Wirtschaftswissenschaftliche Fakultät

Cold Turkey vs. Gradualism –Evidence on Disinflation Strategies from a Laboratory Experiment

Marcus Giamattei

Diskussionsbeitrag Nr. V-67-15

Volkswirtschaftliche Reihe ISSN 1435-3520

PASSAUER DISKUSSIONSPAPIERE

Cold Turkey vs. Gradualism - Evidence on Disinflation Strategies from a Laboratory Experiment

Marcus Giamattei[a,*]

[a]*Innstrasse 27, D-94032 Passau. Phone: 0049 851 509 2553*

Abstract

Disinflation can be implemented gradually or via Cold Turkey - an immediate change of policy - with the latter being mainly recommended by theory and empirical literature. But Cold Turkey may only be superior because it is endogenously selected for favorable environments. To eliminate this endogeneity and to disentangle the credible push through of a disinflation policy from ex-ante credibility, I run an experiment where a central banker has to decide for a disinflationary strategy and four forecasters try to coordinate on it. The design abstracts from any rigidities and provides full information so that Cold Turkey is the Nash equilibrium. But Cold Turkey fails to be the most successful strategy because forecasters react sluggishly due to limited reasoning. Cold Turkey does not speed up learning or increase reasoning, is less successful and is reversed more often.

JEL Classification: E52, E58, C72, C92

Keywords: Disinflation, Credibility, Cold Turkey, Gradualism, Limited Reasoning, Endogenous Treatments

1. Introduction

A long debate in monetary economics has distinguished two[1] possible regimes to conduct a disinflation[2] policy - gradualism versus Cold Turkey (Sargent, 1982; Gordon, 1982). The main contribution of this paper is to disentangle the effect of the policy regime on credibility and successful implementation. In the experiment, a central banker has to push through a disinflation strategy with four forecasters trying to coordinate on that policy. The central banker is punished both for a high inflation and a large deviation of forecasters from equilibrium where the latter is similar to an output target.

Central bankers decided endogenously on a policy and could continue or reverse their policy in each round. Half of the central bankers opted for a Cold Turkey regime. This endogenous regime choice constitutes the treatments forecasters were confronted with. It also induced different levels of credibility as central bankers had to decide if they stick to their choice or abandon it. This also called for human central bankers instead of a computer rule

*Corresponding author. Research Assistant at the Chair in Economic Theory of Prof. Johann Graf Lambsdorff. University of Passau, Germany.

Email address: `marcus.giamattei@uni-passau.de` (Marcus Giamattei)

[1]A third approach not discussed here is called opportunistic disinflation (Orphanides and Wilcox, 2002; Aksoy et al., 2006), an asymmetric strategy to "guard vigorously against any rise in inflation, but wait patiently for the next favorable inflation shock to bring inflation down." (Blinder, 1997, 6).

[2]This problem not only arises for the optimal way to fight inflation but similar policy options are available for establishing a new central bank like the ECB (Illing, 1998), trade liberalization (Rodrik, 1989), stock market liberalization (Kim et al., 2013) or a basic income reform (Spermann, 2006).

in order to determine credibility endogenously. This kind of endogeneity can be also found in empirical data. But in contrast to real central bankers the subjects in the experiments were exogenously matched with new forecasters. They could not build up ex-ante credibility and their choices were completely independent of the knowledge real central bankers may have gained beforehand about the subjects the policy tries to address. So credibility is solely reflected by the fact whether or not a central banker sees his strategy through to the end.

A second contribution is a study of the effects of limited reasoning as a new cause for inflation persistence. The laboratory provides a controlled environment to assess the levels of reasoning of the forecasters (and what the forecasters think of other forecasters and so on) along with the central bank's expectation about that reasoning. The setup desists from any rigidities and builds on complete information, giving Cold Turkey the best chances to be successful. But limited reasoning provides an obstacle to a successful Cold Turkey implementation.

The third contribution, therefore, assesses the success of both regimes in terms of payoff dominance. The experiment thereby allows to meet one of the main critiques already raised against the article of Ball (1994b) that the

> "finding that faster disinflations require less sacrifice raises an important problem of reverse causation. [...] Ball takes monetary [...] as exogenous with respect to the behavior he is investigating. But what if central banks pursue their presumed goal of disinflation more rigorously when they have reason to believe that the short-run trade-offs associated with doing so are more favorable?" (Friedman, 1994, 186).

This can be controlled for in the experiment because central bankers are randomly matched with a group of forecasters. Additionally, the experimental setup gives an insight into disequilibrium behavior, which is difficult to approach theoretically. Haltiwanger and Waldman (1989) show that it is sufficient to have a small share of agents with non-rational beliefs to ensure that disequilibrium behavior persists in a complementary environment because even rational agents may imitate the non-rational behavior to better coordinate. Last, experiments provide an optimal tool to study alternative policies without involving any economic cost by implementing the policy in reality (Cornand and Heinemann, 2014, 170).

The results show that central bankers who try to pursue a Cold Turkey strategy fail due to the large inertia of the forecasters' choices. Forecasters stick to their errors and do not correct them sufficiently. This inertia makes a gradual disinflation the superior strategy. It satisfies coordination better but reduces inflation to a lower extent. Cold Turkey fails to be the payoff dominant strategy and reduces credibility as central bankers reverse their policy. Due to the fact that limited reasoning induces a fail of Cold Turkey even in this simple environment, it may be also an important feature in the real economy.

This paper is organized as follows. Section 2 provides details on the theoretical and empirical debate. Section 3 briefly reviews evidence from related macroeconomic experiments. The experimental design is explained in section 4. Section 5 derives the hypotheses. Section 6 outlines experimental procedures. Data will be described and analyzed in section 7. Section 8 concludes.

2. Theoretical and Empirical Literature

A long debate in the literature has discussed how an optimal disinflation can be implemented. As a standard argument, Cold Turkey is to be preferred because, with rational expectations, an immediate adaption to the new steady state takes place (Mankiw, 2001; Ball, 1994a). Ball (1994b) provides first empirical evidence and shows lower sacrifice ratios for faster disinflation, which is confirmed by other studies (Diana and Sidiropoulos, 2004;

Jordan, 1997; Zhang, 2005)[3]. But the advantages and disadvantages of Cold Turkey are influenced by a range of factors, namely (1) rigidities, (2) limited credibility, (3) limited knowledge and learning and (4) limited rationality.

First, rigidities[4] may hinder immediate adaption and call for a more gradual adjustment[5]. Staggered contracts are one reason for rigidities which cause greater output losses under a Cold Turkey regime (Blanchard, 1985; Taylor, 1983; King, 1996). They create higher inflation persistence (Fuhrer, 1995), which involves higher costs when bringing inflation down abruptly. Fixed costs of price adjustment are another type of rigidities. Only when inflation is high firms are willing to burden these adjustment costs. Therefore, gradualism is optimal for low initial inflation and Cold Turkey for high initial inflation (Ireland, 1997).

Second, limited credibility may influence the optimal regime choice (Ireland, 1995; Erceg and Levin, 2003; Goodfriend and King, 2005). Private agents may believe that a central bank will not push through an announced policy change and may be tempted to return to its old pre-reform policy when economic or political pressure become too high. A part of the literature assumes credibility to be an ex-ante exogenous determinant of the disinflation policy. Some studies, for example, use independence as a proxy and show that credibility is not beneficiary for disinflation (Fischer, 1997; Jordan, 1997, 1999; Posen, 1998; Diana and Sidiropoulos, 2004). Instead, Boschen and Weise (2001) show that ex-ante credibility makes disinflation less costly. Loh (2002) favors Cold Turkey in a menu cost model if credibility is high, but a gradual approach for low credibility. This is due to the fact that again firms incur the cost of price adjustment only when they believe the change is permanent.

But credibility may vary endogenously with the policy pursued. The standard assumption, the so-called credibility hypothesis, states that Cold Turkey enhances credibility (Sargent, 1986; Ball, 1994b). This is due to the fact that it may be considered a regime-shift which goes in line with an expectational shift. Subjects adapt their expectations to the new regime as it may be viewed as a permanent and substantial change. "Gradualism invites speculation about future reversals, or U-turns, in policy" (Sargent, 1986, 150). In contrast thereto, Blanchard argues that "disinflation policies may fail simply because they are expected to fail. Failure is less likely for mild disinflations and this provides some support for gradualist policies" (Blanchard, 1985, 217). These two opinions show a sharp contrast regarding the influence of the policy choice on the central bank's credibility. But credibility is hard to measure (Blinder, 2000, 1421) due to endogeneity problems and other methodological concerns (Agénor and Taylor, 1993).

Third, learning may play a role. In the literature, this is due to the fact that subjects have limited knowledge and are adaptive learners (Orphanides and Williams, 2005; Al-Eyd and Karasulu, 2008; Milani, 2007; Kurz et al., 2013). This creates a policy tradeoff because "a stronger policy response under learning helps to align expectations with those of authorities more quickly reducing the costs of adjustment" (Al-Eyd and Karasulu, 2008, 17). This gives rise to the idea that an advantage of Cold Turkey lies in sending a clear signal on which forecasters can coordinate and with which learning can be enhanced.

[3] Zhang (2005) finds a similar relationship but controlling for initial inflation leaves results inconclusive. The discussion of the right empirical strategy involves discussions on the optimal sample (Durham, 2001), the non-linearity of the Phillips curve (Filardo, 1998) or on the correct estimation of the sacrifice ratio (Cecchetti and Rich, 2001). Andersen and Wascher (1999, 1) highlight that "one should be cautious about drawing strong implications for monetary policy from these kind of estimates". Methodological concerns are also raised by Baltensperger and Kugler (2000).

[4] Another reason could be habits. Collard et al. (2007) show that inflation persistence can be explained with deep habits.

[5] A wide range of theoretical literature discusses the necessary ingredients of New Keynesian models to create realistic disinflationary paths. Several studies discuss different models including Calvo and Rotemberg price mechanisms which yield different output scenarios ranging from boom to recession (Ascari and Rossi, 2011; Blanchard and Galí, 2007; Ascari and Merkl, 2009).

Fourth, the assumption of perfect rationality can be relaxed in another dimension namely with respect to decision making. Subjects may fail at deleting all dominated strategies or may assume others to fail at this elimination. That opens up the question which regime is better at coping with limited rationality. Achieving higher levels of rationality may be endogenously driven because "inflation persistence is not an inherent characteristic of the economy but rather varies with the [...] monetary policy regime" (Erceg and Levin, 2003, 916). Cold Turkey could be seen as "teaching by doing" (King, 1996, 35) as it increases the speed of learning to play the new equilibrium (Schaling and Hoeberichts, 2010; Cogley et al., 2011).

In the experiment, any rigidities like staggered contracts or menu costs are excluded because subjects can adjust their decision without any costs. Limited knowledge is excluded in the experiment, too, as the framework builds on a simple complete information structure. Therefore, Cold Turkey may be easily viable. The only obstacle to a successful implementation may arise due to limited reasoning of the forecasters. Here, my experiment covers new ground in identifying the influence of limited reasoning on disinflation and in determining whether Cold Turkey may help to enhance reasoning. The second novelty is that the credible implementation of a disinflation policy can be disentangled from ex-ante credibility because the latter is excluded by design.

