

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

Muhammed Bulutay, Camille Cornand, Adam Zylbersztejn

▶ **To cite this version:**

Muhammed Bulutay, Camille Cornand, Adam Zylbersztejn. Learning to deal with repeated shocks under strategic complementarity: An experiment. 2020. halshs-02458140v2

HAL Id: halshs-02458140
https://shs.hal.science/halshs-02458140v2

Preprint submitted on 20 Apr 2020

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.

GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 2003 – January 2020, revised April 2020

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

Muhammed Bulutay, Camille Cornand, Adam Zylbersztejn

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

Working Paper

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

Muhammed Bulutay[†a], Camille Cornand[b], and Adam Zylbersztejn[c]

[a]Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
[b]Univ Lyon, CNRS, GATE UMR 5824, F-69130 Ecully, France
[c]Univ Lyon, Université Lumière Lyon 2, GATE UMR 5824, F-69130 Ecully, France

April 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. Notably, refining these rules with similarity-based learning approach improves their predictive power.

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 the editor and three anonymous referees for the valuable comments. This paper has benefited from insightful discussions with Kristen Cooper, Nobuyuki Hanaki, Frank Heinemann, Henry Schneider, Angela Sutan, Michael Waldman and with the participants to the Colloquium Monetary and Experimental Economics at TU-Berlin. We also acknowledge Quentin Thévenet and Maxim Frolov for providing help with the software programming.

[†]Corresponding author.
E-mail addresses: mbulutay@diw.de (M. Bulutay), cornand@gate.cnrs.fr (C. Cornand), zylbersztejn@gate.cnrs.fr (A. Zylbersztejn)

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 is 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 nonrational 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).[1] The main pattern emerging from these studies is that deviations from equilibrium tend to be larger and more persistent under strategic complementarity as compared to strategic substitutability.[2]

Herein, we focus on strategic complementarity which comes as an important feature of various economic contexts including macroeconomic coordination, bank runs, and oligopoly competition.[3] As argued by Hommes (2006), 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 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-Peuker 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 nonrational 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

[1] See Hommes (2011) and Arifovic and Duffy (2018) for an overview of the Learning-to-Forecast literature.
[2] Hanaki et al. (2019) are the first to term this phenomenon as *the strategic environment effect*.
[3] See Milgrom and Roberts (1990) for more examples.

large periodic shocks to the 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 the treatment effect of experiencing nonidentical shocks (i) 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 for the adjustment to occur after further shocks. 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 of the game.

Our contribution to the literature is twofold. Firstly, we document the robustness of 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 find 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 reveals that upgrading expectation rules with similarity-based learning approach improves their predictive power under identical shocks. Notably, the best performing model is a simple nonparametric reformulation of naïve expectations with similarity-based learning (first proposed by Cooper et al., 2017). We discuss its 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 and the way we implement it in the lab. 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 - Convergence in the previous experiments

Study	Type of environment	Shock size (in %)	Convergence period
Fehr and Tyran (2001)[1]	Pricing decision	-67% & +100%	13 & 4
Fehr and Tyran (2008)[2]	Pricing decision	-50%	9
Davis and Korenok (2011)[3]	Monopolistic competition	+100%	21
Petersen and Winn (2014)[1]	Pricing decision	-67% & +92%	8 & 4
Cooper et al. (2017)[4]	Guessing game	-77%	8

[1] Shock as the change in average equilibrium price in the nominal treatment with human opponents.
[2] Shock as the change in average equilibrium price in nominal treatment.
[3] 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.
[4] Shock as the change in the NE guess in first round.

Table 1 summarizes the evidence from experimental studies that investigate the dynamics of convergence following large shocks when actions are strategic complements. Here, the "Convergence period" reported in column 4 indicates the number of periods for the general activity level (price, forecast, guess, etc.) to become statistically indistinguishable (at the 5% level) from the post-shock theoretical equilibrium value.[4] The general pattern in those data is that convergence takes time when actions are strategic complements. In particular, adjustment to the NE tends to be slow after the initial shock, even though acceleration may still occur when markets are repeated (Cooper et al., 2017).[5] We expect to observe the same pattern in a slightly modified environment.[6]

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

[4]We retrieved the information about the convergence period directly from each article. Thus, the table does not account for the differences in experimental designs and statistical methods used across these studies.

[5]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.

