Experimental Economics (2026), 1–24
doi:10.1017/eec.2026.10058

economic science association

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL PAPER

History-dependent monetary policy: a lab experiment and a HENK model

Jasmina Arifovic[1], Isabelle Salle[2,3] and Hung Truong[3,4]

[1]Department of Economics, Simon Fraser University, Canada
[2]Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute, Amsterdam, Netherlands
[3]Department of Economics, University of Ottawa, Ottawa, Canada
[4]VNU University of Economics and Business, Vietnam National University, Hanoi, Vietnam
Corresponding author: Isabelle Salle; Email: i.l.salle@uva.nl

(Received 18 June 2024 UTC; revised 14 May 2025 UTC; accepted 8 July 2026 UTC)

Abstract

Price-level targeting (PLT) is theoretically optimal under the full information rational expectations (FIRE) benchmark but has not been implemented in practice. We use a laboratory experiment to collect data on expectations, inflation and output dynamics under a traditional inflation targeting (IT) framework and a PLT regime with both deflationary and cost-push shocks. A heterogeneous expectation New Keynesian (HENK) model with forward-looking communication can replicate the dynamics observed in the lab. We do not observe the benefits of PLT over IT because both human subjects and HENK agents are unable to learn the history-dependent inflation path, which results in excess macroeconomic volatility. However, once augmented with an inflation guidance consistent with closing the price gap, the advantage of PLT materializes, both in the lab and in the model, and the inflation gap is minimized no matter the type of shocks. The model offers a behavioral narrative for the cross-regime differences observed in the experiment. Combining laboratory data with the HENK framework is a promising addition to the central banks' toolkit since it offers a first test of their policies.

Keywords: Central bank communication; Heterogeneous expectations; Lab experiments; Learning

JEL Classification: E7; E52; E42; C92

1. Introduction

A decade-long episode of persistently below-target inflation at the effective-lower bound (ELB) followed by the steep and broad-based rise in inflation in the aftermath of the COVID-19 pandemic have called forth a rethinking of central bank (CB) policies. Major CBs have emphasized a shift from standard inflation-targeting (IT) regimes to a form of history-dependent policy – such as price-level targeting (PLT).[1] Under PLT, nominal rates are adjusted so that the price level gravitates around the path consistent with a low inflation target while standard IT is only concerned with keeping *current* inflation around this target.

[1]In August 2020, the Fed announced a change to its monetary policy framework to refer to an average inflation target; see Powell et al. (2020). Since June 2021, the ECB has been communicating on an inflation objective over a medium term, with short-run deviations from this target. The Bank of Canada had even considered PLT in its monetary policy framework reviews but fell short of implementing it (Wilkins, 2018).

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

The academic literature on the relative stabilizing benefits of PLT over IT, which we summarize hereafter, is however limited and discrepant. This paper reconciles these contradictory results by combining a lab experiment with a model-based analysis. We provide the first causal evidence that simple forward-looking CB communication makes a striking difference: it helps subjects grasp the history dependence in the policy rule which allows the stabilizing properties of PLT to materialize. We show how a parsimonious heterogeneous expectation New Keynesian (HENK) model with communication replicates our experimental data remarkably well.

Agents' expectations are the primary channel through which monetary policy affects inflation and output because these expectations directly influence current decisions. However, gauging policy-dependent expectations in survey data is difficult, and field experiments in macroeconomic settings are impractical. The laboratory can fill this data gap by providing a controlled environment to gain insights into the expectation channel under different monetary policy regimes. The expectation channel is especially critical under history-dependent monetary policies since future inflation is supposed to offset past deviations from the target. In this respect, the theoretical merits of PLT over IT have been established under the standard FIRE (full information rational expectations) assumption, which presumes that the private sector understands the implications of history-dependence and align their expectations with the PLT policy rule (see Section 2).

To give an example, during an episode of above-target inflation, if consumers understand the history-dependent rule, they expect *below-target* inflation tomorrow to compensate for the current excessive inflation, which results in spending cuts today, and already contributes to limit the inflation surge at the first place. This tailwind is absent under IT since consumers expect inflation to go back on target – not below, which entails higher expectations and calls for higher interest rates to fight the inflation surge than under PLT. However, if the public fails to understand history dependence, their expectations turn into a headwind: while the central bank aims at below-target inflation, consumers' higher inflation expectations call for tighter monetary policy and more severe output costs to bring inflation down than if they had expected lower inflation at the first place. Symmetrically, if inflation falls below the target, higher inflation expectations under PLT helps bring inflation back on target – even more so if interest rate cuts are unavailable due to the ELB – but lower expectations exacerbate the problem. Therefore, not only experts but also the public needs to understand the monetary policy framework for inflation to be more stable under PLT than under IT.

Yet there is no historical experience with PLT,[2] the FIRE assumption falls short in the empirical literature (D'Acunto et al., 2023) and models that relax it generally fail to reproduce the benefits of history-dependent rules (see Section 2). In this respect, forward-looking CB communication, or guidance – which is fully redundant under the FIRE assumption – may help nudge the public towards forming history-dependent expectations and help the properties of PLT materialize with non-FIRE agents. Communication may find a particular rationale once one acknowledges that expectations are also heterogeneous,[3] and the CB also needs to worry about cross-sectional dispersion and coordinating individual beliefs on the target.

We use a forecasting experiment where inflation and output gap expectations are elicited from a group of human subjects and represent the only degree of freedom in an otherwise standard macroeconomic model. Our experimental setting is the first to focus on both deflationary and inflationary environments, in an effort to reproduce the recent course of inflation. Moreover, we modify this class of experiments to come closer to real-world information settings and make past individual data of all forecasters public information. In particular, in surveys of professionals, such as the survey of professional forecasters (SPF), the entire cross-section distribution of forecasts and the (anonymized)

[2]One exception, albeit hardly informative for modern times, is the short Swedish episode between 1931 and 1933.

[3]Earlier survey evidence of heterogeneity in macroeconomic expectations can be found in Branch (2004); see D'Acunto et al. (2023) and the references herein for a recent account.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

identities of the forecasters are published in every period, making peer comparison a relevant element of the forecasting task. Professional forecasts are central to macroeconomic analysis since they are closely monitored by policy institutions and they are the ones used as observables to fit expectations in the standard framework that also underlies our experimental setting.[4] Beyond professionals, social behaviors – such as peer imitation, contagion and news diffusion – have been emphasized when modeling expectation formation of the public; see the work of (Carroll & Wang, 2023) on epidemiological expectations. We provide new evidence of peer imitation in the lab and replicate the experimental dynamics with a HENK model where agents can copy each other's forecasts or learn to follow the CB guidance.

Our main findings are as follows. Both in the lab and in the HENK model, the benefits of PLT over an IT regime obtained under the FIRE assumption fail to materialize. On the contrary, inflation and output under PLT exhibit excess volatility compared to a standard IT regime, both in an inflationary and in a deflationary environment. These wild oscillations result from the commitment to correcting past price gaps in a learning environment, where expectations become unanchored and destabilize further output and inflation. However, the picture is drastically reversed if the CB communicates in each period just a single number, namely the time-varying transitory inflation target that is consistent with closing the price gap. In this case, expectations are better coordinated around the CB target, both deflationary and inflationary shocks result in smaller inflation gaps and shorter ELB episodes under PLT than under IT, in the lab as well as in the HENK model. In particular, the commitment to correct the deviations of the price level brings inflation toward the transitory target which, in turn, confirms the relevance of the CB's guidance to forecast inflation and initiates a credibility-stability loop among the subjects and the HENK agents. We also find that long-run expectations remain closer to the target under PLT with guidance despite the short-run fluctuations in inflation generated by the exogenous shocks than under IT and PLT without additional communication. Both the experiment and the HENK model also produce persistent inflation gaps under IT consistent with the recent inflation course. Importantly, our study contributes to explaining the discrepancy in the literature about the properties of PLT because we find empirical arguments that describe both a PLT regime implying less or more volatility in inflation and output than a traditional IT regime depending on the public's information set.

The rest of the paper proceeds as follows. Section 2 reviews the related literature, Section 3 describes the experimental design, Section 4 draws the main results from the experimental data and Section 5 discusses these data within the context of a HENK model. Section 6 concludes.

2. Related literature

The theoretical literature on history-dependent regimes is somewhat limited and rather discrepant. Woodford (2000) argues that optimal monetary policy rules must involve some degree of history-dependence – which is absent from standard forward-looking IT rules – to reinforce the stabilizing role of expectations; see also Mitra (2003). Under FIRE, PLT is then superior to IT from the point of view of inflation stabilization; see, e.g., Cateau et al. (2009). However, if agents form backward-looking expectations (e.g. under econometric learning), whether PLT remains superior to IT depends crucially on whether agents form history-dependent expectations; see Bodenstein et al. (2022); Honkapohja and McClung (2024); Honkapohja and Mitra (2020).

Which form of expectations would prevail under PLT is ultimately an empirical question that has been tackled in forecasting group experiments where expectations are elicited from human subjects; see Cornand and Heinemann (2014) and Duffy (2016) for surveys.[5] Particularly related to monetary policy are, among others, the experiments of Pfajfar and Zakelj (2016); Pfajfar and Zakelj

[4]See, *inter alia*, Cornea-Madeira et al. (2019); Rychalovska et al. (2024).

[5]These learning-to-forecast experiments, introduced by Marimon et al. (1993), build on the self-referential nature of economic systems that has been long recognized in experimental work; see, notably, Nagel (1995).

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

(2018), Hommes et al. (2019), Mauersberger (2021), Kryvtsov and Petersen (2021), Mokhtarzadeh and Petersen (2021), Rholes and Petersen (2021), Assenza et al. (2021), Kostyshyna et al. (2024) and Hennequin and Hommes (2024). In a nutshell, this literature concludes that *forward-looking* CB communication does not usually have large effects on expectations, most contributions study IT regimes, all focus on ELB dynamics, ignoring the stabilization trade-off between inflation and output arising from severe cost-push shocks, and none provides a unified model of their experimental results. Peer imitation is also left out despite its relevance in professional forecasting contexts.

The seminal contribution of Arifovic and Petersen (2017) sets up an experiment where the CB communicates a time-varying inflation target in each period. They find that the expectations of the subjects become disconnected from such targets as soon as shocks push inflation sufficiently down; in other words, subjects first 'need to see it to believe it.' Closely related is the experiment of Cornand and M'baye (2018) that finds that communicating the inflation target brings only minor stabilization benefits under IT. Salle (2023) establishes a similar result at the ELB. Outside the lab, Del Negro et al. (2023) point out that forward guidance has been less effective than the FIRE models predict (this is the so-called "forward guidance puzzle"); see also Mauersberger et al. (2019) for a behavioral discussion. By contrast, our results bear less dire implications for CB communication than these papers. In our setting, CB guidance is plagued by endogenous credibility issues but the forward-guidance puzzle is clearly ameliorated.

