UvA-DARE (Digital Academic Repository)

Managing self-organization of expectations through monetary policy

A macro experiment

Assenza, T.; Heemeijer, P.; Hommes, C.H.; Massaro, D.

DOI

[10.1016/j.jmoneco.2019.12.005](https://doi.org/10.1016/j.jmoneco.2019.12.005)

Publication date

2021

Document Version

Final published version

Published in

Journal of Monetary Economics

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

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, 170-186. <https://doi.org/10.1016/j.jmoneco.2019.12.005>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Managing self-organization of expectations through monetary policy: A macro experiment

T. Assenza^a, P. Heemeijer^c, C.H. Hommes^{d,e,f,1,*}, D. Massaro^{g,b}

^aToulouse School of Economics, University of Toulouse Capitole, Toulouse, France

^bComplexity Lab in Economics, Italy

^cSymetrics, Netherlands

^dCeNDEF, University of Amsterdam, Netherlands

^eTinbergen Institute, Amsterdam, Netherlands

^fBank of Canada, Canada

^gDepartment of Economics and Finance, Università Cattolica del Sacro Cuore, Milan, Italy

ARTICLE INFO

Article history:

Received 25 October 2014

Revised 16 December 2019

Accepted 17 December 2019

Available online 19 December 2019

JEL classification:

C91

C92

D84

E52

Keywords:

Laboratory experiments

Monetary policy

Expectations

Taylor principle

ABSTRACT

The New Keynesian theory of inflation determination is tested in this paper by means of laboratory experiments. We find that the Taylor principle is a necessary condition to ensure convergence to the inflation target, but it is not sufficient. Using a behavioral model of expectation formation, we show how heterogeneous expectations tend to self-organize on different forecasting strategies depending on monetary policy. Finally, we link the central bank's ability to control inflation to the impact that monetary policy has on the type of feedback –positive or negative– between expectations and realizations of aggregate variables and in turn on the composition of subjects with respect to the type of forecasting rules they use.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The recent literature on inflation dynamics has questioned the ability of the “Taylor principle” to uniquely pin down the inflation path in the baseline rational expectations (RE) New Keynesian (NK) model. The aim of the present paper is to shed new light on this debate by means of laboratory experiments and to empirically test for the effectiveness of the Taylor principle as a device to pin down inflation. The advantage of an experimental approach is that no a-priori assumption needs to be placed on agents' beliefs. Instead, expectations are directly elicited from incentivized human subjects participating in the experiment.

In NK models under rational expectations, inflation control is obtained through monetary policy satisfying the “Taylor principle” (see e.g. [Woodford, 2003](#)). When the nominal interest rate reacts more than one-for-one to deviations of inflation

* Corresponding author.

E-mail address: c.h.hommes@uva.nl (C.H. Hommes).

¹ Roeterstraat 11, 1018 WB Amsterdam.

from its target, there exists a unique non-explosive equilibrium path, also labeled as “forward-stable” (FS) solution (García-Schmidt and Woodford, 2019). The FS-RE solution is then typically selected as the one determining inflation dynamics in the model.

Cochrane (2011), however, shows that there exist other RE solutions that cannot be ruled out by any transversality condition or economic principle. Although the Taylor principle holds, these “non-fundamental” (NF) solutions (Evans and McGough, 2018) are explosive and satisfy all relevant equilibrium conditions. The existence of NF-RE solutions and the ability of the Taylor principle to pin down uniquely inflation dynamics are at the root of the debate on inflation control, surveyed in Section 2. Given the strong linkage, in the NK framework, between inflation dynamics and inflation expectations, the focus has shifted on the ability of central banks to manage expectations via Taylor rules. The literature has then investigated the role of expectation formation in shaping inflation dynamics by considering mild departures from RE (see e.g. Angeletos and Lian, 2018; Coibion and Gorodnichenko, 2015; Evans and McGough, 2018; Farhi and Werning, 2019; Gabaix, 2019; García-Schmidt and Woodford, 2019; Mankiw and Reis, 2002; McCallum, 2009, among others). In this paper we do not impose a-priori any type of expectations, and let them be directly elicited from participants in the experiment. Therefore, an advantage of our approach is that we can study the Taylor principle without taking a stand on the form of expectations.

In our experiment subjects are asked to forecast inflation and the output gap in an artificial NK economy and their rewards depend solely on the accuracy of these forecasts. Forecasts are then aggregated and used as inputs into a computerized NK model, which describes realizations of inflation and the output gap as functions of such forecasts and exogenous disturbances.² This process then repeats itself for a fixed number of periods. Our experimental economic systems are therefore “self-referential” (Marcet and Sargent, 1989) in the sense that expectations affect the data-generating process, which in turn affects expectations. As noted by Eusepi and Preston (2018), expectation errors in such environments, characterized by a dynamic feedback between expectations and realizations of aggregate variables, may propagate through the system, becoming self-fulfilling and causing instability. We use this setup to investigate whether the FS-RE solution emerges as the equilibrium outcome in the experimental economies under different monetary policy regimes by considering different parameterizations of a Taylor-type interest rate rule.

Our contribution is threefold. First, we establish that multiple equilibria do not only emerge in rational or near-rational expectations settings. We also find them in a setup in which expectations are elicited from human subjects participating in the experiment. In other words, we find that the Taylor principle is a necessary, but not sufficient condition for stability and uniqueness of the equilibrium path of inflation. Second, we reframe these results in terms of positive versus negative expectation feedbacks. A positive (negative) feedback between the expectations of a generic variable x and the realization of a generic variable y means that the average forecast of x has a positive (negative) effect on the realization of y .³ In particular, we show that the conditions for the emergence of a FS-RE solution relate to the existence of a strong enough negative feedback from inflation expectations to the output gap through aggressive enough monetary policy. Third, we show that in a heterogeneous expectations setting the convergence to a stable equilibrium is driven by a composition effect. More precisely, convergence to a stable equilibrium obtains when the share of agents adopting an adaptive expectation rule is large enough compared to competing trend-extrapolating rules. A direct policy implication of this result is that the central bank can actually achieve convergence by managing the share of agents using a specific expectation rule, in particular by managing trend-extrapolating behavior. We show that this can be implemented by manipulating the relative size of the negative feedback by tuning the reaction of the policy rule to deviations of inflation from its target. In other words, the central bank can manage the composition of expectation rules adopted by agents, and achieve convergence to the target, by implementing an aggressive monetary policy that in turn increases the “size” of the negative feedback.

The paper is organized as follows. Section 2 relates our work to the existing literature, presents the theoretical framework and describes different monetary policy regimes. Section 3 describes the design of the experiment and shows the experimental results. Section 4 presents the model used to explain self-organization of individual expectations and the emergence of aggregate behaviors observed in the experiment. This section also discusses how the central bank can influence this process through monetary policy in order to achieve convergence to the target equilibrium. Section 5 concludes.

2. Related literature

The aim of this section is twofold. First, we describe the theoretical framework that we use in the experiment and second, we place it in the debate about inflation control via Taylor rules within the NK model.

In the experiment we use the standard New Keynesian workhorse model described by⁴

$$y_t = \bar{y}_{t+1}^e - \varphi(\dot{t}_t - \bar{\pi}_{t+1}^e - \gamma) + g_t \quad (1)$$

² Aggregate outcomes computed in our laboratory economies are consistent with the notion of “temporary equilibria” in the sense that they result from first-order conditions of (computerized) households and firms given subjects’ forecasts (see e.g. Eusepi and Preston, 2018; Farhi and Werning, 2019; García-Schmidt and Woodford, 2019).

³ See Section 2 for a precise definition of positive and negative feedbacks in the NK framework.