3. Previous Experimental Evidence

Laboratory macroeconomic experiments have gained prominence lately, as evidenced in the comprehensive surveys by Ricciuti (2008) and Duffy (2014). Typical experiments in this area implement an economic model in the laboratory but substitute the assumption of a rational representative agent by real subjects who populate the model economy. Cornand and Heinemann (2014) discuss a long list of experiments which explicitly focus on monetary policy and central banking, while Amano et al. (2014) discuss the use as tool in central bank research as done by the Bank of Canada. The following section is divided into the two main building blocks of the subsequent experimental design: full information about the model and the shocks which occur and lab subjects as central bankers.

3.1. Full information about the Model and Shocks

A large number of experiments has focused on limited information. Subjects are not provided with complete information on the quantitative structure of the underlying model. The main challenge in these experiments is to learn to forecast correctly and to learn the parameters of the model. These experiments have thus been labeled Learning to Forecast Experiments (Hommes, 2011). The question that lies at the core of these experiments is whether equilibria can be approached through adaptive learning[6].

In contrast to these studies, the experimental design of my study provides full information about the underlying structure of the economy. Thereby, a closer look can be taken at non-standard decision making and limited reasoning. This is in line with other studies. Lambsdorff et al. (2013) run an experiment on macroeconomic price-setting. They observe that convergence towards equilibrium remains incomplete in a non-stationary environment. Rather than calculating equilibrium values, subjects employ simple heuristics. Giamattei and Lambsdorff (2014) test different regimes of punishing current account imbalances in a similar framework to my study. Akin to my study, subjects are also confronted with a

[6]Another branch of experimental research, which acts under limited information as well, deals with inflation forecasting in complex New Keynesian Dynamic Stochastic General Equilibrium (DSGE) Models (Adam, 2007; Pfajfar and Zakelj, 2009; Assenza et al., 2014a) and the simultaneous determination of prices and quantities in complex economies (Noussair et al., 2014; Petersen, 2011; Davis and Korenok, 2011; Roos and Luhan, 2008, 2013).

monetary shock in the experiment by Fehr and Tyran (2001, 2008). In their experiments, they find evidence for money illusion and the use of anchor-and-adjustment-heuristics which causes a sluggish reaction to the change of the monetary policy circumstances.

Another important strand of research applies the pioneer work of Morris and Shin (2002) to central banking problems. Baeriswyl and Cornand (2014) test the effects of partial publicity - i.e. only giving information to a part of market participants - and partial transparency. They show that both reduce overreactions to public information, but only the first also performs well in the experiment. In these kind of models (which include a similar logic on complementarity as my study), shocks are not known but participants get private and public signals on these shocks. As first discussed by Phelps (1986), the central bank may "face a problem of ensuring that changes in policy become common knowledge" (Stasavage, 2002, 2) to allow agents to coordinate on it. But even if subjects have full information heterogeneous degrees of rationality may hinder common knowledge.

> "When a new disinflationary intention is proclaimed [...] there is no telling what forecasts people will make of the other people's forecasts, *their* forecasts of other' forecasts, and so on ad infinitum (and maybe ad nauseam)." (Phelps, 1986, 36).

Therefore, the experimental design will focus on a simple complementary environment to incorporate these interactions.

3.2. Laboratory Subjects as Central Bankers

In contrast to many experiments which merely involve subjects playing against a computerized central bank, one branch of literature employs experimental subjects in the role of central bankers to model decision making processes in the lab. Blinder and Morgan (2005, 2008) and Lombardelli et al. (2005) were the first to test team decision making and the influence of leadership and the voting process in a monetary policy committee. Engle-Warnick and Turdaliev (2010) test if laboratory central bankers stabilize inflation according to a Taylor rule when they play against a computerized set of equations which characterize the economy. Noussair et al. (2014) implement a whole economy in the lab including three households, firms and central bankers each. Both studies show that subjects are capable of stabilizing inflation with less than optimal weights on inflation. Next to that, subjects show a large preference for interest rate smoothing (although this is not included in their target function). Closely related to my study is the work by Arifovic and Sargent (2003) and Duffy and Heinemann (2014) who study a similar constellation of one central banker and and a group of forecasters. The focus of their studies lies on testing the time-inconsistency hypothesis raised by Kydland and Prescott (1977) and Barro and Gordon (1983) which states that central bankers should be restricted by policy in order to not exploit the positive effect of unexpected inflation. Arifovic and Sargent (2003) cover new ground in an exploratory manner in order to find out which equilibrium prevails. They find evidence that subjects are capable of forecasting inflation quite well and that transitions between the high and low inflation equilibrium are slow. Duffy and Heinemann (2014) test the effects of reputation in contrast to commitment and mechanism to prevent the temptation to cheat on the forecasters.

In difference to these experiments, the experimental design of this study does not involve the distribution of payoffs among players and, therefore, abstracts from any issues of fairness or cooperation. Central bankers in this setting cannot gain from cheating on forecasters' expectation. That allows me to exclude considerations of non-standard or social preferences from the further analysis and focus on non-standard decision making and non-standard beliefs. "The standard approach to disinflation has been to view it as a coordination problem. If wage setters, price setters, and the monetary authority could simply coordinate, then disinflation could proceed at zero cost, leaving all relative prices, including relative wages, unchanged." (Blanchard, 1998, 137). That creates a two-sided coordination problem. First, the private sector has to coordinate itself but also has to coordinate with a central bank's policy. The experimental design, therefore, consists of a pure experimental coordination game with an unique equilibrium which is explained in the following.

4. Experimental Design and Calibration

In the experiment, subjects play in a group of N symmetric inflation forecasters and one subject represents the central bank. Subjects play 12 rounds t of the game, trying to coordinate.

4.1. Forecasters

In each round forecaster n has to decide how to set an integer value $x_{nt} \in [x^{min}; x^{max}]$ in order to come close to a target $\tilde{x}_t$. The target value is given by (1) with x_t being the average value chosen by all forecasters.

$$\tilde{x}_t = p(x_t + i_t) \text{ with } x_t = \frac{1}{N}\sum_{n=1}^{N} x_{nt};\ 0 < p < 1 \tag{1}$$

Hereby i_t is known upfront and announced by the central bank with $i_t \in [i^{min}; i^{max}]$. The forecasters' choices are strategic complements as the own optimal choice positively depends on the choice of other forecasters. p hereby gives the degree of complementarity. The higher p, the higher the influence of the others' choices. Such complementarity can often be found in macroeconomics (Woodford, 2003; Cooper and Haltiwanger, 1996). The forecasters' payoff is given by (2), where the quadratic deviation from the target value is subtracted from a fixed endowment e_n.

$$\Pi_{nt} = e_n - 0.1(x_{nt} - \tilde{x}_t)^2 \tag{2}$$

As in the beauty contest with inner solution (Güth et al., 2002), the Nash equilibrium x_t^* can be calculated by assuming that all players n act in a symmetric way with $x_{nt} = x_t^*\ \forall n$. This gives $x_t = x_t^*$ in order to maximize (2). Then the Nash equilibrium is characterized by (3).

$$x_t^* = p(x_t^* + i_t) \iff x_t^* = \frac{p}{1-p} i_t. \tag{3}$$

This means that playing Nash (conditional on a given i_t) is a proportional rule. Forecasters should take the central banks indicator multiplied by $p/(1-p)$. This equilibrium can also be reached by iterative elimination of (strictly) dominated strategies. The best response function can be calculated by $x_{nt} = \tilde{x}_t$. Denoting x_{-nt} as the average choice of all other players $-n$ and solving (1) yields the best response function (4).

$$\begin{aligned} x_{nt} &= \frac{p}{N}(x_{nt} + (N-1)x_{-nt} + Ni_t) \\ &= \frac{p(N-1)}{N-p} x_{-nt} + \frac{Np}{N-p} i_t \end{aligned} \tag{4}$$

For example, assume an indicator $i_t = 10$ and that all other players (level-0) choose 100 on average. With $N = 4$ and $p = 2/3$ the best response function (4) can be expressed as $x_{nt} = 0.6x_{-nt} + 0.8i_t$. If a forecaster performs one level of reasoning he should set $x_{nt} = 0.6*100+0.8*10 = 68$ (level-1 player). Then level-2 players react on that by assuming all other players are level-1 and best responding with $x_{nt} = 0.6*68+0.8*10 = 48.8$. This continues until level-∞ players set $x_{nt} = 2i_t = 20$.

The role of forecasters can also be thought of as firms in a market with monopolistic competition. x_{nt} can be seen as a firm's price change, which is influenced by all other price changes x_{-nt}. The central bank is assumed to directly control the change for a share of prices i_t (e.g. import prices) which also enter the firm's optimization problem. A simplified model yields the target function (1) and can be found in Appendix A.1.

4.2. Central Bank

In the experiment, the central bank has the task to set its policy instrument i_t which can be thought of as any price changes which are directly controlled by the central bank and which then influence inflation directly[7] like import prices or prices in sectors where firms are very dependent of central bank financing. To implement the need for disinflation within the experiment, i_t was fixed to its maximum value for the first third of the game $i_t = i^{max} \;\; \forall t \leq 4$. From the fifth round on, the central bank could set i_t within its normal range $i_t \in [i^{min}; i^{max}]$.

Central bankers get a fixed endowment e which constitutes the central bank payoff function along with two deductions as shown in (5).

$$\Pi_t = e - i_t - 0.5(x_t - \tilde{x}_t)^2 \tag{5}$$

This function is closely related to the standard central bank loss function with two objectives (Barro and Gordon, 1983). First, the central bank tries to minimize inflation represented by the indicator i_t, which in equilibrium also represents the average inflation rate because $\tilde{x}_t = 2i_t$. Experimental central bankers focus on an operational target (the indicator) which then influences the policy target of low price changes (target values). This simplification enhances the salience of the disinflation goal and helps the central banker as it directly provides the right tool to lower the target without having to think about the influence of the indicator on the forecasters' choices. The deduction $-i_t$ is close to the normal loss functions assumed for central bankers (Blinder, 1997, 4) but differs in the way that deviations are not squared. This simplification was undertaken so that laboratory subjects did not have to deal with two quadratic functions. Another difference is that the deviations are not punished symmetrically, i.e. that the central bank is not punished for negative indicators. But as the range of i_t is restricted to i^{min} this was not necessary and avoided for reasons of simplification.

The second goal is to minimize forecasters' average deviations from the target value[8], which is identical to minimizing their average deviation from the equilibrium. This goal of bringing the average forecasters close to equilibrium can be understood as an output-target of the normal central bank loss function. Appendix A.2 shows that bringing subjects close to equilibrium relates to closing the output gap.

The maximization problem for all rounds t of the central bank can then be written as (6).

$$\Pi = \max_{i_t \forall t > 4} \sum_{t=1}^{T} e - i_t - 0.5(x_t - \tilde{x}_t)^2 \tag{6}$$

If all forecasters are perfect Nash players, then they perfectly hit the target value and $x_t = \tilde{x}_t = x_t^*$. Therefore the second term of the central bank's payoff function $(x_t - \tilde{x}_t)$ is

[7] As noted by Blackburn and Christensen (1989, 11), the central bank directly setting the inflation rate "is clearly a heroic assumption but, for the most part, serves as a useful abstraction." It is also implemented in related studies (Arifovic and Sargent, 2003; Duffy and Heinemann, 2014). The assumption in my study is less demanding because the central bank only determines part of the price changes and not the overall inflation rate.