As for PLT experiments, without communication, Hommes and Makarewicz (2021) find that a sufficiently strong reaction to the price gap can induce convergence back to the target from the ELB more often than under IT in a non-linear model. Salle (2023) finds a similar result in a frictionless model. In a horse-race experiment between different monetary policy regimes, Kostyshyna et al. (2024) show that history-dependent regimes, in particular PLT, are destabilizing compared to IT. An additional treatment (run concurrently with ours) establishes that the publication of medium-run inflation and output projections similar to the ones studied in Mokhtarzadeh and Petersen (2021) decreases the likelihood of deflationary spirals at the ELB. We emphasize a different, more parsimonious communication strategy and, among other differences, we also consider the topical case of inflationary shocks. Beyond the lab, we further provide a model-based interpretation of our results while previous work has solely reported on experimental results.

We therefore also contribute to the growing literature on heterogeneous expectation models by providing experimental data to test these models.[6] Some models feature endogenous switches between different expectation rules in NK models[7] and some include endogenous credibility in IT regimes (Hommes & Lustenhouwer, 2019; Ozden, 2025; Kostyshyna et al., 2024). Other recent related models of heterogeneous expectations include Andrade et al. (2019) and Evans et al. (2025).

3. The experimental design

3.1. The experimental environment

We design a group experiment with a between-subject design. Each experimental session involves a fixed group of six subjects tasked with making point forecasts of output and inflation gaps in the next period. The experimental economy unfolds for 75 periods split into two separate subsessions, one of 40 and one of 35 successive periods. In both subsessions, inflation and output gaps evolve according to a baseline three-equation NK model (see, e.g., Galí 2015). The only difference from the standard textbook implementation is that expectations are provided by the subjects and need not align with RE or any *ad-hoc* learning mechanism commonly used in the theoretical literature.

[6] Our work builds on the literature that uses evolutionary algorithms to model heterogeneity; see, for instance, Anufriev and Hommes (2012); Anufriev et al. (2019).

[7] See, *inter alia*, Branch and McGough (2010); Branch and Evans (2011); Branch and Gasteiger (2019); Busetti et al. (2017); Gasteiger (2014); Massaro (2013).

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

In each period, we aggregate[8] the inflation and output gap expectations of the six subjects denoted by, respectively, $\pi^e_{j,t+1}$ and $y^e_{j,t+1}, j = 1, ..., 6$, using the arithmetic mean, that we denote by $E^*_t(\pi_{t+1})$ and $E^*_t(y_{t+1})$. These aggregate expectations are then inserted into the standard IS and Phillips curves given by:

$$\hat{y}_t = E^*_t \hat{y}_{t+1} - \sigma^{-1}\left(\hat{\imath}_t - E^*_t \hat{\pi}_{t+1} - \bar{r}\right) + g_t, \tag{1}$$

$$\hat{\pi}_t = \beta E^*_t \hat{\pi}_{t+1} + \kappa \hat{y}_t + u_t, \tag{2}$$

where $\hat{y}_t$ is the output gap, $\hat{\pi}_t$ the inflation gap, that is the difference between inflation and the CB target π^T, g_t and u_t exogenous real and cost-push shocks to be specified below, β and σ households' preference parameters, κ the slope of the Phillips curve and $\hat{\imath}_t$ the nominal interest rate set by the CB either with an IT or a PLT rule (depending on the experimental treatments, see Section 3.2), expressed in deviation from its steady-state level $\bar{r} \equiv \pi^T + \rho$. We use the standard parameter values $\beta = 0.994$, $\sigma = 1$, $\kappa = 0.0625$.

Each subsession starts with an initial learning phase for the subjects consisting of small white noise shocks around the targeted steady state. A severe, persistent and unannounced deflationary real shock then disturbs the economy in period 11 of Subsession 1 and a cost-push shock hits the economy in period 9 of Subsession 2.[9] The shock sequences are the same for all sessions and are displayed in Appendix B. In other words, we engineer two recessions in each experimental economy. In Subsession 1, the recession is coupled with low inflation and does not imply any stabilization trade-off but the severity of the initial shock induces an ELB episode in the model under RE independently of the monetary policy rule; see Section 3.2 and Fig. 1 below. By contrast, the recession in Subsession 2 is associated with inflationary pressures that imply a stabilization trade-off. Hence, we collect experimental data on the properties of IT and PLT in distinct economic environments. Additionally, to measure long-run inflation expectation anchoring in each environment, at period 26 of each subsession, we elicit a one-time inflation forecast for the end of the subsession that is, respectively, 15 and 10 periods later.

The role of the subjects is framed as professional forecasters for a statistical bureau. The instructions – reproduced in Appendix F – explain the positive and negative relationships between the main variables in the economy and provide the value of the inflation target. All but a few pieces of information are identical across all treatments: the description of the monetary policy rule differs under IT and PLT, and subjects additionally receive an inflation guidance under the treatment that tests PLT with communication (see Section 3.2 below).[10] At the beginning of every period in all treatments, subjects observe the current and past shocks and all past endogenous variables, including the price level and its path consistent with the inflation target, which is labeled as 'price level if inflation is always 5%' under IT and 'intended price level' in the PLT treatments where the concept of price gap is also introduced in the instructions. All these pieces of information are displayed in a table and in charts on the graphical user interface (GUI), reproduced in Appendix F.2.

Importantly, and in contrast to the existing literature, subjects also observe the most recent forecasts of all other forecasters in their economy along with their payoff. An explicit ranking from the best to the worst forecasters in which each given subject is included (tagged by 'YOU') is updated in

[8] See Arifovic et al. (2025, App. A), on which the HENK model in Section 5 is based, for the micro-foundations of this aggregation process.

[9] Appendix D reports on an additional treatment, denoted `PLT-G-R`, where we reverse the order of the shocks. Our main results are unchanged.

[10] Under IT, the inflation target is constant regardless of past outcomes or economic shocks. While this target is highlighted and repeated to the participants in the instructions and the comprehension quiz, previous studies show that simply announcing a fixed target does little to guide expectations when inflation is stable since it becomes redundant (Cornand & M'baye, 2018), or when inflation moves away from the target, in which case it becomes disconnected from the actual economic outcomes (Salle, 2023). Hence, our experiment focuses on communication under PLT only.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

each period on a table on the GUI.[11] In other words, subjects observe their peers' forecasting performances and can assess their relative performances. This design aligns the information set of the subjects with that of the HENK agents in Section 5.2 and of real-world professional forecasters, and does not affect cross-treatment comparisons. Moreover, the slow convergence under IT and volatile oscillations under PLT at the ELB in Section 4 match prior findings discussed in Section 2, which speaks against peer imitation affecting forecasting incentives.

Finally, subjects accumulate points that reward their forecast accuracy throughout the experiment. In each period, a forecast brings $\frac{100}{1+\text{absolute forecast errors in p.p.}}$ points, so that a perfect forecast brings 100 points. To rule out strategical hedging across the inflation and output forecasts, subjects accumulate a separate amount of forecasting points for each variable and only one of the two is randomly drawn at the end of the experiment to be converted into money at a rate of CA$0.2 per 100 points. To reinforce the salience of the occasional long-run inflation forecasting task, the points corresponding to these two forecasts are converted at a rate of CA$3 per 100 points.[12]

3.2. Treatments and FIRE predictions

We implement three experimental treatments, denoted by `IT`, `PLT`, and `PLT-G`. Tr. `IT` differs from Trs. `PLT` and `PLT-G` only by the monetary rule that sets the nominal interest rate in Eq. (1). Under `IT`, the CB implements a standard contemporaneous Taylor rule subject to the ELB:

$$\textbf{IT}: \hat{\imath}_t = \max\left\{0, \bar{r} + \phi^{\pi}\hat{\pi}_t + \phi^{y}_{IT}\hat{y}_t\right\}, \tag{3}$$

where the ELB is set to 0% and we use $\pi^T = 5\%$ and $\phi^{\pi}_{IT} = 1.5, \phi^{y} = 1$, which ensures determinacy under RE and stability under backward-looking expectations, see App. A.

In Trs. `PLT` and `PLT-G`, the CB uses a PLT monetary policy rule. It is convenient to define the price gap $\hat{P}_t \equiv \frac{P_t - \bar{P}_t}{\bar{P}_t}$, which measures the gap between the current price level P_t and its predetermined targeted level $\bar{P}_t$ that grows at a rate π^T consistent with the CB inflation target. In the log-linearized model, the price gap evolves as $\hat{p}_t = \hat{p}_{t-1} + \hat{\pi}_t$. In Trs. `PLT` and `PLT-G`, the CB sets the interest rate using the following rule:

$$\textbf{PLT}: \hat{\imath}_t = \max\left\{0, \bar{r} + \phi^{P}\hat{p}_t + \phi^{y}\hat{y}_t\right\}, \tag{4}$$

where we use $\phi^P = 1.5$ so that the initial policy responses to the shocks are the same in all treatments (holding everything else equal) and determinacy under RE and stability under learning of the targeted steady state hold; see App. A.

Tr. `PLT-G` differs from Tr. `PLT` only by the provision to the subjects of an inflation guidance. In Tr. `PLT-G`, in each period, subjects receive an inflation guidance that corresponds to the implicit time-varying inflation level that would close the price gap in any period t. (see App. A for detail). Hence, the guidance, that we denote by π^{CB}_t, is model-consistent. Under the FIRE assumption, it is redundant because expectations are already model-consistent, and there is no difference between Trs `PLT` and `PLT-G`. Under real-world expectations however, the guidance may help subjects coordinate their expectations on the history-dependent path. In each period, subjects receive the guidance in a bold and highlighted manner on the GUI, and all previous guidance announcements are displayed on the inflation graph; see App. F.2.

[11] In line with standard ethical rules, the identity of each individual subject remained, of course, anonymous and subjects were labeled by their number in the experiment.

[12] These are the initial 2022 exchange rates. Following the post-COVID-19 lab guidelines at SFU and CREED, for the 2025 additional sessions, we set CA$0.30 and € 0.40 per 100 short-run points, and CA$4 and € 5 for long-run forecasts; see Section 3.3.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

The guidance carries the underlying structure of the PLT regime while keeping the frame of the CB communication in terms of inflation rather than price level, which we conjecture to be an easier-to-grasp and more practical communication regime for the public than a communication in terms of price levels.