⁴ Micro-founded NK models consistent with heterogeneous expectations have been derived by Branch and McGough (2009), Kurz et al. (2013), Massaro (2013), Woodford (2013) and Hommes and Lustenhouwer (2019). System (1) – (3) corresponds to the model developed by Branch and McGough (2009) augmented with demand and supply shocks, or to the model derived in Kurz et al. (2013) in which deviations of average agents’ forecasts of individual future consumption (prices) from average forecast of aggregate consumption (price) enter the error terms.

$$\pi_t = \lambda y_t + \rho \bar{\pi}_{t+1}^e + u_t \quad (2)$$

$$i_t = \text{Max}\{\bar{\pi} + \gamma + \phi_\pi(\pi_t - \bar{\pi}), 0\}. \quad (3)$$

Eq. (1) is the dynamic IS curve, Eq. (2) is the New Keynesian Phillips curve (NKPC) and Eq. (3) is the monetary policy rule, with a zero lower bound (ZLB), implemented by the monetary authority in order to keep inflation at its target value $\bar{\pi}$. Variables y_t and $\bar{y}_{t+1}^e$ denote respectively the actual and average expected output gap, i_t is the nominal interest rate, π_t and $\bar{\pi}_{t+1}^e$ denote respectively the actual and average expected inflation rates, $\bar{\pi}$ is the inflation target. Parameter φ is the intertemporal elasticity of substitution of consumption, λ denotes the slope of the NKPC, ρ is the discount factor, γ is the natural interest rate. The coefficient ϕ_π measures the response of the nominal interest rate i_t to deviations of the inflation rate π_t from its target $\bar{\pi}$. Finally g_t and u_t are exogenous disturbances, which can be thought of as a demand shock and a cost push shock respectively. When the ZLB is not binding, by substituting for the monetary policy rule in Eq. (3), the model (1) – (3) can be reduced to a two variables system and written in matrix form as:

$$z_t = A + M \bar{z}_{t+1}^e + C \epsilon_t, \quad (4)$$

where $z = (y, \pi)'$ is the vector of endogenous variables, $\bar{z}^e = (\bar{y}^e, \bar{\pi}^e)'$ is the vector of average forecasts and $\epsilon = (g, u)'$ is the vector of exogenous disturbances.⁵ The FS-RE solution takes the form:

$$z_t = \Theta^{FS} + C \epsilon_t, \quad (5)$$

with $\Theta^{FS} = (I - M)^{-1}A$, while the form of matrix C depends on the assumptions placed on the observability of the shocks. The NF-RE solution takes instead the form:

$$z_t = \Theta^{NF} + \Phi^{NF} z_{t-1} + C \epsilon_t, \quad (6)$$

with $\Theta^{NF} = (-M)^{-1}A$, and $\Phi^{NF} = M^{-1}$, while the form of matrix C depends on the assumptions placed on the observability of the shocks.

McCallum (2009) proposes “least-squares learnability” as an equilibrium selection device and shows that, when the Taylor principle is satisfied, the NK model with least-squares learning converges to the FS-RE equilibrium. Cochrane (2009) objects to the results derived in McCallum (2009) on the grounds that they hinge on observability of contemporaneous exogenous shocks. Evans and McGough (2018) extend the results of McCallum (2009) to the case of unobservable shocks. In this case NF-RE solutions are never learnable, while the FS-RE equilibrium is learnable provided that the positive feedback from expectations to realizations of the endogenous variable being forecast is not too large, as in the case of a NK model satisfying the Taylor principle.

Our paper is directly related to this debate. In particular, our evaluation of the effectiveness of the Taylor principle for inflation determinacy is consistent with the principle put forward in McCallum (2009) and Evans and McGough (2018): subjects have imperfect information about the exact functioning of the economy they are participating in, but they can nevertheless learn the RE equilibrium through properly designed monetary policy. There are however some important differences. The first obvious difference with the least-squares learnability approach is that we do not postulate any learning mechanism, having instead real human subjects learning in the experimental economies. The second difference concerns the information set available to learning agents. In fact, contrarily to McCallum (2009) and Evans and McGough (2018), contemporaneous realizations of aggregate variables are not available to subjects when forecasting future inflation and output gap.⁶ Our conclusions regarding the effectiveness of the Taylor principle differ from those obtained under least-squares learning since we find that the Taylor principle is not a sufficient condition to ensure convergence to the FS-RE equilibrium.

Given the strong linkage in the NKPC between expectations and inflation dynamics, the role of beliefs formation has been widely investigated. García-Schmidt and Woodford (2019) have developed the concept of “reflective equilibrium” where, given a conjecture about average forecasts, agents refine expectations using their knowledge of the structural equations governing the economy. In this framework issues of indeterminacy are sidestepped as, for a given level of reflection, the equilibrium outcome is unique. Moreover, when the Taylor principle is satisfied, the dynamics of the NK model under the reflective process converge to the FS-RE solution as the degree of reflection increases. Farhi and Werning (2019) adopt a form of bounded rationality based on a discrete deductive procedure rather than continuous, known as “level- k thinking” (see Nagel, 1995). Within the context of a NK model with incomplete markets, they show that the level- k equilibrium converges

⁵ Coefficient matrices A, M and C are defined as follows:

$$A \equiv \begin{pmatrix} \frac{\varphi \bar{\pi} (\phi_\pi - 1)}{1 + \lambda \varphi \phi_\pi} \\ \frac{\lambda \varphi \bar{\pi} (\phi_\pi - 1)}{1 + \lambda \varphi \phi_\pi} \end{pmatrix}, \quad M \equiv \begin{pmatrix} \frac{1}{1 + \lambda \varphi \phi_\pi} & \frac{\varphi (1 - \phi_\pi \rho)}{1 + \lambda \varphi \phi_\pi} \\ \frac{\lambda}{1 + \lambda \varphi \phi_\pi} & \frac{\lambda \varphi + \rho}{1 + \lambda \varphi \phi_\pi} \end{pmatrix}, \quad C \equiv \begin{pmatrix} \frac{1}{1 + \lambda \varphi \phi_\pi} & \frac{-\varphi \phi_\pi}{1 + \lambda \varphi \phi_\pi} \\ \frac{\lambda}{1 + \lambda \varphi \phi_\pi} & \frac{1}{1 + \lambda \varphi \phi_\pi} \end{pmatrix}.$$

⁶ In our experimental implementation we consider unobservable IID exogenous disturbances with zero mean. Moreover, since realizations of endogenous variables z_t in period t depend on expectations $\bar{z}_{t+1}^e$ formed in period t , subjects in the experiment do not observe contemporaneous variables when making forecasts. This addresses the simultaneity issue raised by Cochrane (2009), i.e. how to interpret an equilibrium in which agents are forecasting based on the same endogenous variables being determined.

to the RE with complete markets as k increases only when the Taylor principle is satisfied. The main difference between our approach and both the “reflective” and “level- k thinking” is that the latter assume an iterative reasoning based on knowledge by agents of the correct quantitative specification of the economic structure, while our subjects have imperfect structural knowledge of the economy. Our experimental results show that, even without full information, monetary policy can ensure coordination on the FS-RE equilibrium.

Gabaix (2019) introduces partially myopic agents and shows that, if bounded rationality is strong enough, the NK model exhibits a unique bounded equilibrium even without the Taylor principle. Angeletos and Lian (2018) study the effect of monetary policy focusing on the forward guidance puzzle in a NK model with full rationality and informational frictions, showing how the absence of common knowledge may rationalize the kind of myopia postulated in Gabaix (2019). Mankiw and Reis (2002) propose a framework in which agents receive perfect information infrequently due to slow diffusion of information. In a framework with imperfectly informed firms, Barrdear (2018) shows that a unique bounded equilibrium emerges in the NK model regardless of whether the Taylor principle is satisfied. Our experimental findings show instead that the Taylor principle is a necessary, though not sufficient, condition to observe convergence to the FS-RE equilibrium. In this paper, contrary to this literature, we do not posit a priori a specific form of expectations, instead we rely on laboratory experiments to elicit them (see Section 3). By doing so we do not restrict ourselves to a particular beliefs theory. In this respect our paper relates to the literature on macro experiments in controlled laboratory environments, (see Duffy, 2016, for a recent overview). Our experiment is a Learning-to-Forecast Experiment (LtFE), a design first proposed by Marimon and Sunder (1993) to study expectations dynamics in the laboratory. In recent years a number of LtFEs have been conducted within the NK framework to investigate inflation persistence (Adam, 2007), disinflationary policies (Cornand and M'baye, 2016), the importance of the expectation channel for macroeconomic stabilization (Kryvtsov and Petersen, 2013), inflation-output volatility tradeoff (Hommes et al., 2019a), and monetary and fiscal policy design at the zero lower bound (ZLB) (Arifovic and Petersen, 2017; Hommes et al., 2019b) among other topics. Most closely related to our paper is Pfajfar and Žakelj (2018), who study the stabilization effects of different monetary policy rules by means of LtFEs. Pfajfar and Žakelj (2018) compare inflation variability under contemporaneous vs. forward-looking interest rate rules all satisfying the Taylor principle, finding that the former produces lower inflation variability. We focus instead on different contemporaneous interest rate rules, assessing the role of the Taylor principle for inflation control. Moreover, differently from Pfajfar and Žakelj (2018), participants to our experiment forecast both inflation and the output gap, in accordance to the theoretical NK model.

