

Contents lists available at [ScienceDirect](#)

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jmoneco

Central bank reputation, cheap talk and transparency as substitutes for commitment: Experimental evidence[☆]

John Duffy^a, Frank Heinemann^{b,*}^a Department of Economics, University of California, Irvine, California^b Department of Economics, Technische Universität Berlin, DE-10623 Berlin, Germany

ARTICLE INFO

Article history:

Received 28 November 2018

Revised 22 June 2020

Accepted 29 June 2020

Available online 30 June 2020

JEL Codes:

C73

C92

D83

E52

E58

Keywords:

Monetary policy

Time inconsistency

Repeated games

Central banking

Commitment

Discretion

Cheap talk

Transparency

Experimental Economics

ABSTRACT

We implement a repeated version of the Barro-Gordon monetary policy game in the laboratory and ask whether reputation serves as a substitute for commitment, enabling the central bank to achieve the efficient Ramsey equilibrium and avoid the inefficient, time-inconsistent one-shot Nash equilibrium. We find that reputation is a poor substitute for commitment. We then explore whether central bank cheap talk, policy transparency, both cheap talk and policy transparency, economic transparency or committees of central bankers yield improvements in the direction of the Ramsey equilibrium under the discretionary policy regime. Our findings suggest that these mechanisms have only small or transitory effects on welfare. Surprisingly, the real effects of supply shocks are better mitigated by a commitment regime than by any discretionary policy. Thus, we find that there is no trade-off between flexibility and credibility.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Should central bankers commit to a consistent monetary policy or should they be afforded discretion to alter monetary policy depending on current circumstances? This question, as first posed by [Kydland and Prescott \(1977\)](#) and elaborated upon by [Barro and Gordon \(1983a,b\)](#) continues to be debated. In current terms, the time inconsistency problem arises between commitment to an Odyssean forward guidance policy regarding future interest rates and a discretionary response of setting interest rates in response to changes in inflation or Delphic forward guidance that announces a future policy, but

[☆] We acknowledge financial support for this project from the Dietrich School of Arts and Sciences, the UC Irvine School of Social Sciences and from the Deutsche Forschungsgemeinschaft (DFG) through CRC 649 "Economic Risk" and Priority Program 1578 "Financial Market Imperfections and Macroeconomic Performance"

* Corresponding author.

E-mail addresses: duffy@uci.edu (J. Duffy), frank.heinemann@tu-berlin.de (F. Heinemann).

allows for discretionary deviations from the original announcements.¹ On the one hand, the ability to use monetary policy to flexibly respond to various economic shocks as they arise is the main argument in favor of discretionary policy (e.g., [Blinder \(1998\)](#)). On the other hand, it is well known that the ability to commit to a consistent policy course of action or rule can yield welfare improvements over a pure discretionary policy regime through the effect that the commitment policy has on private sector expectations (e.g., [Taylor \(1999\)](#)). The latter argument hinges on the reputation that central bankers can achieve from consistently applying a low inflation monetary policy and thus envisions a repeated game setting between the policy maker and the private sector.

In this paper we implement a version of the repeated policy game of [Barro and Gordon \(1983a\)](#) in the laboratory with paid human subjects playing the role of the central banker and the private sector. First, we ask whether reputational considerations can serve as a substitute for commitment in a repeated game setting where central bankers lack a commitment device and are free to alter monetary policy each period conditioning on realizations of economic shocks. We compare this flexible, discretionary regime with a commitment regime where central bankers can commit to a course for monetary policy in advance of the formation of private sector expectations, but have no flexibility for stabilizing the economy. Our experimental findings reveal that reputation is indeed a poor substitute for commitment in that inflation is higher and welfare is lower in the discretionary environment as compared with the commitment regime.

However, our baseline model lacks a number of mechanisms (besides reputational considerations) that central bankers, primarily in OECD countries, have come to employ in the discretionary regimes in which they operate. For instance, since the introduction of inflation targeting in the 1990s, central bankers have become considerably less secretive about policy measures and goals. They have taken to communicating more frequently with the public, being transparent about their monetary policy frameworks, targets, and instruments and even sharing information about economic fundamentals (e.g., [Vayid \(2013\)](#)). This trend accelerated following the 2007–08 financial crises; among the unconventional monetary policies resulting from the crisis was *forward guidance*, wherein the central bank (CB) began providing announcements of date-based or outcome-based (or both) criteria for future changes in the policy path. Regarding these announcements, the literature² distinguishes between *Odyssean forward guidance*, by which a CB commits to a future policy rate, from *Delphic forward guidance*, which is essentially cheap talk in a discretionary setting, eventually combined with higher levels of transparency. Together with communication, transparency may have helped central bankers establish or preserve some credibility, while continuing to operate in a discretionary policy, dynamic repeated-game environment. Thus, a second goal of this paper is to evaluate the efficacy of transparency and communication in our discretionary policy regime for enhancing welfare. In particular, we explore the role of non-binding CB communication or “cheap talk” about policy targets as well as transparency about policy actions, both cheap talk and policy transparency, and finally economic transparency as mechanisms for overcoming the inflationary bias under discretionary policy.³ Since many of these mechanisms have also been studied theoretically in the context of the Barro–Gordon model (see, [Geraats \(2002, 2014\)](#) for surveys), we can directly implement these mechanisms in our experimental design. Finally, as a robustness check and a further nod to reality, we also consider whether *committees* of central bankers outperform solo central bankers in the discretionary regime.

Of all the mechanisms we study, we find that cheap talk alone results in some welfare gains relative to the baseline discretionary environment, but that these gains decline with experience. In the end, we conclude that none of the mechanisms we study achieve the welfare levels of the commitment regime, suggesting that there may be *real* welfare-reducing consequences to discretionary monetary policy.

We adopt an experimental approach as it provides us with the control necessary to properly identify how different policy regimes affect policy choices, private sector expectations, and welfare. This is not possible using non-experimental field data, because in the field, changes in policy regimes are often a consequence of insufficient management of expectations and, thus, are endogenous.⁴ Additionally, the indefinitely repeated environments we study admit multiple equilibria. As [Lucas \(1986, p. S421\)](#) has argued, in such settings “it is hard to see what can advance the discussion short of assembling a collection of people, putting them in the situation of interest, and observing what they do.” We further emphasize that the aim of our experiment is not to test comparative static predictions, but to inform theorists and practitioners on likely outcomes under conditions of multiple equilibria.

Our experiment makes use of student subjects to play the role of both central bankers and private sector agents. While ideally, we would have real central bankers make monetary policy choices in our experiment, there are good reasons to think that our experimental findings nevertheless remain externally valid and relevant to the discussion of actual CB practice. As [Cornand and Heinemann \(2014\)](#) emphasize, while quantitative experimental findings might be specific to the laboratory environment and to the chosen parameter values (e.g., the discount rate), the *qualitative* results from comparing treatments with each other and with equilibria are likely to be robust and externally valid.

Two prior experimental studies analyze the relation between commitment and discretion in repeated games. [Van Huyck et al. \(2001\)](#) study a two-player, indefinitely repeated “peasant-dictator” game. Peasants decide how much of their bean

¹ See [Filardo and Hofmann \(2014\)](#) for a discussion of time inconsistency with respect to forward guidance.

² The terms *Odyssean* and *Delphic* forward guidance were coined by [Campbell et al. \(2012\)](#).

³ These changes to the baseline discretionary treatment may be interpreted as different kinds of forward guidance: cheap talk can be viewed at as a form of *Delphic* forward guidance, transparency as forward guidance about data. The commitment regime may be interpreted as *Odyssean* forward guidance.

⁴ Another advantage of conducting a laboratory experiment is that we are able to implement a commitment regime as a theoretical benchmark, while in non-laboratory settings such commitment devices may have credibility problems.

endowment to eat or to plant (invest) yielding new beans. Dictators levy a tax on the post investment bean harvest and can either commit to a tax rate in advance of the peasant's investment (commitment regime) or decide on the rate after investments have been made (discretionary regime). Van Huyck et al. vary the peasants' endowments and the interest rate earned on investments. They report that reputation in the repeated game is an imperfect substitute for commitment and that efficiency under discretion is positively associated with the interest rate earned on investments. Unlike the Barro–Gordon game, in the peasant-dictator game, expectations do not directly affect any outcome variables and the dictator is only concerned with maximization of his own tax revenue. In the Barro–Gordon game, the CB has *two* objectives, stabilization of inflation and unemployment. Further, there is an expectations-based feedback loop: private sector inflation expectations directly affect the CB's trade-off between inflation and output/unemployment, while the CB's policy choice affects actual inflation which affects the accuracy of the private sector's expectations and payoffs. While Van Huyck et al. study commitment and discretion, we go beyond these two regimes and examine the discretionary regime where the CB has several different mechanisms for enhancing welfare including cheap talk, policy transparency and economic transparency.⁵

Arifovic and Sargent (2003) study a version of the Kydland–Prescott (1977) model using a design similar to our own. In their study, one subject in the CB role is matched with 3–5 subjects in the role of private sector forecasters. Private sector forecasters move first seeking to correctly forecast next period's inflation. These expectations enter into a Phillips curve relation that determines the extent to which unemployment departs from its natural rate. The CB moves second. It has noisy control over the actual inflation rate and seeks to minimize its expected loss from the equal weighted sum of the square of the unemployment and inflation rates. Arifovic and Sargent study only a discretionary regime and their treatment variables consist of changes to the variance of shocks to the Phillips curve and the inflation setting policy rule. By contrast with Van Huyck et al., they report that in three fourths of their sessions, subjects learn to coordinate on the first best Ramsey equilibrium, though they also report some “backsliding” toward the less efficient Nash equilibrium associated with the one-shot pure discretionary regime after the Ramsey equilibrium had been achieved for some time.⁶ The Barro–Gordon model that we study in the laboratory differs in certain timing aspects from the model studied by Arifovic and Sargent (2003). In our setting, the CB learns inflation forecasts *in advance* of setting monetary policy, whereas in the Arifovic–Sargent setup, private sector forecasts and monetary policy decisions are made separately and simultaneously. While private sector inflation forecasts affect real outcomes, in the Arifovic–Sargent setup these expectations are only known to policymakers ex-post. Under this timing, the CB cannot purposefully create surprise inflation and this design difference may well explain why Arifovic and Sargent find sustained periods of coordination around the Ramsey solution. Further, it is well established that CBs consult the inflation forecasts of professionals, households as well as market-based measures prior to policymaking, e.g., Bullard (2016). Also differently from Arifovic and Sargent, in our study, both the CB and forecasters have complete information about the model economy and subjects can play both roles within an experimental session.⁷

Finally, our work is related to research on decision-making by committees versus individuals as we consider in one treatment how CB committees make policy decisions in our baseline discretionary environment. Prior experimental studies by Blinder and Morgan (2005) and Cooper and Kagel (2005) have found that groups of subjects may make better decisions than single subjects. Similarly, we find that committees of central bankers reduce the inflation bias compared to the one-period Nash equilibrium, though not enough to achieve higher welfare than in the commitment case.