More generally, we phrase monetary policy in terms of inflation rate under both IT and PLT, even though the price level may be the salient variable under PLT, and eliciting inflation expectations need not be equivalent to eliciting price-level expectations in terms of expectation formation and macroeconomic dynamics. Our choice is motivated by external validity concerns: in practice, policy institutions communicate in terms of inflation rates, not price levels. Significant resources are devoted to measuring and analyzing inflation expectations under the prevailing IT frameworks. Moreover, inflation, not the price level, underpins key economic measures such as real returns or wages, and international comparisons. In most macroeconomic models, nominal variables like the price level are also indeterminate, and the focus is thus placed on inflation dynamics. For these reasons, both academic and policy discussions of PLT are typically framed in terms of inflation rates, and so is our experiment.

Figure 1 provides the dynamics of the IT and PLT experimental economies under the FIRE benchmark given our calibration; see Appendix A for the derivations. In line with the discussion in introduction, it is clear that PLT is superior to IT for a CB primarily concerned with inflation stabilization. PLT better stabilizes inflation and output gaps than IT in the absence of policy trade-off (namely in face of deflationary pressures, see left panel), even though the ELB binds under both regimes (bottom left panel). When facing inflationary pressures (right panel), PLT better limits the rise in inflation but at the price of a deeper recession than under IT (middle right panel). Comparing inflation and output gap dynamics in the three treatments allow us to assess whether this ranking is robust to laboratory expectations.

3.3. Implementation

The experimental software was programmed using `oTree` (Chen et al., 2016). A first batch of six experimental sessions for each of the three treatments was conducted in person at the CRABE lab at Simon Fraser University (SFU) between August and October 2022. Six additional sessions per treatment were added, two at SFU in January and February 2025 and four at the CREED lab of the University of Amsterdam in April of the same year, involving a total of 216 subjects. The robustness treatment `PLT-G-R` in Appendix D was also run with the same subject pool composition in 2025. The average duration of each session was just over two hours. In 2022, SFU participants earned an average of CA$24.8 including a CA$7 participation fee (from a minimum of 15 to a maximum of 38 with a standard deviation of 6.75), and CA$31.16 with a standard deviation of 5.03 in 2025 following the post-COVID-19 SFU lab guidelines. Following the 2025 CREED guidelines, participants earned an average of € 36.9 (with a standard deviation of 7.3), which reflects higher payment expectations given a higher net minimum wage and higher recent inflation in the Netherlands than in Canada.

Upon entering the lab, subjects were provided with the instruction booklet which gives a qualitative overview of the relationships among the variables and clearly explain the tasks.[13] Subjects were then asked to carefully read the instructions and complete a comprehension quiz in order for the experiment to start. The forecasting game started with Subsession 1. All predictions were submitted on the GUI. At the end of the first subsession and before starting Subsession 2, subjects were provided with the updated instruction booklet that highlights the change in the shock, from real to cost-push shocks. We introduced information about the shocks in such a sequential manner to prime subjects towards the particular economic environment implemented while avoiding the cognitive overload that could result from simultaneous information about various types of shocks. Finally, at the end

[13]Standard laboratory procedures apply, e.g. identification, consent acknowledgment, etc.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

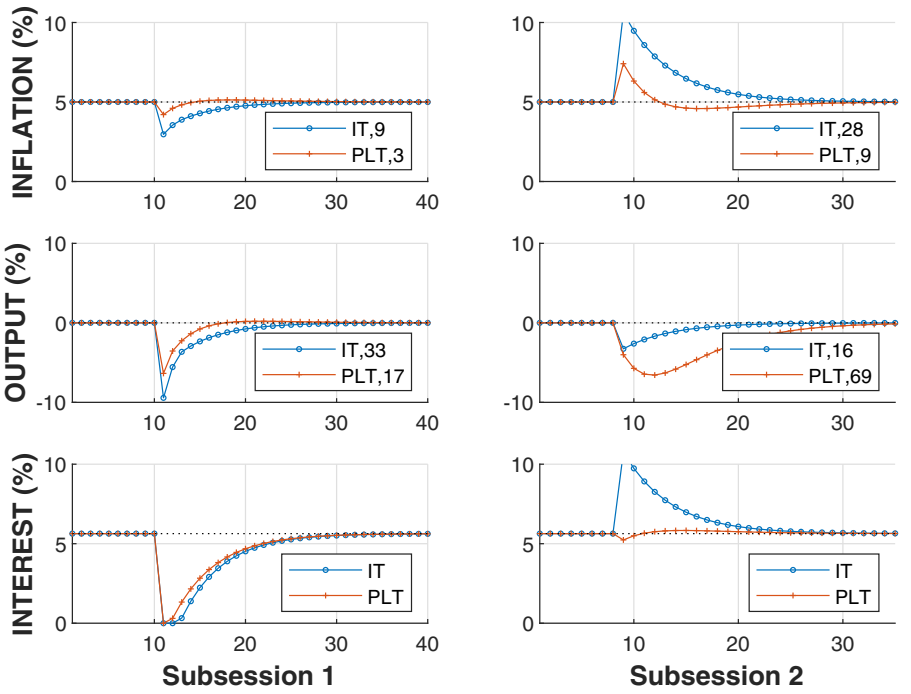


Fig 1. Simulations under RE

Notes: The values in the boxes report the total absolute deviation (in p.p.) for IT (blue circles) and PLT (rec crosses). We use the `occbin` library in `Dynare` to account for the occasionally binding constraint (Guerrieri & Iacoviello, 2015). The shocks in the simulations are an AR(1) process with an autoregressive coefficient of 0.8 which fits the series of three negative shocks in the experiment.

of the second subsession and before leaving the lab, the subjects were asked to complete an exit survey consisting of demographic questions, central banking literacy, and the strategy(ies) used in the experiment, if any.

4. Experimental results

We describe the macroeconomic dynamics and then dig into the individual forecast times series to study cross-sectional dispersion and show evidence of peer imitation.

4.1. Macroeconomic dynamics in the experimental economies

Figs 2-4 report the dynamics of inflation, output gap and interest rate in all experimental economies (dashed blue lines) in, respectively, Trs. `IT`, `PLT`, and `PLT-G`, along with the median realization over the 12 sessions in each treatment (red thick lines). In Tr. `PLT-G`, since we may not directly observe whether subjects use the CB guidance to form their inflation forecasts, we compute a "credibility index," i.e. the distance between individual inflation forecasts and the guidance, averaged across all subjects for each period:

$$Cred_t^{CB} = \frac{1}{6}\sum_{j=1}^{6} exp(-(\pi_{j,t+1}^{e} - \pi_t^{CB})^2), \tag{5}$$

where a higher value indicates a higher CB credibility. The bottom panel of Fig. 4 displays the dynamics of this credibility index in Tr. `PLT-G`.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

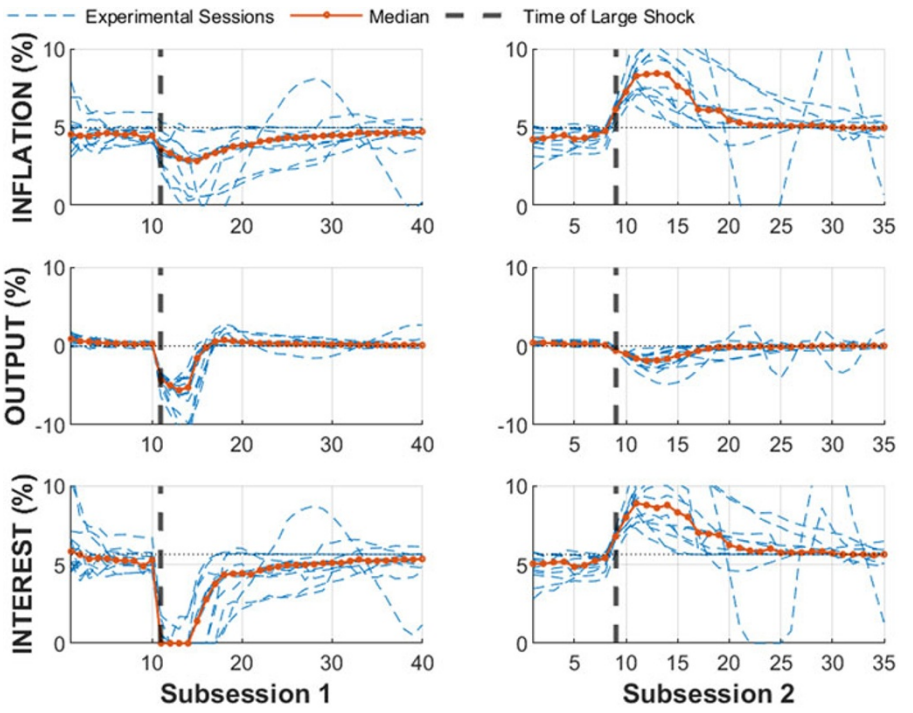


Fig 2. Inflation, output gap and interest rate dynamics in Tr. `IT`

Table 1 reports cross-treatment comparisons regarding inflation and output gap stabilization. To do so, we compute the per-period average of the absolute deviation (AAD) of inflation and output gap from their targets. For each session $s = 1, .., 12$ and subsession $ss = 1, 2$ with a given duration of $T = 40, 35$ periods, the AAD for each variable reads as:

$$AAD^{\pi}_{s,ss} = \frac{1}{T}\sum_{t\in T} |\pi_{s,ss,t} - \pi^{T}| \text{ and } AAD^{y}_{s,ss} = \frac{1}{T}\sum_{t\in T} |y_{s,ss,t}|, \tag{6}$$

where lower values indicate a better stabilization of inflation and output gap. In Table 1, we disentangle pre-shock (top panel) from post-shock periods, which include the shock periods themselves (bottom panel).

Overall, the merits of PLT under FIRE do not extend to the lab environment: experimental economies under PLT are more unstable than under IT. However, once augmented with CB guidance in Tr. `PLT-G`, the stabilization benefits of PLT materialize in the lab and this treatment delivers the smallest inflation gap.