Finally, our paper relates to the literature that studied the role of expectation feedbacks in LtFE. In earlier LtFEs, Heemeijer et al. (2009) and Bao et al. (2012) have shown that, in simple *univariate* environments, the type of expectation feedback is crucial for convergence to the RE equilibrium. In particular, negative feedback experimental markets are rather stable and converge quickly to the unique RE steady state. In contrast, positive feedback markets are rather unstable and typically do not converge, but fluctuate persistently around the RE steady state.⁷ Our experimental environment is *multivariate* and thus more complex than simple *univariate* systems, since realizations of both inflation and the output gap depend on both expectations of future inflation and future output gap. In the remainder of the section we describe the type of expectation feedbacks in the NK model and how they depend on monetary policy.

2.1. Expectation feedbacks in the NK model

The nature of the feedbacks between expectations and realizations of aggregate variables in the NK model is defined by the sign of the entries of matrix M in Eq. (4). The IS curve in Eq. (1) implies that higher expected output gap leads to higher realized output gap. Moreover, since current inflation depends positively on current output gap, the NKPC in Eq. (2) implies that both higher expected inflation and higher expected output gap lead to higher realized inflation. These are all positive feedbacks because the signs of $\partial y_t / \partial \bar{y}_{t+1}^e$ (M_{11} entry of M), $\partial \pi_t / \partial \bar{y}_{t+1}^e$ (M_{21} entry of M) and $\partial \pi_t / \partial \bar{\pi}_{t+1}^e$ (M_{22} entry of M) are all positive, independent from monetary policy. On the other hand, the sign of the feedback between expected future inflation and the realized output gap depends on monetary policy. In particular, if $\phi_\pi < 1/\rho$ then $\partial y_t / \partial \bar{\pi}_{t+1}^e$ (M_{12} entry of M) is positive. In this case the system described by Eq. (4) exhibits only positive feedbacks. If instead $\phi_\pi > 1/\rho$ then $\partial y_t / \partial \bar{\pi}_{t+1}^e$ (M_{12} entry of M) is negative. Hence, there is a negative feedback between inflation expectations and the output gap, so that the system described by Eq. (4) exhibits a mix of positive and negative feedbacks. The only source of negative feedback in the NK model is the monetary policy rule: when the nominal interest rate reacts aggressively enough to inflation, i.e. $\phi_\pi > 1/\rho$, then positive (negative) inflation expectations lead to a negative (positive) effect on the output gap through the real interest rate.⁸ We can therefore distinguish three qualitatively different monetary policy regimes according to i) whether the Taylor principle is satisfied and ii) the implied nature of expectations feedbacks in the economy. In the first regime ($\phi_\pi \leq 1$) the Taylor principle is not satisfied and the economy exhibits purely positive feedbacks. In the second regime ($1 < \phi_\pi < 1/\rho$) the monetary policy rule satisfies the Taylor principle but the model is still characterized by purely

⁷ A LtFE characterized by positive (negative) feedbacks corresponds to an environment where subjects' expectations are strategic complements (substitutes). Fehr and Tyran (2008) show that when agents' actions are strategic complements, aggregate behavior deviates from the predictions of RE models. On the other hand, when agents' actions are strategic substitutes, aggregate outcomes are consistent with RE predictions.

⁸ Notice that the threshold value $1/\rho$ is larger than 1 since parameter $0 < \rho < 1$ denotes the time discount factor. We remark that a higher reaction coefficient ϕ_π also weakens the existing positive feedbacks since all positive entries of matrix M are monotonically decreasing, though rather flat, functions of ϕ_π . Given the assumed parameterization, $M_{11} \in [0.77, 0.69]$, $M_{21} \in [0.23, 0.21]$, and $M_{22} \in [0.99, 0.90]$ for $\phi_\pi \in [1, 1.5]$.

Table 1

Summary of policy regimes and characteristics of RE solutions in different treatments.

Treatment	ϕ_π	Taylor principle	Expectations feedbacks	FS	NF
T1	1	No	Purely Positive	Indeterminate	Explosive
T2	1.005	Yes	Purely Positive	Unique	Explosive
T3	1.015	Yes	Mix Positive/Negative	Unique	Explosive
T4	1.5	Yes	Mix Positive/Negative	Unique	Explosive

positive feedbacks. In the third regime ($\phi_\pi > 1/\rho$) the Taylor principle is satisfied and the system presents a mix of positive and negative feedbacks.

As described below, we experiment with different parameterization of the policy rule in Eq. (3) belonging to these different policy regimes and we link the central bank's ability to control inflation to the impact that monetary policy has on expectation feedbacks. In particular, we show that convergence to the FS-RE equilibrium depends on the strength of negative feedbacks introduced in the system by monetary policy via the effect of interest rate on aggregate demand.⁹

3. Experiment

In our Learning-to-Forecast experiment subjects are asked to predict inflation and the output gap. These forecasts are then used to compute subsequent realizations according to the NK model described in Section 2, with structural parameters set as in Clarida et al. (2000), i.e. $\rho = 0.99$, $\varphi = 1$ and $\lambda = 0.3$. The inflation target is set at $\bar{\pi} = 2\%$, while the natural interest rate is set at $\gamma = 4\%$. Shock g_t and u_t are independent and normally distributed, with mean 0 and standard deviation 0.1. Before describing the experiment in more detail, we first discuss the treatments implemented in our LtFE.

3.1. Treatments

The treatments implemented in the experiment are motivated by the theoretical results on qualitatively different policy regimes described in Section 2. There are four treatments, differing only in the reaction coefficient ϕ_π of the interest rate rule describing monetary policy. By analyzing the experimental results in the four treatments we will be able to investigate both the role of the Taylor principle and of the “size” of the negative feedback in stabilizing our economy. Table 1 summarizes the different treatments implemented in the experiment.

In the first treatment T1 the policy rule coefficient is set to $\phi_\pi = 1$. Monetary policy in T1 belongs therefore to the regime in which the Taylor principle is not satisfied and the system exhibits purely positive feedbacks. With $\phi_\pi = 1$ the determinant of matrix $I - M$ is zero, implying a continuum of steady state solutions so that the FS-RE equilibrium is not unique.¹⁰ Moreover, when $\phi_\pi = 1$, the eigenvalues $|\lambda_1| < |\lambda_2|$ of M^{-1} are such that $|\lambda_1| = 1$ and $|\lambda_2| > 1$, so that the NF-RE solution describes unstable equilibrium paths. We then consider small perturbations around the threshold case $\phi_\pi = 1/\rho$.¹¹ In particular, in the second treatment T2 the policy rule coefficient is set to $\phi_\pi = 1.005$, while in the third treatment T3 the reaction coefficient is set to $\phi_\pi = 1.015$. Treatments T2 and T3 implement both a policy regime in which the Taylor principle is satisfied. They, however, differ in terms of the type of feedback. In T2 the economy exhibits positive feedback only, while in T3 it shows a mix of positive and negative feedbacks. Note that by comparing the outcomes of T1 vs. T2, both characterized by purely positive feedback, we can assess whether a monetary policy rule satisfying the Taylor principle is a necessary and sufficient condition to ensure convergence (if any) to the unique FS-RE equilibrium. While, by comparing the outcomes in T2 vs. T3, characterized by purely positive feedback and a mix of positive and negative feedback respectively, we can determine whether the mere presence of negative feedbacks is enough to ensure convergence (if any) to the unique FS-RE equilibrium. Finally, the last treatment T4 considers the policy parameter originally proposed by Taylor, i.e. $\phi_\pi = 1.5$. Treatments T3 and T4 belong to the same policy regime, with the difference between T3 and T4 being the size of the negative feedback. The feedback from expected inflation to realized output $\partial y_t / \partial \pi_{t+1}^e$ is a decreasing function of ϕ_π , so that the higher ϕ_π the more negative $\partial y_t / \partial \pi_{t+1}^e$. By comparing the outcomes in T3 vs. T4 we can determine whether convergence (if any) to the FS-RE depends on the size of the negative feedback.