[6]The scope of these modifications along with their rationale are described in Section 3.2.

Albeit robust in stationary environments, experience effects are argued to be sensitive to the complexity of the environment. For instance, in the AME and LtFE studies by Kopányi-Peuker and Weber (2018) and Hussam et al. (2008), bubbles do not disappear despite repetition.[7] Hussam et al. (2008) also report that bubbles reignite even with twice-experienced subjects following drastic changes in the environment (e.g., the amount of liquidity in the market). Moreover, 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.[8]

Hypothesis 2. The adjustment accelerates with market repetition at a slower pace in the presence of nonidentical shocks than with identical shocks.

We now turn to the possible explanations of adjustment dynamics. Several studies provide a descriptive explanation of the observed inertia based on nonrational 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.[9] Cooper et al. (2017), in turn, obtain the best fit with heterogenous groups: one rational player and three nonrational players whose expectations follow a version of naïve expectations rule adapted to a repeated shocks design. Other studies point to trend-following expectations (Haruvy et al. 2007) or even RE (Marquardt et al. 2019) as best describing their experimental evidence.

We note, however, that the aformentioned studies either do not compare the fit of their model with other expectation rules, or only consider a relatively narrow set of competing rules.[10] More systematic comparisons exist in the LtFE literature. The design of LtFEs

[7]According to Kopányi-Peuker and Weber (2018), a possible explanation is that this occurs because interactions in their experiment have an indefinite horizon.

[8]Bao et al. (2012) study large nonidentical shocks in LtFEs by introducing two large shocks to the rational expectations (RE) equilibrium. However, their design does not propose a way to test the effect of nonidentical shocks with respect to identical ones. To the best of our knowledge, our study is the first to directly test the impact of the heterogeneity of shocks in a controlled environment.

[9]In their model of *fully adaptive expectations*, players expect the outcome of the last period to reoccur.

[10]For instance, Marquardt et al. (2019) only consider three models: myopic, trend and RE. Moreover,

is particularly well-suited for investigating expectations since the experimental task is to forecast the prices one-period-ahead. Trend-following has been repeatedly shown to outperform all the 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 the RE (simple rules) cannot be rejected for 30-45% (35-50%) of subjects. This finding has been subsequently confirmed by Marquardt et al. (2019). In the context of the evolutionary heuristic switching model (HSM, Anufriev and Hommes 2012), 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 a substantial time variation, they find that 65% of individuals form naïve expectations, the share of which 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 backward-looking expectation models 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 the 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 the previous occurence of the same shock. We denote this rule as *similarity-based naïve expectations* (SBNE). In the same

the parameters of their trend model resemble 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 environments like AMEs.

vein, we extend the adaptive and trend-following models and denote them respectively as *similarity-based adaptive expectations* (SBAE) and *similarity-based trend-following expectations* (SBTF). We conjecture that this class of rules best explains behavior under repeated identical shock.

Hypothesis 4. Under identical shocks, the rules that are augmented with the similarity-based learning approach provide 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. The fundamental difference between these two designs is that while guessing games provide full information on the game structure (including the parameters), LtFEs provide only qualitative information about the market structure. Nevertheless, Sonnemans and Tuinstra (2010) show that convergence dynamics are similar when the feedback parameters are equal.[11]

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 $i = 1, \ldots, N$. 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[12] at period t, a and b are

[11] See Sonnemans and Tuinstra (2010) for a detailed comparison of guessing games and LtFEs.

[12] 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.

positive constant numbers with $b \in (0,1)$, and ξ_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 $\mid y_{i,t} - p_{i,t} \mid$ wins the fixed stage game payoff F. In case of a tie, the payoff is equally split among the winners.

This game has a unique Nash equilibrium which corresponds to an interior solution: $p_t^{NE} = \frac{a+\xi_t}{(1-b)}$.[13] Here, p_t^{NE} is invariant for the first $T/2$ periods, called the *pre-shock periods*. The shock $\bar{\xi}$ shifts the equilibrium downwards at period $T/2+1$. These remaining periods with a new equilibrium are the *post-shock periods*. In addition, 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 calibration of the experimental game is summarized in Table 2. A group of 5 participants play the guessing game for 4 rounds, and each round is composed of 16 periods. This yields a total of 64 guessing decisions per player. In period 9 of each round, a negative shock $\bar{\xi}$ shifts parameter a from 15 to a value that depends on the experimental condition.