In detail, in the first periods of each subsession – i.e. before the shocks, inflation and output appear similarly stabilized around their target under IT and PLT with guidance while displaying wilder oscillations under PLT. This visual impression is confirmed by the cross-treatment comparisons of the AADs (top panel of Table 1). The ADDs of inflation and output are significantly higher in Tr. `PLT` than in Trs. `IT` and `PLT-G`. In the aftermath of the large shocks, in both subsessions, the cross-treatment differences become even more salient: while economies in `IT` and `PLT-G` eventually converge back to the targeted steady-state, economies in Tr. `PLT` exhibit wilder and non-dampening oscillations. The bottom panel of Table 1 reports that the AADs of both inflation and output gap are considerably and significantly higher in Tr. `PLT` than in Trs. `IT` and `PLT-G`. The standard deviations of the AADs

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

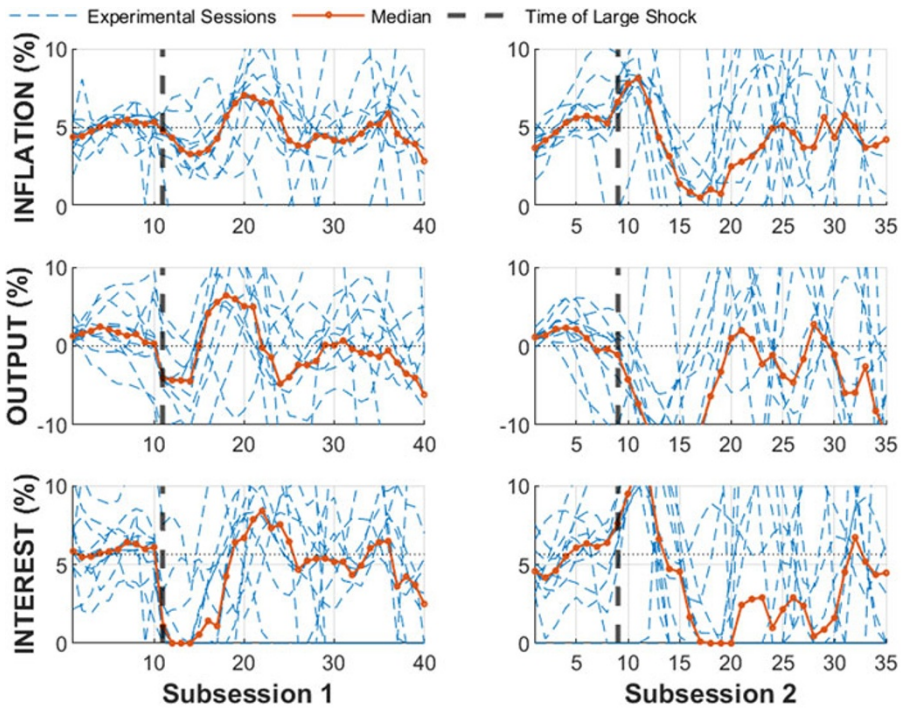


Fig 3. Inflation, output gap and interest rate dynamics in Tr. PLT

under PLT are also strikingly larger than in the two other treatments, which indicates more variation across independent economies.

Even if inflation returns towards the target in both Trs. IT and PLT-G, the inflation gap under IT exhibits more persistence in the aftermath of the recessions. Inflation remains persistently below target in the deflationary shock subsession (left panel) and persistently above target in the cost-push shock subsession (right panel). These observations are consistent with the persistently low inflation experience post-Great Financial Crisis and the persistently high inflation in the post-COVID-19 pandemic. These circumstances result in an unanchoring of inflation expectations, see Section 4.2, consistent with the hoovering of the one-year-ahead inflation forecasts from professionals (SPF) below 2% in the US and the Euro area over the 2010 decade, and the recent landing of both household and professional one-year-ahead inflation forecasts toward 3% in the US (see the SPF and Michigan household survey as of 2024Q4).

By contrast, in Tr. PLT-G, inflation returns to the target in a quicker time frame than under IT with negligible fluctuations, as shown in Fig. 4. This inflation stabilization in Tr. PLT-G is observed under both types of shocks but is most salient after the deflationary shock. In particular, in the wake of the deflationary recession, the ELB binds for a median time of only two periods under Tr. PLT-G, in line with the FIRE benchmark and considerably less than the four periods of ELB in Tr. IT. Looking at the credibility index in Tr. PLT-G (bottom panel of Fig. 4), while credibility substantially drops at the onset of the recessions, it increases back toward one (full credibility) along the recoveries. Finally, in line with the FIRE prediction, a better inflation stabilization in Tr. PLT-G than in Tr. IT comes at the cost of a wider output gap, in particular after a cost-push shock. This is confirmed in the bottom

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

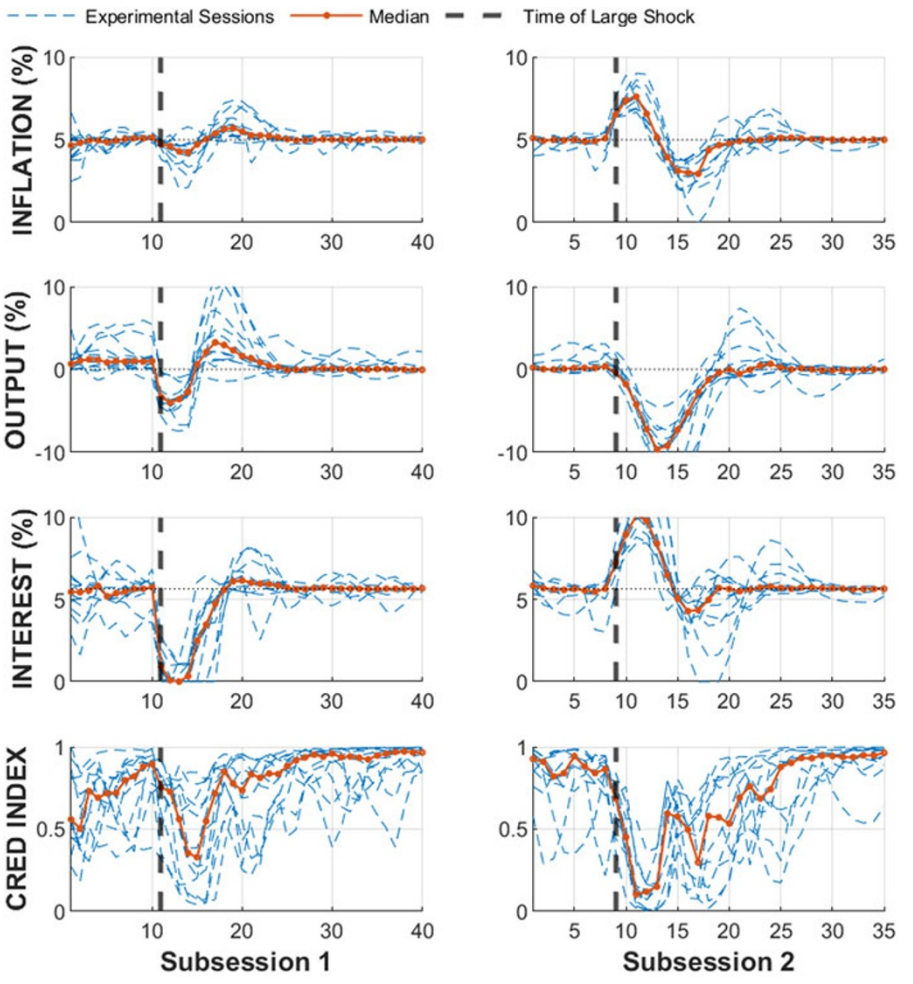


Fig 4. Inflation, output gap and interest rate dynamics in Tr. PLT-G

panel of Table 1: the output gap in Subsession 2 is significantly wider in Tr. PLT-G than under Tr. IT.

We now turn to the analysis of individual forecasts.

4.2. Heterogeneity in individual forecasts

Fig. 5 reports the median short-run forecast dispersion across all 12 sessions of each treatment, measured by the standard deviation of the distribution of forecasts of the six subjects in each period. Table 2 complements the graphical analysis with the corresponding cross-treatment comparative statistics.

Forecast dispersion varies both within treatments, i.e. over time, and between treatments. From Fig. 5, we see that pre-shock forecast dispersion declines in all treatments and subsessions, especially in Trs. IT and PLT-G, which suggests that subjects learn to coordinate on the target in the absence of large shocks in these two treatments. Table 2 (top panel) indicates that pre-shock disagreement is similar across treatments in Subsession 1 but significantly higher in Subsession 2 under Tr. PLT than in the two other treatments.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

Table 1. Cross-treatment comparisons of inflation and output gap in the experiment

		Numerical values			*Associated p-values*		
		IT	PLT	PLT-G	IT *vs* PLT	IT *vs* PLT-G	PLT *vs* PLT-G
Top panel: ***Pre-shock AAD*** *(p.p.)*							
π	**Sub 1**	0.68	0.67	0.32	0.93	0.03	0.03
		(0.39)	(0.36)	(0.18)			
	Sub 2	0.67	2.07	0.28	0.08	0.06	< 0.01
		(0.56)	(3.33)	(0.29)			
$\mathbf{y}$	**Sub 1**	0.47	2.65	1.51	< 0.01	0.02	0.04
		(0.22)	(1.67)	(1.36)			
	Sub 2	0.32	4.55	0.87	< 0.01	0.71	< 0.01
		(0.24)	(6.15)	(1.20)			
Bottom panel: ***Post-shock AAD*** *(p.p.)*							
π	**Sub 1**	1.24	5.14	0.36	0.05	0.02	< 0.01
		(0.81)	(7.47)	(0.29)			
	Sub 2	1.56	10.33	1.18	< 0.01	0.38	< 0.01
		(1.37)	(10.12)	(1.16)			
$\mathbf{y}$	**Sub 1**	1.25	11.10	1.56	< 0.01	0.38	< 0.01
		(0.58)	(12.85)	(0.94)			
	Sub 2	0.76	19.99	3.5	< 0.01	< 0.01	< 0.01
		(0.66)	(17.19)	(3.29)			

Note: The average of the absolute deviation (AAD) of inflation π and output y from their target displayed in the first three columns are computed as per Eq. (6) and are averaged across all 12 sessions of each treatment. Standard deviations are reported between brackets. The p-values displayed in the last three columns correspond to the non-parametric Wilcoxon rank-sum tests for comparison between each pair of treatments performed at the matching group level with $N = 12$ observations per treatment. The p-values are adjusted for multiple comparisons with the Holm (1979) sequentially-rejective procedure.

In the wake of the shocks, forecast dispersion rises in all treatments. In Trs. IT and PLT-G, in line with the better stabilization of the endogenous variables compared to Tr. PLT, forecast dispersion tends to revert back towards its pre-shock level, while it remains persistently elevated in Tr. PLT.[14] The statistical tests reported in the middle panel of Table 2 confirm this pattern: after the shocks, short-run forecasts are significantly more dispersed under Tr. PLT than under Trs. IT and PLT-G. By contrast, forecast dispersion is not significantly different between Trs. IT and PLT-G, except for output forecasts after the cost-push shock in Subsession 2 which are more dispersed under PLT-G than under IT. This difference is consistent with the sharper stabilization trade-off and resulting steeper output drop in the wake of the shock under PLT than under IT discussed above. However, Fig. 5 shows that, by the end of Subsession 2, output forecast dispersion converges to a negligible level in Tr. PLT-G, comparable to the level observed in Tr. IT, while it remains elevated in Tr. PLT.