3.2. Procedures and implementation

The design of the experiment is a between-subjects design with within-session randomization. At the beginning of each session, all participants are randomly divided into groups (experimental economies) of six. Subjects only interact with other

⁹ Interestingly, Cornand and Heinemann (2019) show that, in a NK model with RE, monetary policy affects strategic uncertainty, turning pricing decisions into strategic substitutes when the Taylor principle is satisfied.

¹⁰ Note that $\text{Det}(I - M) = \lambda\varphi(\phi_\pi - 1)/(1 + \lambda\varphi\phi_\pi)$. Given that φ , λ and ϕ_π are positive coefficients, when $\phi_\pi = 1$ then $\text{Det}(I - M) = 0$. On the contrary, whenever $\phi_\pi \neq 1$, matrix $I - M$ is invertible.

¹¹ Note that $1/\rho$ is approximately 1.01 given the calibrated value of the time discount factor $\rho = 0.99$.

subjects in their experimental economy, without knowing who they are. Subjects are assigned the fictitious role of professional forecasters and they are asked to forecast inflation and the output gap. The average forecasts of all subjects in each economy are then used to calculate the realizations of inflation and output gap according to the NK model in Section 2. In each period t subjects make forecasts for period $t + 1$. Their information set (visualized on their screen as numbers and partly also in graphs) is composed of: all realizations of inflation, output gap, and interest rate up to period $t - 1$, their own forecasts of inflation and output gap up to period t and their scores indicating how close their past forecasts were to realized values up to period $t - 1$.¹² Contemporaneous realizations of the small IID shocks are not observable. Moreover, the noise series used in the model equations differed across groups within each treatment, but the sets of noise series used in the four treatments were the same. Fig. B.7 in Online Appendix B displays the computer interface as visualized by the participants in the experiment.

Subjects' rewards depend on their forecasting performance. At the end of the experiment it is randomly determined whether a participant is paid for inflation or output gap forecasting. The final scores for inflation and output gap forecasting are given by the sums of the respective forecasting scores over all periods. The score of subject i for e.g. inflation forecast in period t is computed as $100/(1 + |\pi_{i,t}^e - \pi_t|)$, where $\pi_{i,t}^e$ denotes subject i 's forecast for period t and π_t realized inflation in period t (the score is computed in the same way for the output gap). Therefore rewards decrease with the distance of the realizations from their forecasts. In the instructions, subjects receive a qualitative description of the economy that includes an explanation of the mechanisms that govern the model equations, but they do not know the underlying model equations and have no quantitative information on the exact values of structural parameters, nor on the inflation target $\bar{\pi}$.¹³ The complete instructions can be found in Online Appendix A.

The experiment has been programmed in Java and conducted at the CREED laboratory at the University of Amsterdam. The experiment was conducted with 144 subjects (6 groups of 6 subjects for each of the 4 treatments). After each session, participants filled out a short questionnaire. Participants were primarily undergraduate students and the average age was slightly below 22 years. About half of the participants were female, about 60% were majoring in economics or business, and about 20% were Dutch. During the experiment, participants earned "points" according to the forecasting score mentioned above. Points were then exchanged for euros at the end of each session at an exchange rate of 0.75 euros per 100 points. The experiment lasted around 2 h, and the average earning was about 25 euros.

3.3. Results

Fig. 1 presents an overview of the experimental results. Each line depicts realized inflation (left panels) and output gap (right panels) in a single experimental economy throughout the 50 periods of the experiment. The dashed lines refer to the constant equilibrium level $\bar{\pi}$ and $(1 - \rho)\bar{\pi}/\lambda$ respectively for inflation and the output gap. Before describing the results in more detail, we note that, for practical reasons, we imposed bounds on the forecasts that subjects could input in the computer program. In particular the upper and lower bounds for both inflation and the output gap were respectively +100% and -100%, thus not very restrictive. Subjects were not informed ex-ante about these bounds and a pop-up message would appear on their screens only in case their forecasts were outside the allowed range. We interpret scenarios in which these constraints were binding as laboratory evidence of the possibility of subjects' coordination on explosive paths. The erratic behavior typically observed in experimental economies after subjects reach these bounds is not very meaningful from an economic point of view. Complete data for all groups separately including individual forecasts can be found in Online Appendix C.

The first row of Fig. 1 displays realized inflation and output gap in treatment T1. Inflation and the output gap never converge to the equilibrium defined by the target $\bar{\pi}$. This is not necessarily surprising since the FS-RE is indeterminate in T1. In four out of six economies (groups 2, 4, 5 and 6) we observe explosive dynamics, with inflation forecasts rising to the upper bound on allowed forecasts. Reversal of the trend in these economies typically occurs when participants reach this upper bound.¹⁴ As mentioned before, the ensuing large oscillations do not have a clear economic interpretation. We note that in these economies the output gap does not explode immediately with inflation. In fact, the impact of real interest rate on output is close to zero since $\phi_\pi = 1$. On the other hand, when the upward trend in inflation is reversed and deflationary spirals occur, the nominal interest rate hits the ZLB and the economy enters a severe recession. In one economy (group 3) we observe convergence to a non-fundamental steady state, while in another (group 1) we observe slow oscillations away from the target equilibrium.¹⁵ The second row of Fig. 1 shows the dynamics of inflation and the output gap in treatment T2. Although the Taylor principle is satisfied, we only observe convergence to the unique FS-RE equilibrium in one economy out of six (group 3). All other groups do not converge to the FS-RE equilibrium. One economy (group 2) converges to

¹² Since the information set of subjects in each period t includes realizations up to period $t - 1$, forecasts for period $t + 1$ are actually two-period-ahead forecasts.

¹³ Given that our experiment is a two-period-ahead LtFE, after reading the instructions subjects are asked to enter forecasts for periods 1 and 2 simultaneously. Subjects therefore receive some indication of reasonable values by being told in the instructions that, in economies similar to the one they are participating in, inflation has historically been between -5% and 15% and the output gap between -5% and 5%.

¹⁴ In treatment T1 group 6 the upward trend in inflation is interrupted due to one participant who predicted -100% in the attempt to reverse the trend. Given that inflation rose to about 40% before this event, we interpret it as evidence of explosive behavior.

¹⁵ In treatment T1 group 1, a participant committed a typing error swapping inflation and output gap forecasts. This caused an interruption of the upward trend in inflation. We conjecture that, without the typing error, group 1 would have also experienced explosive dynamics.

Fig. 1. Inflation (left panels) and output gap (right panels) dynamics in different groups and treatments. Each line refers to one experimental economy, numbered from 1 to 6.

an *almost self-fulfilling* stable equilibrium (see Hommes, 2013). The latter is characterized by coordination of expectations around a constant value which, although mathematically not a steady state, is hardly distinguishable from an equilibrium due to an eigenvalue very close to 1 and the presence of exogenous disturbances. Three out of six economies (groups 4, 5 and 6) display the same explosive behavior observed in treatment T1, while one economy (group 1) is characterized by sustained oscillatory behavior away from steady state. The third row of Fig. 1 presents aggregate dynamics in treatment T3. Strikingly, the mere presence of a small negative feedback from expected inflation to realized output gap eliminates coordination of subjects on unstable paths. In fact, we do not observe explosive dynamics in any of the experimental economies. Instead, all economies oscillate much closer to target when compared to treatments T1 and T2, with the exception of one economy (group 6) which stabilizes on an *almost self-fulfilling* equilibrium after about 30 periods of oscillatory behavior. Finally, the last row of Fig. 1 presents the results for treatment T4. The difference with all other treatments is remarkable: all experimental economies converge to the unique FS-RE equilibrium.

