{HistAvg} = \frac{1}{12} \sum_{t=1}^{t=12} \left(\pi_t^{cb} - \pi_t\right)$.

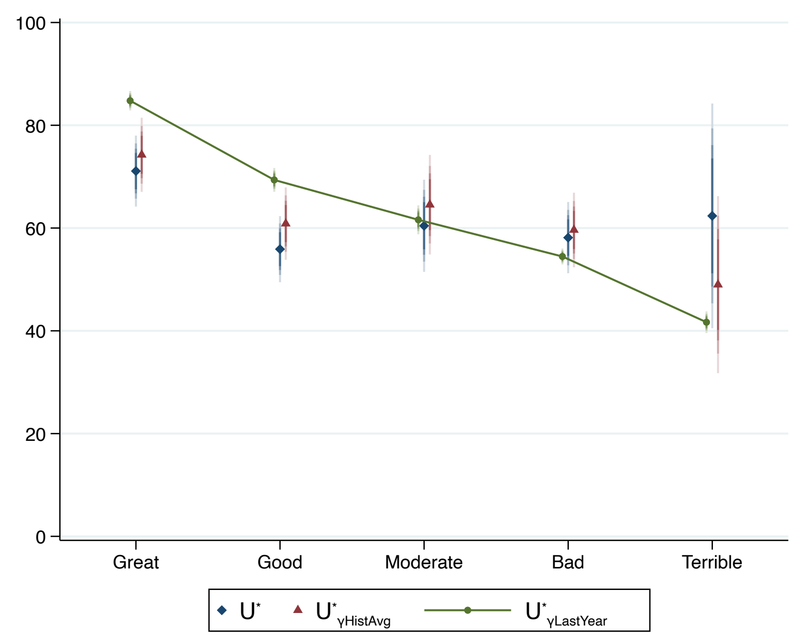

Figure 4: Estimates of central bank forecast credibility in *ForecastPerformance* treatments.

We see that for both assumptions on the behaviour of γ, subjects' perception of the central bank's forecast credibility is decreasing in the bank's historical average absolute forecast error. This constitutes support for hypothesis one – subjects respond like

Bayesian's to changes in historical forecast, at least qualitatively. However, the empirical relationship is flatter than would be predicted by the theory. This means that subjects tend to underweight signals whenever the central bank's forecasts precision is high and overweight signals when the bank's precision is low. This suggests that the cost of large errors is not as high as theory predicts, but also that the benefit of precision was be less than would be warranted.

4.2 Forecast Performance Hypothesis 2

Equation (7) also clarifies that a bank's forecast credibility doesn't depend simply on its own performance. Instead, it depends on the bank's performance relative to a participant's belief about her own forecasting credibility.

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

This hypothesis says that participants who exhibit higher levels of forecast uncertainty in the Initial Forecast will update more. Differentiating u^* with respect to α_i gives $\frac{\partial u_i^*}{\partial \alpha_i} = \frac{-\beta}{(\alpha_i+\beta)^2}$, which implies that perceived credibility should be negatively related to an individual's forecast uncertainty. Put differently, subjects who are highly uncertain of their forecasts should be more forgiving of historical forecast errors than subjects who are more certain about their forecast.

Because we elicit individual-level measures of forecast uncertainty, we can also consider whether subjects correctly incorporate their own forecast uncertainty into their assessments of the central bank's forecast credibility. We show in Figure 5 that there is no relationship between forecast uncertainty and perceived central bank credibility, which clearly violates the logic of Equation (7). This result 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.

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 (7) says – an uncertain agent should more highly value new signals that help her better predict 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 underweights the central bank's forecast relative to the Bayesian benchmark when updating her point forecast of inflation. This implies that α_i is too large in her updating function. Since we measure forecast uncertainty as α_i^{-1}, this implies that the participant undervalues her own forecast uncertainty when considering the central bank's forecast. This would yield $\frac{u_i^*}{\beta_T(1-u_i^*)} - \alpha_i < 0$, which says that the participant's implied uncertainty is smaller than the incentivized measure of forecast uncertainty she provided in her initial forecast.

Figure 5: Scatter plot of individual-level forecast uncertainty and perceived forecast credibility of the central bank.

We show 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.[5]

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 under reacting to own forecast uncertainty in our experiment, which is what we observe in our treatment where the central bank's historical forecast performance is best. However, we also document a sort of underprecision, where subjects over react to forecast uncertainty, thereby treating their private information as overly imprecise when forming a perception of central bank forecast credibility.

5 Does Timing of Forecast Error Matter?

One implication of hypothesis 1 above is that the central 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 surrounding discrete

[5]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.

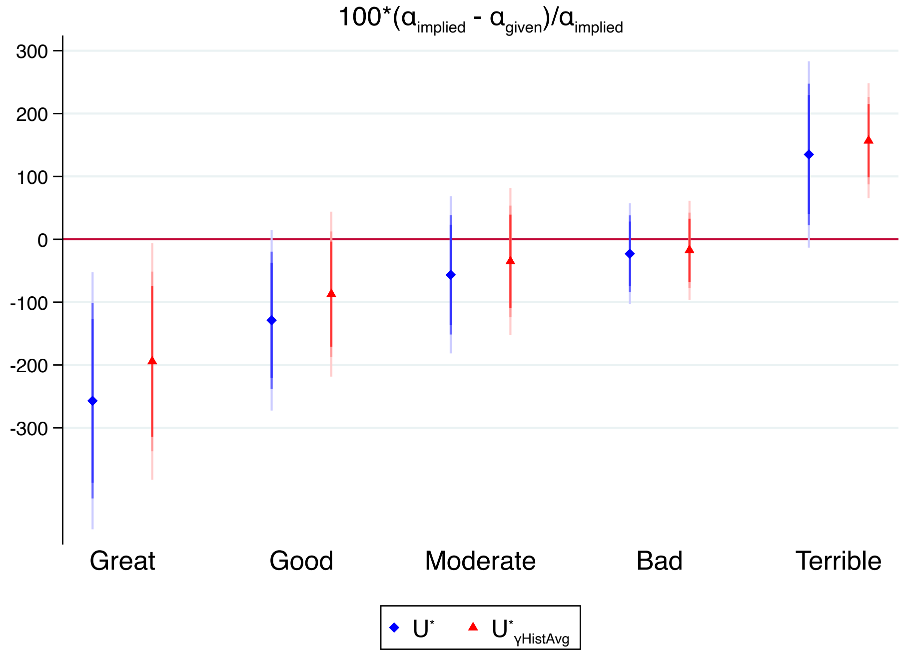

Figure 6: Percentage by which the average participants undervalues (< 0) or overvalues (> 0) her own uncertainty when incorporating the central bank's forecast into his or her updated forecast.