The rest of the paper is organized as follows. Section 2 outlines the one-shot and repeated Barro–Gordon model that we implement in the laboratory. Section 3 describes our experimental design and hypotheses, Section 4 reports on the main findings from our experiment. Section 5 analyzes interaction and learning and Section 6 concludes.

2. The model

The model economy we implement in the laboratory is a version of that used by Barro and Gordon (1983a,b). We begin by explaining the static version before moving to the repeated (dynamic) version. Within the static environment, we consider first the case of discretion and then the case of commitment.

⁵ While we study cheap talk and policy transparency in the context of the repeated Barro–Gordon model, other experimental studies have explored the role of such policies in New Keynesian models. For instance, Kryvstov and Petersen (2013) find that CB announcements (cheap talk) about future interest rates are more destabilizing relative to a regime without such announcements. Cornand and M'baye (2018) find that announcing an inflation target (policy transparency) does not lead to any welfare improvement relative to standard discretionary policy when the CB only cares about inflation stabilization, though there are modest improvements if the CB cares about both inflation and output stabilization.

⁶ Arifovic (2014) studies a version of the Kydland–Prescott model found in Arifovic et al. (2010) where the CB makes cheap talk announcements about inflation in advance of private sector expectation formation and the private sector consists either of non-believers or in a second treatment, non-believers and believers. The only experimental (human) subjects in her study are the non-believers. Both the CB and the believers are automated robot players who either learn over time in an evolutionary manner (the CB) or blindly follow the CB's pronouncements (the believers). Arifovic finds that with non-believers only or non-believers and believers, actual inflation lies below the one-shot Nash equilibrium prediction, though inflation is more volatile in the non-believer, humans-only, treatment.

⁷ An exception is our treatment with CB committees, where subjects stay in the same role for the entire experiment.

2.1. Static model

The unemployment rate, u , is determined according to a Lucas-style aggregate supply function

$$u = u_n - c(\pi - \pi^e) + w,$$

where u_n denotes the non-accelerating inflation rate of unemployment (NAIRU), π denotes the time t inflation rate, π^e denotes the private sector's mean expectation of the inflation rate at time t , c is a constant, and w is a mean zero random supply shock. In the experiment, we integrate u_n and w into one random variable $W = u_n + w$ with a positive mean u_n that resembles the NAIRU. The CB chooses the rate of growth of the money supply, denoted by m , which determines the actual inflation rate according to:

$$\pi = m + v,$$

where v is a policy disturbance term with mean \bar{v} (e.g., due to changes in the velocity of money or an unanticipated demand shock).⁸ The model is closed under the assumption that the private sector has rational expectations, so that $\pi^e = E(\pi)$, and that the CB seeks to minimize the time t loss function

$$EL = E(b(\pi - \pi^*)^2 + (u - u^*)^2),$$

where EL denotes the expected loss, π^* is the CB's desired inflation rate and $u^* < u_n$ denotes the CB's desired unemployment rate which is assumed to be smaller than the NAIRU.

In the discretionary regime, the private sector moves first forming their expectations for inflation, π^e . The CB is informed of these expectations and takes them as given when minimizing EL subject to the expressions for u and π . The CB's reaction function, for given π^e , is:

$$m = \frac{b\pi^* + c(u_n - u^* + w) + c^2\pi^e}{b + c^2} - \bar{v}. \quad (1)$$

The private sector is assumed to have rational expectations about inflation and we distinguish whether or not the private sector is informed about supply shocks when forming those expectations. If the private sector cannot observe supply shocks, we have that $\pi^e = E(\pi) = E(m) + \bar{v} = \frac{b\pi^* + c(u_n - u^*) + c^2\pi^e}{b + c^2}$, or that $\pi^e = \pi^{NE} = \pi^* + \frac{c}{b}(u_n - u^*)$, which implies that the money supply in the Nash equilibrium is:

$$m^{NE} = \pi^* - \bar{v} + \frac{c}{b}(u_n - u^*) + \frac{c}{b + c^2}w.$$

Thus, the policy choice in the Nash equilibrium (NE) of the one-shot discretionary environment involves an average inflation rate, π^{NE} , that is greater than the desired level, π^* , by the amount $\frac{c}{b}(u_n - u^*)$. This difference, $\pi^{NE} - \pi^*$, is known as the *inflation bias* of discretionary policy. The term $\frac{c}{b + c^2}w$ reflects the CB's incentive for stabilizing employment by adjusting the money supply to supply shocks, w .

If the CB releases its information about supply shocks to the private sector before the formation of expectations, then $\pi^e = E(\pi|w) = E(m|w) + \bar{v}$, which implies $\pi^e = \pi^* + \frac{c}{b}(u_n - u^* + w)$ and

$$m^{ET} = \pi^* - \bar{v} + \frac{c}{b}(u_n - u^* + w).$$

We refer to this regime as one of “economic transparency” (ET) following [Geraats \(2002, p. F540\)](#), who writes that “economic transparency focuses on the [openness to the private sector about the] economic information that is used for monetary policy.” In our experiment we will compare welfare under economic transparency with welfare in the discretionary and commitment regimes without economic transparency. As is well known from [Geraats \(2002, 2014\)](#), a discretionary regime with economic transparency combines a lack of credibility with a lack of flexibility since monetary policy responses to supply shocks are perfectly foreseen and therefore cannot affect employment. The inflation bias in the discretionary regime is the same with and without economic transparency, but the monetary policy response to supply shocks is larger when there is economic transparency, causing a higher variation of inflation rates without stabilizing employment. This regime combines time-inconsistent levels of inflation with time-inconsistent responses to shocks.

We next consider the commitment regime. In this environment, the CB moves first and commits to set m in advance of the private sector's formation of inflation expectations, but the CB may be able to condition this decision on realizations of the shock, w . The CB assumes that the private sector forms rational expectations, $\pi^e = E(\pi|m) = m + \bar{v}$. Thus in this setting, the CB's optimization problem is:

$$\min_m E[b(m + v - \pi^*)^2 + (u_n - c(m + v - \pi^e) + w - u^*)^2], \text{ s.t. } \pi^e = m + \bar{v}.$$

The solution, which we refer to as the one-shot commitment (C) equilibrium, is given by $m^C = \pi^* - \bar{v}$.

⁸ [Barro and Gordon \(1983a\)](#) assume that the CB can directly control inflation, while most papers building up on their model allow for demand or transition shocks that affect the relation between the policy instrument and inflation. In the experiment, we assume that $\bar{v} > 0$, because we want to avoid negative random numbers.

Under commitment, the CB cannot create surprise inflation. Hence, it cannot stabilize the real economy and so its best policy is to target the desired inflation rate irrespective of the supply shock. Comparing this case to the discretionary equilibrium, inflation is both stable and at the desired level, avoiding the inflation bias. However, the inability to stabilize the real sector under commitment causes welfare losses that can be avoided under discretion.⁹ This reflects the well-known trade-off between credibility and flexibility of monetary policy. This trade-off can be mitigated in the repeated (dynamic game) model where the private sector can condition its behavior on the CB's past responses to supply shocks, so that reputational considerations come into play.

2.2. Dynamic model

Barro and Gordon (1983b) argue that time inconsistency can be overcome by reputation, if the CB is sufficiently patient. In a repeated-game version of the model described in the previous section, there are multiple equilibria. If the CB's long-run advantages from a reputation for low inflation are higher than the immediate rewards from surprise inflation, it can overcome the time inconsistency problem and achieve an equilibrium that is more efficient than the repeated one-period Nash equilibrium. Our experiment implements such a dynamic version of the game.

In each period of this dynamic game, the CB has the option to exploit low inflationary expectations by surprising the private sector with an unexpected high inflation which reduces unemployment and raises welfare in that period. However, unexpected high inflation can trigger a rise in future expectations about inflation which is to the CB's disadvantage. Thus, the CB has an incentive to keep inflation low, in order to maintain low inflationary expectations. Whether or not long-term reputational considerations for low inflation dominate short-term welfare gains from surprise inflation depends on the CB's discount factor and on the effect of current inflation on future expectations. Here, we focus on conditions under which the efficient linear Ramsey rule can be sustained as an equilibrium of the infinitely repeated game. In the efficient linear "Ramsey" solution, the CB sets

$$m_t = \pi^* + \frac{c}{b + c^2} w_t,$$

avoiding the inflation bias, because $\pi_t^e = \pi^*$, but at the same time having the flexibility to stabilize employment. A formal derivation of the Ramsey solution is provided in Online Appendix A. As this environment involves multiple equilibria with no clear equilibrium selection mechanism, a laboratory experiment can be informative as to which equilibria agents are likely to coordinate on and under what conditions.

3. Experimental design and hypotheses

Our experimental design consists of seven different treatments that vary in the timing of moves and in the information available to participants. However, across all treatments, a number of factors were held constant and we begin with this basic structure.

3.1. Baseline design

Each session of a given treatment involved 20 inexperienced subjects. The experiment was conducted over networked computers and was programmed using the z-Tree software (Fischbacher, 2007). At the start of each session, subjects were randomly divided up into two matching groups of size 10. Subjects in different matching groups never interacted with each other and thus each matching group constitutes an independent observation. A session for each matching group consists of a number of repeated games (supergames) known as "sequences" with each sequence consisting of an unknown number of rounds.