[8] The only difference to the term in the forecasters' payoff function (2) is that central bankers are only concerned about the average overall deviation and are not punished for the variance or extreme choices.

always equal to 0. This reduces (6) to $\max_{i_t} \sum_{t=1}^{T} e_n - i_t$, which yields $i_t^* = i^{min}$. That means the central bank sets i_t to its minimum and all subjects adjust immediately by setting (7). Therefore, the Nash equilibrium is to determine inflation according to Cold Turkey.

$$x_t^* = \frac{p}{1-p} i^{min} \tag{7}$$

4.3. Calibration

Three pilots were used to calibrate the experiment. The range for x_{nt} was set to $x_{nt} \in [0;150]$, while i_t was restricted to $i_t \in [10;50]$. The lower bound was not set to zero in order to avoid the change of complexity with a boundary instead of an inner solution[9]. The number of forecasters was set to $N = 4$ in order to have a large degree of heterogeneity within the laboratory capacities[10]. The p-value was chosen to be $p = 2/3$ in order to introduce enough complementarity between the forecasters' decisions and to allow for convergence within the number of rounds. The endowments were set to $e = 60$ and $e_n = 40$ to meet the salary targets of the laboratory and to account for the fact that the central bankers were subject to higher deductions. The experiment consisted of three repetitions of the 12 rounds, which were named lives in the experiment. Each life was identical, with i_t being fixed in the first four rounds to i^{max}. Lives were clearly separated and each life started with an initial payoff of zero. These lives made it possible for subjects to learn. After each life, a new central banker was matched to a group of forecasters (absolute stranger rematch). Table 1 summarizes the parameter values.

x_{nt}	i_t	N	p	e	e_n	lives
$\in [0;150]$	$\in [10;50]$	4	2/3	60	40	3

Table 1: Parameter Values

5. Hypotheses

Hypotheses can be derived from the experimental design. According to that, forecasters play the Nash equilibrium conditional on a given value for the indicator, i.e. they set $x_{nt} = x_t^* = 2i_t$.

Hypothesis 1 *Forecasters play the Nash equilibrium conditional on a given value for the indicator* $x_{nt} = x_t^* = 2i_t$.

Contrary to the hypothesis, they may fail in reacting rationally to the central bank's announcement of i_t. As evidenced from a large body of literature, limited reasoning may play a role as subjects fail to iteratively delete all dominated strategies or think others will fail to do so (Camerer, 2003, 199-255). Subjects do not react completely to a change in i_t, but only in a smoothed way and stick to past values. Another reason would be the anchor-and-adjustment heuristic used in Fehr and Tyran (2001, 2008). Subjects may remain anchored by the previous equilibrium of the first four rounds ($i_t^* = 100\ \forall t \leq 4$) and fail in sufficiently adjusting to the policy change. An additional anchor would lie in the range of possible choices between 0 and 150 with an average of 75. I therefore expect hypothesis 1 to be rejected.

[9]This can also be thought of a central bank which has a positive inflation target like the ECB.
[10]This is well in line with Arifovic and Sargent (2003) and "reflects the fact that the private sector is considerably larger than the government sector" (Duffy and Heinemann, 2014, 13).

This opens the question on the degree of rationality of the forecasters. If forecasters do not play conditional Nash this turns to an empirical question whether Cold Turkey may be still appropriate. Even with some limits on rationality, Cold Turkey could still be optimal. Deviations from equilibrium in the beginning can be balanced out by a lower indicator in the end if Cold Turkey helps to bring down expectations. Additionally, it may provide better guidance in learning. "A gradual disinflation might retard the rate at which private agents learn, thus prolonging the transition and making it more turbulent" (Cogley et al., 2011, 1). This idea comes from considerations with limited knowledge (Cogley et al., 2011; Schaling and Hoeberichts, 2010; Kurz et al., 2013), but may translate to limited reasoning. If forecasters are anchored by past values, Cold Turkey could help in destroying this old anchor. But this process is bought with high errors in the beginning. Correcting for these errors allows forecasters to learn from them and to advance to equilibrium faster. Similar evidence comes from Learning-To-Forecast Experiments (Assenza et al., 2014b) which call for aggressive monetary policy to stabilize the economy. Errors may help the forecasters in learning the reasoning process because reasoning failures are punished more heavily. This relates to the literature on the strategic nature of the environment. Sutan and Willinger (2009) and Fehr and Tyran (2008) show that with substitutes instead of complements the equilibrium is approached much faster because the alternating reasoning process with substitutes yields more useful information. The same may hold true if a change is highly visible. Then errors are easier to detect and correct and reasoning is increased. This goes in line with arguments by Akerlof et al. (2000) who state that changes have to be detectable to induce changes in behavior.

To illustrate limited reasoning, I use a simplified level-k model (Nagel, 1995; Ho et al., 1998; Güth et al., 2002)[11]. Level-0 players are assumed to take the past target value[12] as their choice (Duffy and Nagel, 1997, 1694). Level-1 players then decide on their value by thinking that all others are level-0. Level-2 players think all others are level-1 and so forth.

As the central bank is only concerned with the aggregate view it is sufficient to consider only the overall reasoning. With a level-k model, the average value of all forecasters always lies between the choice of level-0 players $\tilde{x}_{t-1}$ and the conditional Nash. Therefore, the average value can be represented as the weighted sum of level-0 and the Nash solution and all possible levels of reasoning $k \in [0; \infty]$ can be standardized between 0 and 1. This is shown in (8). The degree of reasoning is denoted by $\varrho \in [0; 1]$ With $\varrho = 0$ all forecasters are level-0 and choose the previous target value. With $\varrho = 1$ all play according to Nash.

$$x_t = (1 - \varrho)\tilde{x}_{t-1} \quad + \varrho x_t^* \tag{8}$$

With this measure of reasoning, the hypothesis is that reasoning may be higher with Cold Turkey ϱ^{CT} than with a gradualist approach ϱ^{G}.

Hypothesis 2 *With high initial errors, Cold Turkey helps in correcting past errors and in increasing the overall reasoning level, so that $\varrho^{CT} > \varrho^{G}$.*

Central bankers may be tempted to overestimate the forecasters' capabilities. But as central bankers can use the first four rounds of every life for observing the forecasters'

[11] Players think that they are able to perform one step of reasoning more than all other players. As Ho et al. (1998, 961) note, everybody's assumption of being one level smarter is "logically impossible, [but] consistent with a large body of psychological evidence showing widespread overconfidence about relative ability". This idea may hold true for forecasters among each other but also for a central banker who thinks of bein one level smarter than forecasters.

[12] An additional assumption for the first round is that level-0 players choose randomly from the whole interval. The choice of level-0 players is normally assumed to be the midpoint of the choice interval (Nagel, 1995). Instead of the past target value, Ho et al. (1998) use an weighted average of previous target values instead.

behavior, this overestimation may be corrected after the first four rounds so that the policy choices are based on correct expectations. Hypothesis 3 refers to this.

Hypothesis 3 *Central bankers do not overestimate the forecasters' degree of reasoning in round* $5-12$.

Taken all previous hypotheses together this suggests a higher performance by implementing Cold Turkey. In round 5, payoffs may be lower than with a gradual regime. But these losses are compensated by higher payoffs in the later rounds because then expectations are well in line with a low indicator. To test this, cumulative payoffs can be assessed because they may be lower in round 5 but rise over the level of the gradual regimes after some rounds. Therefore, in the end, total payoffs may be higher which is stated in hypothesis 4.

Hypothesis 4 *Cold Turkey yields higher total payoffs because it increases coordination and reasoning.*

But with subjects adjusting only sluggishly, implementing Cold Turkey is quite costly in the beginning. That raises doubts about the regime's credible lead through. A central banker who observes that forecasters do not follow her regime may be tempted to reverse her strategy. This may, however, be compensated by correct expectations (hypothesis 3) and faster learning by forecasters (hypothesis 2), so that Cold Turkey can be pushed through.

Hypothesis 5 *If Cold Turkey is chosen, it is pushed through in all rounds.*

6. Experimental Procedures

The experiment was conducted computer-based at the classEx Laboratory at the University of Passau using z-Tree (Fischbacher, 2007) and Orsee (Greiner, 2004). All written and oral instructions can be found in Appendix B. Upon arrival, subjects were randomly seated in the laboratory and publicly instructed about the purpose of the game, its expected length, the conversion of the experimental currency unit Taler ($\mathbb{T}$) into Euro, dos and dont's, the use of a pocket calculator and about (standard) payment and blindness procedures.

In order to increase overall understanding of the rules, the first screens explained the game in a detailed manner for both roles. Subjects had to complete an incentivized test of comprehension, where they could earn 2 $\mathbb{T}$ per question. At the end of the instructions, they were assigned to the role of the central bankers (B) or one of the four forecasters (A). The instructions were framed completely neutral. i_t was named an indicator, while participants A had to simply set a number.

The first four rounds in each life were reserved for learning, thus payoffs in these rounds were hypothetical. Actual payoffs were earned in the following 8 rounds of each life. Additionally to the choice of a number x_{nt}, forecasters were asked after their decision what they think all $N-1$ other forecasters will choose on average. If their belief was right ($\pm 5\%$) they earned 2 $\mathbb{T}$. Also, while forecasters were deciding on their numbers, the central bank's beliefs about the average guess of all N forecasters was elicited in this way.

After the experiment, subjects participated in a cognitive reflection test (Frederick, 2005) and had to answer some questions from an IQ test as well as filling in a questionnaire on demographic variables. The questionnaire can be found in Appendix B.3. At the end of the experiment, one life was randomly drawn to determine the actual payoff. This payoff was converted at an exchange rate of 1 $\mathbb{T} = 10$ Eurocent and paid out to the participants at the end of the experiment by a third person.

The experiment was conducted in two sessions with 20 participants each on May 28$^{\text{th}}$ 2014. Each subject participated only once. Subjects required 25 minutes for the instructions and the test of comprehension and between 15 and 25 minutes for each life. At the end,

Type	Female	Age	Payoff Comprehension Questions [20]	Payoff IQ Test [20]	Cognitive Reflection Test [3]
Central Bankers	0.75	22.38	**19.00**	9.00	0.375
	(0.463)	(1.69)	(2.83)	(3.02)	(0.52)
Forecasters	0.75	22.38	**17.06**	7.63	0.72
	(0.440)	(3.51)	(3.33)	(4.47)	(0.89)

Standard deviation in parentheses, Mann-Whitney-U Test for differences, number in square brackets in the first row denote maximum possible value, **bold** values indicate significant differences at a 5% level.

