



**HAL**  
open science

# Learning to deal with repeated shocks under strategic complementarity: An experiment

Muhammed Bulutay, Camille Cornand, Adam Zylbersztein

► **To cite this version:**

Muhammed Bulutay, Camille Cornand, Adam Zylbersztein. Learning to deal with repeated shocks under strategic complementarity: An experiment. 2020. halshs-02458140v1

**HAL Id: halshs-02458140**

**<https://shs.hal.science/halshs-02458140v1>**

Preprint submitted on 28 Jan 2020 (v1), last revised 20 Apr 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 2003 – January 2020

## Learning to deal with repeated shocks under strategic complementarity: An experiment

Muhammed Bulutay, Camille Cornand, Adam Zylbersztein

### Abstract:

Experimental evidence shows that the rational expectations hypothesis fails to characterize the path to equilibrium after an exogenous shock when actions are strategic complements. Under identical shocks, however, repetition allows adaptive learning, so that inertia in adjustment should fade away with experience. If this finding proves to be robust, inertia in adjustment may be irrelevant among experienced agents. The conjecture in the literature is that inertia would still persist, perhaps indefinitely, in the presence of real-world complications such as nonidentical shocks. Herein, we empirically test the conjecture that the inertia in adjustment is more persistent if the shocks are nonidentical. For both identical and nonidentical shocks, we find persistent inertia and similar patterns of adjustment that can be explained by backward-looking expectation rules. A reformulation of naïve expectations with similarity-based learning approach is found to have a higher predictive power than rational and trend-following rules.

### Keywords:

Strategic complementarities, expectations, adjustment speed, similarity-based learning, guessing games, heuristics switching

### JEL codes:

C72, C73, D83, D84, D91, G41

# Learning to deal with repeated shocks under strategic complementarity: An experiment\*

Muhammed Bulutay<sup>†a</sup>, Camille Cornand<sup>b</sup>, and Adam Zylbersztein<sup>c</sup>

<sup>a</sup>Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

<sup>b</sup>Univ Lyon, CNRS, GATE UMR 5824, F-69130 Ecully, France

<sup>c</sup>Univ Lyon, Université Lumière Lyon 2, GATE UMR 5824, F-69130 Ecully, France

January 2020

## Abstract

Experimental evidence shows that the rational expectations hypothesis fails to characterize the path to equilibrium after an exogenous shock when actions are strategic complements. Under identical shocks, however, repetition allows adaptive learning, so that inertia in adjustment should fade away with experience. If this finding proves to be robust, inertia in adjustment may be irrelevant among experienced agents. The conjecture in the literature is that inertia would still persist, perhaps indefinitely, in the presence of real-world complications such as nonidentical shocks. Herein, we empirically test the conjecture that the inertia in adjustment is more persistent if the shocks are nonidentical. For both identical and nonidentical shocks, we find persistent inertia and similar patterns of adjustment that can be explained by backward-looking expectation rules. A reformulation of naïve expectations with similarity-based learning approach is found to have a higher predictive power than rational and trend-following rules.

**Keywords:** Strategic complementarities, expectations, adjustment speed, similarity-based learning, guessing games, heuristics switching.

**JEL Classification:** C72, C73, D83, D84, D91, G41

---

\*This research was performed within the framework of the LABEX CORTEX (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). The authors are thankful to the ANR-DFG joint grant for financial support (ANR-18-FRAL-0008-01 ASUR). We thank Frank Heinemann, Nobuyuki Hanaki, Angela Sutan and participants of Colloquium Monetary and Experimental Economics at TU-Berlin for stimulating discussions and helpful comments. We also acknowledge Quentin Thévenet and Maxim Frolov for providing help with the software programming.

<sup>†</sup>Corresponding author.

E-mail addresses: mbulutay@diw.de (M. Bulutay), cornand@gate.cnrs.fr (C. Cornand), zylbersztein@gate.cnrs.fr (A. Zylbersztein)

# 1 Introduction

How long would it take for market outcomes to fully adjust to the new equilibrium level in response to an exogenous shock? In a seminal paper on the rational expectations (RE) hypothesis, Muth (1961) demonstrates that convergence to equilibrium would be instantaneous in a frictionless economy if the errors in agents' expectations are not highly correlated as they cancel out at the aggregate level. However, the empirical evidence points to systematic errors due to heuristic-based reasoning under which the aggregate outcomes may exhibit substantial inertia. Whether and how the adjustment would be delayed in the presence of non-rational expectations is a key question for policy-makers – central banks that aim at engineering structural changes – and for actors in markets where equilibrium is frequently shifting due to shocks. If adjustment is sluggish and shocks occur frequently, aggregates may rarely be in accordance with the equilibrium path predictions generated by the impulse-response analyses of RE-based models.

Early experimental evidence from double auctions shows that equilibrium prices emerge within a few periods (Smith, 1962). Convergence occurs even in the presence of zero-intelligence computer traders who submit random bids and asks if these bids are constrained with a budget (Gode and Sunder, 1993). Nonetheless, persistent deviations from equilibrium are reported in different types of competitive markets (e.g. asset market experiments, AMEs henceforth, Smith et al. 1988). Thus, the extent to which limited rationality influences market outcomes depends on the characteristics of the market.

The type of strategic environment governing the market is one of the key characteristics determining the impact of limited rationality on behavior and outcomes. Following the theoretical work of Haltiwanger and Waldman (1985, 1989), Fehr and Tyran (2005, 2008) experimentally test the role of the strategic environment on the adjustment dynamics after a monetary shock. In accordance with the theoretical predictions, the adjustment is immediate when actions are strategic substitutes, and gradual when actions are strategic complements. The role of the strategic environment has been further experimentally investigated in Learning-to-Forecast

Experiments (LtFEs, Heemeijer et al. 2009, Bao et al. 2012), guessing games (Sutan and Willinger 2009, Cooper et al. 2017, Hanaki et al. 2019) and duopoly games (Potters and Suetens, 2009).<sup>1</sup> The main pattern emerging from these studies is that deviations from equilibrium tend to be higher and be more persistent under strategic complementarity as compared to strategic substitutability.<sup>2</sup>

Herein, we focus on strategic complementarity which comes as an important feature of various economic contexts including macroeconomic coordination, bank runs, and oligopoly competition.<sup>3</sup> Hommes (2006) argues that strategic complementarity is crucial for modeling asset markets characterized by a positive feedback mechanism between expectations on asset prices and the realizations of these prices.

The literature still lacks common consensus on how repeated shocks (whether they are identical or not) could affect adjustment under strategic complementarity. On the one hand, the initial deviations from RE may subsequently disappear due to *experience effects*, as commonly reported in AMEs (e.g. Smith et al. 1988). In a recent study, Cooper et al. (2017) show that these results can be extended to guessing games. They introduce three identical shocks into Nash equilibrium (NE) in a periodic manner and report slight acceleration in the adjustment speed over shocks. On the other hand, experimental studies based on AMEs and LtFEs question the robustness of experience effects (Kopányi-Peucker and Weber 2018; Shestakova et al. 2019). Hussam et al. (2008) argue that experience effects critically rely on the stationarity of the environment. Accordingly, both Cooper et al. (2017, p. 207) and Fehr and Tyran (2008, p. 387) conjecture that in case of repeated nonidentical shocks, the impact of non-rational expectations would persist. However, neither paper provides an empirical test of this conjecture. Our work aims at filling this gap.

We experimentally test the conjecture of a relative persistence of nonidentical shocks in a guessing game with strategic complementarity (based on Cooper et al., 2017). We introduce

---

<sup>1</sup>See Hommes (2011) and Arifovic and Duffy (2018) for an overview of the Learning-to-Forecast (LtF) literature.

<sup>2</sup>Hanaki et al. (2019) are the first to term this phenomenon as *the strategic environment effect*.

<sup>3</sup>See Milgrom and Roberts (1990) for more examples.

large periodic shocks to NE and compare adjustment dynamics between two experimental conditions: one where shocks are identical and another where they are not. During the first and last post-shock phases, the NE are the same in both conditions. Through this design, we are able to measure (i) the treatment effect of experiencing nonidentical shocks on the aggregate adjustment speed, and (ii) on the way individuals form expectations. Related to (i), we find that post-shock adjustment accelerates due to repetition. Compared to the initial post-shock adjustment, it takes fewer periods less for the adjustment after further shocks to occur. However, we fail to identify a significant effect of nonidentical shocks on the pace of adjustment. Related to (ii), our results show that experience may not be enough to deplete naïvety, at least not within four repetitions.

Our contribution to the literature is twofold. Firstly, we replicate the findings of Cooper et al. (2017) in the context of identical shocks, and further extend their findings to a more complex environment with nonidentical shocks. In this context, we show that the inertia in adjustment is a robust feature of markets governed by strategic complementarity and that it does not depend on the stationarity of periodic shocks. Secondly, the data on expectations across subjects and over time allow us to study the individual underpinnings of the observed aggregate dynamics. To avoid arbitrariness in model selection, we consider a wide range of backward-looking expectation rules and take their predictions to the experimental data. This novel horse race exercise identifies one heuristic rule that not only outperforms all other rules (including rational and trend-following homogeneous expectations), but also attains the highest impact factor within an evolutionary heuristic switching model (HSM, Anufriev and Hommes 2012). This rule, first proposed by Cooper et al. (2017), is a reformulation of naïve expectations with similarity-based learning. We discuss the behavioral foundations and relate it to the previous literature.

The remainder of this paper is organized as follows. Section 2 outlines our research hypotheses. Section 3 presents our methodology: the guessing game, the experimental design and procedures. Section 4 summarizes the main results which are then discussed in Section

5. Lastly, Section 6 concludes by summarizing the main findings, as well as the implications and limitations of the study.

## 2 Related Literature and Hypotheses

**Table 1 - Adjustment speed in different studies**

Study	Type of environment	Shock size (in %)	Adjustment period
Fehr and Tyran (2001) <sup>1</sup>	Pricing decision	-67%	13
Fehr and Tyran (2008) <sup>2</sup>	Pricing decision	-50%	9
Davis and Korenok (2011) <sup>3</sup>	Monopolistic competition	+100%	21
Petersen and Winn (2014) <sup>1</sup>	Pricing decision	-67% & +92%	8 & 4
Arifovic and Petersen (2017) <sup>4</sup>	New Keynesian LtFE	-47%	12
Cooper et al. (2017) <sup>5</sup>	Guessing game	-77%	9

<sup>1</sup> Shock as the change in average equilibrium price in the nominal treatment with human opponents.

<sup>2</sup> Shock as the change in average equilibrium price in nominal treatment.

<sup>3</sup> Shock as the change in monopolistically competitive prices in the BASE/PUB treatment. Prices remain significantly different than competitive level in the first reported 20 post-shock periods.

<sup>4</sup> Shock as the change in the steady-state nominal interest rate in the constant target treatment.

<sup>5</sup> Shock as the change in NE guess in first round.

As shown in Table 1, the previous investigations of adjustment dynamics following large shocks report that adjustment generally takes time when actions are strategic complements. In particular, adjustment tends to be slow after the initial shock, even though acceleration may still occur when markets are repeated (Cooper et al., 2017). This is also a standard finding across AMEs. For instance, Smith et al. (1988), Dufwenberg et al. (2005) and Haruvy et al. (2007) show that repeating market interactions three times eliminates bubbles.<sup>4</sup> We expect to replicate this main pattern observed in the literature.

**Hypothesis 1.** When the shocks are identical, adjustment is slow and gradual after the initial shock, but accelerates over repetition of the same market.

<sup>4</sup>This type of experience effect constitutes a strong argument for the promoters of *the efficient market hypothesis*. For instance Malkiel (2003) argues that any systematic deviation from the fundamental value would fade away over repetition. Therefore, such deviations from fundamentals as bubbles are only temporary exceptions, rather than long-term features of asset markets.

Albeit robust in stationary environments, experience effects are argued to be sensitive to the complexity of the environment. For instance, in the AMEs and LtFEs studies by Kopányi-Peuker and Weber (2018) and Hussam et al. (2008), bubbles do not disappear despite repetition in contrast to the literature. The former argue that this is due to the indefinite horizon feature of their experiments. Hussam et al. (2008) state that bubbles reignite even with twice-experienced subjects when drastic changes (e.g., the amount of liquidity in the market) occur in the environment. In particular, Cooper et al. (2017) and Fehr and Tyran (2008) conjecture that nonidentical shocks may thwart experience effects which constitutes the basis of our second hypothesis.

**Hypothesis 2.** The rate of acceleration is lower for nonidentical shocks as compared to the identical ones.

We now turn to the possible explanations of the adjustment dynamics. Several studies provide a descriptive explanation of the observed inertia based on non-rational expectations. Yet, they strongly diverge in terms of the best fitting model. For instance, Fehr and Tyran (2008) report that their data are best organized by a model in which all agents exhibit naïve expectations.<sup>5</sup> Cooper et al. (2017), in turn, obtain the best fit with heterogenous groups: one rational player and three non-rational players whose expectations follow a version of naïve expectations rule adapted to a repeated shocks design. Other studies point to trend-following expectations as best describing their experimental evidence (Haruvy et al. 2007), while others yield support for RE (Marquardt et al. 2019).

We note, however, that the aforementioned studies either do not compare the fit of their model with other expectation rules, or focus on narrow comparison sets.<sup>6</sup> More systematic comparisons exist in the LtF literature. The design of LtFEs is particularly well-suited for investigating expectations since the experimental task is to forecast the prices one period

---

<sup>5</sup>In their model of *fully adaptive expectations*, players expect the outcome of the last period to reoccur.