At the start of each new sequence, subjects in each matching group were randomly divided up into two groups of size 5 and group membership remained constant for all rounds of that sequence. Prior to the first round of the sequence, one member of each group was randomly selected to play the role of the central banker, known as the "type A" player, while the other four members were assigned the role of private sector, "type B" players. Subjects remained in the same role in all rounds of a given sequence. At the start of each new sequence, two groups were randomly formed anew and the type A player was again randomly chosen from the membership of each new group, so there is turnover of central bankers in our environment.

To avoid triggering any pre-conceived notions of the proper role or choices to be made by each player type, we used neutral language and a neutral framing of the model as detailed below. Specifically, we avoided any references to CBs, inflation, unemployment etc. as such contextualization may lead to a loss of control over the incentives of the experiment.¹⁰ We wanted the incentive structure of the model to be the main determinant of subjects' decisions as it is in the theory.

⁹ Whether the benefit from a flexible policy response to such shocks outweighs the costs arising from the inflationary bias of discretionary policy depends, of course, on the parameterization of the model.

¹⁰ For instance, subjects might be averse to raising "unemployment" even though according to the incentive structure of the game, it may be payoff maximizing to do so.

We told subjects to imagine that the two variables, u_t and π_t , stand for two “containers” holding varying amounts of water.¹¹ Subjects were instructed that at the start of each round $t = 1, 2, \dots$, Container 1 (unemployment) held W_t “gallons” (“liters”) of water where W_t was publicly known to be an i.i.d. random draw each period from a uniform distribution over the interval $[120, 160]$. The expected value, $E[W_t] = 140$, corresponds to the NAIRU, u_n , in the model, while the mean-zero supply shock $w_t = W_t - E[W_t]$. Thus, in our parameterization one can think of the adverse supply shock, w_t , as an i.i.d. random draw from a uniform distribution with support $[-20, 20]$. The initial amount of water in Container 1 (unemployment) thus consists of both the NAIRU and the adverse supply shock, i.e., $W_t = u_n + w_t$. The timing of when or whether players learned the value of W_t is an important element of our treatments. Subjects were further instructed that Container 2 (inflation) was initially empty.

In our baseline, discretionary policy treatment, the timing of moves was as follows. The four type B players in each economy moved first each submitting a forecast, $\pi_{i,t}^e$, as to how many gallons (liters) of water would be in Container 2 at the end of round t . They did so without knowing the realized value of W_t , though they *did* know that W_t was an i.i.d. random draw from a uniform distribution on the interval $[120, 160]$ and they were told that $E[W_t] = 140$. They were also informed about the player A's objective function (as described below), so they knew the Player A's (CB's) target values for inflation, π^* , and unemployment, u^* . After all four Player Bs had made their forecasts, the computer program calculated the mean forecast $\pi_t^e = \frac{1}{4} \sum_{i=1}^4 \pi_{i,t}^e$ for the economy/group and revealed this value to the group's Player A – this forecast corresponds to π_t^e in the model. Subjects were instructed that this average forecast value would be added to the amount of water that was already in Container 1, so that the amount of water in Container 1 now increased to $W_t + \pi_t^e$. Then, the CB player A alone in each group learned the value of both W_t and π_t^e and the sum $W_t + \pi_t^e$, representing the new total amount of water in Container 1. Player A was then instructed to “move” some amount $m \in [0, 80]$ of water from Container 1 to Container 2. This choice corresponds to the policy choice of m_t for period t . In the baseline discretionary treatment, Player Bs do not observe Player A's choice for m_t but it is public knowledge that $m \in [0, 80]$. In addition, it was public knowledge to both player types that there was a random, uncontrolled flow of water, v_t , from Container 1 to Container 2, corresponding to the policy transmission shock. The value of v_t was publicly known to be an independent random draw each period from a uniform distribution having support $[0, 40]$, and all subjects were instructed that $E[v_t] = 20$. Note that the transmission shock has a positive mean; this choice was made because the policy action space was $m \in [0, 80]$, and we did not want to have inflation $\pi_t = m + v_t$ be negative. Player As do not observe the value of v_t until after they have chosen m_t .

At the end of each period t , the final amount of water in Container 1 is thus given by $W_t + \pi_t^e - \pi_t$, which correlates with the Phillips curve relationship, in which surprise inflation reduces unemployment with the parameter $c = 1$. The final amount of water in Container 2 corresponds to $\pi_t = m_t + v_t$. This simple additive structure made it easy for subjects to understand the rules of the game.

The final amounts of water in Containers 1 and 2 were revealed to all subjects in each economy of size 5 at the end of each period as these amounts determined subjects' payoffs for the round. Specifically, each player type was incentivized to make choices consistent with the objective functions posited by the theory. Type A players' point earnings were given by the formula:

$$\text{Player A Points} = 6000 - 2(\text{Final Container 1 amount} - 120)^2 - (\text{Final Container 2 amount} - 40)^2.$$

Thus, Player As (CBs) had as their policy objectives: $u^* = 120$ and $\pi^* = 40$ and the parameter b was set equal to $\frac{1}{2}$. Type B players' earnings were given by the formula:

$$\text{Player B Points} = 4000 - (\pi_{i,t}^e - \text{Final Container 2 amount})^2.$$

Thus, Player B's had the simpler task of just forecasting the value for π_t , the amount of water in Container 2, correctly. These equations were presented to both player types and for ease of understanding, subjects were given payoff tables showing how their choices would convert into points. The experimental instructions given to subjects including these payoff tables are provided in Online Appendix B.

At the end of each round, all subjects were informed of the final amounts of water in the two containers corresponding to the realizations of u_t and π_t and their point earnings for the round as determined by the expressions given above. In addition, type B players learned the realization of W_t and thereby the realization of the supply shock $w_t = W_t - E(W)$ at the end of the period. Whether type B players learned the central banker's choice for m_t depends on the treatment as detailed below.

Subjects were instructed that at the end of each round the computer would draw a number randomly from the set $\{1, 2, 3, 4, 5, 6\}$ simulating a die roll. If a 6 was drawn, the sequence ended but otherwise the sequence would continue with another round. This constant random continuation probability implements both discounting with factor $\delta = \frac{5}{6}$ and the stationarity associated with an infinite horizon. As shown in Online Appendix A, it also fulfills the conditions for existence of the Ramsey equilibrium. Subjects were instructed that, depending on the time remaining in the session, a new sequence of indefinite length might begin.

¹¹ The idea of framing a monetary policy game in terms of targeting amounts of water or chips in a container has been used first by Engle-Warnick and Turdalić (2010). Phillips (1950) actually describes a hydraulic machine that was built to demonstrate the effects of fiscal and monetary policy in an IS-LM-framework.

Subjects were informed that at the end of the session (which could last up to three hours), two sequences would be chosen from all sequences played and their point earnings from the chosen sequences would be converted into cash at certain rate. Further, at the start of *each* sequence, subjects were endowed with 5000 points to avoid negative payoffs; since two sequences were chosen for payment at the end, the endowment of 2×5000 points served as subjects' show-up payment (equivalent to €5 in Germany, \$5 in U.S.).

3.2. Treatments

Our experiment consists of seven treatments exploring the role of reputation, cheap talk, policy and economic transparency, and committee decision making on welfare in a repeated discretionary environment relative to the case where the CB can pre-commit to a monetary policy. The seven treatments are:

1. **Discretionary policy:** The timing for this baseline treatment is as described in Section 3.1. The private sector, type Bs, move first forming inflationary expectations, $\pi_{i,t}^e$, without knowledge of the supply shock, w_t . Next, the CB (type A) player observes w_t and π_t^e and then chooses m_t . By deciding *after* type B players, type A players can surprise type B players with unexpected inflation. This flexibility allows the CB to accommodate supply shocks. Type Bs never learn the value of m_t or v_t , but do learn the final amounts of water in each container, u_t and π_t , and the values for w_t and π_t^e at the *end* of each period.
2. **Commitment:** In this treatment, the CB type A player moves first, observing the realization of the supply shock, w_t and then choosing m_t prior to the formation of inflationary expectations by the private sector. Thus in this environment, the CB has the ability to respond to supply shocks but also to commit to monetary policy for the period. While theory assumes that a commitment is credible by definition, in the field the credibility of a commitment may be limited, because CBs may renege on their promises. In the experiment, we can easily establish credibility by letting the CB decide before expectations are formed.
3. **Cheap talk (CT):** In this treatment, type A players again move first, observing the value of the supply shock w_t . They send a message to the 4 type B players in their group of the form: "The amount of water I intend to move from Container 1 to Container 2 is ___." In the blank space they entered a value of $m \in [0, 80]$. After receiving this message, the type B players form their forecasts of inflation for the period, $\pi_{i,t}^e$. Since type Bs know that the final amount of water in Container 1, $\pi_t = m_t + v_t$, the CB's announcement for m_t can play a role in coordinating private sector inflationary expectations. To ensure that the message is understood to be cheap talk, subjects are further instructed that "it is up to Player A whether he or she moves as much water as previously announced. Player A can move the announced amount or more or less water." Finally, the Player A sees the average inflation forecast for the period and chooses m_t .
4. **Policy transparency (PT):** This treatment has the same timing as the discretionary treatment. The only difference is that private sector type B players learn the realizations of *both* m_t and v_t at the end of each round, immediately after the CB has chosen m_t thus making it transparent to type Bs whether inflation was high (low) due to the transmission shock or to the CB's choice. This timing is consistent with Geraats's (2002, p. F540) definition of policy transparency as the "prompt announcement and explanation of policy decisions."¹² In theory, a transparent policy such as in this and the next treatment (cheap talk plus policy transparency) makes it easier to sustain the Ramsey solution as an equilibrium of the repeated game.
5. **Cheap talk and policy transparency (CT+PT):** This treatment combines the cheap talk phase of the CT treatment with the information revealed about monetary policy (m_t and v_t) at the end of each round as in the PT treatment. This treatment thus allows the private sector to evaluate the truthfulness of the CB's cheap talk announcements providing a potentially more credible means by which the CB can attempt to manage private sector expectations as compared with the CT or PT treatments by themselves.
6. **Economic transparency (ET):** In this treatment, the private sector type B players learn the value of the supply shock w_t at the same time the CB player A learns it and prior to forming expectations of inflation for the period. The timing and information is otherwise identical to that of the discretionary treatment.
7. **Central bank committees (CB Group):** This treatment is the same as the discretionary treatment except that instead of there being just one type A player (CB), there is a "committee" of 3 CB type A players who confer with one another and jointly decide on m_t . Each of the three player As get the *same* payoff from their joint decision. A further difference is that each 3-player CB committee is kept the *same* across sequences, and is matched with a randomly formed group of 4 player Bs at the start of each new sequence.