Table 2: Selection of Central Bankers and Forecasters

the IQ test and the questionnaire lasted another 15 minutes, which sums up to a total time of about $100 - 120$ minutes. Total payoffs to the participants amounted to 882.40 euros and 22.06 euros per person on average with a minimum of 5 euros and a maximum of 37.40 euros. Additionally 336 euros were paid out for the pretests. 75% of the participants were female (as in the whole subject pool) with an average age of 22.25 ranging from 18 to 36. All subjects were students from a broad range of courses (economics, law, cultural studies and others). As the roles of forecasters were randomly assigned, they should not differ with respect to intelligence and demographics. Table 2 reports the results of this robustness check also with respect to the payoff gained through the comprehension questions and the IQ test. Differences are not significant for any characteristics apart from the payoff for the comprehension questions, where central bankers performed slightly better. The results of the IQ test and the cognitive reflection test show no significant differences. Including these variables as controls in the following regressions does not change the result. They are therefore excluded from further considerations.

7. Results and Regressions

Due to the fact that treatments were chosen endogenously, they have to be identified in the data. The minimum indicator change was used as proxy if a central banker implemented Cold Turkey. It is defined as $\min(i_t - i_{t-1})\forall t$ and ranges between 0 and -40. Central bankers were classified as Cold Turkey with $\min(i_t - i_{t-1}) < -20$, i.e. if the minimum indicator change was smaller than -20, the midpoint of all possible indicator changes. According to that criterion, 46% applied a gradual strategy as can be seen in table 3. While in the first life 88% implemented Cold Turkey, this dropped to 38% in the second and third life. To test if the regime choice is an exogenous variation, the choices of the forecasters in the rounds $1-4$ are compared by treatments. A Mann-Whitney U-Test yields no significant difference for the chosen number by forecasters ($z = -1.501, p = 0.1334$). The same holds true for the central bankers' expectations about a group's decision in rounds $1 - 4$ ($z = -0.705, p = 0.4807$). That means that central bankers did not opt for a policy

Treatment	Criterion	All	Life 1	Life 2	Life 3
Gradualism	$\min(i_t - i_{t-1}) \geq -20$	46%	12%	62%	62%
Cold Turkey	$\min(i_t - i_{t-1}) < -20$	54%	88%	38%	38%

Table 3: Treatments

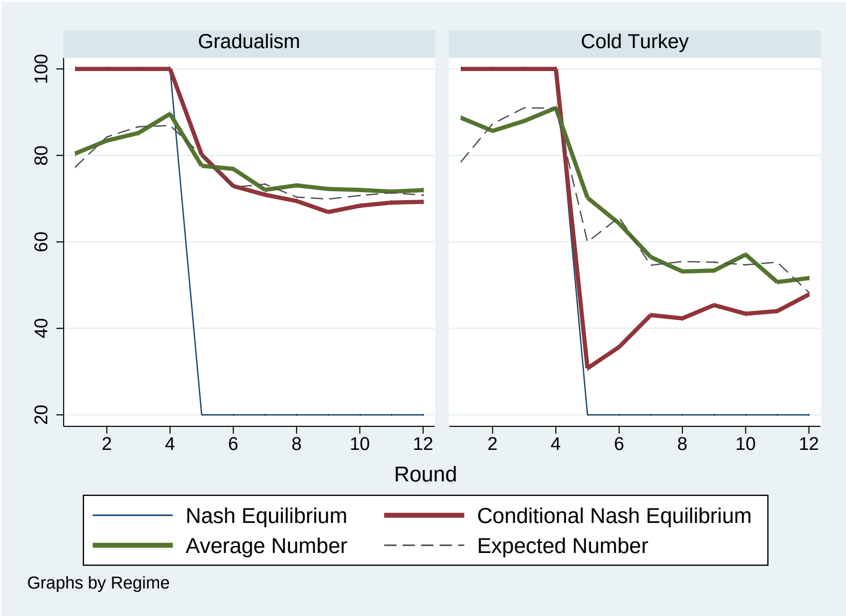

Figure 1: Nash Equilibrium and Choice of Numbers

regime due to different observations or expectations in the first four rounds and treatment effects are not driven by prior differences.

Figure 1 provides a first grasp on the results. For each treatment it shows the Nash equilibrium of playing Cold Turkey (blue thin line). With both forecasters and central bankers being perfectly rational, the forecasters' optimal choice is a number of 100 in the first four rounds and then jumping to 20 for the rounds $8-12$. The red line shows the conditional Nash equilibrium. After the central banker chose her value for i_t, forecasters should choose their number according to $2i_t$. The green line displays the mean number chosen by all forecasters. As can be seen clearly, forecasters deviate substantially from the Nash prediction when confronted with a Cold Turkey regime while they coordinate better in a more gradual regime. Figure 1 also evidences that Cold Turkey is not sustained as indicator choices go up again from round 6 on. The thin dotted line plots the central bankers' expectations, which are well in line with the actual average value most of the time. Only in round 5, Cold Turkey central bankers overestimate the capabilities of their forecasters and correct for that in the subsequent round. It has to be noticed that these central bankers did not expect their forecasters to perfectly jump to the conditional Nash equilibrium, but they expected that forecasters would go only half the way. This yields some evidence that Cold Turkey was not chosen because central bankers expected perfect rational forecasters but to force inertial forecasters to adjust downwards. The first test examines whether forecasters are capable of approaching the conditional Nash equilibrium. This is done by estimating the equilibrium relationship (3) with equation (9).

$$x_{nt} = \beta_1 i_t + \beta_2 + \epsilon_{nt} \tag{9}$$

As derived in (3), the conditional Nash equilibrium would predict a weight of $\beta_1 = 2$ for the indicator and a constant of zero $\beta_2 = 0$. Table 4 shows regression results for the forecasters' behavior. Subjects do not play the conditional Nash equilibrium. Their choice of numbers shows some persistence. The indicator coefficient of 1.276 is significantly different from the equilibrium prediction ($F = 36.92$ $p = 0.000$) and the constant is not zero (see (1)

Number x_{nt}		(1)	(2)	(3)
Rounds 5-12	Equilibrium	All	Gradualism	Cold Turkey
Indicator	2	1.276***	**1.624*****	**1.180*****
		(17.30)	(29.34)	(12.07)
Constant	0	29.48***	**15.85*****	**32.57*****
		(11.18)	(7.59)	(10.24)
Observations		768	352	416
R^2		0.501	0.667	0.341
ADF on residuals		-11.10	-7.93	-8.67

OLS, t statistics in parentheses, clustered by forecaster and group.
Bold values indicate significant differences between (2) and (3) at a 5% level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression for Forecasters' Decision (Nash)

in table 4). The significant constant provides some evidence for the anchor-and-adjustment heuristic. Some players may be anchored by the midpoint of the range or the equilibrium in the first four rounds and do not react to changes in the indicator sufficiently. Subject fail to play the conditional Nash equilibrium in both treatments, while they perform slightly better for the Gradualism treatment. The coefficients are significantly different between treatments. Still, even with a gradual approach, they fail in playing the conditional Nash equilibrium because the coefficient of 1.624 is significantly different from 2 ($F = 46.15$ $p = 0.000$). This allows me to reject hypothesis 1.

Result 1 *Subjects fail to play the Nash equilibrium $x_{nt} = 2i_t$ conditional on a given i_t.*

Given that subjects deviate from Nash in both treatments, it has to be checked which treatment facilitates the correction of past errors to come closer to a long-run equilibrium relationship. To check if coefficients are estimated consistently, first x_{nt} and i_t have to be tested on co-integration. An Augmented Dicky Fuller (ADF) test on the residuals ϵ_{nt} of estimation (9) is performed by estimating (10) where $\Delta\epsilon_{nt} = \epsilon_{nt} - \epsilon_{nt-1}$ and $\Delta\epsilon_{nt-1} = \epsilon_{nt-1} - \epsilon_{nt-2}$.

$$\Delta\epsilon_{nt} = \gamma_1 + \gamma_2\epsilon_{nt-1} + \gamma_3\Delta\epsilon_{nt-1} + \nu_{nt} \tag{10}$$

If γ_2 is significant, residuals fade out over time, which is the condition for a co-integration relationship. The last row of table 4 reports the t-statistics for the coefficient γ_2, which is well below the critical McKinnon value of -3.50 for a 1% error level. This allows to infer that residuals have a unit root and all regressions in table 4 denote a co-integrating relationship. The residuals can therefore be used to estimate an error correction model (ECM)[13] given by (11).

$$\Delta x_{nt} = \phi\Delta i_t + \theta\epsilon_{nt-1} + \nu_{nt} \tag{11}$$

ϵ_{nt-1} is the lagged residual from regression (9). The coefficient ϕ shows the short-term reaction to changes in the indicator and should be equal to 2 in equilibrium. The

[13]The standard formulation of the error correction model would be $\Delta x_{nt} = \phi\Delta i_t + \gamma(x_{nt-1} - \beta_1 - \beta_2 i_{t-1})$ where the last term in brackets denotes the lagged long-run relationship. This bracket can be substituted by the lagged residual ϵ_{t-1} as the main interest is only on the coefficient γ. For a description of the error correction model see e.g. Verbeek (2004, 318-319).

Number Difference Δx_{nt} Rounds 5-12	(1) Gradualism	(2) Cold Turkey
ϕ Indicator Difference Δi_t	**1.463*****	**0.570*****
	(6.05)	(4.98)
γ Lagged Residuals ϵ_{t-1}	*-0.944****	*-0.611****
	(-6.80)	(-7.45)
Observations	352	416
R^2	0.512	0.406

OLS, t statistics in parentheses, Critical McKinnon values. Clustered by forecaster and group. *Italic* values indicate significant differences at a 10% level, **bold** values indicate significant differences at a 5% level between coefficients in (1) and (2).
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Error Correction Model

most interesting coefficient for the hypothesis is θ. The coefficient θ denotes the speed of adjustment. When $|\theta|$ is close to 1, this indicates a fast adjustment to the long-run relation as past errors are corrected immediately. Lower absolute values show that past errors are persistent and corrected only sluggishly. θ can, therefore, also be interpreted as the speed of learning. Table 5 shows the results for both treatments separately. The short-run reaction ϕ is again characterized by an under-evaluation of the indicator changes as both coefficients are below 2, with gradualism performing better. More interestingly, Cold Turkey does not perform better with respect to the correction of past errors. The absolute value of the coefficient θ is lower for Cold Turkey $|\theta^{CT}|$, but only at a 10 % error level. That means it is at least equal or lower than in the Gradualism treatment $|\theta^{G}|$ but not higher as expected.