<sup>6</sup>For instance, Marquardt et al. (2019) make such a comparison but only between three models: myopic, trend and RE. Moreover, the parameters of the trend model that they investigate is closer to what other studies denote as *strong trend following* rule (e.g. Anufriev and Hommes 2012). In Section 5, we discuss why strong trend following may not be a suitable rule for the type of environments as AMEs.



ahead. The trend-following rule has been repeatedly shown to outperform all others under homogeneous expectations (Bao et al. 2012; Anufriev et al. 2013; Heemeijer et al. 2009). Pfajfar and Žakelj (2014) estimate the share of RE and simple expectations in their New Keynesian LtFEs. They arrive to the conclusion that RE (simple rules) cannot be rejected for 30-45% (35-50%) of subjects. This finding has been confirmed by Marquardt et al. (2019). In the context of the HSM, Cornea-Madeira et al. (2019) estimate the weights of naïve and fundamentalist rules in inflation expectations in the U.S. inflation data spanning from 1968:Q4 to 2015:Q2. Despite the substantial time variation, they find that 65% of individuals form naïve expectations. Furthermore, they state that the share of naïve expectations increases in reaction to large inflationary shocks, thus creating self-fulfilling inflation persistence.

Based on this body of empirical literature, we conclude the following. First, the best fitting expectation models vary across different experimental settings. Second, for the experimental settings closest to ours (i.e., guessing games and LtFEs) simple expectation models in backward-looking form outperform RE. This observation leads us to our third hypothesis.

**Hypothesis 3.** Backward-looking expectation rules in form of heuristics fit the data better than RE.

Finally, we provide the first out-of-the-sample test of the relative performance of the expectation rule proposed by Cooper et al. (2017). This rule seems promising in the context of repeated shocks since it echoes the similarity-based learning approach (Gilboa and Schmeidler, 1995; Plonsky et al., 2015). Accordingly, a player expects the outcome of last period to reoccur in stable phases. After observing a shock, the player reviews all the past periods and expects the outcome of the period following previous occurrence of the same shock. We denote this rule as *similarity-based naïve expectations* (SBNE) and conjecture that it best explains behavior under repeated identical shock.

**Hypothesis 4.** Under identical shocks, the SBNE rule provides the best fit to the experimental

data.

### 3 Method

#### 3.1 Guessing Game under Strategic Complementarity

To investigate whether repeating identical shocks improves the speed of adjustment, and whether nonidentical shocks slow down this process, we refer to a repeated guessing game under strategic complementarity that is adapted from Nagel (1995). Our experimental game also resembles those used in LtFEs with positive feedback and provides similar convergence dynamics when the feedback parameters are equal (Sonnemans and Tuinstra, 2010).<sup>7</sup>

In each period  $t \in [1, T]$ , a group of  $N$  players simultaneously choose a number (rounded up to two decimals) from the closed set  $p_{i,t} \in [0, 100]$  where  $i = 1, \dots, N$ . Each player  $i$  has a target number  $y_{i,t}$  that is calculated as

$$y_{i,t} = b\bar{p}_{-i,t} + a + \xi_t, \tag{1}$$

where  $\bar{p}_{-i,t}$  is the average number chosen by the remaining players players<sup>8</sup> at period  $t$ ,  $a$  and  $b$  are positive constant numbers where  $b \in (0, 1)$  and  $\xi_t$  is a deterministic large shock which takes the values

$$\xi_t = \begin{cases} 0, & \text{if } t \leq T/2 \\ \bar{\xi}, & \text{if } t > T/2. \end{cases} \tag{2}$$

The constant term  $b$  generates strategic complementarity among the players' actions. The player with the smallest guessing error  $|y_{i,t} - p_{i,t}|$  wins the fixed stage game payoff  $F$ . In case of tie, the payoff is divided equally among the winners.

---

<sup>7</sup>There is also a fundamental difference between these two designs: while guessing games provide full information on the game structure including the parameters, LtFEs provide only qualitative information about the market structure.

<sup>8</sup>Sutan and Willinger (2009) report that the inclusion of own guesses causes a significant amount of confusion among subjects. Therefore, we opted for excluding player's own guess from the target formula.

This game has an unique NE which corresponds to an interior solution:  $p_t^{NE} = \frac{a+\xi_t}{(1-b)}$ .<sup>9</sup> Here,  $p_t^{NE}$  is stable for the first  $T/2$  periods. We call these periods *pre-shock periods*. The shock  $\bar{\xi}$  shifts the equilibrium downwards at period  $T/2 + 1$ . We call the remaining periods under the new equilibrium *post-shock periods*. In addition, the shocks are repeated: a sequence of  $T$  periods (pre- and post-shock) is repeated over  $R$  rounds.

### 3.2 Experimental Design

Our experimental manipulation consists in varying the value of  $\bar{\xi}_r$  over rounds. For identical shocks (*baseline*),  $\bar{\xi}_r = \bar{\xi}$  for all  $r \in [1, R]$ . For nonidentical shocks (*treatment*), the size of the shock varies across rounds. Importantly, the equilibrium solution outlined above applies to both cases, so that players are always incentivized to play the NE.

The experiment is calibrated as follows. A group of  $N = 5$  participants play the guessing game for  $R = 4$  rounds, and each round is composed of  $T = 16$  periods. This yields a total of 64 guessing decisions per player. In each period, the prize is  $F = 4.40$  euros. In period 9 of each round, a negative shock  $\bar{\xi}$  shifts the constant variable  $a = 15$  to a value that varies across experimental conditions. In the baseline condition, the shock component is equal to  $-9$  and remains constant across rounds. There are two experimental conditions: baseline (identical shocks) in which the shock component is equal to  $-9$  and does not vary across rounds and treatment (nonidentical shocks) in which the shock component is characterized by the sequence  $(-9, -6, -12, -9)$  in rounds  $(1, 2, 3, 4)$ . Thus, post-shock NE are the same during the first and last rounds across baseline and treatment which allows us to identify the treatment effect of experiencing nonidentical shocks by comparing adjustment speeds at round 4.

The design of this experiment closely follows Cooper et al. (2017), although some differences

---

<sup>9</sup>Proof uses the iterative elimination of strictly dominated strategies. See Nagel (1995) for details. This equilibrium is also the rational expectations equilibrium (REE). Note that by taking the expectations on both sides, (1) can be rewritten as  $p_{i,t} = E[y_{i,t}] = (a + \xi_t) + bp_{i,t}^e$ . Bray (1983) shows that when  $b < 1$ , a misspecified expectation rule – ordinary least-squares learning – almost surely converges to the REE. However, she also emphasizes that this does not imply unbiased expectations. As she notes (Bray 1982, pp. 330), “[r]ational expectations are, if anything, a long run rather than a short run phenomenon.”

should be noted. First, in their study, groups are composed of 4 players. Following Hanaki et al. (2019), we increase the number of players to 5 per group.<sup>10</sup> Second, in Cooper et al. (2017) there are three rounds of 20 periods. We decrease the length of each round to 16 periods to be able to add one additional round without extending the duration of the experiment excessively. Finally, since post-shock phases are shorter in our study, the post-shock equilibrium in baseline groups is set at a higher level.

We implement a fixed matching protocol within each round, and a random rematching protocol between rounds. To reduce the scope of session effects due to random rematching, in each session we divide each group of twenty participants into two equal and permanent rematching clusters. Random rematching only occurs within a rematching cluster which makes observations potentially correlated within a cluster, but strictly independent between clusters.

### 3.3 Procedures

Experimental sessions were conducted at the GATE-Lab in Lyon by using z-Tree (Fischbacher, 2007).<sup>11</sup> 120 participants were recruited for 6 sessions in October 2019. Each session has 20 subjects recruited through a between-subjects design and divided into two separate rematching clusters of ten players.<sup>12</sup> This yields six independent clusters of observations per condition, and twelve clusters in total.

Before reading the instructions and playing the game, subjects solve a series of questionnaires which contain cognitive reflection test (CRT, Frederick 2005), reading the mind in the eyes test (RMET, Baron-Cohen et al. 2001) designed to measure one’s capacity to infer the internal emotional states of others, a short-term memory test (STM) adapted from Wechsler digit span test and seven questions designed to measure subjects’ propensity to reason in a heuristic manner. Two (five) of these last seven questions aim at measuring propensity

---

<sup>10</sup>They show that the effect of strategic environment is statistically significant for groups of five or more agents.

<sup>11</sup>All the experimental procedures have been approved by the GATE-Lab Review Board.

<sup>12</sup>See Appendix B for a description of the subject pool and Appendix A for all the experimental materials.

for representativeness (availability) heuristic (Kahneman and Tversky, 1972; Tversky and Kahneman, 1973).

After this phase, subjects are provided with the instructions of the game in paper form that are read aloud by the experimenter. These instructions specify all the rules of the game except the values of shocks. Participants are informed that this value will be displayed on the decision screen, may be subject to variation during the experiment, and that in a given period it remains the same for everyone. Once the instructions are read, subjects are asked to answer nine comprehension questions displayed on their screens. They are also informed about the correct answers with brief explanations.

In the main part of the experiment, each participant makes a series of 64 guessing decisions. Each time, subjects first see a decision screen where they enter their guesses. For a given guess, the computer automatically provides the corresponding *expectation of the average guess of others in their group*. After seeing these expectations, subjects can either revise or confirm their guesses.<sup>13</sup> After each decision screen, subjects pass to the feedback screen where they receive feedback on the realized target number and their own payoffs.

To smoothen the game, in each period the decision screen has a non-binding timer set to 60 seconds (except for the initial period of the game which has 120 seconds).<sup>14</sup> Both screens also display the ongoing period and round, and provide a summary of the previous outcomes: a figure representing the time series of previous guesses and realized target numbers as well as a historical period-by-period table containing own expectations and the actual average guesses of others, as well as own stage game payoffs.

Lastly, subjects reply to a demographic questionnaire. Experiments take 150 minutes on average. Subjects are payed 7 euros for their participation and 14.08 euros on average

---

<sup>13</sup>This method allows us to jointly elicit guesses and expectations in a consistent manner. The reason for not eliciting expectations through a second question is that some studies discover severe inconsistencies between the elicited expectations and the decisions (Costa-Gomes and Weizsäcker, 2008) and introducing both forecasting and optimizing tasks is found to be detrimental to learning and contributing to mispricing in LtFEs and AMEs (Bao et al., 2013; Hanaki et al., 2018).

<sup>14</sup>Kocher and Sutter (2006) show that average decision time in the first period of guessing games is around 50 seconds and decreases gradually until the range 10-15 seconds after 20 periods. Thus, we do not expect that this feature of our design puts participants under excessive time pressure.

for the experimental game. To minimize any potential wealth or end-game effects, the final payoff corresponds to the payoffs accumulated in all periods of a randomly chosen round of the game.

## 4 Results

First, we analyze the group-level deviations from the NE and measure the adjustment speed across rounds and experimental conditions. We rely on the statistical framework previously adopted by Cooper et al. (2017). Second, we compare the descriptive power of various expectation models by their one-period-ahead forecast accuracy, as measured through the root-mean-squared-error (RMSE). We also evaluate the changes in performance of models across rounds by computing their impact factors in the evolutionary learning model HSM. This allows us to investigate whether the observed acceleration is due to an increase in the share of subjects forming RE, or rather due to the adaptive dynamics of simple rules. Lastly, we estimate the relationship between subjects' cognitive abilities and (i) the accuracy of their predictions, and (ii) the predictive power of expectation models.

### 4.1 Aggregate Outcomes

We first turn to the aggregate outcomes. Figure 1 presents the median of group guesses across periods compared to NE benchmarks. Our first observation is that in the pre-shock phase of round 1, median guesses do not fully converge to the NE both in baseline and treatment.<sup>15</sup> The second salient observation is that median guesses tend to converge to the NE in each phase of the experiment with a convex pattern, though with different speeds.

For a formal statistical comparison of the patterns of convergence in both experimental conditions, let us define the speed of adjustment as the number of periods required for full adjustment to the NE in post-shock period  $t^c$  – i.e., a situation in which, for all  $t$  with

---

<sup>15</sup>We note that Cooper et al. (2017) observe the same pattern of adjustment under identical shocks.

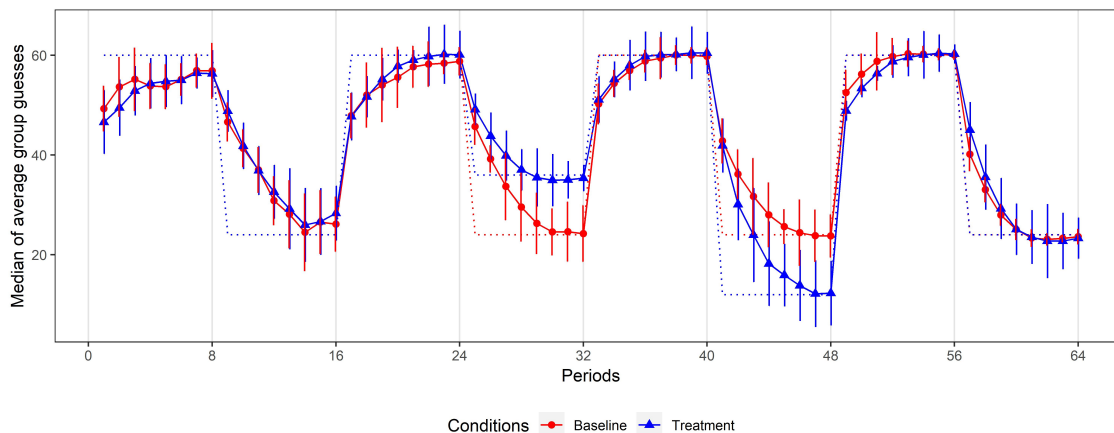


Figure 1: Median of group guesses by experimental conditions. Dots represent NE for experimental conditions. Circles (triangles) represent guesses in baseline (treatment) condition. Whiskers represent the standard deviation of group guesses. Unit of observation is the average guess of a group.