changes in economic conditions (i.e. at the onset of a financial crisis or a global pandemic). To see the extent to which the timing of forecast errors influences participants' perceptions of the central bank's forecast credibility, in *Timing* we fix the central bank's level of historical forecast precision but allow for variation in the pattern of historical forecast errors underlying that level of forecast precision.

5.1 Timing Hypotheses

Given that the presented sample history is only 12-quarters long, it might be expected that subjects use the full history to estimate the central bank's precision, β. If subjects equally weight historical information when deciding on a central bank's forecast credibility, β is calculated according to:

$$\beta^{-1} = \frac{\sum_{j=1}^{j=12} \left|\mathbb{E}_{j-1}^{CB}\left(\pi_j\right) - \pi_j\right|}{12}. \tag{17}$$

Given that the absolute average forecast error is the same across these treatments (Table 2), if Equation (17) holds then a Bayesian agent should form perceptions β that are the same across all of *Early, Late*, and *Consistent* and so the forecast credibility should also be the same across treatments. For this sort of subject, the perception of the central bank credibility ought to be invariant to timing of errors.

We summarise this hypothesis on the effect of forecast-error timing as:

Hypothesis 3. *Subjects weight observed histories equally such that the timing of forecast errors does not matter.*

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 weight differently very large and very small errors. This does not necessarily translate into timing effects however; if they underweight (overweight) large (small) errors, but timing does not matter in and of itself, then we might expect that the effects net out over *Early* and *Late* such that $\beta^{Early} = \beta^{Late} = \beta^{Consistent}$.

There is, however, a 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 were the case, the timing of forecast errors does matter when forming a perception of central bank forecast credibility: $\beta_i \neq \beta_j$ for some set of histories $i, j \in \{Early,\ Late,\ Consistent\}$.

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:

$$\beta^{-1} = \lambda \sum_{j=0}^{j=11} (1-\lambda)^j |\mathbb{E}^{CB}_{t-2-j}\left(\pi_{t-1-j}\right) - \pi_{t-1-j}| \tag{18}$$

where the weighting function exhibits exponential decay in time. Figure 7 depicts the implied weighting functions from Equation (18) 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). In that context, economists typically motivate these models as a way to for an agent to account for structural change in whatever macroeconomic time series an agent is forecasting.

This gives us a potential second hypothesis:

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

5.2 Timing Treatments and Results

To do explore these hypotheses, we use a within-subject design that exposes each participant to *Early, Late,* and *Consistent - Bad*. We provide information about the historical forecast precision over the entire sample (column 4) and within each year (columns 1-3) for these treatments in Table 2. Note that, because we only use *Consistent-Bad* in this wave, we will refer to *Consistent - Bad* as *Consistent* throughout the *Timing* section.

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 4.

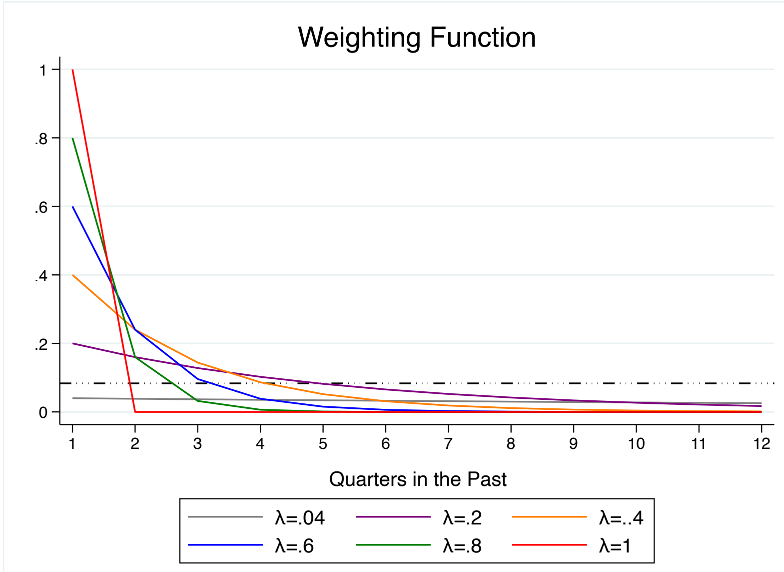

Figure 7: Weighting functions (Equation (18)) with different values of λ

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

Table 4: ADD CAPTION
Notes: This table summarizes our *Timing* treatments. Note that T6 and T9 are the same as T4a and T4b in Table 3.

We first consider central bank credibility in *Early, Consistent, and Late* using u_T^* assuming $\gamma = 0$. Figure 8 reports our treatment-level measure of perceived forecast credibility (as was reported reported for *ForecastPerformance*). This figure reports the optimal level of updating for a Bayesian agent who equally weights all available information when forming a perception of the central bank's forecast credibility (green dots), the estimated updates across treatments with 95% confidence intervals (red) and the implied deviation from the Bayesian benchmark optimal (blue diamonds). The main result from Figure 8 is that participants view the central bank as drastically under precise in *Late*. There is also marginal evidence that the central bank is viewed as overly precise in *Early*. As was the case in Table 3, updates roughly aligned with theoretical predictions for *Consistent* (which is *Consistent-Bad* above). We reject the null of Hypothesis 3.