This result can be further confirmed by looking at the overall degree of reasoning which can also be interpreted as the convergence towards the Nash equilibrium. The degree of overall reasoning ϱ, which was defined in (8) in section 5, can be directly estimated. Equation (8) has to be rearranged to (12) in order to be estimated[14], with ϱ being the estimator for the overall degree of reasoning. β_2 is included because forecasters' choices may also depend on a constant value as seen in table 4.

$$x_t = \beta_1 \tilde{x}_{t-1} + \varrho(x_t^* - \tilde{x}_{t-1}) + \beta_2 + \epsilon_t \tag{12}$$

The upper part of table 6 shows the results of the regression for the forecasters' value. The estimated degree of reasoning is given by a coefficient of 0.298 for all observations. The coefficient means that subjects only move about 30% towards the Nash equilibrium and remain firmly anchored by past target values and do not adjust sufficiently to a policy change. The value of $\varrho = 0.298$ is comparable to observed steps of reasoning in other experiments[15]. Performing the regressions separately for lives or for the first and second

[14]This model is closely related to a normal error correction model of the type $x_{nt} = \gamma_1 i_t + \gamma_2 i_{t-1} + \gamma_3 x_{t-1} + \gamma_4 + \epsilon_t$. This can be transformed into (12). First, as $x_t^* = 2i_t$ the coefficient γ_1 can be expressed as $\gamma_1 = 2\varrho$. Second, by assuming that $\gamma_2 = \gamma_3$, the second and the third term of the error correction model can be rearranged to $\gamma_2 i_{t-1} + \gamma_3 x_{t-1} = \gamma_2(i_{t-1} + x_{t-1}) = \gamma_2 \tilde{x}_{t-1}/p$. It means that the lagged values follow the functional form of the target function. With setting $\beta_1 - \varrho = \gamma_2/p$, the models are identical. Therefore, the only additional restriction in (12) is that $\gamma_2 = \gamma_3$. This relationship can be confirmed by a Wald test ($F = 0.06$ $p = 0.801$) on the equality of $\gamma_2 = \gamma_3$, which shows that this additional restriction should pose no problem for the estimation.

[15] For example, the empirical distribution in Nagel (1995, 1322) with $p = 2/3$ yields a degree of reasoning

	Forecasters			
Avg. Number x_t	(1)	(2)	(3)	(4)
Rounds 5-12	All	Gradualism	Cold Turkey	Round 5-7
β_1 Lagged	0.880***	0.874***	0.934***	0.685***
Target Value $\tilde{x}_{t-1}$	(23.97)	(18.98)	(19.81)	(9.21)
ϱ Nash - Lagged	0.298**	**0.538****	0.323**	*0.212**
Target Value $\tilde{x}_{t-1}$	(3.27)	(3.52)	(3.18)	(2.56)
β_2 Constant	10.02**	10.01*	8.644*	22.22***
	(3.48)	(2.69)	(2.48)	(4.38)
Observations	192	88	104	72
R^2	0.823	0.875	0.770	0.609

	Central Bankers			
Exp. Avg. Number x_t^e	(1)	(2)	(3)	(4)
Rounds 5-12	All	Gradualism	Cold Turkey	Round 5-7
β_1 Lagged	0.875***	0.921***	0.920***	0.675***
Target Value $\tilde{x}_{t-1}$	(21.87)	(24.44)	(14.28)	(8.42)
ϱ^e Nash - Lagged	0.412***	**0.721*****	0.437***	*0.402****
Target Value $\tilde{x}_{t-1}$	(5.58)	(8.93)	(5.88)	(3.87)
β_2 Constant	10.58**	6.234*	10.29*	24.93***
	(3.77)	(2.28)	(2.68)	(4.55)
Observations	192	88	104	72
R^2	0.741	0.957	0.611	0.473

OLS, t statistics in parentheses, clustered by group (and lives in (1) and (5)).
Bold values indicate significant at a 5% level, *italic* at a 10% level between ϱ and ϱ^e.
$^+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Regression for Forecasters' Decision (Level k)

half of a life shows that reasoning does not vary across lives or between rounds 5-8 and 9-12. Interestingly, the coefficients are not significantly different between treatments, which allows to reject hypothesis 2. Cold Turkey does not lead to higher reasoning. The hypothesis that $\varrho^{CT} > \varrho^{G}$ has to be rejected.

Result 2 *Learning is not faster under Cold Turkey. Forecasters only move about 30% towards the Nash equilibrium. This result is identical across treatments.*

This raises an interesting follow-up question on the central bankers' expectations regarding the forecasters' rationality. They are of interest because they show if central bankers base their regime choice upon them. These expectations are depicted in figure 1 as a plot of the mean expectations of central bankers. At the policy switch in round 5, the central bank overestimates the capabilities of the forecasters to perform the downward jump in the Cold Turkey treatment. The central bankers start to correct their wrong expectations. The estimation error can be further analyzed by looking at the lower part of table 6. It shows the expected degree of reasoning ϱ^e. The central bankers' expectations about the average value of the forecasters yield a prediction for the expected degree of reasoning. Mean predictions show that central bankers tend to overestimate the forecasters' capability but this turns out to be only significant in the Gradualism treatment. It can be seen that central bankers' estimations are higher than actual reasoning on a 10% level when looking only at round $5-7$ as in column (5) of table 6. This gives some further evidence on overestimation and hypothesis 3 can be rejected, but only at a low significance level. It has to be noticed that this overestimation can not be made responsible for the choice of Cold Turkey as gradual central bankers are also affected by overestimation.

Result 3 *Central bankers tend to overestimate the forecasters' degree of reasoning, so that $\varrho^e > \varrho$ in rounds $5-7$ and in the Gradualism treatment.*

But if Cold Turkey does not increase learning and reasoning and central bankers tend to overestimate forecasters' capabilities, it can be doubted that the high losses with Cold Turkey in round 5 are compensated in later rounds. Figure 2 plots the cumulative payoffs for both regimes. The dotted lines show the average cumulative payoff for central bankers only, while the solid lines depict the average cumulative payoff for both forecasters and central bankers. The latter provides some measure for a welfare comparison between the treatments. The welfare is captured by the two central bank objectives of a low indicator and output gap. But the latter objectives only focus on the deviation of the average forecasters. Therefore, the payoff losses due to deviations of the single forecasters have to be taken into account, too.

If initial Cold Turkey losses are compensated by higher payoffs in later rounds, the cumulative payoff line for Cold Turkey is expected to intersect the line for Gradualism after some rounds. But the cumulative payoffs are always higher with a gradual regime. While in the first rounds after the policy choice the expected pattern can be seen with initial losses for Cold Turkey, payoffs fail to increase at a higher rate for Cold Turkey. That is why they remain below the payoffs of the gradual regime for all rounds with lower total payoffs in round 12.

This pattern is tested in table 7, where the cumulative payoffs are estimated based on the round and the regime. This is done only for central bankers in regressions (1) and (3) and for both central bankers and forecasters in (2) and (4). With Cold Turkey cumulative

ϱ between 0.3 and 0.5. The study of Sutan and Willinger (2009) with $p = 2/3$ and $i_t = 30$ is even closer to my experimental design. Their level distribution (Sutan and Willinger, 2009, 1128) translates into $\varrho = 0.3$. Another option would be to assume a Poisson distribution of levels $f(k) = e^{-\tau}/k!$ (as done in Camerer et al. (2004)) with the empirically observed $\tau = 1.61$, which yields $\varrho = 0.41$.

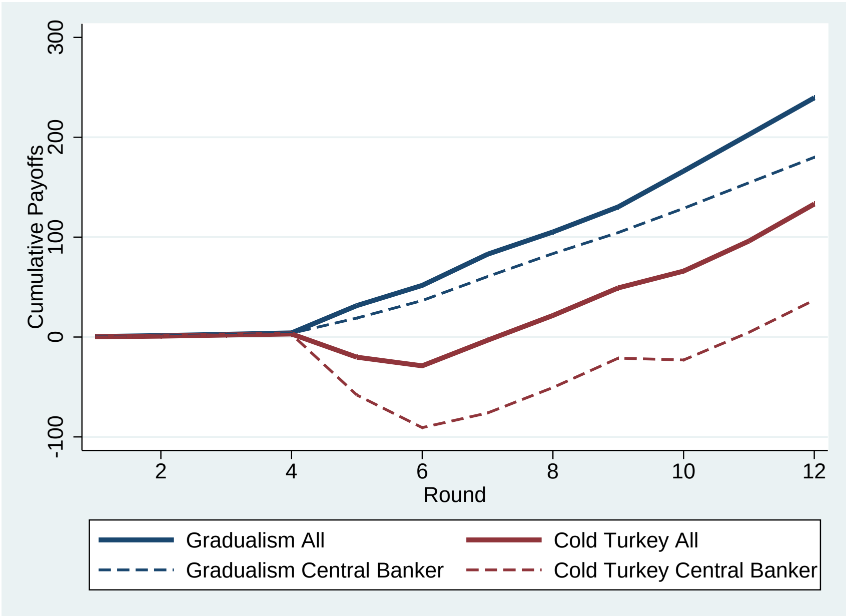

Figure 2: Cumulative Payoffs

Cumulative Payoff	(1) Rounds 5-12 Central Bankers	(2) Rounds 5-12 All	(3) Round 12 Central Bankers	(4) Round 12 All
Cold Turkey	-130.6*	-87.02**	-142.9	-106.3*
	(-2.28)	(-2.91)	(-1.61)	(-2.43)
Round	19.18**	26.03***		
	(3.38)	(8.79)		
Constant	-67.11	-95.00***	180.0***	239.5***
	(-1.41)	(-4.50)	(8.94)	(9.92)
Observations	192	960	24	120
R^2	0.190	0.132	0.091	0.044

OLS, t statistics in parentheses, clustered by central banker and life.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Cumulative Payoffs Across Treatments

Positive Indicator Changes, Rounds 5-12	(1) All	(2) All	(3) Gradualism	(4) Cold Turkey
Cold Turkey	10.17*	5.481*		
	(2.72)	(2.28)		
Lagged Indicator i_{t-1}		0.136	**-0.537****	**0.215**
		(0.63)	(-6.49)	(1.01)
Lagged Deviation $x_{t-1} - \tilde{x}_{t-1}$		1.097***	**0.0983**	**1.211*****
		(4.76)	(1.24)	(6.23)
Constant	3.714***	-4.390	**17.18****	**-1.526**
	(7.93)	(-0.59)	(6.08)	(-0.20)
Sigma	10.66***	7.175***	0.954*	8.160***
	(4.27)	(4.46)	(3.45)	(4.42)
Observations	25	25	7	18
R^2	0.022	0.124	0.332	0.108

Tobit with lower limit 0, t statistics in parentheses, clustered by central banker.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **Bold** values indicate significant differences at a 1% level between the coefficients in (3) and (4).

Table 8: Positive Indicator Changes Across Treatments

payoffs are significantly lower when looking at rounds $5 - 12$ (see (1) and (2)). By looking only at the last round in (3) and (4), this turns out to be only significant when looking at all players, but not when merely looking at the central bank. This is due to the fact that Cold Turkey does not only lower the central bankers' payoff but also causes higher heterogeneity among forecasters with larger individual deviations.

Result 4 *Cold Turkey does not yield higher total payoffs.*

Since central bankers may notice that Cold Turkey yields lower payoffs, they may not push through their strategy. As the policy was clearly incentivized to be disinflating, upward changes in the indicator can be used as proxies for a credible lead through as they signal a U-turn in the policy. As can be seen qualitatively from figure 1, these turns are more pronounced on average in the Cold Turkey treatment. To confirm this statistically, table 8 shows the result of a Tobit regression on the positive indicator changes ($i_t - i_{t-1} > 0$). They are significantly higher at a 5% level for the Cold Turkey treatment which can be seen from the first coefficient in (1) and (2). Controlling for the lagged indicator and the lagged deviation of the average value x_{t-1} from the target value $\tilde{x}_{t-1}$ does not change the results. Cold Turkey is reversed more often. Ex-post, a low indicator in round 5 does not turn out to have been a credible promise with respect to the policy pursued. Looking at the causes of positive indicator changes separately for the treatments yields some further insights (see regression (3) and (4)). In the Gradualism treatment the choice to increase the indicator depends on the lagged indicator. This suggests some continuation in policy. An upward movement is performed if the indicator is too low. In contrast thereto, the positive indicator changes in the Cold Turkey treatment do not depend on the lagged indicator but are the higher, the larger the deviations of the average from the target value.