$t^c \leq t \leq 8$ , we fail to reject the hypothesis:  $H_0 : \bar{p}_t - p_t^{NE} = 0$ .<sup>16</sup> Following Cooper et al. (2017), we use median quantile regression and estimate the following model to identify the adjustment period:<sup>17</sup>

$$\bar{p}_{g,t} - p_t^{NE} = b_t \text{Period}_t + \epsilon_{g,t}, \quad (3)$$

where the dependent variable is the difference between the average group guess and the NE in a given period, while the independent variables are 63 indicator variables for time periods. The coefficient of the final period is dropped to avoid linear dependencies.<sup>18</sup> We run this regression separately for baseline and treatment. Table 2 reports the estimated coefficients for post-shock periods in which the NE remains the same.

<sup>16</sup>This definition echoes the definition of convergence proposed by Hyndman et al. (2012).

<sup>17</sup>This method minimizes the role of outliers. Robustness analyses with FE regressions are provided in Appendix C.1.

<sup>18</sup>Estimated standard errors take into account the possible correlation between guesses through clustering with bootstrap resampling. We followed the recommendations of Davidson and MacKinnon (2000) for choosing the number of bootstrap samples.

**Table 2 - Post-shock deviations from NE**

Post-shock periods	Baseline				Treatment	
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 4
1	22.756***	22.076***	19.310***	18.376***	24.886***	20.862***
2	17.880***	16.130***	13.190***	9.456***	17.950***	12.316***
3	13.932***	10.130***	8.384***	4.570***	13.154***	5.702***
4	7.390***	6.380***	4.656***	1.510***	9.318***	1.644*
5	5.216**	3.016***	2.384***	-0.216	6.670**	0.168
6	1.642	1.142	0.990	-0.540	3.500	-0.822
7	3.404	1.260	0.370	-0.444*	3.984	-0.802*
8	2.648	0.736	0.148	-	3.180	-

Note: Coefficients from median quantile regression of identification (3) are reported. Standard errors are clustered at rematching cluster level (2 clusters per session) and bootstrapped with 1999 replications. \*, \*\*, \*\*\* indicate statistical significance at the 10, 5, 1% levels, respectively.

In accordance with Hypothesis 1, we observe gradual adjustment in the first round of both baseline and treatment conditions. Adjustment accelerates across rounds, but this acceleration is not necessarily monotonic. The adjustment periods  $t^c$  are respectively  $\{6, 6, 6, 5\}$  in baseline and  $\{6, 4\}$  in the first and the last rounds of treatment when a 5% significance level is considered. Initial deviations remain high in each post-shock phase, indicating that experience does not prevent deviations despite a repetition over four rounds.

**Result 1a:** In both experimental conditions, after the initial shock, guesses gradually adjust toward the post-shock NE in a convex manner.

**Result 1b:** In both experimental conditions, full adjustment occurs earlier in response to the last shock (round 4) compared to the initial shock (round 1).

Notwithstanding our Hypothesis 2, these results indicate that the number of periods required for full adjustment in round 4 is similar in both conditions. We also propose another way to test (and eventually reject) Hypothesis 2. First, we investigate the within-period variation of guess between the two conditions. We estimate the following median quantile regression model separately for each of the eight post-shock periods of round 4:

$$p_i = \alpha_0 + \alpha_1 Treatment + \epsilon_i, \tag{4}$$



where the independent variable is an indicator variable that takes the value 1 if observation comes from treatment sessions and the dependent variable is individual guesses ( $N = 120$  per regression). Like before, we employ bootstrapped standard errors clustered at rematching cluster level (999 replications). The coefficients  $\alpha_1$  remains insignificant in each of the eight models, suggesting that the evolution of guesses over time (and thus their gradual adjustment to the NE) in the final periods is statistically indistinguishable in both conditions.

**Result 2:** As compared to identical shocks, nonidentical shocks do not cause a slow down in adjustment.

## 4.2 Expectation Formation

In this section, we consider several expectation models to provide a descriptive explanation for the observed aggregate results. These rules (summarized in Table 3) have been selected for two main reasons. First, they are commonplace in the literature (see Section 2). Second, they are based on backward-looking heuristics (with the exception of rule 12), so that their functional forms are simply calculable. We predetermine the parameters of these rules based on their computational ease as an attempt to imitate different kinds of boundedly rational reasoning.<sup>19</sup>

---

<sup>19</sup>Note that our design does not allow players to form level- $k$  type of expectations as proposed in the rule learning model of Stahl (1996), since there is no common prior through level-0 type can form imitation after the first period. Alternatively, we could assume that level-0 type always selects randomly; however, this would be unreasonable when imitation dynamics and evolution over time are considered.

**Table 3 - Description of selected expectation rules for comparison**

No	Description	Functional form
1	Naïve expectations	$p_{i,t}^e = p_{-i,t-1}$
2	Sticky beliefs	$p_{i,t}^e = p_{i,t-1}^e$
3	Adaptive expectations v1.	$p_{i,t}^e = 0.75p_{-i,t-1} + 0.25p_{i,t-1}^e$
4	Adaptive expectations v2.	$p_{i,t}^e = 0.50p_{-i,t-1} + 0.50p_{i,t-1}^e$
5	Adaptive expectations v3.	$p_{i,t}^e = 0.25p_{-i,t-1} + 0.75p_{i,t-1}^e$
6	Strong trend-following exp.	$p_{i,t}^e = p_{-i,t-1} + 0.75(p_{-i,t-1} - p_{-i,t-2})$
7	Medium trend-following exp.	$p_{i,t}^e = p_{-i,t-1} + 0.50(p_{-i,t-1} - p_{-i,t-2})$
8	Weak trend-following exp.	$p_{i,t}^e = p_{-i,t-1} + 0.25(p_{-i,t-1} - p_{-i,t-2})$
9	Strong contrarian exp.	$p_{i,t}^e = p_{-i,t-1} - 0.75(p_{-i,t-1} - p_{-i,t-2})$
10	Medium contrarian exp.	$p_{i,t}^e = p_{-i,t-1} - 0.50(p_{-i,t-1} - p_{-i,t-2})$
11	Weak contrarian exp.	$p_{i,t}^e = p_{-i,t-1} - 0.25(p_{-i,t-1} - p_{-i,t-2})$
12	Fundamentalist	$p_{i,t}^e = p_t^{NE}$
13	Similarity-based naïve exp.	$p_{i,t}^e = \begin{cases} p_{-i,t-1}, & \text{if } \xi_t = \xi_{t-1} \\ p_{-i,t-m}, & \text{if } \xi_t \neq \xi_{t-1} \end{cases}$

Note: Each description provides a rule for how players expect the average guess of other players one period ahead such that:  $p_{i,t}^e = E_{i,t-1}[p_{-i,t}]$ .

Accordingly, rules 1 and 2 do not require any computation and thus should involve little cognitive effort.<sup>20</sup> Rules 3 to 5 are variations of adaptive expectations heuristic where the weights put on the most recent outcome and player's previous expectation are changing. Rules 6 to 11 are variations of extrapolative expectations where a player tracks the most recent change in the realized outcome. If the coefficient of the change component is positive (negative), then this becomes a trend-following (contrarian) expectation. In rule 12, a player expects the NE outcome to happen. Nevertheless, this rule is not equivalent to RE since it ignores the fact that other players might be non-rational.<sup>21</sup>

We also define one more rule: the SBNE (rule 13). According to this rule, a player expects the outcome of last period to reoccur as in the case of naïve expectations if the parameters in the target formula did not change (i.e., if there were no shock). Once the target formula is changed, the player reviews all the past periods and expects the outcome of the most recent period  $t - m$  where the same change has happened, such that  $\xi_t - \xi_{t-1} = \xi_{t-m} - \xi_{t-m-1}$ . If

<sup>20</sup>Note that the nomenclature of the rules may vary across fields. For instance, rule 1 is equivalent to Cournot play in the context of game theory and to random walk believers in the context of finance.

<sup>21</sup>Yet, fundamentalism matches with RE when one assumes homogeneous expectations and common knowledge of rationality.

there is no such a period, the player expects the outcome of last period nevertheless. This reformulation transforms naïve expectations into a learning rule.

We determine which rule best fits the data by computing one-period ahead forecast deviations through RMSE. A low RMSE indicates a good fit with the data. We use the following RMSE formula:

$$RMSE(p_t^h) = \sqrt{\frac{\sum_{t=3}^{64} \sum_{g=1}^G (p_{g,t}^h - p_{g,t}^e)^2}{62 \times G}}, \quad (5)$$

where  $p_{g,t}^h$  is the prediction of rule  $h \in \{1, \dots, 13\}$  for the average expectation of group  $g$  in period  $t$  and  $p_{g,t}^e$  is the actual average expectation of group  $g$  in period  $t$ . Here, the superscript  $G$  is the scale of RMSE. We measure the RMSE in three different ways: at the cluster level ( $G = 2$ ), for each experimental condition ( $G = 12$ ), and for the pooled data ( $G = 24$ ). We exclude the data from the first two periods since certain rules require at least 2 past observations.

Panel A of Table 4 reports the calculated RMSE for each of the thirteen rules. At the pooled and experimental conditions level, the best fit is achieved with rule 13. At the cluster level, that rule provides the best fit for most (7 out of 12) clusters. Fundamentalism (rule 12) has always the worst fit regardless the scale of RMSE. The second best fitting rule is the weak-trend following rule (rule 8) which best describes 4 of the treatment clusters. The SBNE rule (rule 13) clearly describes the expectations of subjects in baseline clusters better than any other rule in accordance with Hypothesis 4. Moreover, this rule doubles the fit at pooled level when it is compared with fundamentalist expectations and increases the fit by 12% compared to the second-best homogeneous expectation rule (rule 8). As expected, rules 1 and rule 13 provide similar fits among treatment clusters, since only one of the shock repeats itself identically in this condition. However, the use of SBNE enhances the fit on average by 24% compared to regular naïve expectations in the baseline clusters.

These comparisons are done under the assumption that all players refer to the same

expectation rule (i.e., under the homogeneity of expectations). There is, however, a wide range of evidence indicating heterogeneity in expectations (Hommes, 2011). Moreover, expectation rules may vary not only across individuals, but also over time. Time variations in performance of expectation models is highly relevant for our hypotheses: a given rule (say, RE) may initially perform poorly, but then become more relevant thanks to the experience. For this reason, we refer to an evolutionary model of expectations: the HSM of Anufriev and Hommes (2012) that inherits its features from Brock and Hommes (1997). According to the HSM, agents choose the expectation rule from a set of heuristics, evaluate the performance of each heuristic over time and switch to the heuristic that performs best in terms of the forecasting error. Accordingly, the one-period-ahead expectation of the HSM for group  $g$  is

$$p_{t+1}^{HSM} = \sum_{h=1}^H n_{h,t} p_{t+1}^h, \quad (6)$$

where  $n_{h,t}$  is the impact factor of heuristic  $h$  at period  $t$ . This impact factor can be regarded as the weights attributed by agents to different heuristics. The impact factor depends on the performance of the heuristic measured with the current and past squared forecast errors

$$U_{h,t} = -(\bar{p}_{-i,t} - p_t^h)^2 + \eta U_{h,t-1}, \quad (7)$$

where  $\eta \in [0, 1]$  is a free parameter representing the weight assigned to the past performance compared to current. When  $\eta = 0$ , then only the performance in the most recent period matters. The impact of heuristic is updated through a discrete choice model with asynchronous updating described by

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

where the impact of the expectation heuristic  $h$  at period  $t$  depends on its accumulated impact and its relative performance normalized with sum of all competing heuristics. There

are two free parameters in (8). The first one,  $\delta \in [0, 1]$  represents the proportion of agents who do not update their heuristic each period, or the individual inertia in beliefs. The second parameter  $\beta > 0$  represents agents' sensitivity toward differences in performances where  $\beta = 0$  indicates equal impacts regardless of the differences in performances.

To compute expectations with HSM and compare fits, one must determine the set of expectation heuristics for comparison, assign their initial impacts (e.g.  $n_{h,3}$ ) and assign values to the free parameters  $\eta, \delta, \beta$ . We compare four heuristics ( $H = 4$ ) in accordance with the literature. We choose the best performing 3 heuristics under homogeneous expectations (rules 3, 8, 13) and the worst performing heuristic that is fundamentalism (rule 12). We assign the initial impact factors equal as  $n_{h,3} = 0.25$  for all  $h$  and set the free parameters as  $\eta = 0.1, \delta = 0.4, \beta = 0.1$ . By trial and error, we discover that this combination of free parameters fits the data best.<sup>22</sup>

Panel B in Table 4 reports the computed RMSE with the defined HSM. RMSE are computed for periods 4-64 where expectations are endogeneously determined by the HSM, excluding period 3. Results show that the HSM fits better than all other rules under homogeneous expectations at all scales of RMSE. The improvement in fits ranges between 16% and 59% depending on the scale and the improvement at pooled level is as high as 33%.