We also consider a within-subject measure of perceived forecast credibility. We report these measures in Figure 9. These measures present the relative updating weight using

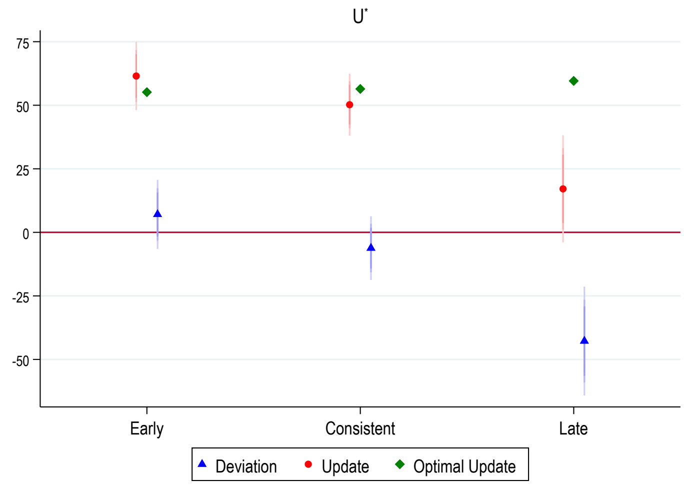

Figure 8: Treatment Average Perceived forecast credibility in *Timing*

updating in *Consistent* as a subject-level baseline. Specifically, for $X \in \{Early,\ Late\}$:

$$u^*_{T,within} = \frac{1}{N_T} \sum_{n \in N} \left(u^*_{n,X} - u^*_{n,C} \right) \tag{19}$$

The within-subject measure acts as a sort of participant-level fixed effect, assuming idiosyncratic biases are invariant to forecast history (for example, participant i faces the same cognitive cost of information processing for each of our three histories). The same pattern of relative updating between *Early* and *Late* shows up in our within-subject measure of perceived forecast credibility. Further, the magnitude by which participants deviate from the Bayesian benchmark is significantly larger in *Late* than in *Early*.

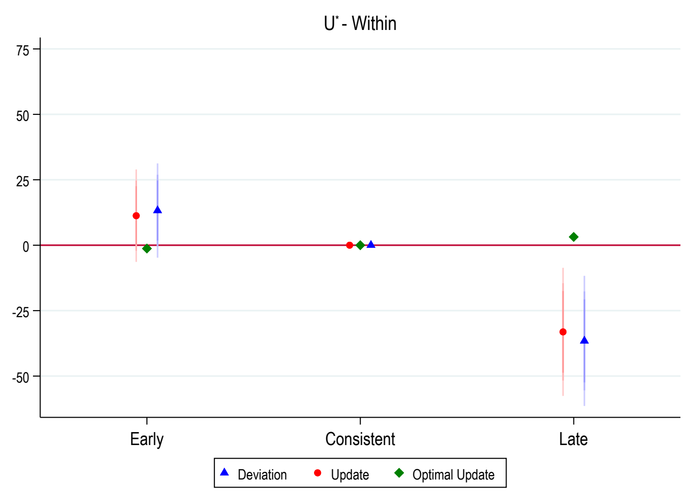

Figure 9: Within Subject Perceived forecast credibility in *Timing*

One potential reason to be more concerned about the role played by bias is that the average absolute forecast errors are larger in *Early* and *Late* as shown in Table 2. Figure 10 corrects for the $\gamma_{HistAvg}$ and shows that subjects are now estimated to under-weight the central bank for both *Early* and *Late*. (The within-subject variation, not shown, is similar.) This is consistent with the results in Figure 4 whereby the subjects tend to under-weight very good performance; the net effect of over-weighting recent performance but underweighting great performance is not, ex-ante, obvious.[6] Nonetheless, the finding that timing matters remains and the estimated under-weighting of *Late* is larger.

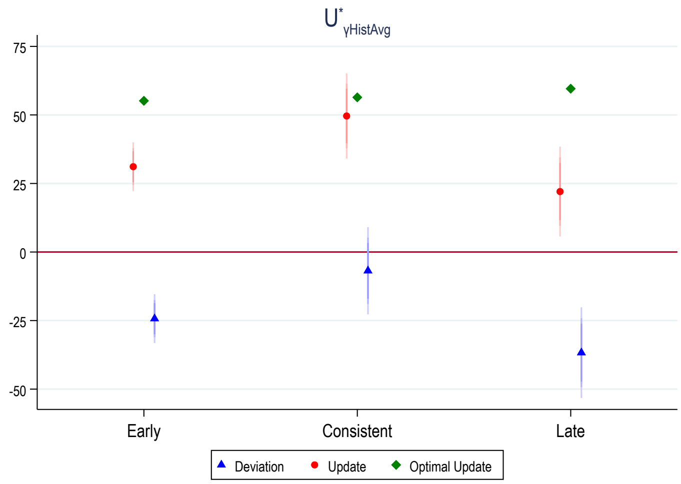

Figure 10: Perceived forecast credibility in *Timing*: $\gamma_{HistAvg}$

5.3 Exploring the Extent of Recency Bias

The identified pattern of updating potentially indicates recency bias in the sense that subjects are more heavily weighting temporally proximal forecasting information provided in the economic histories. To see this, note that in Table 2 that β^{-1} decreases (increases) over time and is considerably smaller (larger) in the final year of the economic history than when measured across full sample in *Early* (in *Late*). Given this, and considering Equation (3), subjects who form a perception of central bank forecast precision using the later years of the economic history in *Early (Late*) ought to update more (less) than the Bayesian agent who forms perceptions of central bank forecast precision according to Equation (17).

This leads us to ask how strong is this recency bias and is it different across economic histories? The difference in the magnitude of deviations from the Bayesian benchmark between *Early* and *Late* could be driven by the differences in the deviations of β^{-1}_{Year3} and $\beta^{-1}_{FullSample}$ across *Early* and *Late*. However, it could also be driven by differences in the amount of recency bias induced by the different histories. To answer this, we estimate λ in Equation (18) using the treatment-averaged results.[7]

[6]This highlights an interesting question for future research: is there an asymmetry in how timing matters (i.e. when forecast performance increases vs decreases).