¹² Geraats also refers to political transparency as "openness about policy objectives and institutional arrangements" – the rules of the game and payoff function of the CB. Since we always provide such information to our subjects, political transparency is present in all of our treatments. Similarly, we also have "operational transparency" in that we reveal the distribution of transmission shocks. We do not consider "procedural transparency" which includes revelation of the CB's strategy, as this would require that subjects submit strategies for policy decisions, and the strategy space is too large to elicit such strategies.

3.3. Experimental hypotheses

As we implement a game theoretic model in the laboratory, we propose to test a number of model-based predictions. Since our model admits multiple equilibria, we cannot test precise comparative statics predictions; rather our main aim is to provide theorists and policymakers with some evidence on the behavior of subjects in the various environments we study and the equilibria they appear to coordinate upon.

Given our parameterization of the model, the commitment solution is for the CB to set $m_t = m = 20$ with forecasts $E[\pi_t] = m + E[v_t] = 40 = \pi^*$. As the CB should not respond to supply shocks, inflation varies only with the transmission shock, $\text{Std}[\pi] = \text{Std}[v] = 11.55$, while the standard deviation of unemployment $\text{Std}[u] = \sqrt{\text{Var}[v] + \text{Var}[w]} = 16.33$. Expected CB welfare is given by

$$6000 - 2E((140 - 120 - w - v)^2) - E(v^2) = 5200 - 2\text{Var}(w) - 3\text{Var}(v) = 4533.3.$$

In the case of discretion, we have multiple equilibria ranging from the Ramsey solution, in which agents expect the optimal inflation rate of 40, to the one-period Nash with an inflation bias of 40. Thus, $E(\pi_t) \in [40, 80]$. While there is a wide range of possible equilibrium responses to supply shocks, both Ramsey and Nash predict the CB to stabilize employment and respond to supply shocks with a coefficient of $\partial m / \partial w = 2/3$. The same coefficient applies to the CB's response to inflation expectations. The optimal response to supply shocks reduces the standard deviation of unemployment to $\text{Std}(u) = \sqrt{\text{Var}(v) + \frac{1}{9}\text{Var}(w)} = 12.17$. It raises the standard deviation of inflation to $\text{Std}(\pi) = \sqrt{\text{Var}(v) + \frac{4}{9}\text{Var}(w)} = 13.88$.

In the Ramsey equilibrium, expected CB welfare is:

$$6000 - 2E\left(\left(140 - 120 - \frac{1}{3}w - v\right)^2\right) - E\left(\left(v + \frac{2}{3}w\right)^2\right) = 5200 - \frac{6}{9}\text{Var}(w) - 3\text{Var}(v) = 4711.1.$$

In the Nash equilibrium, expected welfare is lower, because of the inflation bias, and is given by

$$6000 - 2E\left(\left(140 - 120 - \frac{1}{3}w - v\right)^2\right) - E\left(\left(40 + v + \frac{2}{3}w\right)^2\right) = 3600 - \frac{6}{9}\text{Var}(w) - 3\text{Var}(v) = 3111.1.$$

The discretionary regime admits a range of possible welfare values for the CB, [3111, 4711] that includes the commitment welfare value, 4533, in its interior. The CT, PT and CT+PT treatments have no impact on the range of welfare levels that can be achieved in equilibrium relative to the discretionary regime.

Under economic transparency, on the other hand, the private sector learns the supply shock prior to forming expectations. As laid out in Section 3.1, this raises the equilibrium coefficient by which money supply, inflation, and inflation expectations respond to supply shocks to $c/b = 2$ without stabilizing employment, which results in $\text{Std}(\pi) = \sqrt{\text{Var}(v) + 4\text{Var}(w)} = 25.82$, while $\text{Std}(u) = 16.33$ as in the commitment regime. As in the baseline discretionary treatment, there is a set of equilibria, ranging from the one period Nash to a constrained efficient solution, in which there is no inflation bias, but no stabilization of employment either. Welfare in the one-period Nash is given by

$$6000 - 2E((140 - 120 - w - v)^2) - E((40 + v + 2w)^2) = 3600 - 6\text{Var}(w) - 3\text{Var}(v) = 2400.$$

In the constrained efficient equilibrium, welfare is

$$6000 - 2E((140 - 120 - w - v)^2) - E((v + 2w)^2) = 5200 - 6\text{Var}(w) - 3\text{Var}(v) = 4000.$$

Beside the point predictions from equilibria, we have some more fundamental hypotheses that are related to the comparison of different treatments. Qualitative treatment effects are more robust against changes in the design and, thereby, have a higher external validity than numerical results.

Hypothesis 1. *Repetition in the discretionary regimes serves as a substitute for commitment in containing the inflation bias of monetary policy.*

By this we simply mean that in the repeated discretionary games, average money supply and average inflation are the same as under a commitment regime.

Hypothesis 2. *Private sector forecasts of inflation are on average correct and the distribution of forecast errors is the same across treatments.*

This hypothesis follows from the assumption of rational expectations. We will test whether average forecast errors are null and, if not, whether they differ across treatments. If inflation forecasts systematically deviate from realized inflation, unemployment will deviate from the NAIRU, which affects welfare.

Hypothesis 3. *Repetition in the discretionary regime serves as a substitute for commitment regarding (CB) welfare.*

While the inflation bias in discretionary regimes may be higher than under commitment, the flexibility with which the CB can respond to supply shocks under discretion may reduce employment fluctuations. Thus, there are two opposing effects for the final level of CB welfare. Eventual systematic deviations of unemployment from the NAIRU may go either way. Hence, we will test whether the average payoffs of CB players are the same across treatments.

In the equilibria under commitment and ET, the CB cannot stabilize employment; under commitment the CB should not respond *at all* to supply shocks, while in the Nash and Ramsey equilibria of the regime with ET, the coefficient with which the CB responds to supply shocks is $\frac{c}{b} = 2$, which is much higher than the coefficient $\frac{c}{b+c^2} = \frac{2}{3}$ for the other discretionary treatments. Accordingly, relative to the baseline discretionary treatment, under commitment we should see a lower volatility of inflation and a higher volatility of unemployment while under ET, we should see higher volatility in both variables.

Hypothesis 4. *Compared with discretionary treatments, the standard deviation of inflation is lower under commitment and higher under economic transparency, while output volatility is larger under commitment and economic transparency.*

Finally, we formulate a hypothesis regarding the CB group treatment that was added as a robustness check after analyzing the results from the other treatments. Since the incentives are the same for one or for 3 CB players, theory does not predict any difference in behavior, which defines our null hypothesis. As noted earlier, in the group treatment, subjects stayed in their role for the entire session. Thus, the CB groups gained more experience than any individual CB player in the other treatments. This change should help them to achieve low inflation and high welfare levels. The constant roles of players might also lower the volatility of unemployment and inflation, reduce B-players' forecast errors, and thereby raise their payoffs.

Hypothesis 5. *There is no difference in behavior and payoffs in the CB group treatment as compared to the baseline discretionary treatment.*

We generally use non-parametric tests based on average observations from a matching group to compare levels across treatments. We adopt this conservative approach because the behavior of subjects from the same matching group need not be independent. For point predictions arising from theory, we use the two-sided Wilcoxon matched pairs test, and for comparing different treatments, we use the two-sided Mann-Whitney *U*-test. More precise coefficient predictions are tested on the basis of confidence intervals from panel data regressions, where the unit of observation is matching group, sequence number, and subject ID. In the CB-group treatment, type-B players moved between the two groups of their session randomly at the start of each sequence, so that the data from these two groups are not independent.

3.4. Subjects, sessions and earnings

With the exception of the group CB treatment, each session consisted of 20 inexperienced subjects who were further divided up into two "matching groups" of size 10; the subjects in each matching group never interacted with one another and thus each matching group constitutes an independent observation. We have 8 observations of the six non-committee treatments. One half of these sessions/observations were conducted at the Berlin University of Technology and the other half at the University of Pittsburgh.¹³ As each observation involves 10 subjects, we thus report data from a total of $8 \times 10 \times 6 = 480$ subjects. We did not find substantial differences in behavior between our two subject populations (Berlin and Pittsburgh) and so in the analysis that follows we have pooled the data from all matching groups of a given treatment.

The CB group treatment was a robustness check conducted at the University of California, Irvine. Each session of that treatment consisted of only a single matching group of 14 subjects with 6 type A and 8 type B players. We randomly assigned players to roles as type A and B players at the start of a session, and players remained in those same roles in all rounds of all sequences. The 6 type A players were split into two groups of size 3 and were placed in two separate rooms so that they could talk with one another and their discussions would not be overheard by the type B players, who were in a different room. The type A players were free to discuss the policy choice of m_t , amongst themselves, but all 3 members had to agree on a single choice for m_t as each got the same payoff from their single choice. A further change was that at the start of each new sequence, the computer program randomly formed two groups of size 4 among the 8 type B players, but the groups of three player As remained intact; this design change was made because moving players between rooms was too costly in term of time. This change also had the effect that our CB committees gained more experience than did the individual CB players in all other treatments. Still, there was turnover in the composition of each group, as 4 player Bs were randomly assigned to each group of 3 type As at the start of each new sequence. In all other respects, the experimental design was the same as in the baseline discretionary treatment. We conducted 4 sessions of this "committee" discretionary treatment henceforth (CB Group) involving 8 CB committees and in total $4 \times 14 = 56$ subjects.¹⁴

At the start of each session, subjects were given written instructions that were also read aloud. Online Appendix B provides sample instructions from the baseline discretionary treatment.¹⁵ Subjects then had to answer several quiz questions designed to check their comprehension of the written instructions. Subjects' answers were individually checked for correctness; the experimenter explained to subjects any errors they made and what the correct answers should be. Then subjects played several indefinite-length sequences – they did not know how many would be played – and they were paid in cash at the end of the session. Each session lasted 2–3 h (subjects were always invited for 3 h) and involved 4–10 sequences. No

¹³ Subjects were mainly Bachelor- and Master students of engineering, economics, and other majors at TU Berlin and undergraduates in various fields of study at the University of Pittsburgh.