Result 5 *Cold Turkey is not pushed through but reversed more often.*

8. Conclusion

I investigated whether the strategy of Cold Turkey is more credible than a gradual one and superior in a laboratory setting, where forecasters had to coordinate on a central banker's policy choice of setting an indicator. After four rounds with a high indicator and, therefore, high inflation, central bankers were incentivized to lower the indicator. But they were punished for deviations of forecasters from equilibrium, as these deviations are linked to an output target.

First, subjects react slowly to policy changes despite the absence of rigidities or limited knowledge. Limited reasoning, i.e. the failure to delete all dominated strategies, therefore constitutes another important factor why inflation may be persistent and reducing it is costly. Contrary to the hypotheses, Cold Turkey does not speed up learning or increase the overall degree of reasoning.

Second, the experimental design with endogenous treatments and exogenous matching allowed to disentangle ex-ante credibility from a credible push trough of a regime. With a Cold Turkey regime, central bankers reverse their strategy more often. This provides some support for the quote of Blanchard (1985) in section 1 which stated that failure is less likely for gradual disinflations. Cold Turkey is hard to be pushed through consequently. These two reasons lead to a lower performance of Cold Turkey in terms of payoffs. With Cold Turkey, initial losses are not compensated by higher earning in later rounds and cumulative payoffs remain below the ones of the Gradualism treatment. Furthermore, the reversal impedes successful coordination which leads to lower payoffs for forecasters, too.

To conclude, the issue of external validity has to be addressed. Surely, the results do not quantitatively translate to a real disinflation policy but still may yield some qualitative indications. It is obvious that real-world central banking is a much more complex task. Still, a wide range of studies show that students can successfully compete with e.g. business professionals (Davis and Holt, 1993; Camerer, 2003). The simplified setup in an experiment may put them under similar circumstances like real-world central bankers who have to tackle harder problems, but also have a higher degree of information and support. Such an abstraction seems plausible, because an experimental setup may pose "perplexities that are comparable to those facing real-world central bankers, [...] who [...] know much more, have far more experience, and have abundant staff support" (Blinder and Morgan, 2008, 123). Cornand and Heinemann (2014, 182) provide further arguments for using experimental subjects as central bankers. As the experiment does not allow for habits, staggered prices or wage rigidities as drivers of persistence, Cold Turkey is even more likely to fail in reality if it already fails in this limited environment. Central bankers have to be aware of the inertia of the agents' choice. Heterogeneous degrees of reasoning and a complementary environment are the drivers of this inertial behavior. This experiment provides a first workhorse for a further study of questions regarding the optimal design of disinflation policy. Future experiments may test the credibility of Cold Turkey with pre-announcement or different selection mechanisms for central bankers.

Acknowledgements

The author is grateful to Johann Graf Lambsdorff and his colleagues Susanna Grundmann, Manuel Schubert, Katharina Werner, Isabelle Riviere and Ann-Kathrin Crede for the support, help in the pretest and a lot of useful suggestions and to the participants of the research seminar of the University of Passau for useful comments.

References

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

Agénor, P.-R., Taylor, M. P., 1993. Analysing Credibility in High-Inflation Countries: A New Approach. The Economic Journal 103 (417), 329–336.

Akerlof, G. A., Dickens, W. T., Perry, G. L., Bewley, T. F., Blinder, A. S., 2000. Near-Rational Wage and Price Setting and the Long-Run Phillips Curve. Brookings Papers on Economic Activity 2000 (1), 1–60.

Aksoy, Y., Orphanides, A., Small, D., Wieland, V., Wilcox, D., 2006. A quantitative exploration of the opportunistic approach to disinflation. Journal of Monetary Economics 53 (8), 1877–1893.

Al-Eyd, A., Karasulu, M., 2008. Ambition versus Gradualism in Disinflation Horizons under Bounded Rationality: The Case of Chile. NIESR Discussion Papers (2008-03).

Amano, R., Kryvtsov, O., Petersen, L., 2014. Recent developments in experimental macroeconomics. Bank of Canada Review 2014 (Autumn), 1–11.

Andersen, P. S., Wascher, W. L., 1999. Sacrifice ratios and the conduct of monetary policy in conditions of low inflation. BIS working paper 82.

Arifovic, J., Sargent, T., 2003. Laboratory Experiments with an Expectational Phillips Curve. In: Altig, D., Smith, B. D. (Eds.), Evolution and procedures in central banking. Cambridge University Press, Cambridge, pp. 23–55.

Ascari, G., Merkl, C., 2009. Real wage rigidities and the cost of disinflations. Journal of Money, Credit and Banking 41 (2–3), 417–435.

Ascari, G., Rossi, L., 2011. Real wage rigidities and disinflation dynamics: Calvo vs. Rotemberg pricing. Economics Letters 110 (2), 126–131.

Assenza, T., Bao, T., Hommes, C., Massaro, D., 2014a. Experiments on expectations in macroeconomics and finance. In: Duffy, J. (Ed.), Experiments in macroeconomics. Vol. 17 of Research in Experimental Economics. Emerald Group Publishing Limited, pp. 11–70.

Assenza, T., Heemeijer, P., Hommes, C., Massaro, D., 2014b. Managing Self-organization of Expectations through Monetary Policy: a Macro Experiment. CeNDEF Working Paper 14 (07).

Baeriswyl, R., Cornand, C., 2014. Reducing Overreaction to Central Banks' Disclosures: Theory and Experiment. Journal of the European Economic Association 12 (4), 1087–1126.

Ball, L., 1994a. Credible Disinflation with Staggered Price-Setting. American Economic Review 84 (1), 282–289.

Ball, L., 1994b. What determines the sacrifice ratio? In: Mankiw, N. G. (Ed.), Monetary policy. The University of Chicago Press, pp. 155–193.

Baltensperger, E., Kugler, P., 2000. Central bank independence and sacrifice ratios: Some further considerations. Open Economies Review 11 (2), 111–125.

Barro, R. J., Gordon, D. B., 1983. Rules, Discretion and Reputation in a Model of Monetary Policy. Journal of Monetary Economics 12 (1), 101–121.

Blackburn, K., Christensen, M., 1989. Monetary Policy and Policy Credibility: Theories and Evidence. Journal of Economic Literature 27 (1), 1–45.

Blanchard, O., 1998. Optimal Speed of Disinflation: Hungary. In: Cottarelli, C., Szapary, G., International Monetary Fund, Magyar Nemzeti Bank (Eds.), Moderate Inflation: The Experience of Transition Economies. International Monetary Fund, pp. 132–147.

Blanchard, O., Galí, J., 2007. Real Wage Rigidities and the New Keynesian Model. Journal of Money, Credit and Banking 39 (s1), 35–65.

Blanchard, O. J., 1985. Credibility, disinflation and gradualism. Economics Letters 17 (3), 211–217.

Blinder, A., 2000. Central Bank Credibility: Why Do We Care? How Do We Build It? American Economic Review 90 (5), 1421–1431.

Blinder, A. S., 1997. Distinguished Lecture on Economics in Government: What Central Bankers Could Learn from Academics–and Vice Versa. The Journal of Economic Perspectives 11 (2), 3–19.

Blinder, A. S., Morgan, J., 2005. Are Two Heads Better Than One? Monetary Policy by Committee. Journal of Money, Credit and Banking 37 (5), 798–811.

Blinder, A. S., Morgan, J., 2008. Leadership in Groups: A Monetary Policy Experiment. International Journal of Central Banking 4 (4), 117–150.

Boschen, J. F., Weise, C. L., 2001. The ex ante credibility of disinflation policy and the cost of reducing inflation. Journal of Macroeconomics 23 (3), 323–347.

Camerer, C. F., 2003. Behavioral game theory: Experiments in strategic interaction. The roundtable series in behavioral economics. Princeton University Press, Princeton and N.J.

Camerer, C. F., Ho, T.-h., Chong, J.-K., 2004. A cognitive hierarchy model of games. The Quarterly Journal of Economics 119 (3), 861–898.

Cecchetti, S. G., Rich, R. W., 2001. Structural Estimates of the U.S. Sacrifice Ratio. Journal of Business & Economic Statistics 19 (4), 416–427.

Cogley, T., Matthes, C., Sbordone, A. M., 2011. Optimal disinflation under learning. FRB of New York Staff Report (524).

Collard, F., Fève, P., Matheron, J., 2007. The dynamic effects of disinflation policies. IDEI Working Paper 426.

Cooper, R., Haltiwanger, J., 1996. Evidence on Macroeconomic Complementarities. The Review of Economics and Statistics 78 (1), 78–93.

Cornand, C., Heinemann, F., 2014. Experiments on Monetary Policy and Central Banking. In: Duffy, J.

(Ed.), Experiments in macroeconomics. Vol. 17 of Research in Experimental Economics. Emerald Group Publishing Limited, pp. 167–228.

Davis, D., Korenok, O., 2011. Nominal price shocks in monopolistically competitive markets: An experimental analysis. Journal of Monetary Economics 58 (6), 578–589.

Davis, D. D., Holt, C. A., 1993. Experimental Economics. Princeton Univ. Press, Princeton, NJ.

Diana, G., Sidiropoulos, M., 2004. Central bank independence, speed of disinflation and the sacrifice ratio. Open Economies Review 15 (4), 385–402.

Duffy, J., Heinemann, F., 2014. Central bank reputation, cheap talk and transparency as substitutes for commitment: Experimental Evidence. Working Paper.

Duffy, J. M., 2014. Macroeconomics: A Survey of Laboratory Research. Working Paper.

Duffy, J. M., Nagel, R. C., 1997. On the robustness of behaviour in experimental beauty contest games. The Economic Journal 107 (445), 1684–1700.

Durham, J. B., 2001. Sacrifice ratios and monetary policy credibility: do smaller budget deficits, inflation-indexed debt, and inflation targets lower disinflation costs? Working Paper.

Engle-Warnick, J., Turdaliev, N., 2010. An Experimental Test of Taylor-type Rules with Inexperienced Central Bankers. Experimental Economics 13 (2), 146–166.

Erceg, C. J., Levin, A. T., 2003. Imperfect credibility and inflation persistence. Journal of Monetary Economics 50 (4), 915–944.

Fehr, E., Tyran, J.-R., 2001. Does money illusion matter? American Economic Review 91 (5), 1239–1262.

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.

Filardo, A. J., 1998. New evidence on the output cost of fighting inflation. Economic Review-Federal Reserve Bank of Kansas City 83, 33–62.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

Fischer, A. M., 1997. Do institutional factors matter for the speed of disinflation? Swiss Journal of Economics and Statistics (SJES) 133 (III), 539–556.

Frederick, S., 2005. Cognitive reflection and decision making. Journal of Economic perspectives 19 (4), 25–42.

Friedman, B. M., 1994. Comment. In: Mankiw, N. G. (Ed.), Monetary policy. The University of Chicago Press, pp. 182–188.