[7]Combining $\beta^{-1} = (\alpha \times u^*)^{-1}(1 - u^*)$ with Equation (18), gives:

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

which we solve for λ via numerical approximation.

We depict the results of this estimation exercise in Table 5 along with the corresponding estimated weighting functions in Section 5.3. These results suggest that the average participant displays recency bias for both *Early* and *Late*. These estimated weighting functions, displayed in Section 5.3, clearly differ from the Bayesian benchmark (black dashed line). But they also differ from each other – $\lambda_{Early} < \lambda_{Late}$. This suggests that the recency bias participants exhibit depends critically on the central bank's historical forecast performance. In particular, subjects focus significantly more on temporally proximal forecasting performance whenever that performance is particularly bad.

Table 5: Estimated Values of λ

	γ_0	$\gamma_{HistAvg}$	γ_{Last4}
Early	0.245	0.275	0.224
	(0.0170)	(0.0160)	(0.0171)
Consistent	0.523	0.511	0.615
	(0.022)	(0.022)	(0.022)
Late	0.622	0.560	0.423
	(0.0198)	(0.0222)	(0.0253)

Standard errors in parentheses

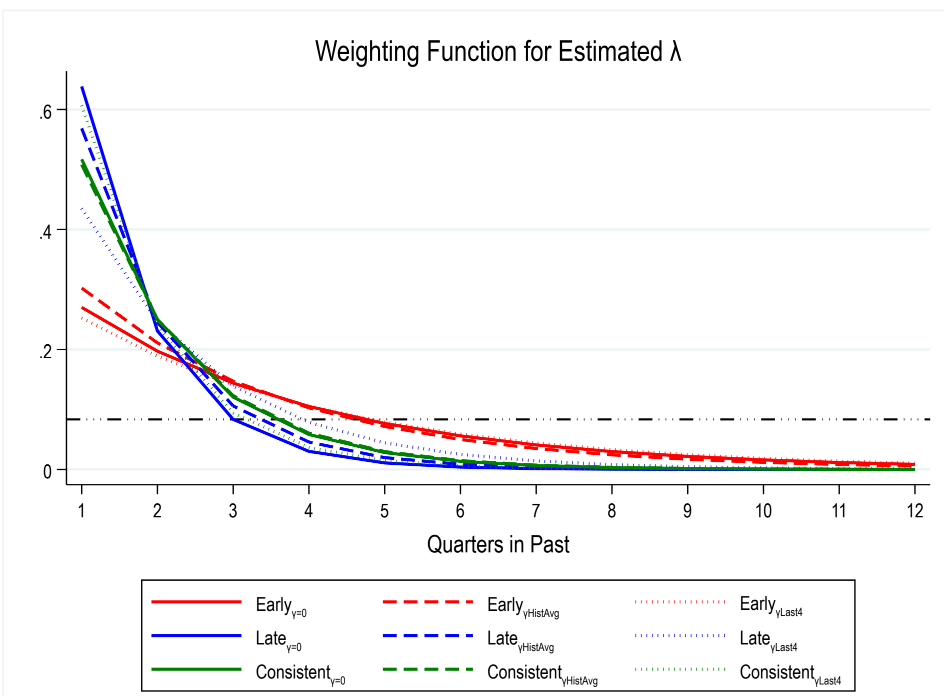

Figure 11: Estimated weighting functions based on λ by history.

The estimated λ values do not seem to be affected too much by the assumed treatment of bias; using $\gamma_{HistAvg}$ raises λ_{Early} and lowers λ_{Late} but the gap between them remains large.

5.4 Recency Bias and Perceived Forecast Bias γ

Based on these estimates, our participant's base their perceptions of the central bank's forecast credibility almost entirely on performance in the last for quarters of the economic histories we provide. If we assume this recency bias also applies to how subjects form perceptions of γ, then we can re-estimate perceived credibility in our *Timing* treatments assuming that $\gamma_{LastYear} = \frac{1}{4}\sum_{t=9}^{t=12}\left(\pi_t^{cb} - \pi_t\right)$. We depict the results of this exercise in, Figure 12.[8]

Considering recency bias in γ are more consistent (and even strengthen) the finding under γ_0 that participants more than fully adjust toward the central bank's forecast in *Early*. Our results on *Late* are qualitatively robust. and not all toward the central bank's forecast in *Late*. Interestingly, we observe very little change in the point estimate of credibility in *Consistent*.

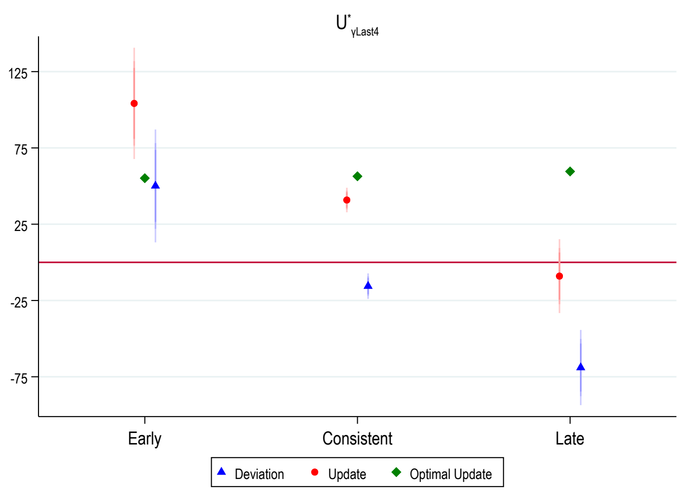

Figure 12: Perceived forecast credibility in *Timing*

5.4.1 Updating *ForecastPerformance* for $\gamma_{LastYear}$

We also consider what assuming this sort of recency bias about γ might imply for our estimates of central bank credibility in *ForecastPerformance*. Thus, we re-estimate credibility for that subset of treatments assuming participants base their perception of γ on the last year of economic history. We provide these estimates in Figure 13 (green dots) alongside previous (blue and red markers) and the corresponding Bayesian benchmark (orange line).