In baseline, shocks are identical and the shock component equals -9 in every rounds. In treatment, shocks are not identical and the shock component is characterized by the

[13] The proof is based on the iterated elimination of dominated strategies. See Nagel (1995) for details. This equilibrium is also a RE equilibrium. Bray (1983) shows that when $b < 1$, a misspecified expectation rule – ordinary least-squares learning – almost surely converges to the RE equilibrium. 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."

sequence $(-9, -6, -12, -9)$ in rounds $(1, 2, 3, 4)$. Thus, the post-shock NE is the same in the first and last rounds in both conditions. This allows us to capture the effect of experiencing nonidentical shocks by comparing adjustment speeds in round 4.

Table 2 - Experimental design parameters

General parameters		
Number of periods per round	$T = 16$	
Number of rounds per session	$R = 4$	
Group size	$N = 5$	
Stage game prize	$F = 4.40$ euros	
The slope of target formula	$b = 0.75$	
Pre-shock value of constant	$a = 15$	
Pre-shock equilibrium	$p^{NE}_{pre} = 60$	
Post-shock equilibrum (p^{NE}_{post})	Baseline	Treatment
Round 1	24	24
Round 2	24	36
Round 3	24	12
Round 4	24	24

The design of this experiment closely follows Cooper et al. (2017), with some noteworthy modifications. 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.[14] 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. Third, since post-shock phases are shorter in our study, the post-shock equilibrium in baseline groups is set at a higher level. Finally, Cooper et al. (2017) elicit expectations of subjects in addition to their guesses while we elicit expectations jointly with the guesses which comes as a remedy against potential inconsistencies.[15]

[14]They show that the effect of strategic environment is statistically significant for groups of five or more agents.

[15]The elicitation method is explained in Section 3.3. *Ex ante*, the design of our baseline condition provides more suitable circumstances to observe rapid adjustment than the one in Cooper et al. (2017), since we have an extra round and a belief elicitation mechanism that emphasizes best replying to one's expectations. However, the data suggest that the patterns of adjustment are similar in both experiments.

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).[16] 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.[17] This yields six independent clusters of observations per condition, and twelve clusters in total.

Subjects are provided with the instructions of the game in paper form that are read aloud by the experimenter.[18] 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.

[16] The experimental procedures have been approved by the GATE-Lab Review Board.

[17] See Appendix A for the experimental materials and Appendix B for a description of the subject pool.

[18] 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), 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. We use these scores to test the sensitivity of our findings with respect to several individual characteristics. This analysis is reported in Appendix C.2. The content and the measurement method of each test are provided in Appendices A.4 and A.6.

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.[19] 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 nonbinding timer set to 60 seconds (except for the initial period of the game which has 120 seconds).[20] 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. To minimize any potential wealth or end-game effects, the final payoff correponds to the payoffs accumulated in all periods of a randomly chosen round of the game. At the end, subjects also reply to a demographic questionnaire.

An experimental session took 150 minutes on average. Subjects were paid 7 euros for their participation and 14.08 euros on average for the experimental 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 investigate the within-period variation of individual guesses across experimental conditions. Lastly, we compare the descriptive power of various expectation models by their one-period-ahead forecast accuracy,

[19]This method provides a consistent way for joint elicitation of guesses and expectations. Some of the previous studies document systematic inconsistencies between expectations and decisions that are elicited separately (Costa-Gomes and Weizsäcker, 2008). Moreover, LtFEs and AMEs studies show that having both forecasting and optimizing tasks may be detrimental to learning and cause mispricing (Bao et al., 2013; Hanaki et al., 2018).

[20]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, this feature of our design should not put participants under excessive time pressure.

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 an evolutionary learning model – the heuristic switching 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 expectation rules.