¹⁴ Subjects were again undergraduate students majoring in a variety of subjects.

¹⁵ Instructions for all treatments are available at: <http://www.socsci.uci.edu/~duffy/CBExperiment/>.

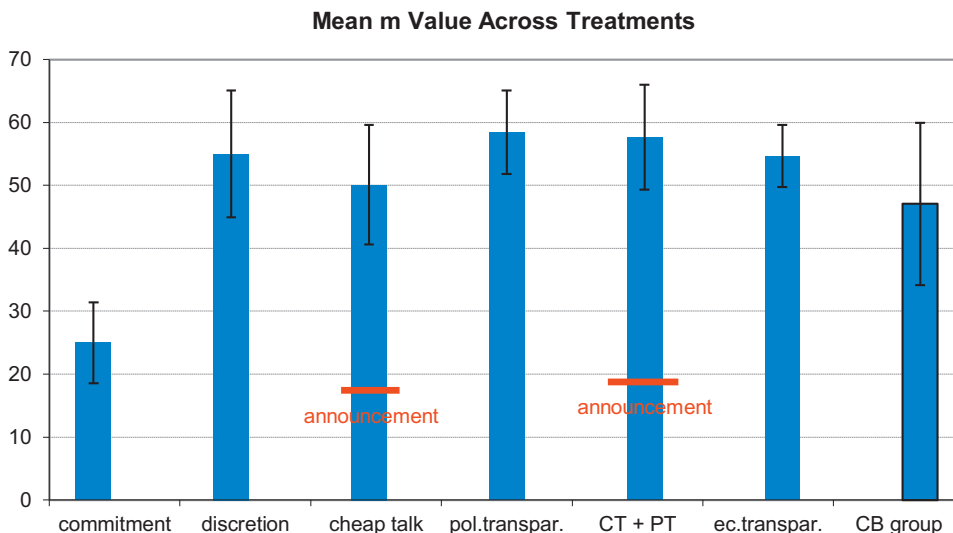


Fig. 1. Mean m Value Across Treatments.

session had to be continued, that is, all sessions finished within the 3 h time frame for which subjects had been recruited. Subjects were paid their earnings from 2 sequences; one of these sequences was the one in which the subject had earned the highest payoff and one was chosen randomly from among the other sequences.¹⁶ On average, each session involved 5.96 sequences with 34.1 rounds in total. Individual payoffs ranged from USD \$13.30 to \$56.46 with an average of about \$37.5.

4. Experimental Results

Means values of all our main outcome variables by treatment and by session are found in Online Appendix C. We report our experimental results as a number of different findings which address Hypotheses 1–5. Regarding Hypothesis 1, Fig. 1 shows the mean choice of m over all sequences of all sessions of each of our seven treatments. Also included is a one-standard error bar and the mean announced value of m in the two treatments involving pre-game communication (CT and CT+PT). Fig. 1 clearly reveals that the mean choice of m in the commitment treatment is indistinguishable from the Ramsey solution with $E(m) = 20$, whereas the mean value of m in the other treatments is significantly greater than 20.

Finding 1. *Inconsistent with Hypothesis 1, reputation does not serve as a substitute for commitment in any of the five discretionary regimes. In the discretionary regimes, the money supply and average inflation are close to the predictions of the one-period Nash equilibrium and higher than under commitment.*

Two-sided Mann-Whitney U-tests show that the money supply under commitment is smaller than in any of the other treatments ($p < 1\%$), while there are no significant differences in mean m between these other treatments ($p > 5\%$). A two-sided Wilcoxon matched pairs test cannot reject that $m = 20$ under commitment ($p = 15\%$), but rejects this efficient value for all other treatments ($p < 1\%$). The hypothesis that average money supply is equal to the point prediction of the one-period Nash equilibrium ($m = 60$) can be rejected for the CT ($p = 3.5\%$) ET ($p = 1.6\%$), and CB group ($p = 1.6\%$) treatments, but not for the other three discretionary treatments ($p > 10\%$). These results suggest that the one-shot Nash equilibrium and its predicted inflation bias may be highly relevant in a repeated game. Furthermore, neither cheap talk, nor policy transparency, nor economic transparency are effective in reducing the inflation bias.

Finding 2. *Inconsistent with Hypothesis 2, the private sector systematically under-predicts inflation in the five discretionary treatments, though not in the commitment nor in the CB group treatment.*

Evidence in support of Finding 2 is presented in Fig. 2. Since inflation is equal to the money supply plus a transmission shock with an average value of 20, rational forecasts should be distributed around $m + 20$. While this hypothesis cannot be rejected for the commitment treatment ($p = 7.8\%$) in which subjects know m when stating their forecasts, we can clearly reject it for all discretionary treatments ($p < 1\%$), except CB groups ($p > 10\%$). As Fig. 2 reveals, subjects in the discretionary treatments under-predict inflation. Comparing the different treatments, we find that under commitment, this systematic component of forecast errors is smaller than in all discretionary treatments ($p < 1\%$), except CB groups ($p > 10\%$). In the CT treatment, the under-prediction of inflation is more pronounced than in the other discretionary treatments ($p < 2\%$). The PT,

¹⁶ We selected the sequence with the highest payoff, because longer sequences typically lead to higher profits. Thus, type A had high incentives to build up a reputation that pays off in long sequences.

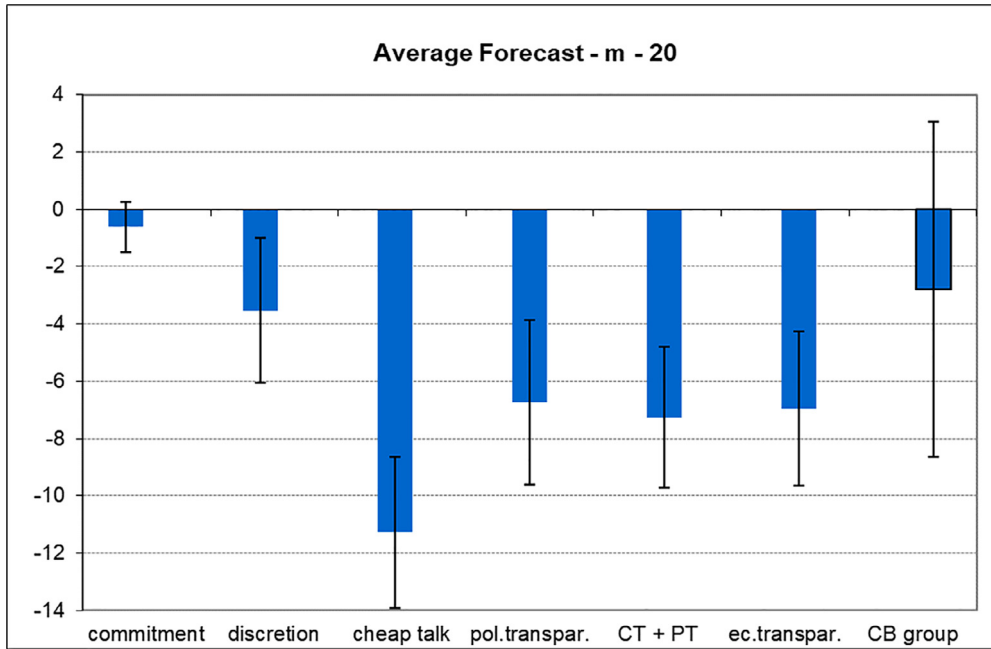


Fig. 2. Average Forecast - m - 20.

CT+PT, and ET treatments all have similar systematic forecast errors ($p > 40\%$) and these are all larger than under baseline discretion ($p < 3\%$). The existence, direction and size of these systematic forecast errors are surprising, and we will discuss some possible explanations below. Note that even though subjects under-predict inflation, in all discretionary treatments the average inflation forecast is significantly higher than the target value π^* ($p < 1\%$) and also higher than in the commitment regime ($p < 2\%$).

Since private sector expectations fall short of average inflation rates, the unemployment rate deviates from the natural rate towards the CBs' target rate (except for commitment treatment). Under cheap talk, the effect is so strong that average unemployment is closer to target than to the natural rate. There are two explanations that may contribute to these results. (i) Forecasting type B subjects are irrational in the sense that they under-predict how the CB will respond to expectations, (ii) the CB player tends to raise the money supply above expectations to gain from the short-run effect of lower unemployment.

While theory predicts differences in welfare between the different treatments arising from the different inflation biases and the different abilities of CBs to stabilize employment, the systematic deviations of average unemployment from the natural rate provide a third and unexpected factor influencing (CB) welfare. For testing Hypothesis 3, we compare the actual payoffs of our CB players as a measure of welfare.

Finding 3. *In most discretionary treatments, average welfare is closer to the one-shot Nash equilibrium and significantly smaller than under commitment counter to Hypothesis 3. Only cheap talk works as a substitute for commitment regarding welfare.*

Fig. 3 shows that average welfare is below the level associated with optimal policy in all treatments ($p < 4\%$). However, since our subject central bankers are not perfect, they also make mistakes under commitment, so that the payoffs of CB players in this treatment are smaller than predicted by equilibrium ($p < 1\%$). If we compare the achieved payoffs between different treatments, we cannot reject Hypothesis 3 for the CT treatment. Here, the achieved CB payoffs are not smaller than under commitment ($p = 13\%$), while they are significantly smaller for the other discretionary treatments including CB-groups ($p < 4\%$). Comparing the payoffs for the CT treatment with the other discretionary treatments directly, the evidence is mixed: CB payoffs are higher under CT than for PT ($p = 2.1\%$) and ET ($p = 5.0\%$), but not significantly different from baseline discretion ($p = 8.3\%$) and CT+PT ($p = 13\%$). Comparing welfare with the predictions of the one-period Nash equilibrium, we can clearly reject the null of no difference for the CT, ET, and CB group treatments ($p = 1.6\%$ each), but *not* for the other discretionary treatments ($p > 80\%$). The main reason for the observed differences in welfare is the inflation bias of the discretionary treatments. The resulting welfare loss is mitigated to some extent by the private sector's under-prediction of inflation, which leads to average employment above the NAIRU. Volatility of output (unemployment) and inflation also affect welfare, and so we next turn to evaluating Hypothesis 4 which pertains to volatilities in these two measures.