The bottom panel of Table 2 reports the average long-run inflation forecasts (namely for the final period of each subsession) across all subjects, which speaks to cross-treatment differences in the level of long-run expectations in each subsession. Fig. A.2 plots their distributions, which illustrates their dispersion. While we cannot find any statistically significant difference across the three treatments in

[14]More forecast dispersion also means greater confusion and poorer forecasting performances amid the greater macroeconomic volatility in Tr. PLT than in the two other treatments. Table A.2 in App. C confirms that the payoff of the subjects is about one-third to one-half smaller in Tr. PLT than in the two other treatments.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

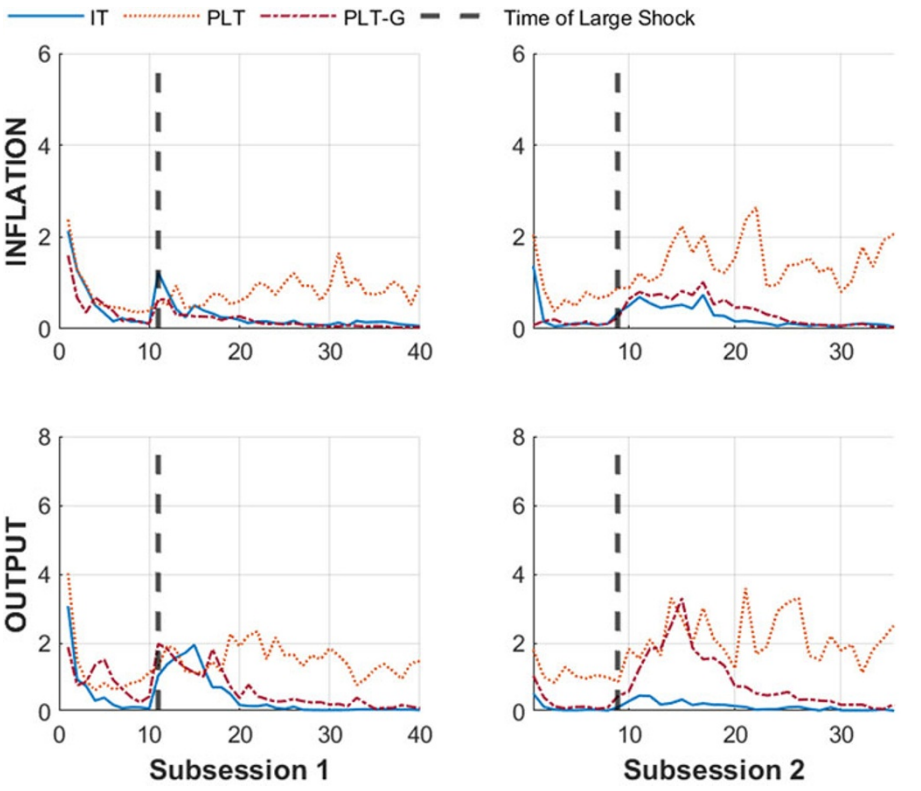


Fig 5. Between-subject disagreement in short-run forecasts
Note: Median over time (x-axis) across all 12 sessions in each treatment in p.p. (y-axis).

their level, their distribution differs sharply: under Tr. PLT, long-run forecasts are strikingly more dispersed, especially in Subsession 2, whereas they remain more tightly concentrated around the target under Trs. IT and PLT-G.

To gain further insights into long-run anchoring of expectations, Table 3 reports the estimated asymptotic values of the short-run forecasts, in deviation from the target, using the method of Duffy (2016). Short-run forecasts converge to the target in a strong sense only in Tr. PLT-G, with the asymptote being the largest in Tr. PLT.[15]

To conclude, we highlight the two main experimental results:

(IT>PLT) Contrary to the FIRE predictions, inflation and output gaps are better stabilized around their target under IT than under PLT, no matter the type of shocks. Short-run expectations are also more homogeneous under IT than under PLT, especially after shocks, and inflation forecasts are better anchored at the target.

(PLT-G>IT) Once the CB communicates in each period the inflation rate that is consistent with closing the price gap, the superior stabilization properties of PLT under FIRE materialize: shocks have less persistent effects, the ELB binds for fewer periods, and expectations are better anchored

[15]The method of Noussair et al. (1995), as in e.g. Riedl and Van Winden (2012), also reveals that inflation forecasts do not converge to the target in Tr. PLT, where the estimated asymptotic deviation from the target is −3.78, with a robust standard error of 1.15, but do so in Trs. IT and PLT-G.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

Table 2. Cross-treatment comparisons of subjects' forecasts in the experiment

		Numerical values			*Associated p-values*		
		IT	PLT	PLT-G	IT *vs* PLT	IT *vs* PLT-G	PLT *vs* PLT-G
Top panel: ***Pre-shock disagreement in short-run forecasts*** (p.p.)							
π	**Sub 1**	0.73	1.14	0.91	0.29	0.80	0.41
		(0.34)	(0.82)	(0.93)			
	Sub 2	0.35	1.19	0.32	< 0.01	0.48	< 0.01
		(0.24)	(0.88)	(0.29)			
y	**Sub 1**	0.80	1.91	1.46	0.04	0.18	0.32
		(0.42)	(1.55)	(1.30)			
	Sub 2	0.20	1.87	0.44	< 0.01	0.02	< 0.01
		(0.14)	(1.87)	(0.27)			
Middle panel: ***Post-shock disagreement in short-run forecasts*** (p.p.)							
π	**Sub 1**	0.43	1.34	0.33	< 0.01	0.48	< 0.01
		(0.31)	(0.69)	(0.26)			
	Sub 2	0.54	2.20	0.59	< 0.01	0.24	< 0.01
		(0.68)	(1.28)	(0.51)			
y	**Sub 1**	0.58	2.61	0.91	< 0.01	0.11	< 0.01
		(0.37)	(1.52)	(0.54)			
	Sub 2	0.43	3.15	1.31	< 0.01	< 0.01	< 0.01
		(0.41)	(1.98)	(1.12)			
Bottom panel: ***Long-run inflation forecasts*** (p.p.)							
π	**Sub 1**	4.37	4.54	5.39	0.12	0.03	0.71
		(0.92)	(5.18)	(1.06)			
	Sub 2	5.40	1.39	5.42	> 0.99	> 0.99	> 0.99
		(0.83)	(11.11)	(1.17)			

Note: See Table 1. Forecast disagreement (top and middle panel) is computed as the standard deviation across the six forecasts in each period and is averaged over the 12 sessions of each treatment. The bottom panel reports the average of the long-run inflation forecasts, elicited once per subsession for the final period of the subsession. The associated p-values report pairwise cross-treatment comparisons.

Table 3. Asymptote estimates of the deviation of the average short-run inflation forecasts from the target

	IT	PLT	PLT-G
Asymptote estimate (sd.)	−0.15 (1.39)	−5.79 (11.36)	-0.003^{*} (0.33)
Lower and upper p-values	(0.40, 0.32)	(0.68, 0.29)	(0.06, 0.06)

Note: Estimated mean of an AR(1) process on short-run inflation forecasts, averaged per period at the treatment level; see Duffy (2016). The p-values refer to the equivalence test using the two one-sided test (TOST) procedure where the null hypothesis is a deviation of the asymptotic forecast value from a 10%-radius of the 5% target, and the alternative hypothesis is the convergence of the long-run forecasts within 10% of the target. ***, ** and * indicate that non-convergence is rejected at 1, 5, and 10%.

around the target under PLT with guidance than under IT, while remaining equally coordinated under both regimes.

We now dig further into individual forecast data.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

4.3. Expectation formation and evidence of peer imitation

We first use panel models at the subject-period level to identify which pieces of the available information significantly relate to the inflation and output forecasts in each treatment. The details of the models are deferred to Table A.1 in App. C. The main takeaway is that the subjects react to the most recent inflation and output gap values, their own latest forecasts, the best forecasts in their group and, to a lesser extent, the shocks. They never react to the past price gap, which indicates that expectations are not model-consistent in Tr. PLT. Moreover, consistent with prior experimental findings (e.g., Assenza et al., 2021), inflation and output gap forecasts appear decoupled: inflation forecasts respond primarily to past inflation and past inflation forecasts, while output gap forecasts are linked to their own respective histories, with minimal cross-influence.

To refine our description of the forecasts and uncover evidence of peer imitation, we estimate heuristics on the forecast time series for each subject. We modify the set of heuristics that usually approximate well the forecasting behaviors of the subjects in this class of experiments[16] to include the forecasts of the most accurate forecaster of the group in any given period. We denote this additional piece of information by $\pi^e_{BF,t}$. For each subject, we estimate the following inflation forecasting models:

$$\pi^e_{t+1} = f(\pi^e_{BF,t}) + \epsilon_t, \tag{7}$$

$$\pi^e_{t+1} = f(\pi^e_{BF,t}, \pi^e_t) + \epsilon_t, \tag{8}$$

$$\pi^e_{t+1} = f(\pi_{t-1}, \pi_{t-2}) + \epsilon_t, \tag{9}$$

$$\pi^e_{t+1} = f(\pi_{t-1}, \pi_{t-2}, \pi^e_{BF,t}) + \epsilon_t, \tag{10}$$

$$\pi^e_{t+1} = f(\pi_{t-1}, \pi_{t-2}, \pi^e_{BF,t}, \pi^e_t, F_t) + \epsilon_t, \tag{11}$$

$$\pi^e_{t+1} = f(\pi^e_t, \pi_{t-1}) + \epsilon_t, \tag{12}$$

$$\pi^e_{t+1} = f(F_t) + \epsilon_t, \tag{13}$$

$$\pi^e_{t+1} = f(\pi_{t-1}, \pi_{t-2}, F_t) + \epsilon_t, \tag{14}$$

where $f(\cdot)$ is a linear function in its arguments. Models (7), (8), (10) and (11) make use of the best forecasts and correspond to "imitation models." Model (7) is a "copy-cat" strategy that consists in adopting the best forecast. Model (8) also relies on the individual previous forecasts and may therefore be defined as an "adaptive imitator." Model (10) corresponds to an "anchoring-and-adjustment" rule where the anchor is the best forecast and the adjustments take place in the direction of the most recent inflation trend. Model (11) is a mixed model where all these pieces of information are used to make predictions. Models (9), (12) (13) and (14) are the four benchmarks discussed in Footnote 16. We include these standard 'non-imitation' models to allow for the possibility that subjects do not significantly imitate. We use the Akaike Information Criterion (AIC) to classify each subject across one of these forecasting heuristics. The results are presented in Fig. 6 for inflation forecasts and in Fig. A.4

[16] See, *inter alia*, Anufriev and Hommes (2012); Assenza et al. (2021); Mokhtarzadeh and Petersen (2021) where Models (9), (12), (13) and (14) which deliver, respectively, trend-following expectations, adaptive expectations (including the limiting case of naive expectations), 'fundamentalist' expectations, and a mixture of these models are the benchmarks to describe forecasting behaviors. The fundamentals F_t are the shocks in t, plus the past price gap in Tr. PLT and the CB guidance (which is a linear function of the past price gap) in Tr. PLT-G. See App. C for the output gap forecasting models.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press



Fig 6. Proportions of imitation, mixed and non-imitation heuristics among all subjects

for output gap forecasts, where we show the fraction of 'imitators' (defined as subjects assigned to Models (7), (8) or (10)), 'mixed imitators' (assigned to Model (11)) and non-imitators (assigned to any other models) over all sessions, per treatment and subsession.