Interestingly, these updated estimates suggest that subjects over respond to precise forecasts and under-respond to imprecise forecasts. What does this mean in terms of

[8]For ease of comparison, we also report credibility estimates from *Timing* treatments assuming subjects estimate γ equally weighting all historical information. Figure A-2 in the appendix presents the within-subject results.

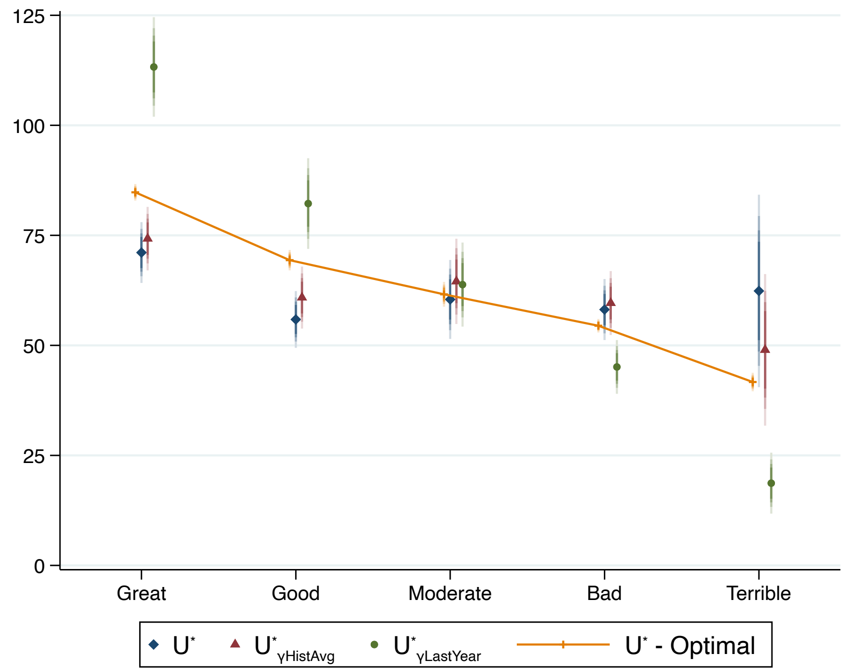

Figure 13: Updating forecast credibility estimates from *ForecastPerformance*

overprecision and underprecision? We still find evidence of misperceptions of precision in both directions. However, estimates that assume recency bias in γ suggest that subjects are underprecise whenever the central bank's historical forecast precision is high and overprecise whenever the central bank's historical forecast precision is very low.

Understanding how subjects estimate γ is an important step for our future research.

5.5 Dynamics of Perceived Credibility

We can also exploit our experimental design to gain some insight into the dynamics of perceived forecast credibility. Recall 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*. Similarly, the central bank's historical forecast precision in *Great* aligns with that from the final year of *Early*. By comparing perceived forecast credibility across these treatments, we can learn something about how quickly perceived forecast credibility builds and erodes.

We first compare perceived credibility measures from *Terrible* and *Late*, which we depict as sample densities in Figure 14. First we note that the mean level of perceived credibility is not statistically different across treatments ($p = .704$, two-sample t-test).[9] 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.

[9]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*.

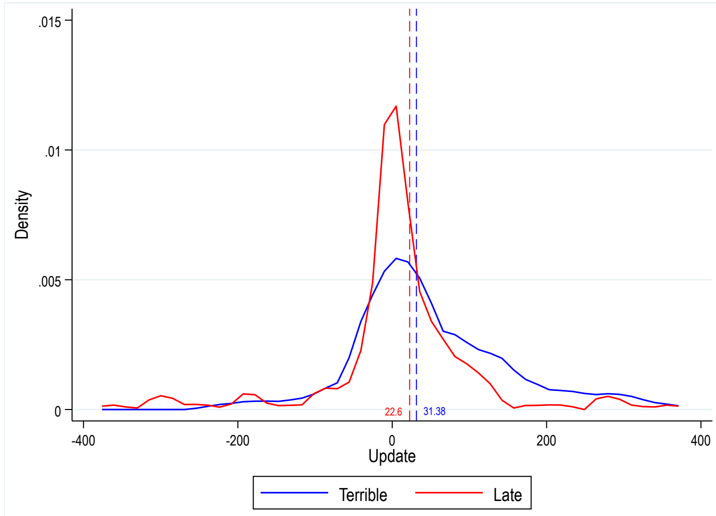

Figure 14: *Consistent-Terrible* vs. *Late*

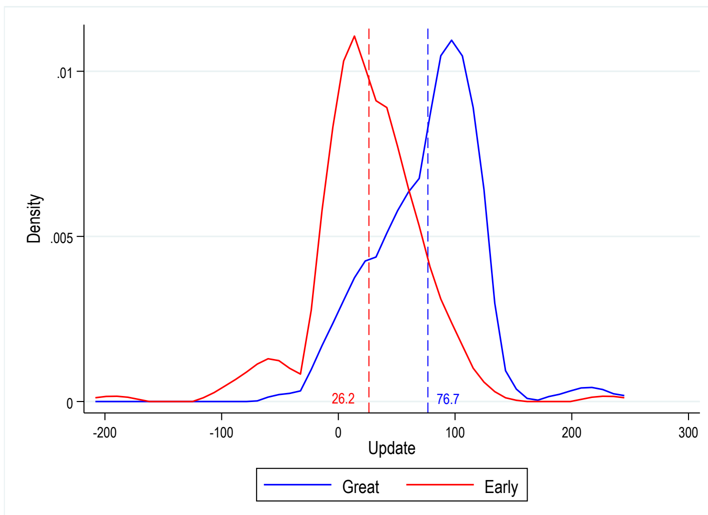

Figure 15: *Consistent-Great* vs. *Early*

However, this does not hold when comparing *Great* and *Early* in Figure 15. 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 accurately predict