Lastly, we evaluate the impact factors of heuristics across time. Figure 2 reports these impact factors measured at both baseline and treatment scales. Both subfigures show that fundamentalist expectations attract very low weight in the first two rounds and gain slight improvement toward the last round. The SBNE rule attracts very large weights subsequent to changes in target formula during the last two rounds. Moreover, it peaks similarly in both baseline and treatment conditions, but its average weight during the post-shock phases are higher in baseline compared to treatment after the first round.

**Result 3:** Backward-looking expectation rules describe the expectations better than RE.

---

<sup>22</sup>We compute RMSE with various combinations of free parameters as robustness check and provide the results in Appendix C.2. While numerical values differ, the ranks of rules do not vary excessively. In fact when the  $\eta$  parameter increases to 0.7, the share of SBNE jumps to 1 in the beginning of round 3 and remains there afterwards.

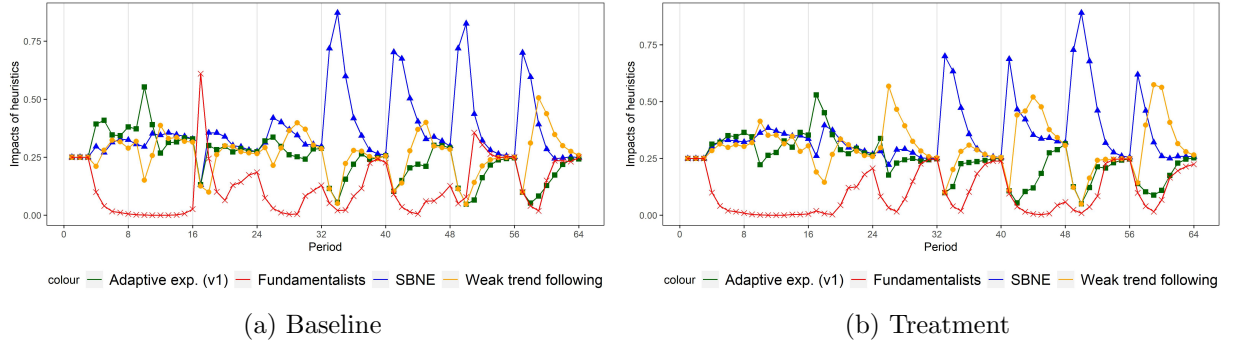


Figure 2: Impact factors of the HSM across periods and experimental conditions. The symbols square, cross, triangle and circle represent respectively adaptive expectations (v1), fundamentalist, SBNE and weak trend following expectations.

**Result 4:** The SBNE rule has the best fit among all the expectation rules for the pooled data. The SBNE rule (weak-trend following rule) best describes the outcomes in all the (4 out of 6) clusters in the identical (nonidentical) shocks condition.

### 4.3 Cognitive Skills and Individual Expectations

In this section, we ask whether individual scores on cognitive skills tests predict the accuracy of expectations and the best fitting expectation rule. For this sake, we compute the average relative prediction error (ARPE) in a given round

$$\frac{1}{16} \sum_{t=1}^{T=16} \frac{|p_{i,t}^e - \bar{p}_{-i,t}|}{\bar{p}_{-i,t}}, \quad (9)$$

as well as the goodness of fit of the different expectation models per subject, as measured through RMSE. We then regress each of these measures on the set of subjects' test scores. All test scores are standardized as  $\frac{Score_i - \min(Score)}{\max(Score) - \min(Score)}$ . Round and treatment dummies, as well as their interactions, are also included in the model. There are 4 observations per subject, and 480 observations in total. We use a random effects specification.

**Table 4 - RMSE of expectation rules**

Rule	Baseline: cluster-by-cluster						Treatment: cluster-by-cluster						All-cluster		
	1	2	3	4	5	6	1	2	3	4	5	6	B	T	B+T
	Panel A - Homogeneous expectation rules														
1	6.45	6.59	5.71	7.24	5.80	8.18	6.71	6.23	6.53	6.50	7.21	6.56	6.71	6.63	6.67
2	7.55	7.02	6.49	8.14	7.14	8.28	7.55	7.48	7.86	7.56	7.80	7.57	7.47	7.64	7.55
3	6.65	6.58	5.78	7.31	6.03	8.05	6.83	6.42	6.74	6.66	<b>7.12</b>	6.66	6.78	6.74	6.76
4	6.91	6.65	5.93	7.49	6.34	8.02	7.01	6.71	7.04	6.90	7.20	6.87	6.93	6.96	6.94
5	7.21	6.80	6.17	7.77	6.71	8.10	7.26	7.06	7.42	7.20	7.43	7.17	7.16	7.26	7.21
6	7.64	8.33	7.43	9.30	6.97	9.85	8.40	7.75	6.84	7.69	9.39	7.78	8.32	8.01	8.17
7	6.68	7.33	6.45	8.24	6.04	8.83	7.32	6.77	6.15	6.74	8.34	6.87	7.33	7.07	7.20
8	6.25	6.71	5.83	7.52	5.62	8.24	6.72	<b>6.23</b>	<b>6.04</b>	<b>6.31</b>	7.58	<b>6.44</b>	6.76	6.57	6.67
9	9.87	9.00	8.07	9.12	9.09	10.85	9.74	9.02	10.32	9.90	8.68	9.59	9.37	9.56	9.47
10	8.41	7.84	6.94	8.10	7.69	9.58	8.36	7.75	8.81	8.45	7.80	8.26	8.13	8.25	8.19
11	7.22	6.99	6.11	7.44	6.55	8.65	7.30	6.77	7.51	7.26	7.29	7.21	7.20	7.23	7.22
12	10.02	10.10	11.25	14.29	11.55	13.80	10.68	12.32	12.47	11.07	17.07	13.65	11.95	13.05	12.51
13	<b>4.61</b>	<b>4.58</b>	<b>4.00</b>	<b>6.62</b>	<b>4.08</b>	<b>6.13</b>	<b>6.14</b>	6.23	6.47	6.58	7.26	6.47	<b>5.10</b>	<b>6.54</b>	<b>5.86</b>
	Panel B - Heterogeneous expectation model														
HSM	2.45	3.45	3.37	5.46	2.56	5.08	2.52	3.53	4.33	2.59	5.42	4.29	3.90	3.92	3.91
$\Delta$ SBNE	26%	30%	30%	9%	27%	24%	8%	0%	-7%	-4%	-2%	-1%	24%	1%	12%
$\Delta$ HSM	47%	25%	16%	18%	37%	17%	59%	43%	28%	59%	24%	33%	24%	40%	33%

Note: RMSE at different scales reported. The first column indicates the number assigned to the rule in Table 3 for panel A. Columns 2 to 7 are cluster numbers that were in baseline condition and columns 8 to 13 are cluster numbers in the treatment sessions. The last three columns indicate the experimental conditions where "B" refers to baseline ( $G = 6$ ), "T" refers to treatment ( $G = 6$ ) and "B+T" refers to pooled data ( $G = 12$ ). Bolded values in panel A indicate the lowest RMSE among homogeneous expectation rules. HSM in panel B refers to RMSE with the HSM.  $\Delta$ SBNE reports the improvement in fit by using SBNE compared to the best alternative homogeneous expectation rule. Negative  $\Delta$ SBNE implies losses in fit.  $\Delta$ HSM reports the improvement in fit by using the HSM compared to the best fitting homogeneous expectation rule.

**Table 5 - Prediction errors and RMSE explained by test scores**

Variables	ARPE	RMSE of rule			
		Rule 13	Rule 8	Rule 3	Rule 12
Intercept	0.224*** (5.39)	10.524*** (3.13)	9.348** (2.41)	8.943** (2.52)	19.144*** (6.77)
Baccalaureate grade	0.016 (0.85)	2.990** (2.44)	3.373** (2.32)	3.092** (2.22)	2.347* (1.78)
CRT score	-0.020 (-1.36)	-1.531 (-1.62)	-1.546 (-1.60)	-1.493* (-1.73)	-0.811 (-1.09)
RMET score	-0.071** (-2.30)	-2.258 (-1.16)	-1.969 (-0.98)	-1.327 (-0.69)	-5.411*** (-2.72)
Memory score	-0.021 (-0.56)	-2.947 (-0.86)	-1.260 (-0.35)	-0.958 (-0.28)	-2.547 (-0.72)
Represent. score	0.026 (1.61)	0.812 (0.66)	0.479 (0.40)	0.114 (0.10)	2.888*** (3.24)
Availability score	-0.018 (-0.79)	-0.751 (-0.36)	-0.703 (-0.32)	-0.244 (-0.12)	-0.489 (-0.36)
Treatment	0.002 (0.07)	-0.753 (-0.53)	-0.908 (-0.60)	-0.652 (-0.47)	0.024 (0.02)
Round 2	-0.043** (-2.28)	0.744*** (0.63)	0.640 (0.47)	0.586 (0.46)	-2.258** (-2.28)
Round 3	-0.066*** (-2.64)	-1.695 (-1.48)	0.507 (0.41)	0.458 (0.38)	-4.384*** (-4.59)
Round 4	-0.107*** (-4.97)	-3.672*** (-2.86)	-0.308 (-0.25)	-0.172 (-0.15)	-8.056*** (-9.21)
Round 2 × Treatment	-0.023 (-1.11)	-0.179 (-0.14)	0.000 (0.00)	-0.129 (-0.09)	-2.977*** (-2.56)
Round 3 × Treatment	0.065* (1.82)	2.285 (1.47)	-0.141 (-0.09)	0.104 (0.07)	4.382** (2.56)
Round 4 × Treatment	0.039 (1.58)	6.471*** (4.62)	3.205** (2.25)	2.972** (2.23)	3.243*** (2.61)
$R^2$ within	0.224	0.089	0.033	0.031	0.327

Note: In this table, the coefficients of random effects regressions are reported. The dependent variable is the ARPE in the first model and RMSE of the specified rule per subject in the models 2-5. The independent variables are standardized test scores, round and treatment dummy variables and their interactions.  $z$ -statistics in parentheses. Robust standard errors clustered at rematching cluster level (2 clusters per session) are reported. \*, \*\*, \*\*\* indicate statistical significance at the 10, 5, 1% levels, respectively.

Table 5 reports the corresponding estimates. Overall, the included test scores only weakly explain the variation in either measure of interest. Higher RMET score predicts lower prediction error and better fit for fundamentalist rule. By contrast, in groups where subjects are more prone to representativeness heuristic, Nash play is less likely to occur.



The impact of round dummies are more systematical on explaining the dependent variables. In round 4 compared to round 1, ARPE is likely to be lower and the fits of the SBNE rule and the fundamentalism improve. Note that even with the improvement in fit at round 4, the fundamentalist rule still performs worse than the other rules. The treatment dummy has no significant impact on any variable, but its interaction with round 4 worsens the fit of all rules. So, the improvement in the goodness of fit over time is attenuated under nonidentical shocks. The coefficient of this interaction term is also positive for ARPE (in line with Hypothesis 2), but not statistically significant. Lastly, the RMSEs of rules 3 and 8 seem to be unaffected by experience.

**Result 5a:** Cognitive abilities do not capture major variation in the performance of rules and the prediction accuracy. Higher RMET (representativeness) score predicts better (worse) fit for the fundamentalist rule. Higher RMET also indicates more accurate expectations.

**Result 5b:** Descriptive power of the SBNE and fundamentalist rules, as well as the prediction accuracy improve from the first round to the last round. However, these improvements are weaker in nonidentical shocks condition.

## 5 Discussion

The results in Section 4.2 reveal that the SBNE rule explains the expectation process well both in homogeneous and heterogeneous expectations models, especially when the shocks are identical. Under this rule, a player expects the last period outcome to reoccur in stationary phases as it is the case in naïve expectations. In case of a change in the environment, this expectation rule points to the outcome of the most recent period where the same change has happened. This kind of rule is consistent with the Gilboa and Schmeidler (1995) case-based decision theory and the Plonsky et al. (2015) similarity-based learning model, both of which suggest that agents choose the action which generated the best outcome under similar circumstances that an agent can recall from the past. Thus, the SBNE rule

can be viewed as a combination of the similarity-based learning process with naïve heuristic that is applied to the domain of expectations formation.

This framework proposes a potential explanation of why the trend following rule with a strong extrapolation parameter fits the data well in LtFEs and poorly in guessing games like ours. The main difference between LtFEs and guessing games is that while the LtFEs inform subjects only qualitatively on the data generating process, guessing games provide quantitative information by disclosing the target formula. If there are also unexpected large shocks as in the LtFEs of Bao et al. (2012), the last period is less likely to be perceived as the most similar state. A stronger extrapolation of recent changes may thus help detect the arrival of a shock. This strong extrapolation also creates a self-fulfilling prophecy, since it endogenously generates large oscillations around equilibrium. In guessing games, players may judge the similarity with certainty which decreases the necessity of extrapolation. Naïve and weak trend following rules thus perform better. This may explain why a trend model with extrapolation factor 1 performed poorly in the AMEs of Marquardt et al. (2019). Weak trend following and naïve heuristics create slow convergence toward NE similar to what we observe in our experiment.

## 6 Conclusion

We investigate the evolution of adjustment speed across repeated identical and nonidentical shocks. We investigate whether adjustment accelerates over repetitions and whether this acceleration varies across the different types of shocks. We find that adjustment accelerates thank to repetition, yet only slightly: despite four repetitions, converge speeds up by only two periods at best. Nonidentical shocks do not impact adjustment speed, and adjustment acceleration remains weak regardless the type of shock. A descriptive analysis of the expectation formation process reveals that backward-looking rules organize the data well and the SBNE rule, a learning rule proposed by Cooper et al. (2017), best fits the data. Individual measures

of cognitive skills do not capture much of the variation in the goodness of fit of these rules of subjects' expectations.