Overall, we find evidence of peer imitation, and to a similar extent in both inflation and output gap forecasts and across treatments.[17] Imitation may arise due to the complexity of the environment or strategic complementarities in the NK models, where aligning with others' expectations is payoff-enhancing. Salient reference points – like the best forecaster's predictions or the CB guidance – can serve as coordination devices since they are self-fulfilling.

Finally, Fig. A.3 plots the distributions of the estimated coefficients of the imitation models. The coefficient values are consistent with the interpretation of each rule. In the "copycat" model (7), the estimated coefficients associated with the best forecasts are distributed around 1 (see the top-left panel of Fig. A.3). In the "anchoring-and-adjustment" rule (10), the coefficients associated with the past two inflation lags are of the same magnitude on the unit interval but of opposite signs, which is consistent with an adjustment in line with the latest inflation trend (see top-right panel). The "adaptive imitator" model (8) reflects a weighted average between the best forecast and the latest inflation (see bottom-left panel). The estimated coefficients associated with these pieces of information remain of the same order of magnitude in the mixed model (see bottom-right panel).

We now use a behavioral model with peer imitation to rationalize the lab data.

5. A parsimonious model of the experimental data

5.1. Overview of the HENK model

We describe here expectations in the HENK model, based on a simplified version of Arifovic et al. (2025), and refer to Appendix E for detail. Each agent j enters each period with their own inflation and output gap forecasts $\xi^x_{j,t-1}$, $x = \pi, y$. They then receive private news $\iota^x_{j,t}$ which modifies their forecasts into: $m^x_{j,t} \equiv \xi^x_{j,t-1} + \iota^x_{j,t}$, $x = \{y, \pi\}$. Each updated belief m is evaluated using its '*F*oregone' error, denoted by $F^x_{j,t} \equiv (\hat{x}_{t-1} - m^x_{j,t})^2$, which represents the forecast error that the agent would have made in $t-1$ had they used m^x as a forecast for x. Finally, agents are paired for peer imitation, and the one with the higher foregone error adopts the other's beliefs, which sets $\xi^x_{j,t}$. This process repeats itself in the next period.

[17] The p-values of the Chi-squared proportion tests are 0.55 between Trs. `IT` and `PLT`, 0.99 between Trs. `IT` and `PLT-G`, and 0.63 between Trs. `PLT` and `PLT-G` for inflation forecasts, and 0.27, 0.72 and 0.11 respectively for the output forecasts.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

With CB guidance, agents have a third element, a "credibility" level, denoted by $c_{j,t} \in [0,1]$, that is their probability of aligning their inflation forecast with the guidance, instead of using their own belief. This component is also modified by news. The foregone inflation forecast error is adjusted to reflect the weighted average of the two errors under the two forecasting strategies, i.e. $F^{\pi}_{j,t} = c_{j,t}\left(\widehat{\pi}_{t-1} - \pi^{CB}_{t}\right)^2 + \left(1 - c_{j,t}\right)\left(\widehat{\pi}_{t-1} - m^{\pi}_{j,t}\right)^2$. The CB guidance does not affect output gap expectations.

Expectations are initialized around the steady state so that the news are the only processes to be specified, which ensures parsimony. We use independent Gaussian draws $\iota^{x}_{j,t} \sim \mathcal{N}(\theta^{x}(g_t, u_t), \sigma_x)$, $x = y, \pi, c$, where the 'average news' θ^{x} about variable x linearly depends on the latest observable shocks so that the news convey some fundamental information while keeping an unconditional zero mean. Since the instructions qualitatively describe the effects of the shocks, we impose $\frac{\partial\theta^{\pi}}{\partial g}, \frac{\partial\theta^{\pi}}{\partial u}, \frac{\partial\theta^{y}}{\partial g}, \frac{\partial\theta^{c}}{\partial g} \geq 0$ and $\frac{\partial\theta^{y}}{\partial u}, \frac{\partial\theta^{c}}{\partial u} \leq 0$, which means that in the model, the negative real shock g in Subsession 1 tends to decrease inflation expectations, credibility and output gap outlooks, and a positive cost-push shock in Subsession 2 tends to bring inflationary and recessionary prospects with a lower CB credibility.

The model is simulated without any input from the experiment besides the time series of the shocks and the parameter values of the New Keynesian model. The information set of the HENK agents is consistent with that of the experimental subjects and the model nests the RE benchmark, which avoids taking an ex-ante stand on the nature of expectations that generates the experimental dynamics.

5.2. Cross-treatment comparisons

Figs 7-9 plot the simulated counterparts of Figs 2-4 under, respectively, `IT`, `PLT` and `PLT-G`, along with the median realizations in the experiment. A first glance at the figures reveals that the model can account for the main five stylized facts of the experiment.

First, inflation and output exhibit wilder and non-dampening oscillations under `PLT` compared to the stabilization observed towards the end of the subsessions under `IT` and `PLT-G`. Second, inflation returns to the target more quickly under `PLT-G`, along mild, symmetric oscillations, than under `IT`, where the inflation gap is more persistent – negative in the first subsession and positive in the second. Third, this persistence translates into a fairly long episode of low nominal interest rates in Subsession 1 under IT, which echoes the 'low-for-long' narrative up until 2021. It is different under `PLT-G`, where the nominal rate converges back to steady state after a few periods only. Fourth, even under `PLT-G`, the macroeconomic oscillations contrast with the RE dynamics. Fifth, the recession after the cost-push shock in Subsession 2 under `PLT-G` is larger than under `IT`.

These cross-treatment differences are confirmed by statistical tests in Table 4. At least qualitatively, Table 4 replicates its counterpart Table 1. Before the shocks (top panel), `IT` and `PLT-G` regimes exhibit similar AAD-levels but the picture changes after the shocks (middle panel). `PLT-G` has the lowest inflation AAD and `IT` has the lowest output AAD, implying a quicker closure of the inflation gap under `PLT-G`, but at a higher stabilization cost in terms of output than under IT. Without guidance, PLT exhibits the largest AADs. Inflation differences are less pronounced than for output because the large oscillations under `PLT` emerge later in the simulations than in the experiment. Extrapolating the model for a few more periods reveals increasing volatility, consistent with the experiment (see Fig. A.5).

Additionally, the model captures the cross-treatment dispersion in individual forecasts; see the bottom panel of Table 4 and its experimental counterpart in Table 2. The best coordination among inflation forecasts (i.e. the lowest dispersion) is obtained in Tr. `PLT-G`, which reflects subjects' coordination on the CB's guidance. Inflation and output forecasts are most heterogeneous under Tr. `PLT`, in the experiment and in the model.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

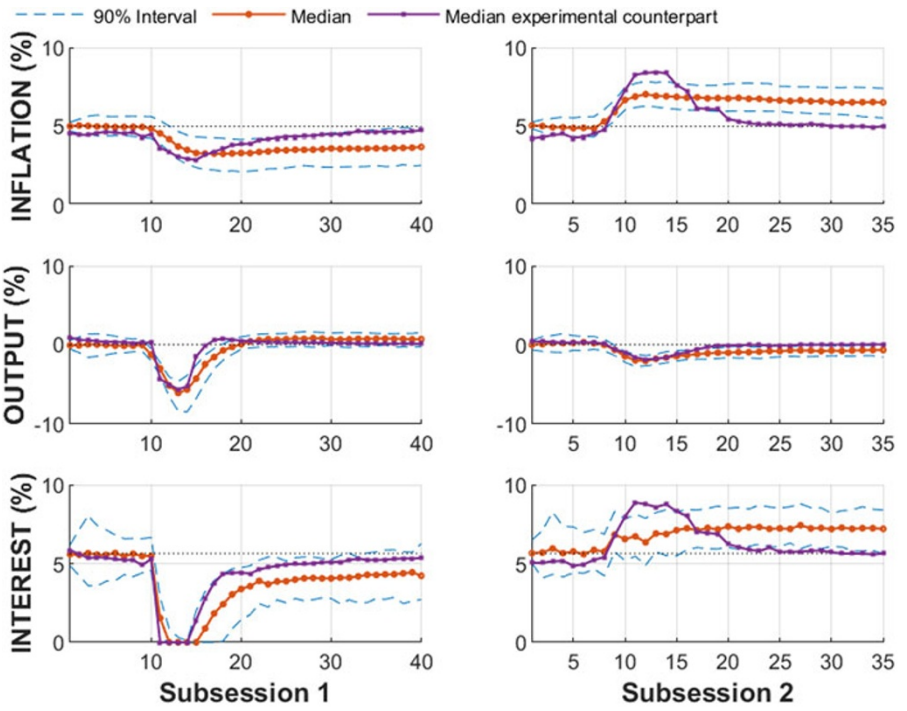


Fig 7. Inflation, output gap and interest rate dynamics under IT in the HENK model
Notes: 100 Monte Carlo simulations of the HENK model, see Appendix E for detail.

We find that the HENK model replicates six experimental stylized facts and the cross-treatment differences without using the experimental data. The model is parsimonious because only the news process has to be specified, and the simulations are robust to variations in this specification since news are screened out via performance-based peer imitation before affecting the economy.

5.3. Discussion

Unraveling the expectation dynamics in the HENK model helps describe the mechanisms at play in the experiment. Persistent expectations are a known feature of these models (Bullard, 2025). Under IT, persistent inflation gaps arise from pessimistic deflationary (in Subsession 1) and inflationary (in Subsession 2) outlooks at the onset of the recessions. These outlooks quickly become contagious because inflation expectations are almost self-fulfilling per the Phillips curve (2). This mechanism generates long-lasting coordination on stable but off-target expectations, both in the experiment and in the model. In this respect, persistence is stronger in the model where agents imitate in each period, than in the experiment where non-imitation is also prominent; see, again, Fig. 6.