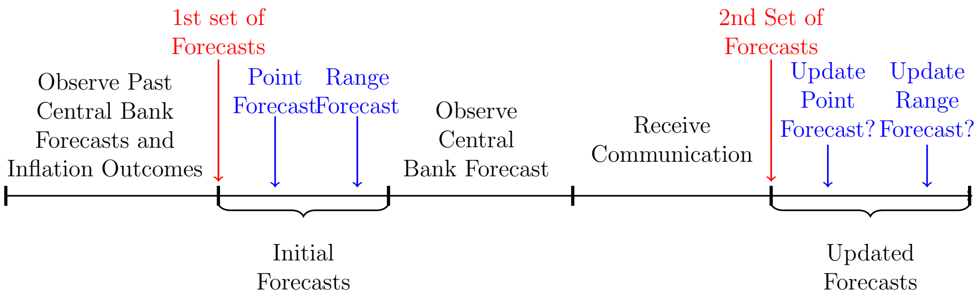

Figure 16: Timing of decision period in *Communication* treatments

inflation.

Overall, our *Timing* results indicate that forecast credibility erodes much more quickly than it builds. An optimistic caveat is that forecast credibility, once lost, can be rebuilt.

6 Does Communication Alter the Effect of Forecast Performance?

As a last step, we consider whether high-frequency communication plays a role in shaping perceptions of the central bank's forecast credibility. From a practical perspective, this question is important because most central banks devote considerable resources to crafting and publishing high-frequency communication - often alongside projections - to rationalize and contextualize policy actions and outlooks. 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?

6.1 Communication Treatments

To answer these questions, we incorporate high-frequency communication into *T7* from our *Timing* treatments. Specifically, alongside the the final history (*Late)*, we publish the high-frequency communication 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 high-frequency communication.[10] We summarize the timing of decision periods in *Communication* in Figure 16.

High-frequency communication in our experiment takes the form of written commentary that deal with the central bank's outlook, whether the source of poor historical forecast performance in *Late* is endogenous or exogenous, and whether the central bank is performing better or worse than peer forecasting institutions. Additionally, we include a control text that provides generic information about the central bank. To the extent

[10]We could also have based the analysis on *T11* but since we found little difference in results from the order of the first two histories, we proceeded with *T7*.

that such text is uninformative, this allows us to examine the extent to which the mere presence of communication can enhance the central bank's forecast credibility:

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 graphical forecast and adds no new information for participants. This allows us to discern whether reinforcing graphical information via text can better covey 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.

Importantly, we are careful to control for the complexity of communication across treatments to avoid introducing a potential confound We summarize these treatments in Table 6 while the full text is provided in Section A1.3.

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 6: Treatment summary for *Communication*

6.2 Finding on the Effects of Communication

Though we have not resolved perfectly how to treat the implied bias in the results γ, this should be less of a concern here since all treatments are based off *Late* and so face the same error pattern. As such, different choices of how to treat γ should only shift the estimated distributions but not drive the relative effects of communication. We therefore present below the results for $\gamma_{HistAvg}$ in Figure 17.

Three results stand out. First, we first note that treatment-average estimates of central bank forecast credibility in *Control* are not statistically different from our equivalent no-text baseline reported in Figure 10. Specifically, *T7* for $U^*_{\gamma_{HistAvg}}$ ($p \approx .69$).

Second, communication can increase perceptions of the central bank's forecast credibility. Even the addition of the outlook text, which is not meant to provide new information

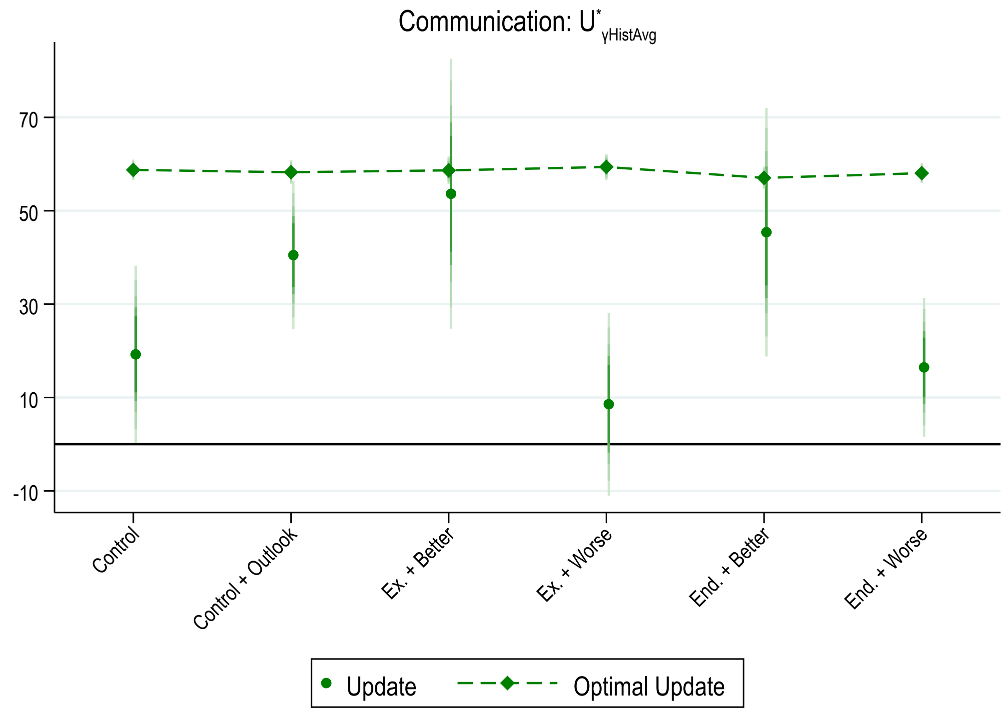

Figure 17: *Communication* Update: $\gamma = \frac{1}{12}\sum_{t=1}^{t=12}\left(\pi_t^{cb} - \pi_t\right)$

relative to the graphical forecast, shifts the centre of the distribution up. This might happen because some people are better at extracting qualitative/narrative information from the 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. Or 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).