Our experiment successfully replicates the finding of Cooper et al. (2017) from a guessing game with strategic complementarity: a gradual and convex adjustment in response to identical shocks and acceleration over repetition. The evidence strengthens the empirical validity of the strategic environment effect, in line with an early conjecture by Haltiwanger and Waldman (1989). The patterns of learning to play equilibrium under strategic complementarity remain unchanged in a more complex environment in which we introduce time-varying shocks into the equilibrium. By simultaneously comparing a large set of expectation rules that were proposed to be explanatory in the literature, we are able to provide an individual-based explanation to the observed aggregate dynamics. The SBNE rule, a recent learning rule that was not previously studied in a comparative analysis, is found to outperform all other rules and arise as a promising behavioral expectation model in non-stationary environments.

The main implication of these findings is that inertia in adjustment is not a phenomenon that fades away with repetition, but may rather persist over time. The fact that the type of shock has not affected behavioral dynamics suggests that sluggishness is inherent to strategic complementarity, rather than being a case-specific effect. Our design also does not include any market frictions that are usually considered as the main drivers of sticky behavior. This, in turn, suggests that cognitive frictions such as non-rational expectations suffice to create stickiness, and potentially opens doors for policy implications. Although an experimental testbed for policy instruments is beyond the scope of this study, we note that monetary policy interventions may prove to be effective.<sup>23</sup>

A second implication rises from the fact that the SBNE rule explains the data well in our study. Similarity-based reasoning implies that agents review the past and form expectations

---

<sup>23</sup>Cornand and Heinemann (2019) show that in a New Keynesian framework, monetary policy obeying the Taylor principle decreases the degree of complementarity between pricing decisions of firms and even turns them into strategic substitutes if its effect on aggregate demand is sufficiently strong. In a similar vein, Assenza et al. (2019) show through their New Keynesian LtFEs that the Taylor principle with sufficiently strong interest rate rule ( $\xi_\pi = 1.5$  in their experiment) manages convergence to the forward stable solution.

that are continuation of the most similar past periods. In case of large shocks, if judging the similarity between changes is possible, more experience would result in faster adjustment according to this reasoning. On the other hand, if agents cannot judge the similarity due to underprovision of information, referring to a heuristic which takes the moving average of longer length (e.g., least squares learning) or extrapolating changes more strongly (e.g., strong trend following) can be more tempting for coping with uncertainty. We leave the test of this argument for future research: if this were warranted, then emphasizing the similarities between shocks may prove a fruitful policy tool.

Despite its virtues, our study may also have certain limitations. Here, we set the number of shocks to four and one may claim that this is not enough for a major acceleration. While this might be a limitation of experiments in general, four rounds should be sufficient for observing accelerated adjustment since the environment is a lot simpler than the real market phenomena. Another design decision is to rematch subjects in the beginning of each new round to control for factors that are accumulating across rounds such as the degree of strategic uncertainty. This random rematching mechanism might partially be the reason for limited acceleration and it may stand at odds with certain real-world environments, such as asset markets. Lastly, we only look at negative shocks. Even though the sign of a shock should not matter in guessing games, it may matter in a pricing context and the set of shocks we study may not be sufficiently heterogeneous. We believe that this constitutes a possible agenda for future studies.

# Appendices

## A Experimental Material

### A.1 Instructions and Comprehension Questions in Original Language

#### Instructions Générales

L'expérience comprend **4 tours** et chaque tour comprend **16 périodes**, soit un total de 64 périodes. Au début de l'expérience, vous serez assignés au hasard en **groupes de cinq**. Vous n'interagirez qu'avec les autres joueurs de votre groupe. Au début de chaque nouveau tour, les groupes seront reconstitués de façon aléatoire. Cela signifie que vous jouerez au sein du même groupe pendant un tour, et que la composition de votre groupe variera aléatoirement d'un tour à l'autre.

#### Votre Tâche

Au début de chaque période, il sera demandé à chaque participant de choisir un nombre entre 0 et 100, inclus. Ce nombre peut comporter jusqu'à deux décimales, par exemple 11.35 ou 95.23. Aucun participant ne pourra voir le nombre choisi par un autre participant.

À chaque période, chaque joueur a un "Nombre cible". À la fin de chaque période, le joueur de votre groupe **dont le nombre choisi est le plus proche de son nombre cible** remportera le gain de 4.40 euros pour cette période. Les autres joueurs gagneront 0 euro pour cette période. Si plusieurs joueurs se trouvent à la même distance de leur nombre cible, le gain de 4.40 euros est divisé à part égales entre ces joueurs, tandis que les autres gagnent 0 euro.

Le nombre cible de chaque joueur est calculé à l'aide de la formule suivante :

Nombre cible =  $0.75 \times$  (moyenne des nombres choisis par les autres joueurs de votre groupe) + une constante

Ici, "la moyenne des nombres choisis par les autres joueurs de votre groupe" est égale à la somme des nombres choisis par les autres joueurs de votre groupe divisé par quatre. Cette

moyenne est calculée de la même façon pour tous les participants à l'expérience. Tous les participants seront informés sur la constante par le biais de leur écran de décision. Cette constante est la même pour tous les participants mais peut changer au cours de l'expérience. Lorsqu'un changement survient, ce changement est annoncé à tous les participants à l'écran. Veuillez vérifier la formule à chaque période.

### **Écran de Décision**

La formule du Nombre cible vous est indiquée à l'écran. Sur cet écran, vous pouvez entrer votre décision dans une cellule. Lorsque vous cliquez sur le bouton "OK", le programme vous montre la "Moyenne des nombres choisis par les autres joueurs de votre groupe" pour laquelle votre décision correspondrait au nombre cible. Après avoir vu cette information, vous pouvez modifier votre décision autant de fois que vous le souhaitez. Une fois que vous avez cliqué sur "Confirmer" votre décision pour cette période sera définitive.

Notez qu'il y a une **limite de temps** pour les décisions à chaque période et vous pouvez suivre le temps restant sur votre écran. Vous aurez 120 secondes pour votre première décision et **60 secondes** pour les décisions de toutes les périodes restantes. Un tableau et une figure vous permettent également de suivre **vos décisions précédentes** et **la décision moyenne précédente des autres joueurs** de votre groupe.

### **Paiement**

A la fin de l'expérience, l'ordinateur sélectionnera au hasard l'un des tours joués, et votre paiement final sera basé sur les gains que vous avez accumulés au cours de ce tour, plus  $5 + 2 = 7$  euros pour la participation et le questionnaire auquel vous venez de répondre.

### **Questionnaire de Compréhension**

Vous allez maintenant répondre à un questionnaire de compréhension des règles du jeu. Un bouton se situant au milieu de l'écran vous permettra d'accéder à une calculatrice lorsque vous en aurez besoin.

### **True of False Questions:**

Question 1: Il y a 4 autres joueurs dans mon groupe.

Question 2: Je joue avec le même groupe de joueurs pendant toute l'expérience.

Question 3: La formule pour le nombre cible peut changer au cours de l'expérience.

Question 4: Tous les joueurs ont leur propre formule pour le nombre cible.

Question 5: Je serai payé en fonction de mes gains accumulés au cours d'un tour choisi au hasard.

### **Questions Based on an Example:**

Imaginez que la formule pour le nombre cible soit égale à

$$\text{Nombre cible} = 0.75 \times (\text{Moyenne des nombres choisis par les autres joueurs}) + 15$$

Question 6: Si les autres joueurs dans le groupe ont choisi 10, 30, 35, 85 comme décisions pour le nombre cible, quel est selon vous la "Moyenne des nombres choisis par les autres joueurs de votre groupe" ?

Question 7: À quoi votre nombre cible serait-il égal dans cette situation?

Question 8: Imaginez que vous avez choisi le nombre 55 comme décision pour cette période. Quelle est la distance entre le nombre cible et votre décision?

Question 9: Dans cet exemple, les distances entre les nombres choisis et les nombres cibles pour les autres joueurs sont respectivement: 47.18, 17.5, 0.31 et 41.87.

Dans cet exemple, êtes-vous le gagnant?

### **Answers and Explanations Provided to Subjects:**

Question 1: Vrai.

Explication: Il y a 5 joueurs dans chaque groupe et 4 autres quand vous êtes exclu.

Question 2: Faux.

Explication: Au début de chaque nouveau tour (17ème, 33ème et 49ème périodes), les groupes seront reconstitués d'une manière aléatoire. Cela signifie que vous jouerez au sein du même groupe pendant un tour, et que la composition de votre groupe variera aléatoirement d'un tour à l'autre.

Question 3: Vrai.

Explication: La formule pour le nombre cible peut changer. Veuillez prêter attention au

début de chaque période.

Question 4: Faux.

Explication: La formule pour le nombre cible pour une période donnée est la même pour tous les joueurs.

Question 5: Vrai.

Explication: À la fin de l'expérience, un tour parmi 4 sera sélectionné au hasard et vous obtiendrez vos gains cumulés pour ce tour.

Question 6: 40.

Explication: La réponse correcte est 40. C'est la moyenne des autres joueurs du groupe, soit dans cet exemple :  $(10+30+35+85)/4 = 40$ .

Question 7: 45.

Explication: La réponse correcte est 45. Le nombre cible est calculé avec la formule pour le nombre cible, soit dans exemple :  $0.75 \times 40 + 15 = 45$ .

Question 8: 10.

Explication: La réponse correcte est 10. Le nombre cible est 45 et vous avez choisi 55. La distance entre ces deux nombres est égale à 10.

Question 9: Non.

Explication: Votre distance (10) n'est pas la plus petite. 0.31 est la plus petite distance dans ce groupe.

## A.2 Instructions and Comprehension Questions Translated to English

### General Instructions

The experiment has **4 rounds** and each round has **16 periods**, a total of 64 periods. At the beginning of the experiment, you will be randomly assigned in **groups of five**. You will only interact with other players in your group. At the beginning of each new round, the groups will be reconstituted in a random manner. This means that you will play in the same group during a round, and that the composition of your group will vary randomly from one



round to another.

### **Your Task**

At the beginning of each period, each participant will be asked to choose a number between 0 and 100, inclusive. This number can be up to two decimals, such as 11.35 or 95.23. No participant will be able to see the number chosen by another participant.

In each period, each player has a "Target Number". At the end of each period, the player in your group **whose chosen number is closest to his or her target number** will win the prize of 4.40 Euros for that period. The other players will earn 0 euro for this period. If several players have the same distance from their target number, the prize of 4.40 euros will be divided equally between these players, while the others would win 0 euros.

The target number of each player is calculated using the following formula:

Target number =  $0.75 \times$  (average of the numbers chosen by the other players in your group) + a constant

Here, "the average of the numbers chosen by the other players in your group" is equal to the sum of the numbers chosen by the other players in your group divided by four. This average is calculated in the same way for all participants in the experiment. All participants will be informed about the constant through their decision screen. This constant is the same for all participants but may change during the experience. When a change occurs, this change will be announced to all participants on their screen. Please check the formula for each period.

### **Decision Screen**

The target number formula is going to be shown on the screen. On this screen, you can enter your decision in a cell. When you click on the "OK" button, the program will show you the "Average of the numbers chosen by the other players in your group" for which your guessing decision corresponds to. After seeing this information, you can change your decision as many times as you want. Once you click on "Confirm" button your decision for this period will be final.

Note that there is **limited time** for decisions in each period and you can track the remaining time on your screen. You will have 120 seconds for your first decision and 60 seconds for each decisions of the remaining periods. A table and a figure also allow you to follow **your previous decisions** and **the previous average decision of the other players** in your group.

### **Payment**

At the end of the experiment, the computer will randomly select one of the rounds played, and your final payment will be based on the payoffs that you have accumulated during this round, plus  $5 + 2 = 7$  euros for participation and the questionnaire you answered.

### **Comprehension Questions**

You will now answer several questions designed to check whether you understood the rules of the game. The button in the middle of the screen will allow you to access a calculator when you need it.

### **True of False Questions:**

Question 1: There are 4 other players in my group.

Question 2: I play with the same group of players throughout the experience.

Question 3: The formula for the target number may change during the experiment.

Question 4: All players have their own formula for the target number.

Question 5: I will be paid based on my accumulated winnings during a randomly chosen round.

### **Questions Based on an Example:**

Imagine that the formula for the target number is equal to

Target number =  $0.75 \times (\text{Average of the numbers chosen by the other players}) + 15$

Question 6: If the other players in the group chose 10, 30, 35, 85 as decisions for the target number, what do you think is the "Average of the numbers chosen by the other players in your group"?

Question 7: What would your target number be equal to in this situation?

Question 8: Imagine that you chose number 55 as the decision for this period. What is the distance between the target number and your decision?

Question 9: In this example, the distances between the chosen numbers and the target numbers for the other players are respectively: 47.18, 17.5, 0.31 and 41.87.

In this example, are you the winner?

**Answers and Explanations Provided to the Subjects:**

Question 1: True.

Explanation: There are 5 players in each group and 4 others when you are excluded.

Question 2: False.

Explanation: At the beginning of each new round (17th, 33rd and 49th periods), the groups will be reconstituted in a random manner. This means that you will play with same group members during a round, and that the composition of your group will vary from round to round.

Question 3: True.

Explanation: The formula for the target number may change. Please pay attention in each period.

Question 4: False.

Explanation: The formula for the target number for a given period is the same for all players.

Question 5: True.

Explanation: At the end of the experiment, one of the four rounds will be randomly selected and you will get your winnings that are accumulated during this round.

Question 6: 40.