When expectations are unanchored, pressures from aggregate demand can alternatively return inflation on target. Due to the nominal rigidities in the NK economy (i.e. a small slope κ in Eq. (2)), a sufficiently large change in output is required to turn around inflation, which unfolds slowly since the IT-CB also responds to the output gap. Post-recession, the economy enters a "new normal," where both inflation and interest rates settle on an off-target path. Consequently, neither HENK agents nor the subjects will coordinate on a different inflation expectation path unless actual inflation leads the way.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

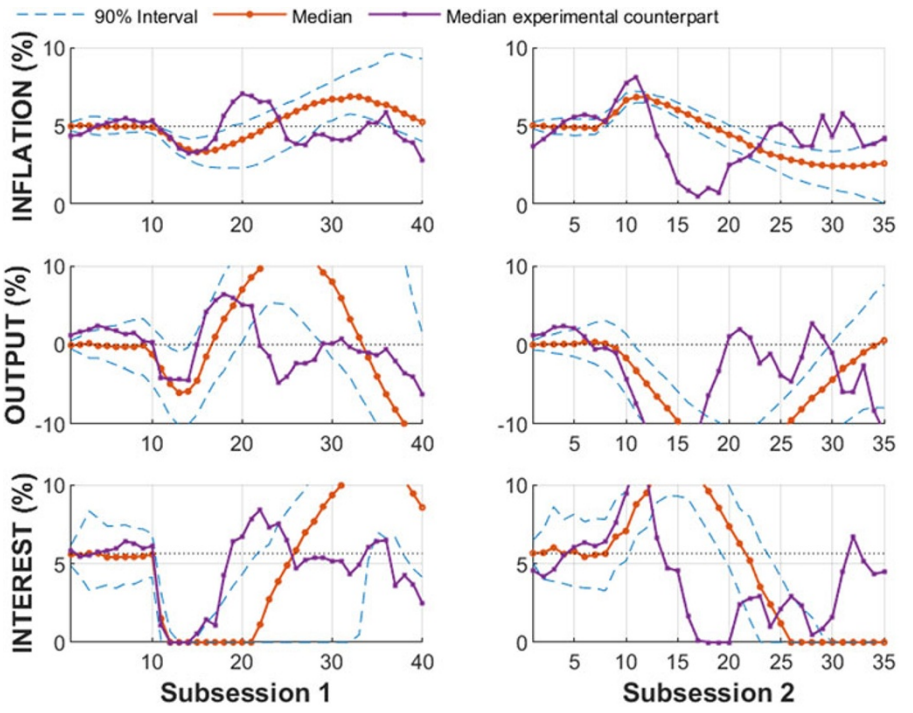


Fig 8. Inflation, output gap and interest rate dynamics under PLT in the HENK model
Notes: see Fig. 7

By contrast, the PLT-CB is more accommodating (in Subsession 1) or more restrictive (in Subsession 2) than under IT to close the price gap in face of shocks. This stronger reaction pressurizes prices and output up (or down) to escape the off-target region, which explains the deeper recession in Subsession 2 in Tr. `PLT-G` than in Tr. `IT`. This reaction also prompts agents to quickly discard off-target inflation forecasts and edge close to the inflation guidance. If initial credibility is sufficiently high, it persists because of self-fulfilling expectations and peer imitation, which limits the movements in inflation expectations caused by the shocks, and ensures a reinforcing loop between credibility and inflation stability. This loop is illustrated in the bottom panel of Fig. 9: while the average credibility briefly dips at the onset of the recessions, it largely recovers by the end of the subsessions.

Without guidance under `PLT`, expectations track recent inflation rather than reacting to the past price gap, which is key to forming history-dependent expectations. As a result, inflation plunges deeper following the deflationary shock in Subsession 1 and rises more following the cost-push shock in Subsession 2 than under `PLT-G`. These larger deviations accumulate in the price gap which delays policy normalization compared to IT. Hence, in Subsession 1, the ELB episode is longer than in the other treatments, in an attempt to boost inflation to close the price gap but at the price of a large overheating of the economy. The resulting rising inflation and output expectations further feed the boom. The PLT-CB eventually raises rates to offset the positive output gap but with a long delay due to the persistent negative price gap inherited from the deflationary shock. By the time the PLT-CB reverses its policy (namely abandons its ELB policy in Subsession 1 or its high-rate policy in Subsession 2), the output gap is particularly large (positive in Subsession 1 and negative in Subsession 2), which

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

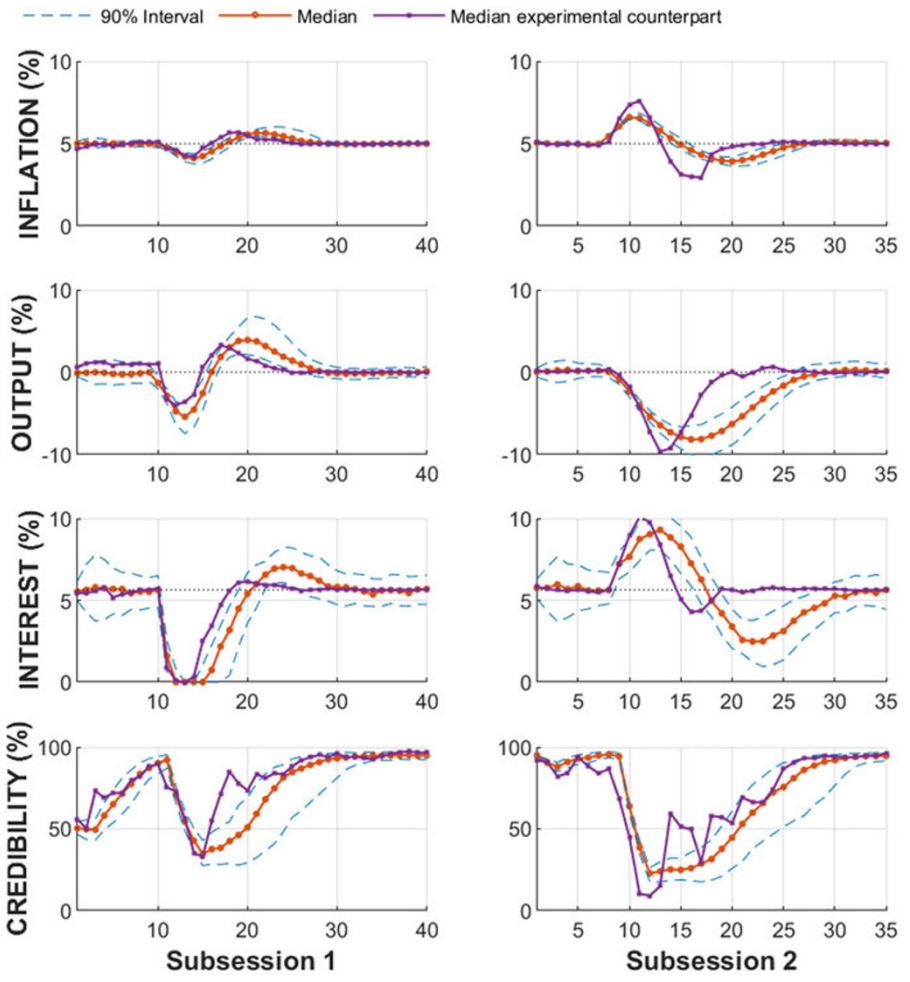


Fig 9. Inflation, output gap and interest rate under PLT-G in the HENK model
Notes: see Fig. 7. The aggregate credibility $\bar{c}_t$ of the CB (last row) is defined as the average of the individual components $c_{j,t}$ across the agents in each period.

prompts abrupt rate adjustments. These brutal adjustments, amplified by unanchored and contagious expectations, generate the large unsettling oscillations observed in both the experiment and the model.

6. Conclusion

The contribution of this paper is twofold. First, we empirically find supporting evidence for the merits of the history-dependent PLT policy – a monetary framework without historical experience so far – in stabilizing inflation and output gaps in face of both real and cost-push shocks. Second, we show how a parsimonious and micro-founded model of heterogeneous expectations replicates remarkably well the cross-regime differences observed in the lab and offers a behavioral description of these differences.

Admittedly, positive results in the lab may not warrant a successful empirical implementation. We rather take a 'wind tunnel' approach: the lab provides a first test of a policy within a highly controlled environment where internal validity allows the researchers to draw causal inference. Theories that

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

Table 4. Cross-treatment comparisons in the simulations

		Numerical values			*Associated p-values*		
		IT	PLT	PLT-G	IT *vs* PLT	IT *vs* PLT-G	PLT *vs* PLT-G
Top panel: ***Pre-shock AAD*** *(p.p.)*							
π	**Sub 1**	0.36	0.27	0.13	0.01	< 0.01	< 0.01
		(0.23)	(0.14)	(0.06)			
	Sub 2	0.37	0.31	0.10	0.13	< 0.01	< 0.01
		(0.25)	(0.20)	(0.02)			
y	**Sub 1**	0.70	1.55	0.79	< 0.01	< 0.01	< 0.01
		(0.28)	(0.87)	(0.30)			
	Sub 2	0.58	1.12	0.57	< 0.01	0.74	< 0.01
		(0.33)	(0.70)	(0.27)			
Middle panel: ***Post-shock AAD*** *(p.p.)*							
π	**Sub 1**	1.48	1.45	0.31	0.55	< 0.01	< 0.01
		(0.71)	(0.74)	(0.10)			
	Sub 2	1.72	1.74	0.58	0.95	< 0.01	< 0.01
		(0.61)	(0.60)	(0.15)			
y	**Sub 1**	1.57	8.74	1.88	< 0.01	< 0.01	< 0.01
		(0.50)	(2.94)	(0.61)			
	Sub 2	1.10	8.63	3.60	< 0.01	< 0.01	< 0.01
		(0.29)	(2.69)	(0.77)			
Bottom panel: ***Short-run forecasts disagreement*** *(p.p.)*							
π	**Sub 1**	1.63	1.64	0.90	0.13	< 0.01	< 0.01
		(0.05)	(0.05)	(0.09)			
	Sub 2	1.62	1.63	0.89	0.10	< 0.01	< 0.01
		(0.05)	(0.04)	(0.09)			
y	**Sub 1**	8.05	8.21	8.08	< 0.01	0.45	< 0.01
		(0.23)	(0.23)	(0.23)			
	Sub 2	8.00	8.16	8.02	< 0.01	0.40	< 0.01
		(0.22)	(0.23)	(0.23)			

Note: See Tables 1 and 2. Each pair-comparison involves 100 independent simulations.

have no explanatory power in the simple lab environment are unlikely to perform well in the (much) more complex real economies. In this respect, our results are encouraging for history-dependent policies and CB communication. We show how just one additional piece in the information set of the public may matter a great deal for expectation coordination. Importantly, our findings suggest that this benefit is robust to the type of shocks, i.e. the CB guidance coordinate expectations under both deflationary or inflationary pressures. Of course, ultimately, the decisions that agents make conditional on their expectations determine macroeconomic outcomes, not the expectations themselves. However, if agents fail to align their expectations with the monetary policy regime, one may question that their economic behaviors would.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

More broadly, this paper shows the added value of combining lab evidence with a HENK model to analyze monetary policy. This framework is an interesting complement to the CB's toolkit because it may easily be deployed to provide a first test of alternative communication contents and policy rules.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2026.10058.