In what follows we investigate further differences between treatments. As argued in Section 3.1, by comparing the outcomes of T1 vs. T2 we can test whether the Taylor principle is a necessary and sufficient condition to ensure convergence to the unique FS-RE equilibrium. To this end, we compute the mean squared deviations (MSE) of inflation and the output gap from the target equilibrium in both T1 and T2 and perform a Wilcoxon rank-sum test.¹⁶ According to the standard NK theory on inflation control, one would expect a significant difference between the two treatments, since monetary policy in T2 does satisfy the Taylor principle. The test does not reject the null that MSE in T1 is equal to MSE in T2 for both inflation and the output gap (p -values equal to 0.47 and 0.65 respectively), confirming the graphical evidence presented in Fig. 1 that the Taylor principle is not a sufficient condition for convergence to the FS-RE equilibrium.¹⁷ We then compare experimental outcomes in T2 vs. T3 to assess whether by simply adding small negative feedback in the system, monetary policy can ensure convergence to the target. The Wilcoxon rank-sum test rejects the null of equal MSE for the output gap in T2 and T3 (p -value equal to 0.01), though the result is not as clear-cut for inflation (p -value equal to 0.06). Although aggregate dynamics are much closer to target in T3 when compared to T1 and T2, the presence of negative feedback in the system is not a sufficient condition for convergence to the FS-RE equilibrium. The Wilcoxon signed-rank test rejects the null that average inflation in T3 is equal to target (p -value equal to 0.03), while it does not reject it for the output gap (p -value equal to 0.09).¹⁸ Finally, we compare T3 vs. T4 to verify whether convergence to the FS-RE depends on the strength of negative feedbacks. Realizations of aggregate variables in treatment T4 are clearly centered around the FS-RE equilibrium. This is largely confirmed by a Wilcoxon signed-rank test (p -values equal to 0.44 and 0.69 respectively for inflation and the output gap). We therefore conclude that, for the FS-RE equilibrium to emerge as the unique outcome, not only monetary policy has to satisfy the Taylor principle, but the negative feedback introduced in the system by the interest rate rule has to be strong enough. Moreover, the Wilcoxon rank-sum test rejects the null of equal MSE for inflation in T3 and T4 (p -value 0.001), while it does not reject it for the output gap (p -value 0.15).¹⁹

4. Monetary policy and self-organization of expectations

The experimental economies presented in Section 3.3 show different types of aggregate behavior, namely explosive dynamics, persistent oscillations and convergence to (some) equilibrium. The goal of this section is to characterize individual forecasting behavior using a simple behavioral model of learning and explain how the emergence of different aggregate patterns depends on monetary policy.

4.1. Heuristics switching model of expectation formation

The fact that different types of aggregate behavior arise in our experiments, both within and between treatments, suggests that heterogeneous expectations play an important role in determining aggregate outcomes. A first result emerging from the analysis of individual forecasts is that subjects tend to coordinate on a common prediction strategy, although participants in different groups may coordinate on different strategies. Coordination is, however, not perfect and heterogeneity in individual forecasts within groups persists throughout the experiment (see Online Appendix C). Another interesting result that emerges from experimental data is that individual forecasting behavior entails a learning process taking the form of switching from one prediction strategy to another (see Online Appendix D).

In light of this empirical evidence we use a heuristics switching model (HSM), which features evolutionary selection among different forecasting strategies, to characterize expectations dynamics and explain emergent aggregate behavior. Denoting by $\mathcal{H}$ a set of H forecasting heuristics for variable x , aggregate expectations in each period t are given by a weighted

¹⁶ In all treatments' comparisons we allow for an initial learning phase and consider data starting from period 15.

¹⁷ Strictly speaking, the Wilcoxon rank-sum test tests the null-hypothesis that the distribution does not change against the alternative that it shifts between treatments.

¹⁸ Technically, the Wilcoxon signed-rank test tests the null hypothesis that the distribution of average inflation or output gap is centered around the target.

¹⁹ One may wonder whether differences across treatments can be explained by differences in subjects' prior beliefs. This is unlikely since subjects are randomly assigned to different treatments. As a further check, we compare the distributions of subjects' initial forecasts and we find that in general there are no significant differences across treatments.

Table 2
Set of heuristics.

ADA	adaptive rule	$x_{1,t+1}^e = 0.65x_{t-1} + 0.35x_{1,t}^e$
WTR	weak trend-extrapolating rule	$x_{2,t+1}^e = x_{t-1} + 0.4(x_{t-1} - x_{t-2})$
STR	strong trend-extrapolating rule	$x_{3,t+1}^e = x_{t-1} + 1.3(x_{t-1} - x_{t-2})$
LAA	anchoring and adjustment rule	$x_{4,t+1}^e = 0.5(\bar{x}_{t-1} + x_{t-1}) + (x_{t-1} - x_{t-2})$

Note: The term $\bar{x}_{t-1}$ denotes the average of all observations up to time $t - 1$.

average of the forecasts resulting from these heuristics. In the context of the NK model x denotes either inflation or the output gap. The key ingredient of the model is that the weight of each heuristic $h \in \mathcal{H}$ evolves over time as a function of past forecasting performance. In particular the measure of past performance of heuristic h denoted as U_h is defined as

$$U_{h,t-1} = F(x_{t-1} - x_{h,t-1}^e) + \eta U_{h,t-2}, \quad (7)$$

where F is a generic function of the forecast error of heuristic h , and $0 \leq \eta \leq 1$ is a memory parameter measuring the relative weight attached to past errors of heuristic h . Performance uniquely depends on the most recent forecasting error when $\eta = 0$, while it is determined by all past prediction errors with exponentially declining weights when $0 < \eta < 1$, or equal weights when $\eta = 1$. Given the performance measure in Eq. (7), the weight attached to each heuristic h at time t is defined as

$$n_{h,t} = \delta n_{h,t-1} + (1 - \delta) \frac{\exp(\beta U_{h,t-1})}{Z_{t-1}}, \quad (8)$$

where $Z_{t-1} = \sum_{h=1}^H \exp(\beta U_{h,t-1})$ is a normalization factor. Parameter $0 \leq \delta \leq 1$ describes inertia in the evolution of weights, while parameter $\beta \geq 0$ represents the intensity of choice, measuring the sensitivity to differences in heuristics performances. The model described by Eqs. (7)–(8) has been developed by Anufriev and Hommes (2012), along the lines of Brock and Hommes (1997), to explain different types of aggregate behavior as well as individual expectations in the asset pricing LtFE of Hommes et al. (2005).²⁰ The model is also related to reinforcement learning models developed in game-theoretical frameworks, (see e.g. Camerer and Ho, 1999), and to rational inattention models, (see e.g. Matějka and McKay, 2015).

In order to use the HSM for policy analysis, specific assumptions have to be made about the forecast error function F and the types of forecasting heuristics to include in set $\mathcal{H}$. In our implementation of the model we use the same forecast error function used to incentivize subjects in the experiment, i.e. $F(x - x^e) = 100/(1 + |x^e - x|)$. Moreover, we discipline the choice of the set of heuristics $\mathcal{H}$ using experimental data. In particular, we consider heuristics describing qualitatively different types of forecasting behavior emerging from data on individual predictions. To keep the model simple, we restrict our attention to a set of four heuristics described in Table 2. Details on the analysis of individual forecasts time series are given in Online Appendix E.

These four heuristics are quite common in the literature. Adaptive expectations may be viewed as a simple form of adaptive learning of a steady state with constant gain parameter (Evans et al., 2008). Trend-extrapolating rules have been found e.g. in survey data of macroeconomic forecasting (Bordalo et al., 2018). Finally, the anchor and adjustment heuristic plays a prominent role in psychology (Tversky and Kahneman, 1974). The parameterization of the heuristics in Table 2 follows Anufriev and Hommes (2012) and it is consistent with estimated values using our experimental data (see Online Appendix E). Based upon the calibration in their paper, we set the model parameters $\beta = 0.4$, $\eta = 0.7$, $\delta = 0.9$.²¹ We adopt therefore the same 4-type HSM that has successfully been used by Anufriev and Hommes (2012) to explain different price patterns emerged in the asset pricing experiment of Hommes et al. (2005). This illustrates the robustness of the HSM across different experimental settings.

As shown in Online Appendix F, different homogeneous expectations models, i.e. economies where all subjects use one of the forecasting heuristics in Table 2 to predict inflation and the output gap, can explain different observed patterns in aggregate variables. For example, coordination on forecasting rules strongly extrapolating past trends (STR) leads to explosive dynamics under all considered policy regimes, while coordination on adaptive rules (ADA) has a stabilizing effect under all considered policy regimes. However, homogeneous expectations models do not answer the question *why* coordination on certain prediction strategies emerge under different policy regimes. Our goal is to explain why subjects coordinate on a certain forecasting rule depending on monetary policy and how this leads to the emergence of different aggregate behavior.

²⁰ In the original approach of Brock and Hommes (1997) the individual heuristics' choice in each period is random, with probability of selecting predictor h given by Eq. (8) with $\delta = 0$. With a continuum of agents and independent decisions Eq. (8) gives the proportion of agents using heuristics h . Given that each experimental economy consists of a small number of subjects, we interpret the weights in Eq. (8) as the weights attributed by subjects to different forecasting rules.