4.1 Adjustment Dynamics

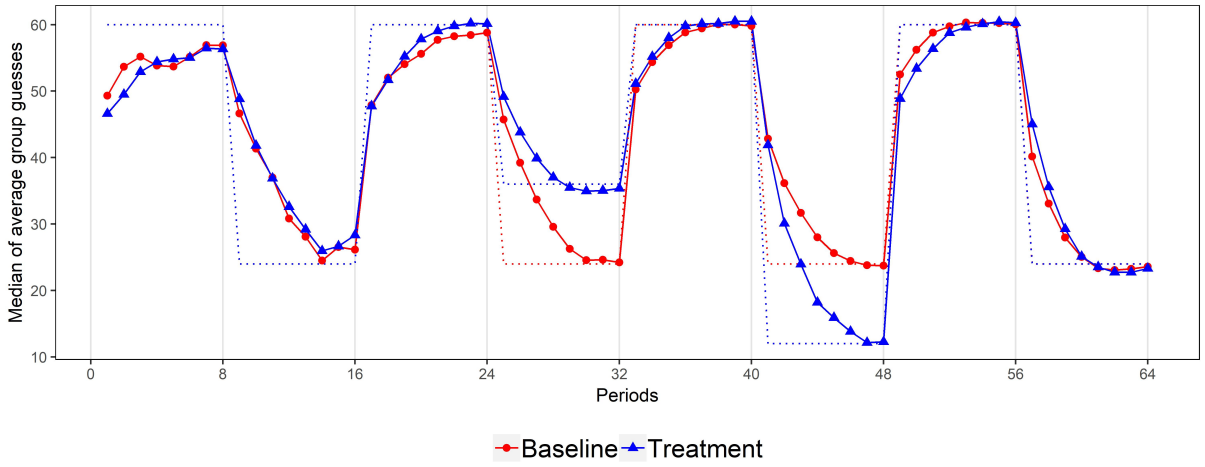


Figure 1: Dotted lines represent NE. In each round, a dot (triangle) corresponds to the median value of the average group guesses in the baseline (treatment) condition.

We first turn to the aggregate outcomes and look at the evolution of the average group guesses over time. Figure 1 summarizes the observed median values of average group guesses (with 12 groups per experimental condition) across rounds and periods. As expected, round 1 – in which the environment remains strictly identical in both experimental conditions – generates the same patterns in the data in baseline and treatment. In particular, in the pre-shock phase of round 1 this median never fully converges to the NE level.[21] The second salient observation is that the convergence to the NE systematically exhibits a convex pattern, but happens at varying speed.[22]

[21] We note that Cooper et al. (2017) observe the same pattern of adjustment under identical shocks.

[22] The evolution of guesses in the pre-shock periods of rounds 3 and 4 also suggests that the path of shocks experienced in the past does not affect *per se* the adjustment to the NE. Despite the different sizes of shocks

Table 3 - Post-shock deviations from NE

Post-shock	Baseline				Treatment	
periods	Round 1	Round 2	Round 3	Round 4	Round 1	Round 4
1	22.76***	22.08***	19.31***	18.38***	24.89***	20.86***
2	17.88***	16.13***	13.19***	9.46***	17.95***	12.32***
3	13.93***	10.13***	8.38***	4.57***	13.15***	5.70***
4	7.39***	6.38***	4.66***	1.51***	9.32***	1.64*
5	5.22**	3.02***	2.38***	-0.22	6.67**	0.17
6	1.64	1.14	0.99	-0.54	3.50	-0.82
7	3.40	1.26	0.37	-0.44*	3.98	-0.80*
8	2.65	0.74	0.15	*-0.38*	3.18	*-0.42*

Note. Coefficients a_t and intercept term a_0 (in italics) from a median quantile regression model specified in (3). Standard errors are clustered at rematching cluster level (6 clusters per condition) and bootstrapped with 1999 replications. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively.

For a formal statistical comparison of the patterns of convergence in both experimental conditions, we use median quantile regression to estimate the following model:[23]

$$\bar{p}_{g,t} - p_t^{NE} = a_0 + a_t 1[Period = t] + \epsilon_{g,t}, \tag{3}$$

where the dependent variable is the difference between the average group guess ($\bar{p}_{g,t}$) and the NE guess (p^{NE}) in a given period t, while the independent variables are 63 period indicators ($1[Period = t]$ equals 1 for period t, and 0 otherwise). The coefficient of the final period is dropped to avoid linear dependencies.[24] We run this regression separately for baseline and treatment. Table 3 reports a subset of the estimated coefficients corresponding to the post-shock periods in all four rounds of the baseline and in rounds 1 and 4 of the treatment (in which the NE is the same as in the baseline).

This model allows us to analyze the patterns of adjustment to the NE in two steps. First, the intercept a_0 provides an empirical benchmark for convergence, i.e. the ability to

in round 2 between baseline and treatment, the adjustment to the NE in the pre-shock phase of the following round is identical in both conditions. The same holds for the adjustment to the NE in the pre-shock phase in round 4.