Replication package. The full replication package, including the experimental software and the raw data, is freely accessible on *Mendeley data* at https://data.mendeley.com/datasets/2xpjr3bycr/2 (DOI: *10.17632/2xpjr3bycr.2*)

Acknowledgements. This project has been initiated within the context of the Fellowship Award of the Bank of Canada that Jasmina Arifovic received in 2020. Jasmina sadly passed away early 2022, before the completion of the present work. However, given this history, the two remaining co-authors opted for keeping her as a co-author but the views ultimately expressed in this paper may only be attributed to them. The study has received ethical approval from the SFU Office of Research Ethics numbered 30000994 and the FEB ethical committee at the University of Amsterdam under reference number EB-13597, and is registered under No. AEARCTR-0011128 on the AEA RCT Registry. Isabelle Salle acknowledges support from the Canada Research Chair program and the Research Priorities Area (RPA) 'Complex Human Systems Lab' from the Faculty of Economics and Business of the University of Amsterdam. The present work has benefited from fruitful discussions with Luba Petersen. We also thank the participants of the 56th Canadian Economics Association Conference in Ottawa on June 2022, the 1st PhD Conference on Expectations in Macroeconomics, held virtually by the Barcelona School of Economics on October 2022, the 8th Atlantic Canada Economics Association in Halifax on October 2022, the 15th Vietnam Economist Annual Meeting in Hanoi on November 2022, 1st CIREQ's Interdisciplinary Conference on Big Data and Artificial Intelligence in Montreal on June 2023. None of the above are responsible for potential errors in the paper.

Funding statement. Open access funding provided by University of Amsterdam.

References

Andrade, P., Gaballo, G., Mengus, E., & Mojon, B. (2019). Forward guidance and heterogeneous beliefs. *American Economic Journal: Macroeconomics*, *11*(3), 1–29.

Anufriev, M., & Hommes, C. (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics*, *4*(4), 35–64.

Anufriev, M., Hommes, C., & Makarewicz, T. (2019). Simple forecasting heuristics that make us smart: Evidence from different market experiments. *Journal of the European Economic Association*, *17*(5), 1538–1584.

Arifovic, J., Grimaud, A., Salle, I., & Vermandel, G. (2025). Social learning and monetary policy at the effective lower bound. *Journal of Money, Credit and Banking*, *57*(2-3), 439–475.

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

Assenza, T., Heemeijer, P., Hommes, C. H., & Massaro, D. (2021). Managing self-organization of expectations through monetary policy: A macro experiment. *Journal of Monetary Economics*, *117*(C), 170–186.

Bodenstein, M., Hebden, J., & Winkler, F. (2022). Learning and misperception of makeup strategies. *Journal of Economic Dynamics and Control*, *139*(C), 104417.

Branch, W. A. (2004). The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations. *The Economic Journal*, *114*(497), 592–621.

Branch, W. A., & Evans, G. W. (2011). Monetary policy and heterogeneous expectations. *Economic Theory*, *47*(2–3), 365–393.

Branch, W. A., & Gasteiger, E. (2019) Endogenously (Non-)ricardian beliefs, ECON WPS - working papers in economic theory and policy 03/2019, TU Wien, Institute of Statistics and Mathematical Methods in Economics, Economics Research Unit.

Branch, W. A., & McGough, B. (2010). Dynamic predictor selection in a new Keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control*, *34*(8), 1492–1508.

Bullard, J. (2025). Social learning for the masses. *Journal of Economic Dynamics and Control*, *172*(C), 104983.

Busetti, F., Delle Monache, D., Gerali, A., & Locarno, A. (2017) Trust, but verify. De-anchoring of Inflation Expectations under learning and heterogeneity. Technical Report 1994, ECB working paper.

Carroll, C., & Wang, T. (2023). Epidemiological expectations. In *Handbook of Economic Expectations*. (pp.779–806) Elsevier.

Cateau, G., Kryvtsov, O., Shukayev, M., & Ueberfeldt, A. (2009) Adopting price-level targeting under imperfect credibility in ToTEM, Staff Working Papers 09-17. Bank of Canada.

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

Cornand, C., & Heinemann, F. (2014). Experiments on monetary policy and central banking. In *Experiments In Macroeconomics, Vol. 17 of Research In Experimental Economics*. (pp. 167–227) Emerald Publishing Ltd.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

Cornand, C., & M'baye, C. K. (2018). Does inflation targeting matter? An experimental investigation. *Macroeconomic Dynamics*, *22*(2), 362–401.
Cornea-Madeira, A., Hommes, C., & Massaro, D. (2019). Behavioral heterogeneity in U.S. inflation dynamics. *Journal of Business & Economic Statistics*, *37*(2), 288–300.
D'Acunto, F., Malmendier, U., & Weber, M. (2023). What do the data tell us about inflation expectations? In *Handbook of Economic Expectations*. (pp. 133–161) Elsevier.
Del Negro, M., Giannoni, M. P., & Patterson, C. (2023). The forward guidance puzzle. *Journal of Political Economy Macroeconomics*, *1*(1), 000–000.
Duffy, J. (2016). Macroeconomics: A survey of laboratory research. In J. H. Kagel, & A. E. Roth (eds), *The Handbook of Experimental Economics*. (Vol 2). Introductory Chapters, Princeton University Press.
Evans, G., Gibbs, C., & McGough, B. (2025). A unified model of learning to forecast. *American Economic Journal: Macroeconomics*, *17*(2), 101–33.
Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton University Press.
Gasteiger, E. (2014). Heterogeneous expectations, optimal monetary policy, and the merit of policy inertia. *Journal of Money, Credit and Banking*, *46*(7), 1535–1554.
Guerrieri, L., & Iacoviello, M. (2015). OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily. *Journal of Monetary Economics*, *70*(C), 22–38.
Hennequin, M., & Hommes, C. (2024). Managing bubbles in experimental asset markets with monetary policy. *Journal of Money, Credit and Banking*, *56*(2-3), 429–454.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.
Hommes, C., & Lustenhouwer, J. (2019). Inflation targeting and liquidity traps under endogenous credibility. *Journal of Monetary Economics*, *107*(C), 48–62.
Hommes, C., & Makarewicz, T. (2021). Price level versus inflation targeting under heterogeneous expectations: A laboratory experiment. *Journal of Economic Behavior & Organization*, *182*(C), 39–82.
Hommes, C., Massaro, D., & Weber, M. (2019). Monetary policy under behavioral expectations: Theory and experiment. *European Economic Review*, *118*(C), 193–212.
Honkapohja, S., & McClung, N. (2024). On robustness of average inflation targeting. *European Economic Review, 183*, Article 105217. Available at SSRN 4021712.
Honkapohja, S., & Mitra, K. (2020). Price level targeting with evolving credibility. *Journal of Monetary Economics*, *116*(C), 88–103.
Kostyshyna, O., Ozden, T., & Zhang, Y. (2024) Endogenous credibility and wage-price spirals. Bank of Canada Working Staff Working Paper, Technical Report 2014-14.
Kostyshyna, O., Petersen, L., & Yang, J. (2024). *Why don't history-dependent monetary policies work*. Simon Fraser University, mimeo.
Kryvtsov, O., & Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, *117*(C), 760–780.
Marimon, R., Spear, S. E., & Sunder, S. (1993). Expectationally driven market volatility: An experimental study. *Journal of Economic Theory*, *61*(1), 74–103.
Massaro, D. (2013). Heterogeneous expectations in monetary DSGE models. *Journal of Economic Dynamics and Control*, *37*(3), 680–692.
Mauersberger, F. (2021). Monetary policy rules in a non-rational world: A macroeconomic experiment. *Journal of Economic Theory*, *197*(C), 105203.
Mauersberger, F., Nagel, R., & Bühren, C. (2019). *Bounded Rationality in Keynesian Beauty contests: A Lesson for Central bankers?, Economics Discussion Papers 2019-53*. Kiel Institute for the World Economy (IfW Kiel).
Mitra, K. (2003). Desirability of nominal GDP targeting under adaptive learning. *Journal of Money, Credit and Banking*, *35*(2), 197–220.
Mokhtarzadeh, F., & Petersen, L. (2021). Coordinating expectations through central bank projections. *Experimental Economics*, *24*(3), 883–918.
Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, *85*(5), 1313–1326.
Noussair, C. N., Plott, C. R., & Riezman, R. G. (1995). An experimental investigation of the patterns of international trade. *American Economic Review*, *85*(3), 462–491.
Ozden, T. (2025). Heterogeneous expectations and the business cycle. *International Journal of Central Banking*, *21*(1), 273–330.
Pfajfar, D., & Zakelj, B. (2016). Uncertainty in forecasting inflation and monetary policy design: Evidence from the laboratory. *International Journal of Forecasting*, *32*(3), 849–864.
Pfajfar, D., & Zakelj, B. (2018). Inflation expectations and monetary policy design: Evidence from the laboratory. *Macroeconomic Dynamics*, *22*(4), 1035–1075.
Powell, J. H. et al (2020). *New Economic Challenges and the Fed's Monetary Policy Review: A Speech at "Navigating the Decade Ahead: Implications for Monetary Policy*. Jackson Hole, Wyoming, August 27, 2020.

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press

Rholes, R., & Petersen, L. (2021). Should central banks communicate uncertainty in their projections?. *Journal of Economic Behavior & Organization*, *183*(C), 320–341.
Riedl, A., & Van Winden, F. (2012). Input versus output taxation in an experimental international economy. *European Economic Review*, *56*(2), 216–232.
Rychalovska, Y., Slobodyan, S., & Wouters, R. (2024), Survey expectations, adaptive learning and inflation dynamics, Technical Report 781, CERGE-EI.
Salle, I. (2023). What to target? Insights from a lab experiment. *Journal of Economic Behavior & Organization*, *212*(C), 514–533.
Wilkins, C. A. (2018). Choosing the best monetary policy framework for Canada. In Remarks at the Conference "Toward 2021: Reviewing the Monetary Policy Framework", Max Bell School of Public Policy, McGill University.
Woodford, M. (2000). Pitfalls of forward-looking monetary policy. *American Economic Review*, *90*(2), 100–104.

Cite this article: Arifovic, J., Salle, I., & Truong, H. (2026). History-dependent monetary policy: a lab experiment and a HENK model. *Experimental Economics*, 1–24. https://doi.org/10.1017/eec.2026.10058

https://doi.org/10.1017/eec.2026.10058 Published online by Cambridge University Press