²¹ We remark that the model is not very sensitive to these parameter values (see also Anufriev and Hommes, 2012), and for different choices of the coefficients of the four heuristics in Table 2 we obtain similar results to those presented in Section 4.2.

Table 3
MSE of one-step-ahead simulations for different models of expectation formation.

	Treatment T1		Treatment T2		Treatment T3		Treatment T4	
	π	y	π	y	π	y	π	y
HSM	3.410	0.098	5.851	8.886	0.714	0.425	0.070	0.083
ADA	61.700	0.323	47.989	18.917	4.350	1.009	0.371	0.482
WTR	19.168	0.152	12.149	10.608	1.586	0.524	0.091	0.149
STR	1.161	0.133	3.599	4.579	2.049	0.690	0.212	0.349
LAA	58.794	0.271	30.221	12.992	2.355	0.559	0.195	0.110

Note: The MSE is computed over periods 5 to 49 in order to minimize the impacts of initial conditions on heuristics' weights and of "ending effects" in individual forecasts observed in several groups.

4.2. Self-organization of heterogeneous expectations

In this section we discuss the performance of the HSM in describing experimental results and illustrate how the model explains the emergence of different aggregate behaviors. For each group, we compute one-step-ahead predictions of the HSM described in Section 4.1, and then compare them with experimental outcomes. Simulations are initialized using the first two realizations for inflation and the output gap, i.e. $\{\pi_1, y_1\}$ and $\{\pi_2, y_2\}$, with equal initial weights $n_h = 1/4$ for all heuristics. Using equal weights for periods 3 and 4 and the heuristics forecasts, we compute $\{\pi_3, y_3\}$ and $\{\pi_4, y_4\}$. Starting from period 5 dynamics are well defined and HSM forecasts are obtained using the same information available to subjects in the experiment. Table 3 reports the mean squared prediction errors averaged across groups in each treatment. We remark that simulations were truncated whenever bounds on individual predictions were reached or subjects tried to strategically reverse explosive trends.²²

The results show that the HSM is a better predictor than any of the four heuristics alone in almost all cases. The only exceptions are the unstable economies in T1 and T2 in which the strong trend-following rule performs better in predicting the explosive behavior of aggregate variables. In fact, although the HSM encompasses the STR prediction strategy, the weights of the four rules are updated with some inertia due to a positive δ and a finite intensity of choice β . This result suggests that in situation of high instability, subjects coordinate faster, i.e. $\delta \rightarrow 0$ and $\beta \rightarrow \infty$, on forecasting rules that strongly extrapolate observed trends. The relatively high MSE registered for all models regarding output gap expectations in T2 is due to predictions of one participants in group 5 which, before hitting the upper bound in period 9, were consistently above the average of all other predictions (almost four times higher on average). Removing this one subject from the sample yields much lower MSE values but it does not change the models' ranking in terms of predicting power.

Figs. 2–4 illustrate how the HSM explains the emergence of different aggregate behaviors observed in the experiment, namely explosive dynamics, persistent oscillations and convergence to (some) equilibrium. Fig. 2 refers to group 4 in T1 as an example of explosive dynamics, Fig. 3 refers to group 5 in T3 as an example of persistent oscillations, while Fig. 4 refers to group 2 in T4 as an example of convergence to equilibrium. Results for other economies displaying the same type of aggregate behavior are qualitatively similar (see Figs. G.23 and G.24 in Online Appendix G) reporting results for all experimental economies). Left panels in Figs. 2–4 display experimental data together with the one-step-ahead predictions under the HSM. Overall, the one-step-ahead forecasts closely track experimental data and the model is able to reproduce qualitatively all different types of aggregate behavior.²³ Right panels in Figs. 2–4 depict the evolution over time of the weights of the four considered heuristics. In different groups different heuristics gain more weight after starting from a uniform distribution. In fact, the evolutionary learning process described by the HSM self-organizes into coordination on one of the four rules, which then determine (long-run) aggregate behavior.

In treatment T1 group 4 (Fig. 2) inflation follows an upward trend in the early stage of the experiment, triggering increasing coordination on trend-following behavior. The increasing trend in inflation is amplified by coordination on the STR forecasting rule, whose weight reaches about 90% by the end of the simulation. As noted in Section 3.3, the output gap does not explode immediately with inflation because the impact of real interest rate on output is close to zero when $\phi_\pi = 1$. Therefore, as long as the output gap remains stable, the weights of the four heuristics are similar. However, the sharp increase of the output gap towards the end of the considered time period, caused by rising inflation expectations, leads to increasing coordination on the STR rule. The emergence of explosive dynamics is thus explained by coordination of individual expectations on forecasting strategies that strongly extrapolate trends observed in the data. This behavior is consistent with the theoretical benchmark derived under homogeneous STR expectations in T1, i.e. explosive dynamics due to real eigenvalues outside the unit circle (see Online Appendix F for details).

²² In particular, groups 2, 4, 5, and 6 in T1 were simulated respectively until periods 19, 25, 22, and 18, while groups 1, 4, 5, and 6 in T2 were simulated respectively until periods 11, 11, 9, and 21. Removing these groups from the sample does not change our qualitative results, though the level of MSE in T1 and T2 is obviously much lower when unstable economies are not considered in the analysis.

²³ We also test for the null hypothesis of equality between observed and simulated mean and standard deviation of inflation and output gap using a Wilcoxon rank-sum test. In all cases we never reject the null using a 5% significance level.

Fig. 2. Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for T1 group 4. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for T3 group 5. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Realized and simulated inflation and output gap (left panels) with corresponding weights of 4 heuristics for T4 group 2. In the left panels, blue circles refer to experimental data while red squares refer to simulated data. In the right panels, ADA, WTR, STR and LAA refer respectively to the adaptive rule, the weak trend-following rule, the strong trend-following rule and the anchoring and adjustment rule. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In treatment T3 group 5 (Fig. 3) aggregate dynamics are characterized by persistent oscillations. The HSM explains sustained oscillatory behavior by coordination of most agents on a learning-anchor and adjustment (LAA) rule. The observed trends in inflation and the output gap in the beginning of the experiment cause an initial coordination on trend-following behavior. However, reversal of the trend favors the LAA rule in the evolutionary competition among heuristics. In fact, in the presence of cyclical oscillations, the purely extrapolative rules WTR and STR tend to overshoot the trend reversal. On the other hand, the LAA rule uses an anchor which is given by a weighted average of the sample mean and the last observation, and therefore it performs better at turning points of the trend. For both inflation and the output gap, the LAA rule dominates reaching a peak weight of about 90% towards the end of the experiment, which slowly decreases afterwards as the amplitude of oscillations decreases in the last few periods. Oscillatory non-explosive behavior is consistent with the theoretical benchmark derived under homogeneous LAA expectations in T3, i.e. sustained non-explosive oscillations due to stable complex eigenvalues close to the unit circle (see Online Appendix F for details).

In treatment T4 group 2 (Fig. 4) dynamics converge to the FS-RE equilibrium. The initial part of the experiment is characterized by coordination on the LAA forecasting rule due to the continuous reversal of trends in aggregate variables. However, as oscillations gradually dampen, the weight of the ADA rule gradually increases. In fact, adaptive rules perform better in converging paths as they do not extrapolate past trends in observed variables. Convergence with progressively dampened oscillations is consistent with the theoretical benchmark derived under homogeneous ADA expectations in T4, i.e. oscillatory convergence due to complex eigenvalues within the unit circle (see Online Appendix F for details).

The one-step-ahead simulations show that initially heterogeneous expectations tend to self-organize on common predictions strategies. A salient result is that the proportion of agents using (strong) trend extrapolation rules plays an important role for the stability of aggregate variables. Groups in which the weight of STR rules is lower are more stable than groups with a higher impact of trend extrapolating behavior. Instead, having more agents that follow adaptive expectations has a stabilizing effect on aggregate dynamics, while oscillatory behavior is associated with anchoring and adjustment heuristics.²⁴ In the following section we discuss how monetary policy can influence the process of self-organization of expectations, preventing coordination on destabilizing trend-extrapolating behavior and ensuring convergence to the FS-RE equilibrium.

²⁴ Interestingly, Pfajfar and Zäkelj (2018) reach a similar conclusion and note that a higher proportion of trend extrapolation increases the standard deviation of inflation while having more agents behaving according to adaptive expectations decreases the standard deviation of inflation.