Explanation: The correct answer is 40. This is the average number of the other players in the group, in this example:  $(10 + 30 + 35 + 85) / 4 = 40$ .

Question 7: 45.

Explanation: The correct answer is 45. The target number is calculated using the formula

for the target number:  $0.75 \times 40 + 15 = 45$ .

Question 8: 10.

Explanation: The correct answer is 10. The target number is 45 and you have chosen 55.

The distance between these two numbers is 10.

Question 9: No.

Explanation: Your distance (10) is not the smallest. 0.31 is the smallest distance in this group.

### A.3 Experimental Interface

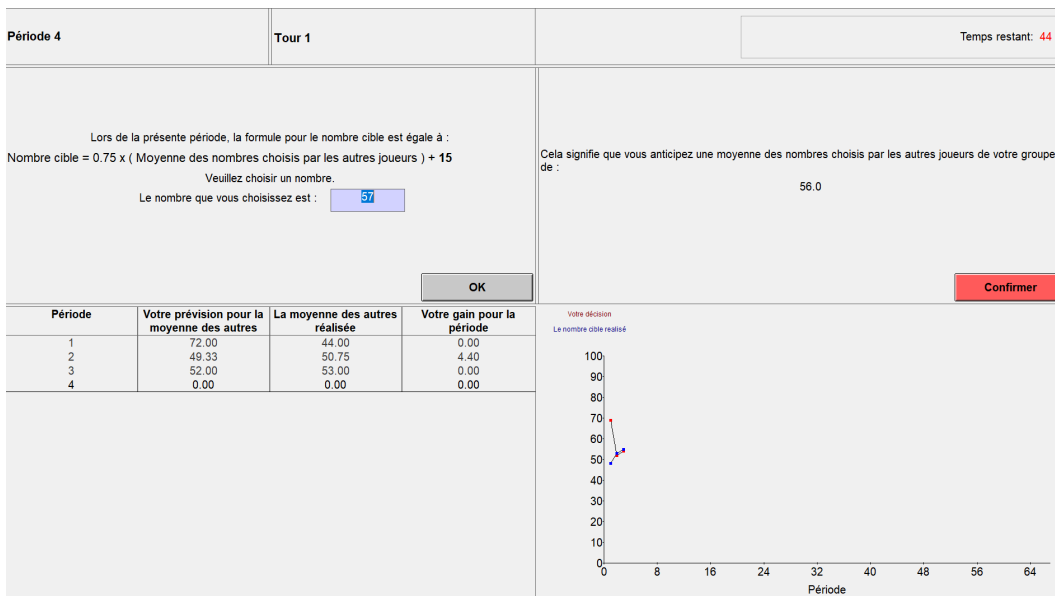


Figure 3: An (arbitrary) example for the decision screen that subjects used to enter their guess

Subject first had to enter their guess decision in the box on the upper left panel of decision screen and see their expectation that this guess corresponds to on the upper right panel. Once they click on "Confirmer", their decision becomes irreversible. For the codes and the translation of the material into English, contact one of the authors.

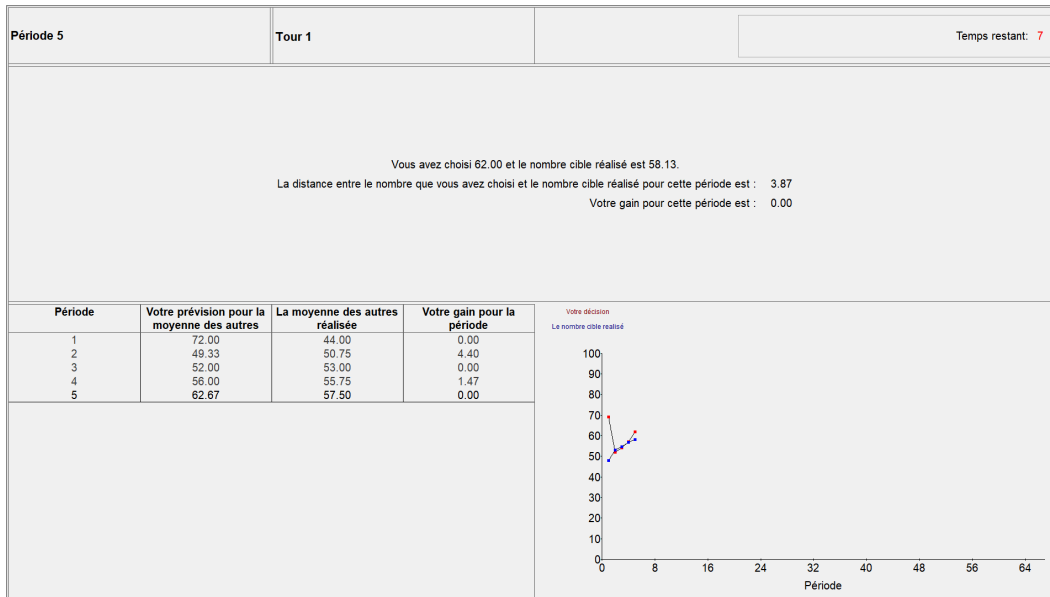


Figure 4: An (arbitrary) example for the feedback screen that subjects used to see the results of the period

## A.4 Tests in Original Language

Note that subjects solve these tests before the main part of the experiment. So, instructions presented here are the initial instructions that subjects see.

### Initial Instructions

Bienvenu!

Vous allez participer à une expérience économique. Pendant cette expérience, vous n'êtes pas autorisé à communiquer avec les autres participants. Si vous avez un téléphone portable, veuillez l'éteindre. Si vous avez une question, appuyez sur le bouton rouge à votre gauche ou levez la main, l'expérimentateur viendra vous voir; ne posez pas votre question à voix haute. Si la question est pertinente pour tous les participants, nous la répéterons et y répondrons à voix haute. Si vous ne respectez pas ces règles, nous devons vous exclure de l'expérience et donc du paiement.

Toutes les informations que vous fournissez, ainsi que le montant de vos gains lors de cette expérience, resteront strictement confidentielles et anonymes. La participation à cette expérience vous fera gagner de l'argent. Vos gains vous seront versés en espèces de manière

privée à la fin de l'expérience. Vous gagnez 5 euros pour vous présenter à temps, 2 euros pour répondre à une série de questions et un montant supplémentaire qui varie entre 0 et 70 euros. Le paiement supplémentaire dépendra de vos décisions et peut également être influencé par les décisions prises par les autres.

Tout d'abord, avant de commencer l'expérience à proprement parler, nous vous demandons de répondre à une série de questions préliminaires. Vous répondrez à ces questions à l'aide de l'interface sur votre écran d'ordinateur.

### **CRT Questions**

1) Un carnet et un crayon coûtent 1,10 Euros. Le carnet coûte 1 Euro de plus que le crayon. Combien de centimes coûte le crayon ? (correct answer: 0.05 cents)

2) En supposant qu'il faut 5 minutes à 5 machines pour fabriquer 5 stylos, combien de temps faudrait-il à 100 machines pour fabriquer 100 stylos? (correct answer: 5)

3) Sur un étang pousse un banc de nénuphars. La taille du banc double chaque jour. En supposant qu'il faille 48 jours pour que le banc recouvre tout l'étang, combien de temps faudrait-il pour que le banc recouvre la moitié de l'étang? (correct answer:47)

### **Representativeness Heuristic Questions**

Veillez lire les descriptions ci-dessous et répondre aux questions.

Description 1: Linda a 31 ans, elle est célibataire, franche et très brillante. Elle possède une maîtrise de philosophie. Étudiante, elle se montrait très préoccupée par les questions de discrimination et de justice sociale, elle participait aussi à des manifestations antinucléaires.

Veillez classer les énoncés suivants en fonction de leur probabilité, en utilisant 1 pour le plus probable et 7 pour le moins probable :

- 1) Linda est enseignante à l'école primaire.
- 2) Linda travaille dans une librairie et prend des cours de yoga.
- 3) Linda est active dans le mouvement féministe.
- 4) Linda est guichetière dans une banque.
- 5) Linda est assistante sociale en milieu psychiatrique.

6) Linda est vendeuse d'assurances.

7) Linda est guichetière dans une banque et est active dans le mouvement féministe.

Correct answer: If option 7 is judged more probable than option 3 and 4, then the answer is correct.

Description 2: Une certaine ville est desservie par deux hôpitaux. Environ 45 bébés naissent chaque jour dans le grand hôpital et environ 15 bébés naissent chaque jour dans le petit hôpital. Comme vous le savez, environ 50% des bébés sont des garçons. Le pourcentage exact de bébés garçons varie toutefois d'un jour à l'autre. Parfois, il peut être supérieur à 50%, parfois inférieur.

Pendant une période d'un an, chaque hôpital a enregistré les jours où plus de 60% des bébés nés étaient des garçons. Selon vous, dans quel hôpital le nombre de tels jours était-il le plus élevé?

A) Le grand hôpital

B) Le petit hôpital

C) Pareil pour les deux hôpitaux.

Correct answer: Option B is the correct answer.

### **Availability Heuristic Questions**

Ci-dessous, chaque élément comprend deux causes possibles de décès. La question à laquelle vous devez répondre est la suivante : parmi les deux causes possibles de décès, quelle est la plus fréquente, en général, en France ? Pour chaque paire de causes possibles de décès, (a) et (b), nous souhaitons que vous choisissiez la cause qui vous semble la plus fréquente.

Paire 1 (a) Accidents de la route (b) Diabète (correct answer: b)

Paire 2 (a) Homicide (b) Suicide (correct answer: b)

Paire 3 (a) Accident vasculaire cérébral (AVC) (b) Tous les accidents (correct answer: a)

Paire 4 (a) Des chutes (b) L'usage de drogues (correct answer: a)

Paire 5 (a) Foudroiement (b) Empoisonnement (correct answer: b)

### **Reading the Mind in the Eyes Test**

Vous allez voir maintenant une série d'images présentant des paires d'yeux, ainsi que 4 mots. Pour chaque paire d'yeux, choisissez le mot qui décrit le mieux ce que la personne sur l'image pense ou ressent. Vous pouvez estimer que plusieurs mots peuvent s'appliquer mais ne choisissez que le mot que vous considérez le plus approprié. Avant de faire votre choix, assurez-vous que vous avez bien lu les 4 propositions.

Il y aura au total 36 questions auxquelles vous devrez répondre dans un délai de 10 minutes. Essayez de répondre aussi rapidement et précisément que possible. Pour répondre, sélectionnez l'un des choix affichés sous l'image puis cliquez sur OK en bas. Veuillez noter qu'une fois que vous aurez cliqué sur OK, vous ne pourrez pas revenir aux questions précédentes.

Cliquez sur OK pour passer à la question d'essai.

Training question:



Options: 1:jaloux, 2:paniqué, 3:arrogant, 4:haineux

Correct answer and explanations:

Parmi les choix : "jaloux", "paniqué", "arrogant", "haineux", la bonne réponse était : "paniqué". Cliquez sur OK pour commencer les autres questions. A partir de maintenant, les bonnes réponses ne seront plus montrées.

Pour rappel, il y aura au total 36 questions auxquelles vous devrez répondre dans un délai de 10 minutes. Essayez de répondre aussi rapidement et précisément que possible.

### **Short Term Memory Test**

Vous allez maintenant regarder plusieurs séries des diapositives. Au début de chaque série, vous verrez le mot "Prêt ?" puis une séquence des numéros s'affichant les uns après les



autres. A la fin de chaque série, vous entrerez dans la zone de saisie la séquence de nombres que vous aurez observée.

Les séquences seront de plus en plus longues à mesure que vous entrez des réponses correctes. Votre objectif est d'aller le plus loin possible. Vous aurez deux essais. Par exemple, si dans une série vous voyez 1, puis 3 et puis 5, vous devez entrer 135.

First sequence: 69, 929, 1021, 34634, 943453, 7374865, 69358267, 690875725, 6457803021, 26456897198, 601518340985, 1285246589042.

Second sequence: 25, 217, 8618, 48629, 727240, 1203439, 32904142, 750572970, 1720378975, 62617825067, 609295956490, 1678606889148.

## A.5 Tests Translated to English

Note that subjects solve these tests before the main part of the experiment. So, instructions presented here are the initial instructions that subjects see.

### **Initial Instructions**

Welcome!

You will participate in an economic experiment. During this experiment, you are not allowed to communicate with other participants. If you have a cell phone, please turn it off. If you have a question, press the red button on your left or raise your hand, the experimenter will come to see you; don't ask your question out loud. If the question is relevant to all participants, we will repeat it and answer it out loud. If you do not respect these rules, we will have to exclude you from the experiment and therefore from the payment.

All the information you provide, as well as the amount of your payoffs during this experiment, will be kept strictly confidential and anonymous. Participating in this experiment will gain you money. Your winnings will be paid to you privately at the end of the experiment. You earn 5 euros for showing up on time, 2 euros for answering a series of questions and an additional amount that varies between 0 and 70 euros. The additional payment depend on your decisions and may also be influenced by decisions made by others.

First of all, before starting the actual experiment, we ask you to answer a series of preliminary questions. You will answer these questions using the interface on your computer screen.

### **CRT Questions**

1) A notebook and a pencil cost 1.10 Euros. The notebook costs 1 Euro more than the pencil. How many cents does the pencil cost? (correct answer: 0.05 cents)

2) Assuming that 5 machines take 5 minutes to make 5 pens, how long would it take 100 machines to make 100 pens? (correct answer: 5)

3) 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? (correct answer:47)

### **Representativeness Heuristic Questions**

Please read the descriptions below and answer the questions.