[23]This approach closely follows Cooper et al. (2017) and minimizes the role of potential outliers.

[24]The estimated standard errors are based on clustering with bootstrap resampling to take into account the possible correlation between guesses within a rematching cluster. The number of bootstrap samples follows Davidson and MacKinnon (2000).

adjust guesses to the NE play as measured by the degree of convergence to the NE that can be reached after 63 periods and with a full scope of experience accumulation and learning (that we further investigate in Section 4.2) in the experimental game. Testing whether this empirical benchmark differs from the formal prediction based on the NE ($H_0 : \bar{p}_{64} - p_{64}^{NE} = 0$) boils down to testing for the statistical significance of a_0. For both baseline and treatment, the estimated values of a_0 are similar and very close to zero (-0.38 and -0.42, respectively); in neither case we reject their nullity ($p = 0.133$ and $p = 0.285$, respectively). Second, we build on this first step and define the speed of adjustment as the earliest period in which the outcome attains (in statistical terms) our empirical benchmark of convergence. This period (denoted t^c) indicates the point of reaching the adjustment benchmark: for each period $t \geq t^c$ of a given phase we fail to reject $H_0 : a_t = 0$.[25]

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 repetition over four rounds.

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

Result 1b: In both experimental conditions, 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 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 guesses between the two conditions. We estimate the following median quantile

[25] This definition echoes the definition of convergence proposed by Hyndman et al. (2012).

regression model separately for each of the eight post-shock periods of round 4:

$$p_i = b_0 + b_1 1[Treatment] + \epsilon_i, \tag{4}$$

where the independent variable $1[Treatment] = 1$ for observations coming from the treatment sessions (and 0 otherwise), and the dependent variable is the individual guess ($N = 120$ per regression). Like before, we employ bootstrapped standard errors clustered at rematching cluster level (999 replications). The coefficients b_1 (reported in Table 4) remain 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. Figure 2 provides a visual representation of this pattern.

Table 4 - The effect of heterogeneity in shocks on guesses of last post-shock phase

Periods	57	58	59	60	61	62	63	64
b_0	39.70***	32.89***	28.40***	25.10***	23.49***	23.07***	23.21***	23.60***
	(1.54)	(1.08)	(0.50)	(0.47)	(0.44)	(0.31)	(0.11)	(0.11)
b_1	4.60*	3.21	1.35	-0.10	0.15	-0.67	-0.41	-0.35
	(2.70)	(2.02)	(1.19)	(0.94)	(1.09)	(0.79)	(0.88)	(0.71)

Coefficients from median quantile regression models specified in (4). Each coefficient comes from one regression ($N = 120$ per regression). Below coefficients standard errors are reported in parentheses. Standard errors are clustered at the rematching cluster level (6 clusters per condition) and bootstrapped with 999 replications. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively.

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

4.2 Expectation Formation

In this section, we exploit the data on expectations retrieved from (and consistent with) the guesses, as previously explained in Section 3.3. We consider a set of expectation rules to provide a descriptive explanation for the observed aggregate results.[26] These rules,

[26]Note that by taking the conditional expectation on both sides, equation (1) can be rewritten as $p_{i,t} = E_{i,t-1}[p_{i,t}] = (a + \xi_t) + bE^h_{i,t-1}[p_{-i,t}]$. Thus, subjects choose their guesses $p_{i,t}$ as best response to their

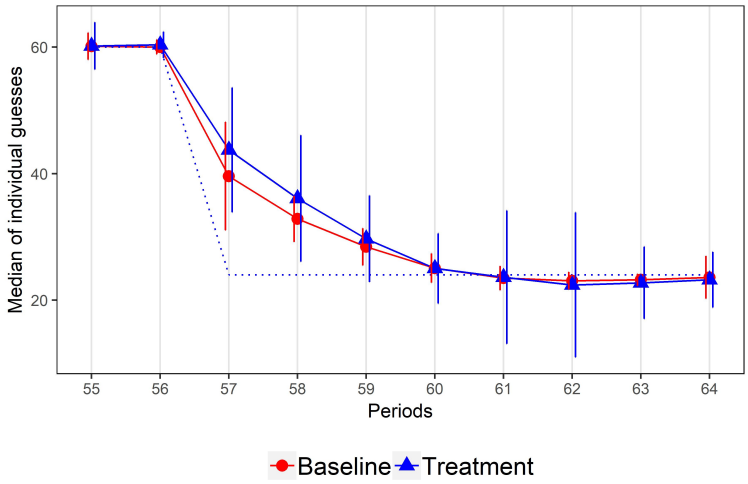


Figure 2: Median individual guess by experimental conditions for periods 55-64. Dots represent the NE. Circles (triangles) represent guesses in baseline (treatment) condition. Whiskers denote standard deviations. Unit of observation is the individual guesses.