Table 1 displays the standard deviation of inflation and unemployment, averaged over all matching groups for each treatment along with equilibrium predictions, repeated from Section 3.4. As Table 1 reveals, the volatility of inflation is about the same across all treatments except CB groups. With the exception of the ET treatment and the CB groups, inflation volatility

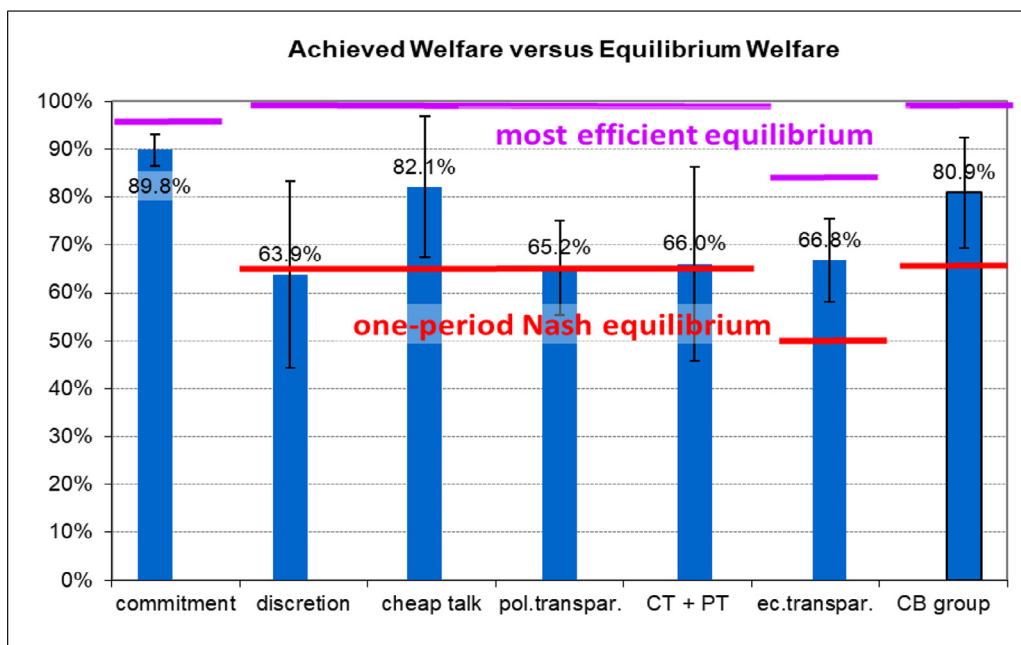


Fig. 3. Achieved Welfare versus Equilibrium Welfare.

Table 1

Standard Deviation of Unemployment and Inflation.

	Commitment	Discretion	Cheap Talk	Policy Transparency	CT + PT	Economic Transparency	CB Group
Mean of π_t	45.13	74.83	69.44	78.60	77.91	74.75	66.77
St.Dev. of π_t	18.86	19.94	19.60	18.44	17.93	20.28	14.86
(Std. error)	(2.79)	(2.10)	(2.69)	(2.63)	(1.14)	(2.13)	(2.26)
Equilibrium st. dev. of π	11.55	13.88	13.88	13.88	13.88	25.82	13.88
Mean of u_t	140.21	135.87	129.43	132.24	132.29	133.22	136.97
St.Dev. of u_t	16.56	19.48	17.73	18.66	19.03	20.55	14.69
(Std. error)	(1.35)	(2.58)	(2.51)	(3.79)	(2.94)	(2.15)	(1.78)
Equilibrium st. dev. of u	16.33	12.17	12.17	12.17	12.17	16.33	12.17

Note: Std. error is the standard error of the "standard deviation" across matching groups.

is significantly greater than in the rational expectations equilibria ($p < 1\%$); under ET, inflation volatility is smaller than predicted ($p < 1\%$). The reason for this is that CB players in the ET treatment do not respond more strongly to supply shocks (as theory would predict) than in other discretionary treatments. The overall high levels of inflation volatility can be explained by eventual changes in CB strategies, in particular when CB-players are changed at the start of a new sequence. In the CB group treatment, we did not have such role changes. Here, inflation volatility is smaller than in all other treatments ($p < 2\%$) and not significantly greater than the equilibrium prediction ($p > 10\%$).

In contrast to Hypothesis 4, we find that volatility in unemployment is higher in the discretionary treatments than under commitment (significant for baseline discretion, $p = 2.1\%$, and for ET, $p < 1\%$). In fact, our human subject CB players even contribute to fluctuations in unemployment, because standard deviations are higher than they would be for a constant money supply, 16.33. Two-sided Wilcoxon matched pairs tests show that fluctuations in unemployment are significantly larger than 16.33 for the baseline discretionary treatment ($p = 2.3\%$), for the ET treatment ($p < 1\%$), and for the CT+PT treatment ($p = 3.9\%$). For all discretionary treatments (including CB groups), fluctuations of unemployment are higher than equilibrium predictions ($p < 5\%$). The reasons for these high fluctuations are: (i) CBs respond to supply shocks in the right direction but with coefficients that are suboptimally small, as will be shown in the next subsection, (ii) changing inflation forecasts directly affect the inflation-unemployment trade-off, while theory predicts constant forecasts, and (iii) changing CB players with changing strategies. Strategy changes should have a smaller effect in the CB group treatment, where CB players stayed in the same role for all sequences. Indeed, we see that fluctuations in unemployment are significantly smaller in the CB group treatment than in any other treatment ($p < 5\%$) and even marginally smaller than 16.33 ($p < 10\%$). Thus, constant CB groups can partially stabilize unemployment in a discretionary environment, though not optimally.

Finding 4. Contrary to Hypothesis 4, there is little difference in the volatility of inflation across treatments with a single CB player. Unemployment (output) volatility is not significantly larger in the commitment and economic transparency treatments

Table 2
Central Bank Behavior.

Parameter	Commitment	Discretion	Cheap Talk	Policy Transparency	CT + PT	Economic Transparency	CB Group
α	−40.06*** (9.414)	42.40** (17.51)	−44.12*** (17.07)	11.82 (24.31)	−19.54 (16.59)	−30.60** (12.92)	56.25*** (11.18)
δ_1	0.44*** (0.06)	0.28*** (0.06)	0.50*** (0.06)	0.44*** (0.08)	0.43*** (0.07)	0.45*** (0.08)	0.52*** (0.05)
δ_2	–	−0.30 (0.20)	0.45*** (0.14)	−0.15 (0.22)	0.31*** (0.10)	0.30*** (0.11)	0.41*** (0.10)
β_1	0.12 (0.08)	−0.06 (0.09)	−0.06 (0.07)	−0.06 (0.12)	−0.08 (0.05)	0.03 (0.07)	0.11 (0.08)
Pr > χ^2 [†]	.000	.000	.000	.000	.000	.000	.000
Observations	432	408	360	384	300	344	172

Notes: *** significant at 1%, ** significant at 5%, * significant at 10% levels; robust standard errors in parentheses.

[†] p-values from a Wald test of the hypothesis that the coefficient estimates on the independent variables other than the intercept are all zero.

as compared with the other discretionary treatments. The volatilities of inflation and output in the discretionary treatments is generally higher than predicted.

The finding that discretionary policy leads to higher fluctuations of employment than commitment was surprising to us. It provides a strong case for rule-based monetary policy. However, our CB group treatment, to which we now turn, shows that this result may not be robust. With constant groups of CB players, inflation volatility is smaller than for changing single CB players, and CB groups also achieve a partial stabilization of unemployment. Finally, we address whether central banking by a committee leads to any improvement in the baseline discretionary treatment regarding average inflation or welfare. Except for CB group 1 in session 1, which achieved something close to the Ramsey solution, the other 7 CB committees made policy choices for m that are in line with our baseline discretionary policy treatment involving a single CB player. Yet, m is significantly smaller than in the one-period Nash equilibrium ($p=1.6\%$) and CB committees achieved marginally higher payoffs than single CBs in the discretionary treatment ($p=8.4\%$). The payoffs to CB committees are about the same as for lone CB players in the cheap-talk treatment ($p=96\%$) but remain lower than for CB players in the commitment regime ($p=3.2\%$). Thus, while committees can outperform individual policymakers, the inefficiency of the discretionary environment persists even when policy is determined by committees of size 3 with more experience than the lone policy makers in the other treatments. The most remarkable observation in the CB group treatment is that the volatilities of unemployment (14.69) and inflation (14.86) are smaller than in any other treatment ($p<5\%$). We attribute this difference to the constancy of subjects in the CB role; in all other treatments, policy makers were changing with each new sequence and did not deliberate with others on a policy choice. We conclude that the implementation of different strategies by different policy makers may be responsible for the high volatility of outcomes we observed in our single CB treatments. Nevertheless, forecasters' payoffs are the same as in the baseline discretionary treatment ($p=27\%$), and smaller than under commitment ($p<1\%$).

Finding 5. While committees of CB players can outperform solo CB players, they are not a substitute for commitment in the discretionary setting. The overall volatility of output and inflation is smaller in the CB group treatment than in any other treatment and forecasters do not under-predict inflation in the CB group treatment, but payoffs to forecasters do not differ significantly from those under baseline discretion.