Description 1: Linda is 31 years old, she is single, frank and very bright. She has a master's degree in philosophy. As a student, she was very concerned about issues of discrimination and social justice, she also participated in anti-nuclear demonstrations.

Please rank the following statements based on their likelihood, using 1 for the most likely and 7 for the least likely:

- 1) Linda is a primary school teacher.
- 2) Linda works in a bookstore and takes yoga classes.
- 3) Linda is active in the feminist movement.
- 4) Linda is a bank teller.
- 5) Linda is a social worker in a psychiatric environment.
- 6) Linda is an insurance salesperson.
- 7) Linda is a bank teller and is active in the feminist movement.

Correct answer: If option 7 is judged more probable than options 3 and 4, then the answer is correct.

Description 2: A certain city is served by two hospitals. About 45 babies are born every day in the big hospital and about 15 babies are born every day in the small hospital. As you know, about 50% of new born babies are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it can be more than 50%, sometimes less.

For a period of one year, each hospital recorded the days when more than 60% of the babies born were boys. Which hospital do you think had the highest number of such days?

- A) The big hospital
- B) The small hospital
- C) Same for the two hospitals.

Correct answer: Option B is the correct answer.

### **Availability Heuristic Questions**

Below, each item includes two possible causes of death. The question to which you must answer is the following: among the two possible causes of death, which is the most frequent, in general, in France? For each pair of possible causes of death, (a) and (b), we want you to choose the cause that you think is the most common.

Pair 1 (a) Road accidents (b) Diabetes (correct answer: b)

Pair 2 (a) Homicide (b) Suicide (correct answer: b)

Pair 3 (a) Stroke (b) All accidents (correct answer: a)

Pair 4 (a) Falls (b) Drug use (correct answer: a)

Pair 5 (a) Lightning strike (b) Poisoning (correct answer: b)

### **Reading the Mind in the Eyes Test**

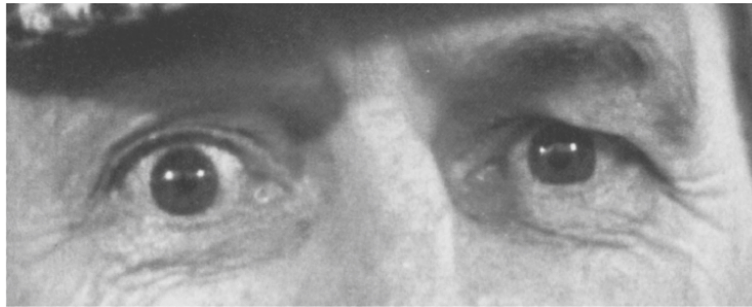
You will now see a series of images presenting pairs of eyes, as well as 4 words. For each pair of eyes, choose the word that best describes what the person in the image thinks or feels. You may feel that more than one word may apply, but choose only the word that you consider most appropriate. Before making your choice, make sure that you have read the 4 propositions correctly.

There will be a total of 36 questions to be answered within 10 minutes. Try to respond

as quickly and accurately as possible. To respond, select one of the choices displayed below the image and click OK at the bottom. Please note that once you click OK, you will not be able to return to the previous questions.

Click OK to proceed to the test question.

Training question:



Options: 1:jealous, 2:panicked, 3:arrogant, 4:hateful

Correct answer and explanations:

Among the choices: "jealous", "panicked", "arrogant", "hateful", the correct answer was: "panicked". Click OK to start the other questions. From now on, correct answers will no longer be shown.

As a reminder, there will be a total of 36 questions which you will have to answer within 10 minutes. Try to respond as quickly and accurately as possible.

### **Short Term Memory Test**

You will now look at several sets of slides. At the start of each series, you will see the word "Ready?" then a sequence of numbers appearing one after the other. At the end of each series, you will enter in the input zone the sequence of numbers that you will have observed.

The sequences will become longer as you enter correct answers. Your goal is to go as far as possible. You will have two tries. For example, if in a series you see 1, then 3 and then 5, you would enter 135.

First sequence: 69, 929, 1021, 34634, 943453, 7374865, 69358267, 690875725, 6457803021, 26456897198, 601518340985, 1285246589042.

Second sequence: 25, 217, 8618, 48629, 727240, 1203439, 32904142, 750572970, 1720378975, 62617825067, 609295956490, 1678606889148.

## **A.6 Score Measurement, Procedures and References for Tests**

The cognitive reflection task is retrieved from Frederick (2005) and adapted to French. Each correct answer is considered as one point in the score calculation. There is a non-binding time limitation which is 30 seconds for this part.

The heuristic questions are retrieved and adapted to the French population from the studies of Kahneman and Tversky (1972), Tversky and Kahneman (1973) and Fischhoff et al. (1977). The correct answers are determined from the data of World Health Organization's (WHO) report on global health estimates between 2000-2016 (World Health Organization, 2018). Each wrong answer is considered as one point in the score calculation.

The French version of The Reading the Mind in the Eyes Test is retrieved from Prevost et al. (2014). Each correct answer is considered as one point in the score calculation. There is a binding time limitation which is 10 minutes for this part.

The short-term memory test is retrieved from Wechsler digit span test. The score is calculated as the maximum of the number of digits accurately remembered in the both sequences. Each number stays in the screen for 2 seconds and the box where subjects type the number appears with 2 seconds delay after the last number.

## A.7 IRB Approval



Groupe d'Analyse et de Théorie Economique - UMR 5824

Ecully, September 26, 2019

Dear Dr. Cornand,

Your research project "Convergence and Adjustment under Strategic Complementarity: An Experimental Investigation of Repeated Shocks" has been evaluated by the GATE-Lab Review Board for ethical standards in research, under the reference code 2019-006. The Review Board made the following recommendation:

Full approval

Approval pending written notification to the Review Board

Rejection, proposal to be revised and re-evaluated

Rejection

Sincerely,

A handwritten signature in black ink, appearing to read "M. Villeval".

Marie Claire Villeval

Chair of the GATE-Lab Review Board for ethical standards in research



93, chemin des Mouilles - BP 167  
69131 Ecully cedex (France)  
Tel: +33 (0)4 72 86 60 60 - Fax: +33 (0)4 72 86 60 90  
gate@gate.cnrs.fr - <http://www.gate.cnrs.fr>



## B Additional Figures and Tables

### Appendix B1 - Description of subject pool

	Pooled	Baseline	Treatment
Age	21.43 (1.91)	21.83 (2.32)	21.03 (1.26)
Share of women	42% (0.49)	38% (0.49)	47% (0.50)
Baccalaureate grade	15.97 (2.03)	15.76 (1.99)	16.18 (2.04)
CRT score	1.49 (1.14)	1.4 (1.07)	1.58 (1.20)
Eyes score	27.2 (3.92)	26.97 (4.30)	27.43 (3.48)
Memory score	7.05 (1.67)	7.02 (1.95)	7.08 (1.33)
Represent. score	1.29 (0.65)	1.30 (0.64)	1.28 (0.66)
Availability score	2.13 (1.02)	2.00 (1.05)	2.27 (0.98)

Note: Below averages, the standard deviations are reported in parentheses.

### Appendix B2 - Deviations as round averages

Rounds	Baseline				Treatment			
	1	2	3	4	1	2	3	4
$\bar{AD}$	10.341 (7.127)	8.636 (6.926)	7.434 (6.787)	4.372 (5.472)	10.361 (7.494)	4.489 (4.445)	11.124 (10.111)	6.648 (6.861)
$R\bar{AD}$	0.431 (0.297)	0.360 (0.288)	0.310 (0.283)	0.182 (0.228)	0.432 (0.312)	0.125 (0.123)	0.927 (0.843)	0.277 (0.286)

Note:  $\bar{AD}$  refer to the mean absolute deviations during each post-shock phase measured as  $\bar{AD}_r = \frac{\sum_{t=9}^{16} \sum_{c=1}^6 |(\bar{p}_{c,t} - p_t^{NE})|}{48}$ .  $R\bar{AD}$  refer to the mean relative absolute deviations during each post-shock

phase measured as  $R\bar{AD}_r = \frac{\sum_{t=9}^{16} \sum_{c=1}^6 \left| \frac{(\bar{p}_{c,t} - p_t^{NE})}{p_t^{NE}} \right|}{48}$  so the deviations are normalized through their distance from NE. Below averages, the standard deviations are reported in parentheses.

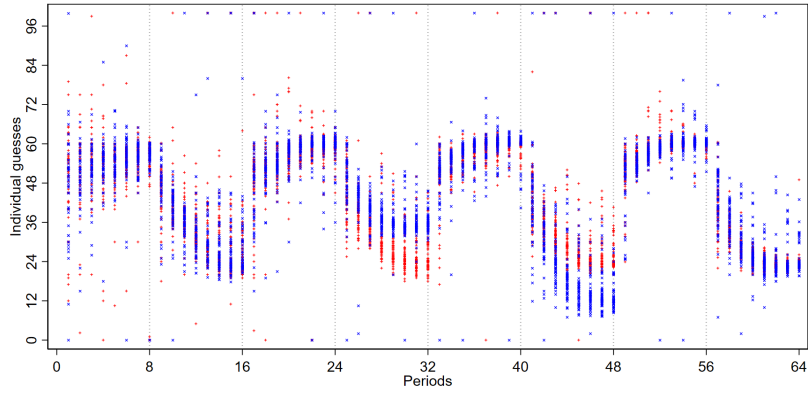


Figure 5: Distribution of individual guesses across periods. 120 subjects per period. Plus signs (cross signs) represent baseline (treatment) conditions.

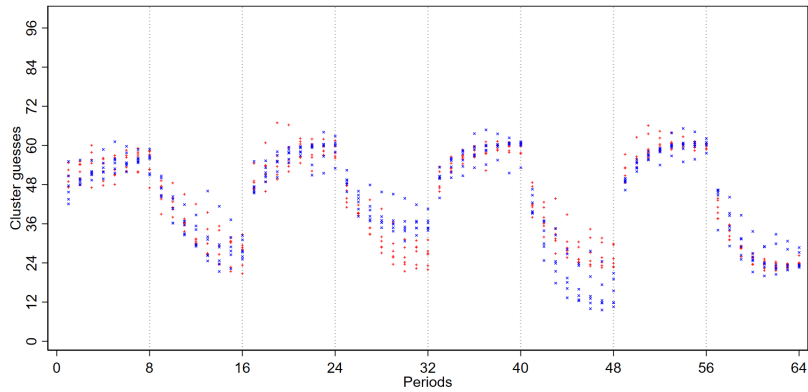


Figure 6: Distribution of average guesses of a cluster across periods. 12 clusters per period. Plus signs (cross signs) represent baseline (treatment) clusters.

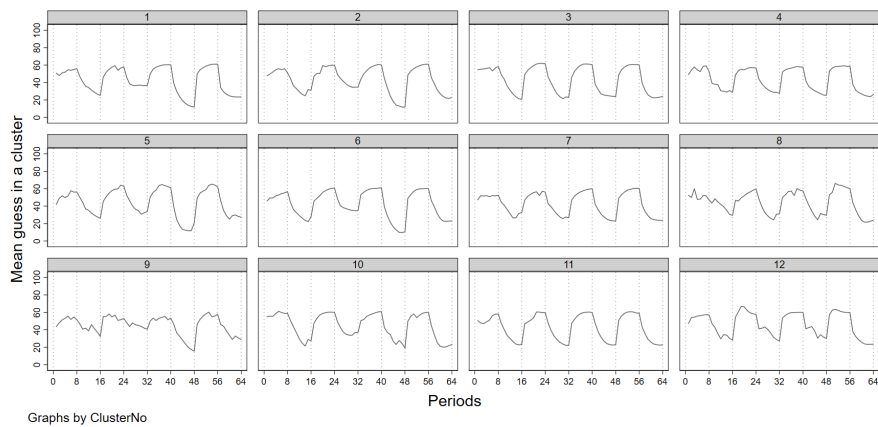


Figure 7: Average guess of clusters by periods. The identification number of clusters are denoted above each figure. Clusters 3, 4, 7, 8, 11, 12 belong to baseline sessions and clusters 1, 2, 5, 6, 9, 10 belong to treatment sessions.



## C Robustness Analyses

### C.1 Adjustment Period Measured with Fixed Effects Regressions

When the period of adjustment is estimated through FE regressions, the results are qualitatively unchanged except the fact that the period of adjustment in treatment condition is earlier ( $t^c = 4$ ). This happens partially because we omit the last period from our regressions to avoid linear dependency and the average deviation of last period in treatment is high ( $b_{64} = 0.733$ ) and it is as low as 0.069 in baseline condition. When we omit another period, for instance period 31 which has one of the smallest deviation, all the estimated coefficients for treatment condition inflate with the amount 0.54. The period of adjustment  $t^c$  becomes 5, the same result as median quantile regressions.

**Table C1 - Post-shock deviations from NE with FE**

Post-shock periods	Baseline				Treatment	
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 4
1	21.130***	20.852***	19.083***	16.366***	22.555***	19.151***
2	18.349***	15.455***	13.417***	9.135***	15.583***	12.053***
3	13.311***	11.134***	9.284***	4.234***	11.414***	5.612***
4	8.382***	7.269***	5.640**	1.112**	8.052***	1.516
5	6.572***	4.143**	2.275***	-0.532	6.380***	0.069
6	4.648**	1.767	3.339*	-0.960**	3.560	0.514
7	3.807**	2.907**	2.188	-0.859***	4.353***	0.074
8	3.101**	2.169*	1.634	-	3.575***	-

Note: Coefficients from Fixed Effects regression of model (3) are reported. Standard errors are clustered at subsession level (2 groups) and bootstrapped with 1999 replications. \*, \*\*, \*\*\* indicate statistical significance at the 10, 5, 1 percent levels, respectively.