In particular, for some of the messaging the delivery of communication shifts the update so much that it is not statistically different to the optimal rate. However, it is worth noting that under $\gamma = 0$, the pattern holds but the impact of communication does not get as high as the optimal.

Third, there were no noticeable differences between the discussion of exogenous and endogenous sources. However, acknowledging that the errors were worse than other forecasters is quite detrimental. In fact, it is worse than simply the control of no communication. This effect is most pronounced in *Exogenous + Worse*, though statistically the effects are similar for *Endogenous + Worse*.

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.

7 Linking to 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?[11].

Our identification strategy involves projecting changes in real yields of different maturities that occur in the 24-hour window surrounding IR releases onto conditioning 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 got 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.[12] Following from

[11] For examples of this identification approach, see Cook and Hahn (1989), Kuttner (2001), Gürkaynak et al. (2004), 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.

[12] We use a 24-hour window following Hansen et al. (2019), who argue the longer window is nec-

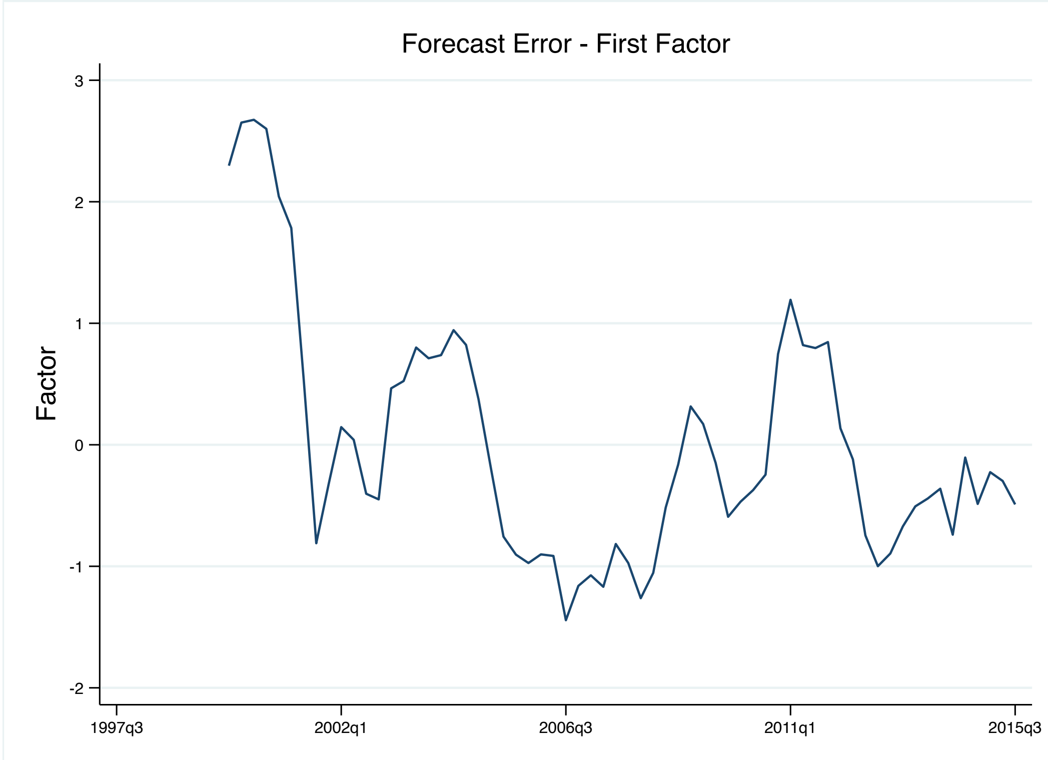

Figure 18: Bank of England Forecast Performance: Factor Summary Variable

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:

$$|\Delta y_i| = \alpha_i + \beta_l \mathbb{I}_{l,t} + \sum_{x,j} \psi_{x,j,i} \Delta PC_{x,j,t} + \eta_{1,i} FTSE_{t-1} + \chi_i X_{i,t} + \eta_{2,i} VIX_{t-1} + \epsilon_{i,t} \quad (20)$$

where $\Delta_{i,t}$ captures changes in the 1,3, and 5yr gilt that occur in the 24-hour window around the IR release, EF_t is the error factor, $\Delta PC_{x,j,t}$ 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 indicator capturing whether the BoE's forecast performance has exceeded its sample average for the last $l = \{1, 2, 3, 4\}$ 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 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

essary 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.

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 4 lags.[13]

We report $\hat{\beta}_i$ in Figure 19, where the top panel depicts estimates using $|\Delta y_{1,t}|$ as our outcome of interest, to middle panel using $|\Delta y_{3,t}|$, and the bottom panel using $|\Delta y_{5,t}|$. Each of these three panels reports 4 different coefficients that correspond to the number of preceding quarters that the BoE's inflation forecast accuracy exceeds the sample average. We see that, in general, markets respond more strongly to the information contained in the inflation report whenever the BoE's forecast credibility, proxied by its forecast accuracy, is relatively high. This effect is increasing in the duration of high forecast credibility but eventually stabilizes, which is consistent with our experimental finding of recency bias in *Timing*. Additionally, this effect is more pronounced for our shortest maturity and attenuates further into the yield curve.