4.3. Managing coordination on trend-extrapolating behavior through monetary policy

All experimental economies start away from, typically above, the target equilibrium.²⁵ By its impact on the feedback between expectations and realizations of aggregate variables, monetary policy can influence the adjustment process towards the target. When the NK model exhibits purely positive feedbacks (treatments *T1* and *T2*), indeterminacy arises because monetary policy is not able to correct drifts in expectations which then may become self-fulfilling. When the policy rule reacts aggressively enough to inflation, it introduces a negative feedback in the system (treatments *T3* and *T4*), which has a stabilizing effect through the impact of real interest rate on the output gap. In order to appreciate the stabilizing effect of this negative feedback, it is instructive to look at cross-correlations among realized and expected aggregate variables in the experiment, as reported in Figs. 5–6. Note that in Figs. 5–6, the notation $\bar{\pi}^e$ and $\bar{y}^e$ refers to expectations formed in period t about inflation and the output gap in $t + 1$, so that e.g. $\text{corr}(y, \bar{\pi}^e)$ refers to correlation between y_t and $\bar{\pi}_{t+1}^e$.

Fig. 5 displays cross-correlations at different leads and lags, averaged across groups, for treatments *T1* and *T2* characterized by purely positive feedbacks.

From Fig. 5(a), treatment *T1*, the first thing that one notices is that correlations are positive across the board. For example, correlation between realized inflation (output gap) and expected future inflation (output gap) is positive not only contemporaneously, but also at several leads/lags. Autocorrelations of expected inflation (output gap) are also positive for several lags. In fact, initial trends in aggregate variables are never reversed because monetary policy responds too weakly. In particular, the positive correlation between the output gap and expected inflation ($\text{corr}(y, \bar{\pi}^e) > 0$) implies that there is no reduction in the output gap, via real interest rate, when inflation expectations are above target because the nominal interest rate does not react strong enough to inflation. Absent the correction mechanism of expectations via monetary policy, deviations from the target are either reinforced by coordination on forecasting rules that extrapolate the direction of change, hence resulting in explosive paths (see Fig. 2), or they stabilize around one of the multiple steady states. Results are very similar in treatment *T2* as correlations in Fig. 5(b) are generally positive across variables. In fact, even if the Taylor principle is satisfied, the system exhibits purely positive feedbacks. Drifts in expectations away from the target are, in general, not corrected by the interest rate rule towards the FS-RE equilibrium and dynamics may either explode or converge to an almost self-fulfilling equilibrium.²⁶

Fig. 6 shows cross-correlations for treatments *T3* and *T4*, characterized instead by a mix of positive and negative feedback. We first discuss results for treatment *T4* and then examine treatment *T3*.

From Fig. 6(b), it is clear that the presence of negative feedback in the system significantly changes the correlation structure among variables when compared to treatments *T1* and *T2*. As in other treatments, inflation expectations above target cause realized inflation to be above target as well. In this case, however, the strong reaction of the nominal interest rate causes an increase in the real interest rate so that the output gap decreases ($\text{corr}(y, \bar{\pi}^e) < 0$), curbing therefore the inflationary pressure. Output gap expectations follow the decreasing trend in the output gap ($\text{corr}(y, \bar{y}_{+1}^e) > 0$), further reducing inflation and subsequently inflation expectations ($\text{corr}(\bar{y}^e, \bar{\pi}_{+1}^e) > 0$). Decreasing inflation and output gap expectations cause inflation to fall and eventually undershoot the target. This leads to lower real interest rate which in turn stimulates aggregate demand. This continuous trend reversal, driven by the strong effect of monetary policy on aggregate demand, is reflected e.g. in the observed negative autocorrelation of output gap expectations after the first lag ($\text{corr}(\bar{y}^e, \bar{y}_i^e) < 0$ for $i < -1$). In this environment destabilizing trend-extrapolating strategies perform poorly, and they are driven out by stabilizing adaptive expectations in the evolutionary competition among predictors (see Fig. 4). As the weight of trend-extrapolating strategies decreases, oscillations in aggregate variables progressively dampen and the system eventually converges to the FS-RE equilibrium.

In treatment *T3* the policy reaction also introduces (weak) negative feedback in the system, which is reflected in the negative correlation between expected inflation and current output ($\text{corr}(y, \bar{\pi}^e) < 0$) in Fig. 6(a). In fact, as in treatment *T4*, we observe reversal of initial trends in inflation via the impact of real interest rate on aggregate demand, so that coordination on forecasting strategies that strongly extrapolate past trends is prevented (see Fig. 3). However, the impact on aggregate demand is not strong enough to quickly revert drifts in inflation expectations. In fact, although output gap expectations follow the decreasing trend in the output gap due to inflation expectations above target ($\text{corr}(y, \bar{y}_{+1}^e) > 0$), their impact on realized inflation is mild, so that inflation expectations may still increase despite a negative trend in output gap expectations ($\text{corr}(\bar{y}^e, \bar{\pi}_{+1}^e) < 0$). In other words, the signals that subjects receive are not strong enough to promptly correct their expectations. The sluggish dynamics observed in treatment *T3* are reflected in the observed positive autocorrelation of e.g. output gap expectations until the third lag ($\text{corr}(\bar{y}^e, \bar{y}_i^e) > 0$ for $-4 < i < 0$).

How can monetary policy manage the self-organization process of expectations and ensure determinacy of the FS-RE equilibrium? Our results show that, in the presence of imperfect information, obeying the Taylor principle does not necessarily lead to convergence to the target. In fact, even if monetary policy reacts more than point-to-point to inflation, the NK model may still exhibit purely positive feedbacks. Results from treatment *T2* show that in such an environment, when a majority of individuals use a trend-extrapolating strategy, other individuals have an incentive to use such strategy too,

²⁵ This is due to the fact that at the beginning of the experiment, when no realizations of aggregate variables are observed yet, forecasts tend to cluster around the midpoint of the interval of historical values given to subjects in the instructions. In the experiment the midpoints of these intervals are 5% and 0% respectively for inflation and the output gap.

²⁶ There is only one experimental economy that oscillates around the target equilibrium in *T2*.

(a) Treatment $T1$ (b) Treatment $T2$ **Fig. 5.** Correlations in experimental data – Purely positive feedbacks.

(a) Treatment $T3$ (b) Treatment $T4$ **Fig. 6.** Correlations in experimental data – Mix positive/negative feedbacks.

thus reinforcing trends in aggregate variables. An insight emerging from our analysis is that the introduction of negative feedback via monetary policy is a necessary condition to prevent coordination on trend-extrapolating behavior. However, the mere presence of negative feedback is not sufficient for the FS-RE to emerge as the unique outcome in the experimental economies, as shown in treatment T3. To ensure convergence to the desired equilibrium, monetary policy has to be aggressive enough to quickly correct drifts in expectations towards the target. How aggressive then should monetary policy be to control inflation? It is important to note that, as long as matrix M in (4), mapping expectations into realizations of aggregate variables, is close to having an eigenvalue equal to 1, the system exhibits sluggish adjustment dynamics and it may converge to almost self-fulfilling equilibria. This is in fact the case for treatment T3, in which the absolute value of largest eigenvalue is about 0.98. For subjects to learn the FS-RE equilibrium from data generated by the economic system, the eigenvalues of matrix M have to be well within the unit circle. Results from treatment T4 suggest that a reaction coefficient $\phi_\pi = 1.5$, leading to a largest eigenvalue of about 0.83, is sufficient to ensure convergence to the target.

5. Conclusions

Laboratory experiments have been used in this paper to test the New Keynesian theory of inflation determination. Our results suggest that the Taylor principle does not ensure convergence to the inflation target. Using a behavioral model of expectation formation, we explain how different aggregate outcomes emerge out of a self-organization process of heterogeneous expectations driven by their relative forecasting performance. We illustrate how monetary policy can prevent coordination on explosive non-fundamental equilibria and steer expectations towards the target. In particular, by introducing a strong enough negative feedback between expected inflation and aggregate demand, the central bank can avoid coordination on trend-following behavior and prevent expectation errors from becoming (almost) self-fulfilling. Our experiment focuses on short-run forecasts. However, recent literature on forward guidance about future central bank actions has highlighted the importance of expectations at far horizons for inflation control. Future experiments within NK economies should also incorporate elicitation of long-run forecasts. Moreover, our study focuses on an heuristic switching model of expectation formation. Recent works have proposed several alternative models of expectations within the NK framework, see e.g. least-squares learning (Evans and McGough, 2018), sticky information (Mankiw and Reis, 2002), sparsity-based bounded rationality (Gabaix, 2019), rational inattention (Maćkowiak and Wiederholt, 2015), reflective equilibrium (García-Schmidt and Woodford, 2019), level- k thinking (Farhi and Werning, 2019), and imperfect information (Angeletos and Lian, 2018) among others.