### C.2 HSM with Different Free Parameters

The HSM has three free parameters where each has its own behavioral implication. To avoid arbitrariness, we provide a robustness analysis by comparing our benchmark HSM, denoted as HSM 1, with the benchmark HSM of Bao et al. (2012), denoted as HSM 4. For analytical tractability, we add two other HSMs where in each step one free parameter gets closer to the one in Bao et al. (2012). Table C3 reports the results. Results show that

the fit worsens when any of the parameter increases *ceteris paribus* and the RMSE of the HSM 4 is worse than some homogeneous expectation rules. Nonetheless, the fits of HSMs 2 and 3 are still better than any of the homogeneous rule. Step-wise changes imply that the HSM applied to our data is not much sensitive to changes in parameters  $\eta$  and  $\beta$  but it is somehow sensitive to the abrupt changes in  $\delta$ , at least for treatment condition. This parameter represents the proportion of agents who do not update their impact factor each period and a value of 0.9 is behaviorally hard to justify. In conclusion, our benchmark results are robust to the medium-level changes in parameters.

**Table C2 - RMSE of HSMs with various free parameters**

Comparison Level	HSM 1	HSM 2	HSM 3	HSM 4
Baseline	3.90	4.43	4.54	5.10
Treatment	3.92	5.07	5.13	7.75
Pool	3.91	4.76	4.84	6.56
$\eta$	0.1	0.7	0.7	0.7
$\delta$	0.4	0.4	0.4	0.9
$\beta$	0.1	0.1	0.4	0.4

Note: HSM 1 is equivalent to the our benchmark HSM reported in Table 4. HSM 4 has the benchmark parameters combination from Bao et al. (2012) and HSM 2 and 3 serve the role of step-wise change in parameters.

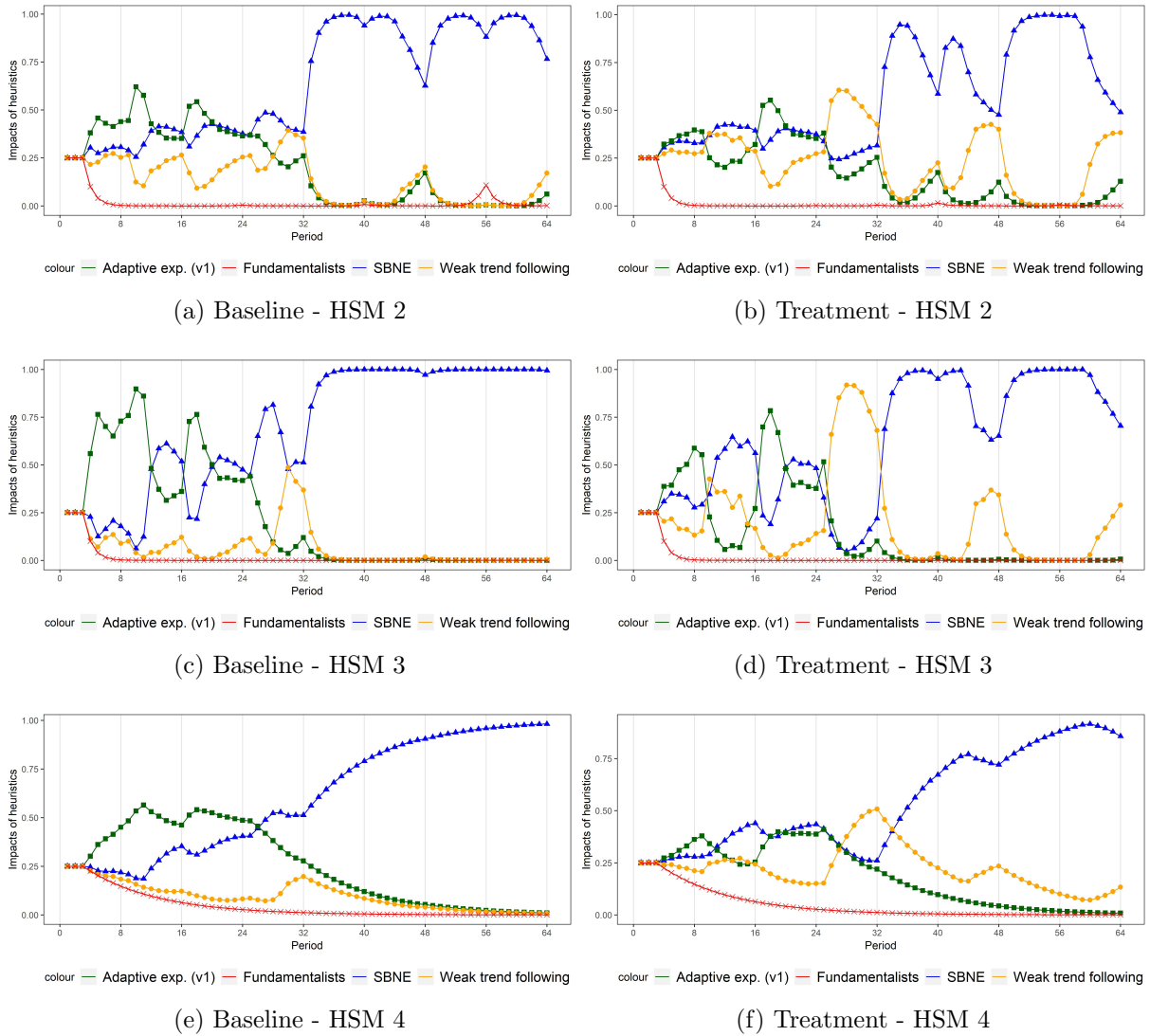


Figure 8: Impact factors calculated with the HSM 2, 3, 4 across periods and experimental conditions. The symbols square, cross, triangle and circle represent respectively adaptive expectations (v1), fundamentalist, SBNE and weak trend following expectations.

## References

- Anufriev, M. and C. Hommes (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics* 4(4), 35–64.
- Anufriev, M., C. H. Hommes, and R. H. Philipse (2013). Evolutionary selection of expectations in positive and negative feedback markets. *Journal of Evolutionary Economics* 23(3), 663–688.
- Arifovic, J. and J. Duffy (2018). Heterogeneous agent modeling: Experimental evidence. In C. Hommes and B. LeBaron (Eds.), *Handbook of Computational Economics*, Volume 4, pp. 491–540. Elsevier.
- Arifovic, J. and L. Petersen (2017). Stabilizing expectations at the zero lower bound: Experimental evidence. *Journal of Economic Dynamics and Control* 82, 21–43.
- Assenza, T., P. Heemeijer, C. Hommes, and D. Massaro (2019). Managing self-organization of expectations through monetary policy: A macro experiment. *Journal of Monetary Economics*, forthcoming.
- Bao, T., J. Duffy, and C. Hommes (2013). Learning, forecasting and optimizing: An experimental study. *European Economic Review* 61, 186–204.
- Bao, T., C. Hommes, J. Sonnemans, and J. Tuinstra (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control* 36(8), 1101–1120.
- Baron-Cohen, S., S. Wheelwright, J. Hill, Y. Raste, and I. Plumb (2001). The “reading the mind in the eyes” test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42(2), 241–251.

- Bray, M. (1982). Learning, estimation, and the stability of rational expectations. *Journal of Economic Theory* 26(2), 318–339.
- Bray, M. (1983). Convergence to rational expectations equilibrium. In R. Frydman and E. S. Phelps (Eds.), *Individual Forecasting and Aggregate Outcomes*, pp. 123–137. Cambridge: Cambridge University Press.
- Brock, W. A. and C. H. Hommes (1997). A rational route to randomness. *Econometrica* 65(5), 1059–1095.
- Cooper, K. B., H. S. Schneider, and M. Waldman (2017). Limited rationality and the strategic environment: Further theory and experimental evidence. *Games and Economic Behavior* 106, 188–208.
- Cornand, C. and F. Heinemann (2019). Monetary policy obeying the taylor principle turns prices into strategic substitutes. *Journal of Economic Behavior & Organization*, forthcoming.
- Cornea-Madeira, A., C. Hommes, and D. Massaro (2019). Behavioral heterogeneity in us inflation dynamics. *Journal of Business & Economic Statistics* 37(2), 288–300.
- Costa-Gomes, M. A. and G. Weizsäcker (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies* 75(3), 729–762.
- Davidson, R. and J. G. MacKinnon (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews* 19(1), 55–68.
- Davis, D. and O. Korenok (2011). Nominal shocks in monopolistically competitive markets: An experiment. *Journal of Monetary Economics* 58(6-8), 578–589.
- Dufwenberg, M., T. Lindqvist, and E. Moore (2005). Bubbles and experience: An experiment. *American Economic Review* 95(5), 1731–1737.

- Fehr, E. and J.-R. Tyran (2001). Does money illusion matter? *American Economic Review* 91(5), 1239–1262.
- Fehr, E. and J.-R. Tyran (2005). Individual irrationality and aggregate outcomes. *Journal of Economic Perspectives* 19(4), 43–66.
- Fehr, E. and J.-R. Tyran (2008). Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. *Econometrica* 76(2), 353–394.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischhoff, B., P. Slovic, and S. Lichtenstein (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance* 3(4), 552.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives* 19(4), 25–42.
- Gilboa, I. and D. Schmeidler (1995). Case-based decision theory. *The Quarterly Journal of Economics* 110(3), 605–639.
- Gode, D. K. and S. Sunder (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy* 101(1), 119–137.
- Haltiwanger, J. and M. Waldman (1985). Rational expectations and the limits of rationality: An analysis of heterogeneity. *American Economic Review* 75(3), 326–340.
- Haltiwanger, J. and M. Waldman (1989). Limited rationality and strategic complements: the implications for macroeconomics. *The Quarterly Journal of Economics* 104(3), 463–483.

- Hanaki, N., E. Akiyama, and R. Ishikawa (2018). Effects of different ways of incentivizing price forecasts on market dynamics and individual decisions in asset market experiments. *Journal of Economic Dynamics and Control* 88, 51–69.
- Hanaki, N., Y. Koriyama, A. Sutan, and M. Willinger (2019). The strategic environment effect in beauty contest games. *Games and Economic Behavior* 113, 587–610.
- Haruvy, E., Y. Lahav, and C. N. Noussair (2007). Traders' expectations in asset markets: experimental evidence. *American Economic Review* 97(5), 1901–1920.
- Heemeijer, P., C. Hommes, J. Sonnemans, and J. Tuinstra (2009). Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics and Control* 33(5), 1052–1072.
- Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control* 35(1), 1–24.
- Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. Judd (Eds.), *Handbook of Computational Economics*, Volume 2, pp. 1109–1186. Elsevier.
- Hussam, R. N., D. Porter, and V. L. Smith (2008). Thar she blows: Can bubbles be rekindled with experienced subjects? *American Economic Review* 98(3), 924–37.
- Hyndman, K., E. Y. Ozbay, A. Schotter, and W. Z. Ehrblatt (2012). Convergence: an experimental study of teaching and learning in repeated games. *Journal of the European Economic Association* 10(3), 573–604.
- Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3), 430–454.
- Kocher, M. G. and M. Sutter (2006). Time is money—time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization* 61(3), 375–392.

- Kopányi-Peuker, A. and M. Weber (2018). Experience does not eliminate bubbles: Experimental evidence. Tinbergen Institute Discussion Paper 2018-092/II.
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives* 17(1), 59–82.
- Marquardt, P., C. N. Noussair, and M. Weber (2019). Rational expectations in an experimental asset market with shocks to market trends. *European Economic Review* 114, 116–140.
- Milgrom, P. and J. Roberts (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* 58(6), 1255–1277.
- Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica* 29(3), 315–335.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review* 85(5), 1313–1326.
- Petersen, L. and A. Winn (2014). Does money illusion matter? comment. *American Economic Review* 104(3), 1047–62.
- Pfajfar, D. and B. Žakelj (2014). Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control* 44, 147–168.
- Plonsky, O., K. Teodorescu, and I. Erev (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review* 122(4), 621–647.
- Potters, J. and S. Suetens (2009). Cooperation in experimental games of strategic complements and substitutes. *The Review of Economic Studies* 76(3), 1125–1147.
- Prevost, M., M.-E. Carrier, G. Chowne, P. Zekowitz, L. Joseph, and I. Gold (2014). The reading the mind in the eyes test: Validation of a french version and exploration of cultural variations in a multi-ethnic city. *Cognitive Neuropsychiatry* 19(3), 189–204.



- Shestakova, N., O. Powell, and D. Gladyrev (2019). Bubbles, experience and success. *Journal of Behavioral and Experimental Finance* 22, 206–213.
- Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy* 70(2), 111–137.
- Smith, V. L., G. L. Suchanek, and A. W. Williams (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica* 56(5), 1119–1151.
- Sonnemans, J. and J. Tuinstra (2010). Positive expectations feedback experiments and number guessing games as models of financial markets. *Journal of Economic Psychology* 31(6), 964–984.
- Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior* 16(2), 303–330.
- Sutan, A. and M. Willinger (2009). Guessing with negative feedback: An experiment. *Journal of Economic Dynamics and Control* 33(5), 1123–1133.
- Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5(2), 207–232.
- World Health Organization (2018). Who methods and data sources for country-level causes of death 2000–2016. Technical report, Global Health Estimates Technical Paper WHO/HMM/IER/GHE/2018.1, WHO, Geneva.