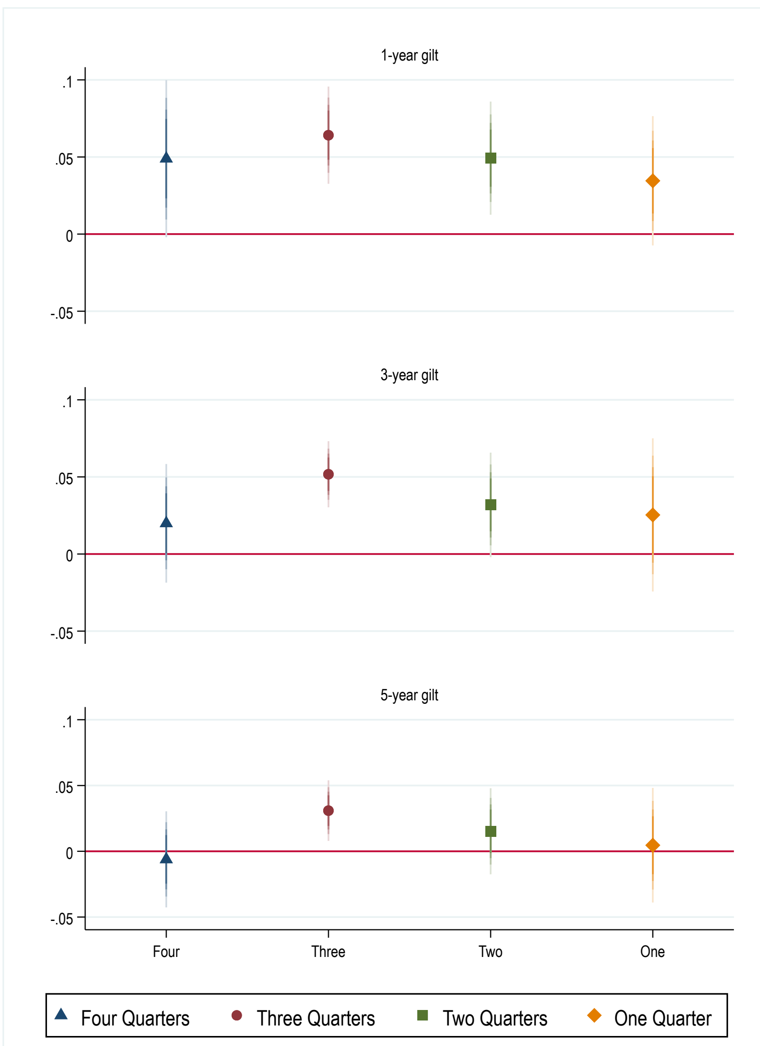

Figure 19: Caption

[13]Lag selection is based on $T^{(1/4)}$), following Newey and West (1987) and Greene (2003).

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 a static feature and should not be thought of or modeled theoretically as a static feature. Instead, central banks can both win and lose credibility. Though this implies that banks can lose their ability to manage expectations whenever unexpected economic shocks lead to poor forecast performance, it also implies that banks can rebuild that credibility. However, we show that these dynamics of credibility are seemingly 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.

Bibliography

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

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

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

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.

Burke, M. A. and M. Manz (2014). Economic literacy and inflation expectations: evidence from a laboratory experiment. *Journal of Money, Credit and Banking 46*(7), 1421–1456.

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.

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.

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.

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

Gürkaynak, R. S., B. P. Sack, and E. T. Swanson (2004). Do actions speak louder than words? the response of asset prices to monetary policy actions and statements. *The Response of Asset Prices to Monetary Policy Actions and Statements (November 2004)*.

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 (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics 108*, 185–202.

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.

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

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

A1.1 Instructions

Experimental Instructions

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

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

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

Your Objective in the Experiment

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

You will provide two types of inflation forecasts:

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

Additional Definitions:

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

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

The experiment:

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

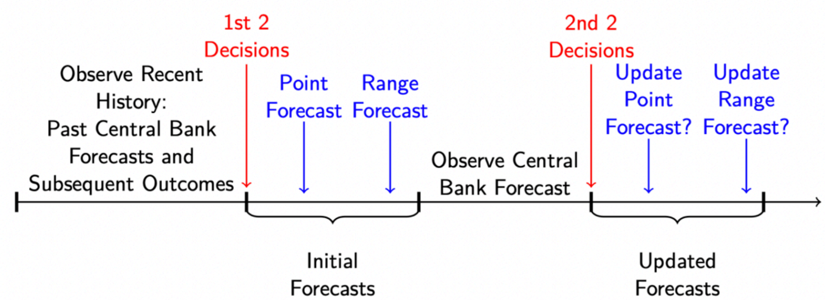

Figure: Experimental Timeline: A single decision period

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

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

How our software scores your performance:

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

Suppose that actual inflation turns out to be 2.5%

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

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

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

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

Interacting with the data and inputting your forecasts:

The historical data:

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

Providing your Point Forecast:

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

Providing your Range Forecast:

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

TimingUpdateBiasHistAvg.pdf

Figure A-1: *Timing* Update: $\gamma_{HistAvg} = \frac{1}{12}\sum_{t=1}^{t=12}\left(\pi_t^{cb} - \pi_t\right)$

A1.2 Tables and Figures

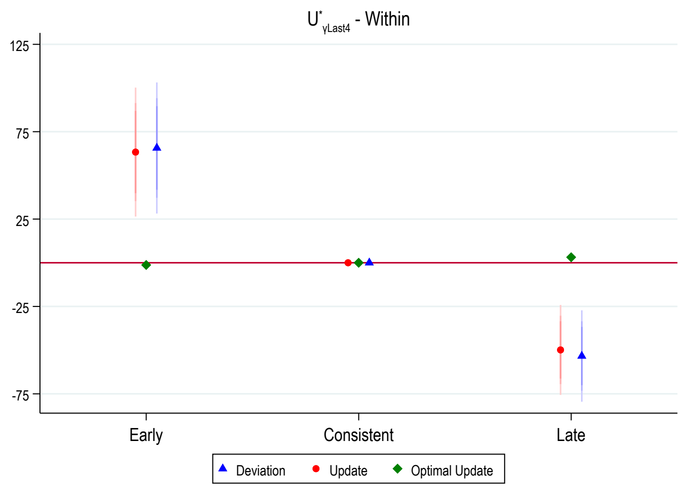

Figure A-2: Perceived forecast credibility in *Timing*

A1.3 Central Bank Messages in *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.