CRediT authorship contribution statement

T. Assenza: Methodology, Validation, Investigation, Formal analysis, Writing - original draft. **P. Heemeijer:** Methodology, Software. **C.H. Hommes:** Supervision, Conceptualization, Writing - review & editing. **D. Massaro:** Methodology, Software, Validation, Investigation, Formal analysis, Writing - original draft.

Acknowledgments

We thank the editor Ricardo Reis and six anonymous reviewers for detailed comments and suggestions. We also thank Klaus Adam, Jess Benhabib, Jim Bullard, Fabrice Collard, Domenico Delli Gatti, John Duffy, George Evans, Patrick Fève, Seppo Honkapohja, Albert Marcet, Argia Sbordone, Shyam Sunder, Matthias Weber, John Williams and Mike Woodford for stimulating discussions and helpful comments. Tiziana Assenza acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). Domenico Massaro acknowledges financial support from the Ministry of Education, Universities and Research of Italy (MIUR), program SIR (grant n. RBSI144KWH). None of the above are responsible for errors in this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmoneco.2019.12.005.

References

- Adam, K., 2007. Experimental evidence on the persistence of output and inflation. *Econ. J.* 117 (520), 603–636.
- Angeletos, G.-M., Lian, C., 2018. Forward guidance without common knowledge. *Am. Econ. Rev.* 108 (9), 2477–2512.
- Anufriev, M., Hommes, C.H., 2012. Evolutionary selection of individual expectations and aggregate outcomes. *American Economic Journal: Microeconomics* 4, 35–64.
- Arifovic, J., Petersen, L., 2017. Stabilizing expectations at the zero lower bound: experimental evidence. *J. Econ. Dyn. Control* 82, 21–43.
- Bao, M., Hommes, C.H., Sonnemans, J., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. *J. Econ. Dyn. Control* 36, 1101–1120.
- Barrdear, J., 2018. The Calm Policymaker: Imperfect Common Knowledge in New Keynesian Models. Working Paper. Bank of England.
- Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2018. Over-reaction in Macroeconomic Expectations. Working Paper. National Bureau of Economic Research.
- Branch, W., McGough, B., 2009. A new keynesian model with heterogeneous expectations. *J. Econ. Dyn. Control* 33, 1036–1051.
- Brock, W.A., Hommes, C.H., 1997. A rational route to randomness. *Econometrica* 65 (5), 1059–1095.

- Camerer, C.F., Ho, T.-H., 1999. Experienced-weighted attraction learning in normal form games. *Econometrica* 67 (4), 827–874.
- Clarida, R., Gali, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. *Q. J. Econ.* 115 (1), 147–180.
- Cochrane, J.H., 2009. Can learnability save new-Keynesian models? *J. Monet. Econ.* 56 (8), 1109–1113.
- Cochrane, J.H., 2011. Determinacy and identification with taylor rules. *J. Political Econ.* 119 (3), 565–615.
- Coibion, O., Gorodnichenko, Y., 2015. Information rigidity and the expectations formation process: a simple framework and new facts. *Am. Econ. Rev.* 105 (8), 2644–2678.
- Cornand, C., Heinemann, F., 2019. Monetary Policy obeying the Taylor Principle Turns Prices into Strategic Substitutes. *J. Econ. Behav. Organ.* in press.
- Cornand, C., M'baye, C.K., 2016. Does inflation targeting matter? An experimental investigation. *Macroecon. Dyn.* 22 (2), 362–401.
- Duffy, J., 2016. Macroeconomics: A Survey of Laboratory Research. In: Kagel, J.H., Roth, A.E. (Eds.), *The Handbook of Experimental Economics*, Volume 2. Princeton University Press, pp. 1–90.
- Eusepi, S., Preston, B., 2018. The science of monetary policy: an imperfect knowledge perspective. *J. Econ. Lit.* 56 (1), 3–59.
- Evans, G.W., Guse, E., Honkapohja, S., 2008. Liquidity traps, learning and stagnation. *Eur. Econ. Rev.* 52 (8), 1438–1463.
- Evans, G.W., McGough, B., 2018. Equilibrium selection, observability and backward-stable solutions. *J. Monet. Econ.* 98, 1–10.
- Farhi, E., Werning, I., 2019. Monetary Policy, Bounded Rationality, and Incomplete Markets. *Am. Econ. Rev.* 109 (11), 3887–3928.
- Fehr, E., Tyran, J.-R., 2008. Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. *Econometrica* 76 (2), 353–394.
- Gabaix, X., 2019. A behavioral new keynesian model. Working Paper. National Bureau of Economic Research.
- García-Schmidt, M., Woodford, M., 2019. Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis. *Am. Econ. Rev.* 109 (1), 86–120.
- Heemeijer, P., Hommes, C.H., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedbacks: an experimental investigation. *J. Econ. Dyn. Control* 33, 1052–1072.
- Hommes, C., Massaro, D., Weber, M., 2019. Monetary policy under behavioral expectations: theory and experiment. *Eur. Econ. Rev.* 118, 193–212.
- Hommes, C.H., 2013. Reflexivity, expectations feedback and almost self-fulfilling equilibria: economic theory, empirical evidence and laboratory experiments. *J. Econ. Methodol.* 20 (4), 406–419.
- Hommes, C.H., Lustenhouwer, J., 2019. Inflation targeting and liquidity traps under endogenous credibility. *J. Monet. Econ.* 107, 48–62.
- Hommes, C.H., Massaro, D., Salle, I.L., 2019. Monetary and fiscal policy design at the zero-lower-bound: evidence from the lab. *Econ. Inquiry* 57, 1120–1140.
- Hommes, C.H., Sonnemans, J., Tuinstra, J., van de Velden, H., 2005. Coordination of expectations in asset pricing experiments. *Rev. Financ. Stud.* 18 (3), 955–980.
- Kryvtsov, O., Petersen, L., 2013. Expectations and Monetary Policy: Experimental Evidence. Working Papers. Bank of Canada.
- Kurz, M., Piccillo, G., Wu, H., 2013. Modeling diverse expectations in an aggregated new keynesian model. *J. Econ. Dyn. Control* 37, 1403–1433.
- Maćkowiak, B., Wiederholt, M., 2015. Business cycle dynamics under rational inattention. *Rev. Econ. Stud.* 82 (4), 1502–1532.
- Mankiw, N.G., Reis, R., 2002. Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. *Q. J. Econ.* 117 (4), 1295–1328.
- Marcet, A., Sargent, T.J., 1989. Convergence of least squares learning in environments with hidden state variables and private information. *J. Political Econ.* 97, 1306–1322.
- Marimon, R., Sunder, S., 1993. Indeterminacy of equilibria in a hyperinflationary world: experimental evidence. *Econometrica* 61 (5), 1073–1107.
- Massaro, D., 2013. Heterogeneous expectations in monetary DSGE models. *J. Econ. Dyn. Control* 37 (3), 680–692.
- Matějka, F., McKay, A., 2015. Rational inattention to discrete choices: a new foundation for the multinomial logit model. *Am. Econ. Rev.* 105 (1), 272–298.
- McCallum, B.T., 2009. Inflation determination with taylor rules: is new-Keynesian analysis critically flawed? *J. Monet. Econ.* 56 (8), 1101–1108.
- Nagel, R., 1995. Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85, 1313–1326.
- Pfajfar, D., Zäkelj, B., 2018. Inflation expectations and monetary policy design: evidence from the laboratory. *Macroecon. Dyn.* 22 (4), 1035–1075.
- Tversky, A., Kahneman, D., 1974. Judgement under uncertainty: heuristics and biases. *Science* 185, 1124–1130.
- Woodford, M., 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.
- Woodford, M., 2013. Macroeconomic analysis without the rational expectations hypothesis. *Annu. Rev. Econ.* 5 (1), 303–346.