Fuhrer, J. C., 1995. The persistence of inflation and the cost of disinflation. New England Economic Review, 3–16.

Giamattei, M., Lambsdorff, J. G., 2014. Balancing the current account: experimental evidence on underconsumption. Experimental Economics, 1–27.

Goodfriend, M., King, R. G., 2005. The incredible Volcker disinflation. Journal of Monetary Economics 52 (5), 981–1015.

Gordon, R. J., 1982. Why stopping inflation may be costly: Evidence from fourteen historical episodes. In: Hall, R. E. (Ed.), Inflation: Causes and Effects. University of Chicago Press, pp. 11–40.

Greiner, B., 2004. An online recruitment system for economic experiments. Working Paper.

Güth, W., Kocher, M., Sutter, M., 2002. Experimental 'beauty contest' with homogeneous and heterogeneous players and with interior and boundary equilibria. Economic Letters 74 (2), 219–228.

Haltiwanger, J. C., Waldman, M., 1989. Limited rationality and strategic complements: The implications for macroeconomics. The Quarterly Journal of Economics 104 (3), 463–483.

Ho, T.-H., Camerer, C. F., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental p-beauty contests. American Economic Review 88 (4), 947–969.

Hommes, C., 2011. The heterogeneous expectations hypothesis: Some evidence from the lab. Journal of Economic Dynamics & Control 35 (1), 1–24.

Illing, G., 1998. Gradualism vs Cold Turkey. Working Paper.

Ireland, P. N., 1995. Optimal disinflationary paths. Journal of Economic Dynamics and Control 19 (8), 1429–1448.

Ireland, P. N., 1997. Stopping inflations, big and small. Journal of Money, Credit, and Banking 29 (4), 759–775.

Jordan, T. J., 1997. Disinflation costs, accelerating inflation gains, and central bank independence. Weltwirtschaftliches Archiv 133 (1), 1–21.

Jordan, T. J., 1999. Central bank independence and the sacrifice ratio. European Journal of Political Economy 15 (2), 229–255.

Kim, T.-h., Kim, J. I., Sung, T. Y., 2013. Stock market liberalization and price response: Gradualism versus cold turkey. Applied economics 45 (3), 273–285.

King, M., 1996. How should central banks reduce inflation? Conceptual issues. Economic Review-Federal Reserve Bank of Kansas City 81, 25–52.

Kurz, M., Piccillo, G., Wu, H., 2013. Modeling diverse expectations in an aggregated New Keynesian Model. Journal of Economic Dynamics and Control 37 (8), 1403–1433.

Kydland, F., Prescott, E., 1977. Rules Rather Than Discretion: The Inconsistency of Optimal Plans. The Journal of Political Economy 85 (3), 473–491.

Lambsdorff, J. G., Giamattei, M., Schubert, M., 2015. An Experimental Search for Calvo. Working Paper.

Lambsdorff, J. G., Schubert, M., Giamattei, M., 2013. On the Role of Heuristics – Experimental Evidence on Inflation Dynamics. Journal of Economic Dynamics and Control 37 (6), 1213–1229.

Loh, S., 2002. A Cold-Turkey versus a Gradualist Approach in a Menu Cost Model. The BE Journal of Macroeconomics 2 (1), 1–19.

Lombardelli, C., Proudman, J., Talbot, J., 2005. Committees Versus Individuals: An Experimental Analysis of Monetary Policy Decision-Making. International Journal of Central Banking 1 (1).

Mankiw, N. G., 2001. The inexorable and mysterious tradeoff between inflation and unemployment. The Economic Journal 111 (471), 45–61.

Milani, F., 2007. Expectations, learning and macroeconomic persistence. Journal of Monetary Economics 54 (7), 2065–2082.

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

Nagel, R. C., 1995. Unraveling in guessing games: An experimental study. American Economic Review 85 (5), 1313–1326.

Noussair, C. N., Pfajfar, D., Zsiros, J., 2014. Persistence of Shocks in an Experimental Dynamic Stochastic General Equilibrium Economy. In: Duffy, J. (Ed.), Experiments in macroeconomics. Vol. 17 of Research in Experimental Economics. Emerald Group Publishing Limited, pp. 71–108.

Orphanides, A., Wilcox, D. W., 2002. The opportunistic approach to disinflation. International Finance 5 (1), 47–71.

Orphanides, A., Williams, J., 2005. Imperfect Knowledge, Inflation Expectations, and Monetary Policy. In: Bernanke, B., Woodford, M. (Eds.), The inflation-targeting debate. Vol. v. 32 of National Bureau of Economic Research–studies in business cycles. University of Chicago Press, Chicago, pp. 201–246.

Petersen, L., 2011. Nonneutrality of Money, Preferences and Expectations in Laboratory New Keynesian Economies. Working Paper.

Pfajfar, D., Zakelj, B., 2009. Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory. mimeo.

Phelps, E. S., 1986. The trouble with rational expectations and the problem of inflation stabilization. In: Frydman, R., Phelps, E. S. (Eds.), Individual forecasting and aggregate outcomes. Cambridge Univ. Press, Cambridge, pp. 31–41.

Posen, A., 1998. Central bank independence and disinflationary credibility: a missing link? Oxford Economic Papers 50 (3), 335–359.

Ricciuti, R., 2008. Bringing Macroeconomics into the Lab. Journal of Macroeconomics 30 (1), 216–237.

Rodrik, D., 1989. Credibility of trade reform—a policy maker's guide. The World Economy 12 (1), 1–16.

Roos, M., Luhan, W., 2008. Are expectations formed by the anchoring-and-adjustment heuristic? An experimental investigation. Ruhr Economic Papers (54).

Roos, M., Luhan, W., 2013. Information, Learning and Expectations in an Experimental Model Economy. Economica 80 (319), 513–531.

Sargent, T. J., 1982. The ends of four big inflations. In: Hall, R. E. (Ed.), Inflation: Causes and Effects. University of Chicago Press, pp. 41–98.

Sargent, T. J., 1986. Stopping Moderate Inflations: The Methods of Poincare and Thatcher. In: Sargent, T. J. (Ed.), Rational expectations and inflation. Harper & Row, New York, pp. 110–157.

Schaling, E., Hoeberichts, M., 2010. Why Speed Doesn't Kill: Learning to Believe in Disinflation. De Economist 158 (1), 23–42.

Spermann, A., 2006. Basic Income Reform in Germany: Better Gradualism than Cold Turkey. Institute for the Study of Labor Discussion Paper (2398), 06–064.

Stasavage, D., 2002. Communication, Coordination and Common Knowledge in Monetary Policy: Implications for the Euro Zone. unpublished paper, LSE.

Sutan, A., Willinger, M., 2009. Guessing with negative feedback: An experiment. Journal of Economic Dynamics & Control 33 (5), 1123–1133.

Taylor, J. B., 1983. Union Wage Settlements during a Disinflation. American Economic Review 73 (5), 981–993.

Verbeek, M., 2004. A guide to modern econometrics, 2nd Edition. John Wiley & Sons, Southern Gate, Chichester, West Sussex, England and Hoboken, NJ.

Woodford, M., 2003. Interest and prices: Foundations of a theory of monetary policy. Princeton Univ. Press, Princeton and NJ.

Zhang, L. H., 2005. Sacrifice Ratios with Long–Lived Effects. International Finance 8 (2), 231–262.

Appendix A. Theoretical Model

Appendix A.1. Forecasters

As in Lambsdorff et al. (2015) the target value (1) can be derived from a reduced form model of monopolistic competition. Assume such a market where firms compete for customers but also supply intermediate products to each other. In (A.1) the demand depends negatively on the log of the own price X_{nt} and positively on the log price of others X_{-nt}. Additionally, firms have to buy raw materials form abroad with I_t being the log price of these imports.

$$Q_{nt} = \frac{40(1.6I_t - X_{nt} + 0.6X_{-nt})}{(0.8I_t)^2} \tag{A.1}$$

Due to the role of other players as suppliers of intermediate goods,the marginal costs of production depend on the others' price according to $0.6X_{-nt}$. Therefore, the firm's profits are given by (A.2).

$$\begin{aligned}\Theta_{nt} &= Q_{nt}(X_{nt} - 0.6X_{-nt}) = \\ &= \frac{40}{(0.8I_t)^2}(1.6I_t - X_{nt} + 0.6X_{-nt})(X_{nt} - 0.6X_{-nt})\end{aligned} \tag{A.2}$$

Maximizing a firm's profits requires $\partial\Theta_{nt}/\partial X_{nt} \overset{!}{=} 0$.

$$\begin{aligned}\frac{\partial\Theta_{nt}}{\partial X_{nt}} = \frac{40}{I_t^2}(1.6I_t - 2X_{nt} + (0.6 + 0.6)X_{-nt}) \overset{!}{=} 0 \\ 0.8I_t + 0.6X_{-nt} = X_{nt}\end{aligned} \tag{A.3}$$

Due to the fact that all prices are logs, price changes are given by $x_{nt} = X_{nt} - X_{nt-1}$ wich is the price change of the own firm, $x_{-nt} = X_{-nt} - X_{-nt-1}$ which is the price change of others and $i_t = I_t - I_{t-1}$ which is the price change of import prices. The best response function (4) can, therefore, be calculated by $X_{nt} - X_{nt-1} = 0.6X_{-nt} + 0.8X_t - 0.6X_{-nt-1} + 0.8I_{t-1} = 0.6x_{-nt} + 0.8i_t = x_{nt}$.

If prices are set according to this function, the firm's profit is given by $\Theta_{nt} = 40$ as in (2). If firms deviate from the best response function by ϵ, payoffs are solved to (A.4). It differs slightly from the payoff function (2) with $\Pi_{nt} = 40 - 0.1\epsilon^2$.

$$\begin{aligned}\Theta_{nt} =& \frac{40(1.6I_t - 0.6X_{-nt} - 0.8I_t - \epsilon + 0.6X_{-nt})}{(0.8I_t)^2} \\ &* (0.6X_{-nt} + 0.8I_t + \epsilon - 0.6X_{-nt}) = \\ &= \frac{40(0.8I_t - \epsilon)(0.8I_t + \epsilon)}{(0.8I_t)^2} = 40 - \frac{40\epsilon^2}{(0.8I_t)^2}\end{aligned} \tag{A.4}$$

Appendix A.2. Central Bankers

The central bank's payoff function is given by (5). As explained in the main text, the first term represents a simplified inflation target as the central bank tries to minimize part of the inflation i_t, which it can directly control.