5. Interaction Effects

The aggregate results can to a large extent be explained by CB and private sector players' interactions with one another. We first consider CB reactions to private sector forecasts and supply shocks. We report on treatment-specific, panel data regressions examining policy choices by CB type A subject (or committee) i of matching group j in sequence k . Our main CB regression specification is:

$$m_t^i = \alpha + \delta_1 w_t + \delta_2 avgf_t + \beta_1 m_{t-1}^i,$$

where m_t^i is i 's time t money supply choice, and w_t and $avgf_t$ are, respectively, the supply shock and average inflation forecast that CB player i faced at time t . We impose the additional restriction that $\delta_2 = 0$ for the commitment treatment since the CB had to decide on m before knowing the private sector's average inflation forecast. Since our specification also includes lagged policy, m_{t-1}^i , we estimate our dynamic panel data model by GMM using the Arellano and Bond (1991) estimator. The results are reported in Table 2.

The estimated coefficient on the supply shock, δ_1 , is significantly positive in all treatments ($p<1\%$) indicating that CBs respond to these shocks. For the commitment treatment, this finding is at odds with the prediction that CBs should ignore the supply shock (provided that forecasters are rational), and explains why inflation volatility is higher than predicted in this treatment. In the discretionary treatments (including CB groups), the estimates for δ_1 are significantly less than the prediction of $2/3$ ($p<5\%$). Thus, our human CBs were attempting to stabilize employment but their responses are, on average, less than optimal.

Table 3
Private Sector Forecasts.

Parameter	Commitment	Discretion	Cheap Talk (1)	Cheap Talk (2)	Policy Transparency	CT + PT (1)	CT + PT (2)	Economic Transparency	CB Group
α	21.66*** (1.03)	37.97*** (3.52)	52.57*** (3.38)	56.14*** (6.30)	65.64*** (6.66)	69.36*** (4.51)	74.43*** (6.26)	3.38 (8.15)	42.88 (5.02)
δ_1	0.94*** (0.02)	–	–	–	–	–	–	–	–
δ_2	–	–	0.41*** (0.04)	–	–	0.21*** (0.03)	–	–	–
δ_3	–	–	–	0.08** (0.03)	–	–	0.04 (0.03)	0.49*** (0.05)	–
β_1	0.00 (0.11)	0.17 (0.13)	0.19* (0.10)	0.33*** (0.13)	0.11 (0.12)	–0.15** (0.06)	–0.10 (0.07)	0.06 (0.08)	0.24*** (0.12)
β_2	0.06*** (0.01)	0.23*** (0.02)	0.09*** (0.02)	0.07*** (0.03)	0.11*** (0.03)	0.07*** (0.02)	0.04** (0.02)	0.09*** (0.02)	0.18*** (0.03)
β_3	–0.08 (0.10)	0.08 (0.12)	–0.29** (0.12)	–0.55*** (0.14)	–0.14 (0.11)	0.05 (0.08)	–0.07 (0.08)	–0.21*** (0.07)	–0.09 (0.12)
$\Pr > \chi^2 \dagger$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Obs.	1728	1632	1440	1440	1536	1200	1200	1376	693

Notes: *** significant at 1%, ** significant at 5%, * significant at 10% levels; robust standard errors in parentheses.

† p-values from a Wald test of the hypothesis that the coefficient estimates on the independent variables other than the intercept are all zero.

Finding 6. CB players' responses to supply shocks are significantly positive in all treatments, contrary to the equilibrium prediction for the commitment regime. They are significantly lower than equilibrium predictions in all discretionary environments.

We next consider how private sector Type B players react to CB policy and announcements. Our regression model of individual inflation forecasts, f_t^i , uses treatment-specific information available to type B players when forming forecasts, including the money supply, m_t , the CB's announcement, ann_t , the supply shock, w_t , lagged forecasts, the lagged inflation rate, and the lagged average forecast. The coefficient on the lagged average forecast provides a measure of convergence in forecasts. Specifically, we estimate the following:

$$f_t^i = \alpha + \delta_1 m_t + \delta_2 ann_t + \delta_3 w_t + \beta_1 f_{t-1}^i + \beta_2 \pi_{t-1} + \beta_3 avgf_{t-1},$$

where the individual unit corresponds to type B subject i of matching group j in sequence k . Regression results, again using the Arellano and Bond dynamic panel data estimator, are reported in Table 3

As Table 3 reveals, CB policy affects private sector expectations directly in the commitment and CT regimes and indirectly via lagged inflation; the estimated β_2 coefficient on lagged inflation is always positive and significant, indicating adaptive learning. In the Commitment treatment, theory predicts that $\delta_1 = 1$, and while we can reject this prediction ($p < 5\%$), the actual estimate (0.94) is close to 1. For both CT treatments, the response of inflation forecasts to CB announcements is positive and significant as indicated by the estimate for δ_2 . If CT is combined with policy transparency (CT+PT), the announcement coefficient is reduced by more than 50% which explains why subjects' under-prediction of inflation is less pronounced in the CT+PT treatment compared with CT only. The observation that CT affects forecasts in particular, when CT messages cannot be falsified may explain why we see the largest under-prediction of inflation in the CT only treatment. In a second specification (2) for the two CT treatments in Table 3, we explore whether unobserved supply shocks affect individual forecasts via CB announcements. We find that in the CT treatment without policy transparency, the supply shock does affect forecasts, as evidenced by the significantly positive estimate for δ_3 suggesting that the CB is communicating this information via its announcements. In the CT+PT treatment, δ_3 is no longer significant. In the ET treatment, the supply shock coefficient, $\delta_3 = 0.49$, which is significantly smaller than the theoretical prediction of 2 ($p < 1\%$) and almost the same as the CB's response to these shocks, which is 0.55 (see Table 2). Summarizing we have:

Finding 7. Central bank policy and the institutional regime both matter for inflation forecasts. In all treatments, subjects' forecasts respond to past inflation. Consistent with theory, under commitment, forecasts are closely aligned with the monetary policy choice of the CB. In the CT and CT+PT treatments, forecasts respond to policy announcements. Without policy transparency, these announcements reveal information about supply shocks. Under ET, subjects condition their inflation forecasts on supply shocks.

If subjects are learning, as indicated by Finding 7, the systematic under-prediction of inflation reported earlier in Finding 2 should decrease over time. To look for evidence of such learning, we analyze the data separately for the first and last 15 periods of each treatment,¹⁷ test whether the systematic forecast errors get significantly smaller, and how these changes affect welfare.

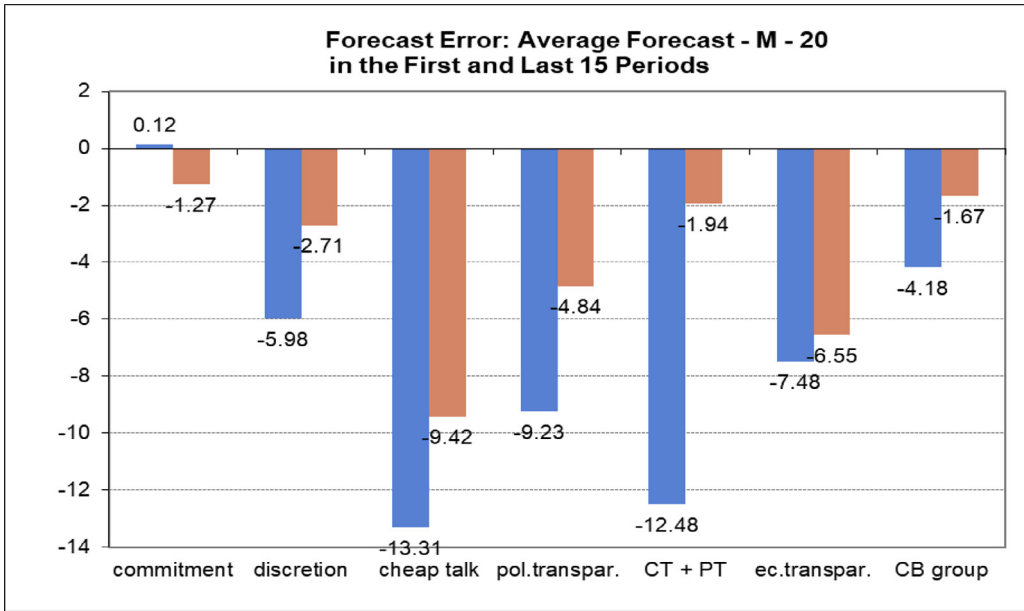


Fig. 4. Systematic Forecast Error: Average Forecast - m - 20 in the First and Last 15 Period.
 Note: In CT+PT we compared the averages from the first and last 12 periods.

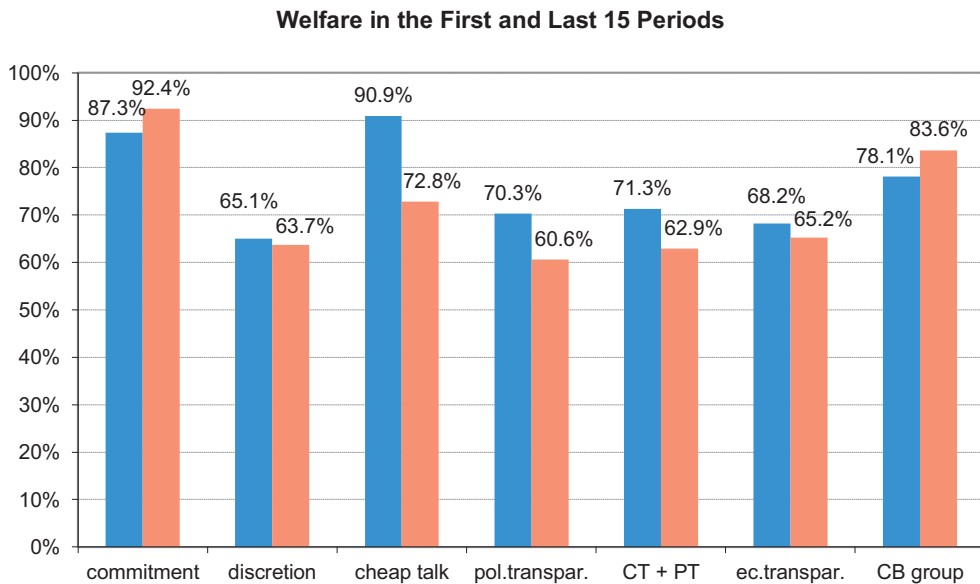


Fig. 5. Welfare in the First and Last 15 Periods.
 Note: In CT+PT we compared the averages from the first and last 12 periods.

This analysis is presented in Figs. 4 (forecast errors) and 5 (welfare).¹⁸ We find that the systematic inflation forecast error is indeed smaller in the later periods than in the early periods in all discretionary treatments (see Fig. 4). This difference is significant for the baseline discretionary treatment and for the CT+PT treatment ($p < 2\%$; in all other treatments: $p > 10\%$). In the CT treatment (and only there), the average money supply is significantly higher in the later periods than in the early periods ($p < 1\%$). Rising average inflation and unemployment reduce the achieved welfare levels in the CT treatment (see

¹⁷ In CT+PT sessions, we had only 25 to 32 periods, because they took a bit longer. Thus, for the CT+PT sessions, we took the first and last 12 periods instead. The other sessions had 30 to 49 periods.

¹⁸ In Figures 4 and 5, the left (blue) bars visualize results from the first periods, the right (brown) bars visualize results from the last periods of each treatment.

Fig. 5): the average welfare level in the last 15 periods is at 72.8% of the first best. This is smaller than under commitment ($p=8.3\%$) and the differences between CT and the other discretionary treatments or the one-shot Nash equilibrium are not significant anymore ($p>19\%$). For CT, the change in welfare is highly significant ($p<1\%$). For all other treatments, the differences are not significant ($p>10\%$). Summarizing, our main result from examining expectations and welfare over time is:

Finding 8. *The forecast error in favor of low inflation diminishes with experience, significantly so in 2 of the 6 discretionary treatments. Under cheap talk, the money supply rises over time and welfare decreases to levels comparable to the other discretionary treatments.*

Why do type B subjects not use cheap-talk announcements or information about supply shocks to better coordinate their expectations than in the baseline treatment without this information? The data indicate that subjects disagree in how actual money supply is related to announcements or supply shocks. Different perceptions of CB credibility create a dispersion in forecasts. Further, under economic transparency, subjects can have different expectations about the CB's responses to supply shocks. If these expectations diverge, the posterior beliefs after learning the supply shock may be more dispersed than prior beliefs without such information. While in theory, economic transparency raises forecasters' payoffs compared to baseline discretion, we observe just the opposite in our experimental data. This is merely the result of strategic uncertainty.

6. Conclusion

Central bankers operate in a discretionary world where they face a trade-off between credibility in stabilizing inflationary expectations on the one hand and flexibility in response to economic shocks on the other. In this paper we have posed the question of whether a balance can be found between these twin objectives in a repeated game setting where reputational concerns might serve as a substitute for commitment so that welfare under discretionary policy might approximate or even exceed the levels attainable under a full commitment regime. We address this question using a version of the Barro-Gordon (1983a,b) monetary policy game and controlled laboratory experiments with paid human subjects serving in the role of private sector agents forecasting inflation or as central bankers facing the policy trade-off between credibility and flexibility.

Our evidence suggests that achieving the first best, Ramsey solution in the discretionary setting is difficult. We find important policy and welfare differences between the discretionary and commitment regimes. In addition, we have considered several augmented versions of our baseline discretionary policy regime that allow CB cheap talk about policy intentions, ex-post transparency about policy actions, both cheap talk and policy transparency, economic transparency about shocks, and committees of central bankers. In all of these regimes, we find evidence that CB behavior affects private sector inflation forecasts. However, among these various discretionary environments, only those with cheap talk served to raise welfare to levels approximating those achieved under the commitment regime, as central bankers tried to convince private sector agents that they would keep inflation low. Still, this effect diminished over time, as private sector agents learned to ignore the CB's cheap talk messages. We conclude that the discretionary regimes of our laboratory study are indeed welfare reducing relative to the commitment regime. This welfare conclusion is further supported by the observation that private sector forecasters earn the highest payoffs under a commitment regime.

It is remarkable that we do not find a trade-off between credibility and flexibility. Indeed, we find that discretionary policy regimes which allow for the stabilization of employment actually lead to *higher* variation in employment than under a perfectly credible commitment regime where stabilization is not possible, although central bankers *do* make an attempt to stabilize the impact of supply shocks in all regimes. As we have seen, there are several reasons for this divergence between theory and outcomes. First, CB responses to shocks are smaller than optimal in all of our discretionary treatments, diminishing the stabilizing role of those more flexible discretionary regimes. Second, changes in strategies by different central bankers provide an additional source of strategic uncertainty that is not addressed by the theory. As this strategic uncertainty is likely to carry over to the field, we conclude that CBs would do well to follow long-run strategies and make them common knowledge to private sector forecasters. Any changes in CB strategies should be communicated thoroughly and before they are implemented, to minimize strategic uncertainty. Third, systematic forecast errors are greatest in the CT treatment, where subjects over-rely on the announcements of CB players and thus persistently under-predict inflation. Forecasters are only successful in correcting this error over time when they receive the additional information provided in the regime with policy transparency.

We expect that our experimental findings carry over to the "real world" since the incentives and uncertainties that our subjects face also approximate those faced by real central bankers and private sector agents. For instance, we note that there is corroborating empirical evidence that CBs publish inflation forecasts to strategically manipulate private inflation forecasts just as our human subject central bankers do in the treatments with cheap talk (Gomez-Barrero and Parra-Polania, 2014). While it is possible that well-intentioned, real world central bankers, aware of the time inconsistency problem of monetary policy, can learn to implement the optimal policy – in the words of McCallum (1995), they "just need to do it" – the long debate about rules versus discretion in CB policy suggests that there are also doubts about the ability of such real world central bankers to effectively manage the trade-off between credibility and flexibility. The experimental evidence that we have presented in this paper provides further evidence that such doubts may be warranted.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jmoneco.2020.06.006](https://doi.org/10.1016/j.jmoneco.2020.06.006).

References

- Arifovic, J., 2014. Evolving better strategies for central bank communication: evidence from the laboratory. In: Duffy, J. (Ed.), *Experiments in Macroeconomics*. Emerald, Bingley, U.K, pp. 229–258.
- Arifovic, J., Dawid, H., Deissenberg, C., Kostyshyna, O., 2010. Learning benevolent leadership in a heterogeneous agents economy. *J. Econ. Dyn. Control.* 34, 1768–1790.
- Arifovic, J., Sargent, T.J., 2003. Laboratory experiments with an expectational Phillips curve. In: Altig, D.E., Smith, B.D. (Eds.), *Evolution and Procedures in Central Banking*. Cambridge Univ. Press, Cambridge, UK, pp. 23–55.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58, 277–297.
- Barro, R.J., Gordon, D.B., 1983a. A positive theory of monetary policy in a natural rate model. *J. Polit. Econ.* 91, 589–610.
- Barro, R.J., Gordon, D.B., 1983b. Rules, discretion and reputation in a model of monetary policy. *J. Monetary Econ.* 12, 101–121.
- Blinder, A.S., 1998. *Central Banking in Theory and Practice*. MIT Press, Cambridge, MA.
- Blinder, A.S., Morgan, J., 2005. Are two heads better than one? Monetary policy by committee. *J. Money Credit Bank* 37, 798–811.
- Bullard, J., 2016. Inflation expectations are important to central bankers, too. In: *Fed. Res. Bank St. Louis Reg. Economist*, p. 3.
- Campbell, J.R., Evans, C.L., Fisher, J.D.M., Justiniano, A., 2012. Macroeconomic effects of federal reserve forward guidance. In: *Brookings Pap. Eco. Ac.*, Spring, 2012, pp. 1–80.
- Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. *Am. Econ. Rev.* 95, 477–509.
- Cornand, C., Heinemann, F., 2014. Experiments on monetary policy and central banking. In: Duffy, J. (Ed.), *Experiments in Macroeconomics*. Emerald, Bingley, U.K, pp. 167–227.
- Cornand, C., M'baye, C.K., 2018. Does inflation targeting matter? An experimental investigation. *Macroecon. Dyn.* 22, 362–401.
- Engle-Warnick, J., Turdaliev, N., 2010. An experimental test of Taylor-type rules with inexperienced central bankers. *Exp. Econ.* 13, 146–166.
- Filardo, A., Hofmann, B., 2014. Forward guidance at the zero lower bound. *BIS Q. Rev.*, March 37–53.
- Fischbacher, U., 2007. Z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178.
- Geraats, P.M., 2002. Central bank transparency. *Econ. J.* 112, F532–F565.
- Geraats, P.M., 2014. Monetary policy transparency. In: Forssbæck, J., Oxelheim, L. (Eds.), *The Oxford Handbook of Economic and Institutional Transparency*. Oxford Univ. Press, Oxford, UK Chapter 3.
- Gomez-Barrero, S., Parra-Polania, J.A., 2014. Central bank strategic forecasting. *Contemp. Econ. Policy* 32, 802–810.
- Kryvtsov, O., Petersen, L., 2013. Expectations and monetary policy: experimental evidence. *Bank of Canada Working Paper No. 2013-44*.
- Kydland, F., Prescott, E.C., 1977. Rules rather than discretion: the inconsistency of optimal plans. *J. Polit. Econ.* 85, 473–491.
- Lucas Jr., R.E., 1986. Adaptive behavior and economic theory. *J. Bus.* 59, S401–S426.
- McCallum, B.T., 1995. Two fallacies concerning central bank independence. *Am. Econ. Rev.* 85, 207–211.
- Phillips, A.W., 1950. Mechanical models in economic dynamics. *Economica* 17, 283–305.
- Taylor, J.B., 1999. In: Taylor, J.B. (Ed.), *Monetary Policy Rules*. Univ. of Chicago Press, Chicago, pp. 319–348.
- Van Huyck, J.B., Battalio, R., Walters, M.F., 2001. Is reputation a substitute for commitment in the peasant-dictator game? Working paper. Texas A&M University.
- Vayid, I., 2013. Central bank communications before, during and after the crisis: from open-market operations to open-mouth policy. *Bank of Canada Working Paper No. 2013-41*.