The second part of the payoff function punishes deviations of the average choice from the target value. This is identical to minimizing the forecasters' average deviation from the equilibrium $p/(1-p)i_t$. This can be seen by rearranging the second term of (5).

$$\begin{aligned}\min_{i_t} 0.5(x_t - \tilde{x}_t)^2 &= \min_{i_t} 0.5(x_t - p(x_t + i_t))^2 \\ &= \min_{i_t}(1-p)^2 0.5(x_t - \frac{p}{1-p}i_t)^2\end{aligned} \tag{A.5}$$

This deviation can be related to an output target. The potential output can be defined as equilibrium output where all players optimally set their prices according to (A.3). Inserting this into the demand function (A.1) yields the potential output $\bar{Q} = 32I_t/(0.8I_t)^2$. Again, if firms deviate by ϵ from the target function (4) by setting $X_{nt} = 0.6X_{-nt} + 0.8I_t + \epsilon$, the squared output gap can be calculated by (A.6).

$$\begin{aligned} \min(Q_t - \bar{Q})^2 &= \min\left(\frac{40(0.8I_t - \epsilon)}{(0.8I_t)^2} - \frac{32I_t}{(0.8I_t)^2}\right)^2 = \\ &= \min\left(\frac{40\epsilon}{(0.8I_t)^2}\right)^2 = \min\frac{\sqrt{40}}{0.8I_t}\epsilon^2 \end{aligned} \tag{A.6}$$

So minimizing the deviations from equilibrium ϵ also minimizes the output gap. Again, the weighting factor is held constant and does not vary with I_t for reasons of simplicity.

Appendix B. Instructions and Questionnaire (English translation)

Appendix B.1. Oral Instructions

A very warm welcome to this experiment and thank you very much for your participation. I will read some general instructions for the experiment. Please listen carefully and only click the button "Start Experiment" after you have listened to these instructions. All participants of this experiment are in this room and are participating at the same experiment. With the experiment, we want to gain insights into human behavior. The game will last about 90-120 minutes. Your earnings in this experiment depend on your decisions and the decisions of the other participants. In the experiment, the payoff will be calculated in the currency Taler. 1 Taler will be converted into 10 cents. At the end of the experiment, you will get 10-15 euros per hour on average, but at least 3 euros per hour.

You interact anonymously and cannot communicate with others. The payoff will be paid out anonymously, too. No other participant will see how much you earn. Also, the persons who conduct the experiment will not be informed about that. Not the experimenter but another person will hand out the payoff. This person can not infer your behavior from your payoff. During the experiment you may have to wait for other participants. This may also take some minutes. Please remain seated patiently. Use the waiting time in order to recalculate the examples and to think about your decisions during the experiment.

After the experiment, you will be asked to leave the laboratory on your own. Outside of the laboratory you get your payoff. You will find all instructions and explanations on the following screens. Please read all information carefully before leaving a screen. You cannot go back to screens you have left. If you want, you can take notes. Pen and paper are lying at your desk. You also have a pocket calculator on your desk. Please remain seated quietly at your desk. Please do not talk at all. If you do not comply with these rules you may be excluded from the further participation at the experiment. If you have any questions, please raise your hand. We will then come to your desk. Please now click on the button "Start Experiment".

Appendix B.2. Written Instructions

A very warm welcome to the experiment! At the beginning of the experiment you are instructed about the general procedures in the lab. These will be read aloud by the experimenter. Please only click on "Start experiment" if you are told to do so.

Screen 1. Please read the instructions of the experiment carefully. At the end of the instructions, you will be asked some comprehension questions. For each answer you get right, you earn 2 Taler. On your desk, you find an overview over the flow of the experiment, which will be explained in the following.

In this experiment you have 3 lives. In each life you interact in a group of 5 persons for 12 rounds. The group consists of 4 participants A and 1 participant B and remains constant within a life. After each life, participant B will be matched with another group of participants A, i.e. all participants A will interact with a different participant B in each life. At the beginning of the experiment, you will be told whether you are participant A or B.

Screen 2. All participants A have the same task in each round and life, participant B has a different task. Also, participant B has the same task in all rounds and lives. In the following, the tasks of participants A and participant B will be explained. Each life thereby follows the same structure.

At the end of the experiment, one of the three lives will be chosen randomly. For all participants the same life will be chosen. The chosen life determines your payoff and the other participants' payoff. Not chosen lives have no impact on your payoff. You do not know which life will be chosen until the end of the experiment. Each life can be the one, that determines your payoff.

Screen 3 - Task of participants A. Every life consists of 12 rounds. All 4 participants A have the same task. Every participant A has to set an integer number between 0 and 150 in every round. With the number, participant A tries to get close to a target value. The target value is determined by: **target value = 2/3 * (average number + indicator)**. The average number is the average of all numbers of all participants A in a group (i.e. including the own number). The indicator (a value between 10 and 50) is announced to all participants at the beginning of each round. All relevant formula are stated later in the experiment, too.

Screen 4 - Task of participants A continued. In each round, participant A receives an initial payoff of 40 Taler, from which 0.1 * the **squared** deviation between his number and the target value will be deducted. The payoff in one round can also be negative. Please notice that the deviation will be squared first and will then be deducted, i.e. the round payoff of participant A is: **Payoff A** $= 40 - 0.1 * \textbf{deviation}^2 = 40 - 0.1 * (\textbf{number} - \textbf{target value})^2$. The round payoff from round 5 onwards will be summed up over all rounds. The first four rounds of every life are trial rounds. If the life is chosen, the life payoff will be paid out in euros at the end of the experiment.

Screen 5 - Comprehension questions for task of participants A. *[Comprehension questions had to be filled in correctly to continue. After each wrong question participants got additional help to complete the question.]* For every correct answer you earn 2 Taler if you provide the correct answer in the first place. Taler you earn for the comprehension questions are paid out in euros at the end of the experiment.

Question 1

Four participants A and one participant B ...

- always interact with the same participants A, but in each life with a different participant B.
- interact in each life and each round with the same participants A and B.
- are matched to an new group in each round.

Question 2

At the end of the experiment ...

- the first life will always be chosen to be paid out.
- the payoff of all lives will be summed up and paid out.
- one life will be randomly chosen to be paid out.

Question 3

If the average number is 130 and the indicator is 50, the target value is ...

Question 4

If the average number is 10 and the indicator is 50, the target value is ...

Question 5

The indicator ...

- will be announced to participants A before they decide on a number.
- will be announced to participants A after they decided on a number.
- can have no values between 10 and 50.

Screen 6 - Task of player B. Participant B has a different task than participants A. In the first four rounds participant B does not have to make a decision and gets no round payoff. In the first four rounds the indicator is automatically set to 50 and participant B can only observe the average number chosen by participants A. From round 5 on, participant B sets the indicator in the range between 10 and 50. Participant B sets the indicator **before** participants A set their number. The indicator will be told to participants A.

Screen 7 - Task of player B continued. Participant B gets an initial payoff of 60 from round 5 on. From this payoff two terms will be deducted: $\textbf{Payoff B} = 60 - 0.5 * (\textbf{average number} - \textbf{target value})^2 - \textbf{indicator}$. First, 0.5 * the squared average deviation of participants A will be deducted, i.e. the deviation of the average number from the target value will be calculated and then squared. This implies that the round payoff of participant B will be reduced if participants A deviate from the target value. Second, the value of the indicator will be deducted from the initial payoff. The payoff in one round can also be negative. For participant B, the round payoff will be summed up over the rounds 5-12. If the life is chosen, the life payoff will be paid out in euros at the end of the experiment.

Screen 8 - Comprehension questions for task of participant B. For every correct answer you earn 2 Taler if you provide the correct answer in the first place. Taler you earn for the comprehension questions are paid out in euros at the end of the experiment.

Question 1

Participant B ...

- always interacts with the same participants A, but in each life with a different participant B.
- interacts in each life and each round with the same participants A and B.
- is matched to an new group in each round.

Question 2

Participant B ...

- sets the indicator in the range from 10 to 50 from round 5 on.
- sets the indicator in the range from 10 to 150 from round 5 on.
- does not set the indicator.

Question 3

If the average number deviates from the target value by 4 and the indicator is 50, the round payoff of participant B is ...

Question 4

If the average number deviates from the target value by 0 and the indicator is 50, the round payoff of participant B is ...

Question 5

Participant B ...

- does not have to make a decision in the first four rounds and can only observe the behavior of participants A.
- gets the same round payoff as participants A.
- gets more payoff if the indicator is higher.

Screen 9A - Final instruction [Player A]. You are one of the four participants A, i.e. you decide on a number in each round, after you are told the value of the indicator. Additionally, you are asked, what you believe which number the other participants A will pick on average. If your belief is equal to the correct average number ±5%, then you get 2 Taler extra. Additionally, during the whole experiment a table with all values will be shown. Start Experiment.

Screen 9B - Final instruction [Player B]. You are participant B. The four others in your group are participants A, i.e. from round 5 on you decide on the indicator, before participants A set their numbers. While participants A decide on their numbers, you are asked what you believe which number participants A will pick on average. If your belief is equal to the correct average number ±5%, then you get 2 Taler extra. Additionally, during the whole experiment a table with all values and a diagram with the indicator and the average number will be shown. Start Experiment.

Appendix B.3. Questionnaire

The main part of the experiment is over now. Before the end of the experiment, you are going to be asked some questions. By answering them correctly you can earn additional Taler. For each question you answer correctly, you get 2 Taler. The time for answering a question is restricted and is shown to you in the upper right corner of the screen. When you click on Start, the questions begin.

Part 1 of 3. You have 90 seconds for these questions.

- A bat and a ball cost 1.10 euro in total. The bat costs 1 euro more than the ball. How much does the ball cost?
- In a clothing factory it takes 5 machines 5 minutes to make 5 shirts, how long would it take 100 machines to make 100 shirts?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

??	+	??	=	3
+		+		
??	+	??	=	3
=		=		
2		4		

Example

??	x	??	=	21		??	+	??	=	13
+		+				x		-		
??	:	??	=	2		??	+	??	=	3
=		=				=		=		
9		10				20		5		

Matrix 1 and 2

Part 2 of 3. You have 180 seconds for these questions.

- A seller of lemons sells a car for 6000 euro. This is 2/3 of the original price. How much did the car cost originally? (in euros).
- As an entrepreneur you should set the price P dependent on the average price DP according to the function P=1/2*DP+20. Which price do you set if all other entrepreneurs calculate the same way that you do and, therefore, D equals DP?
- In order to paint the walls of a quadratic office, you need 3 liters of color. How many liters do you need for a quadratic living room, where the floor space is four times the floor space of the office? Hint: Ignore doors and windows. Only the walls are painted not the ceiling nor the floor.
- A cube has an edge length of 9 centimeters an a weight of 162 grams. How much would the cube weight if the edge length was 3 centimeters?

Part 3 of 3 - Instruction. Three mathematical equation systems are presented to you now, which are readable from top to bottom and from left to right (see Table Example). You have to fill in the empty slots with numbers between 1 and 9, so that all equations can be solved consistently. Take care of both the horizontal and vertical reading direction. For this task you have 180 seconds. For every equation system you solve correctly,you get 2 Taler.

Part 3 of 3. *[Matrices 1-3 were displayed.]*

??	+	??	+	5	=	20
-		+		x		
??	x	??	-	1	=	20
+		+		+		
11	-	3	+	2	=	10
=		=		=		
15		18		7		

Matrix 3

Statistical Questions. Please answer the following questions. They are only used for statistical purposes.

Gender
Age
Course

- Business Administration and Economics
- Governance and Public Policy
- International Cultural and Business Studies / European Studies
- Media & Communication / Language Studies
- Internet Computing
- Teaching
- Law
- Other

Risk

How do you think of yourself: Are a risk seeking person or do you try to avoid risks? *Likert Scale: 0="not risk seeking at all"; 10="very risk seeking"*