summarized in Table 5, are mainly derived from two classes of learning models: adaptive (rules 1 to 6) and extrapolative (rules 7 to 13). Under the data generation process described in equation (1), the adaptive learning can be represented in a recursive form with the following formula:

$$p^e_{i,t} = p^e_{i,t-1} + w(p_{-i,t-1} - p^e_{i,t-1}), \tag{5}$$

where a player expects the weighted average of the most recent outcome and his/her own previous expectation. In the extrapolative expectations, a player tracks the most recent change in the realized outcome in the following manner:

$$p^e_{i,t} = p_{-i,t-1} + \gamma(p_{-i,t-1} - p_{-i,t-2}). \tag{6}$$

The coefficients w and γ are learning parameters. We predetermine these parameters based on their computational ease as an attempt to imitate different kinds of boundedly

one-period-ahead expectation about the average guess of other players in the group that is $E^h_{i,t-1}[p_{-i,t}]$. The superscript h is placed to indicate that the process of expectation formation is more general than RE and may be based on any rule h. For the sake of simplicity, throughout the paper we use the notation $p^e_{i,t}$ instead of $E^h_{i,t-1}[p_{-i,t}]$.

rational reasoning. We also include two models (rules 6 and 13) where parameters w and γ are estimated from the individual expectations data with fixed effects regressions and one equilibrium model (rule 14) where a player's expectation corresponds to the NE.[27]

The next three rules (15 to 17) reformulate rules 1, 3 and 9 (respectively) with similarity-based learning. Under rule 15, for instance, 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 has changed, the player reviews all the past periods and the new expectation now coincides with the outcome of the most recent period involving an analogous change $t - m$, such that $\Delta\xi_t = \Delta\xi_{t-m}$ where m refers to the distance between current period and its most recent analogue from the past. If there is no such analogue, the player simply expects the outcome of the previous period. Rules 16 and 17 apply the same logic to adaptive (v1.) and weak trend-following rules.[28]

These rules have been selected for two main reasons. First, they are commonplace in the literature (see Section 2).[29] Second, they are based on backward-looking heuristics (with the exception of the fitted rules and rule 14), so that their functional forms are easy to compute (e.g., rule 1). Finally, we have excluded level-k type of expectations in the vein of the rule learning model of Stahl (1996) and the cognitive hierarchy model of Camerer et al. (2004), since there is no common prior through which level-0 type can form imitation after the first period.[30]

[27]Note that fundamentalism is not equivalent to RE since it ignores the fact that other players might be nonrational. For these two to be equivalent, one needs to assume homogeneous expectations and common knowledge of rationality.

[28]For the sake of illustration, suppose that a subject uses rule 17. In period 25 of the baseline game, s/he should extrapolate the change between periods 8 and 9 rather than between 23 and 24 (as it would the case in normal trend rule). Following this logic, the similarity-based reformulation changes expectations in periods 25, 33, 41, 49, 57 for identical shocks, and for period 57 for the nonidentical ones.

[29]Note that the nomenclature used to describe these rules may vary across fields. For instance, our rule 1 is equivalent to Cournot play in standard game theory and to the random-walk-believing in finance.

[30]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 5 - Description of selected expectation rules for comparison

No	Description	Functional form
1	Naïve exp.	$p^e_{i,t} = p_{-i,t-1}$
2	Obstinacy	$p^e_{i,t} = p^e_{i,t-1}$
3	Adaptive exp. v1.	$p^e_{i,t} = 0.75p_{-i,t-1} + 0.25p^e_{i,t-1}$
4	Adaptive exp. v2.	$p^e_{i,t} = 0.50p_{-i,t-1} + 0.50p^e_{i,t-1}$
5	Adaptive exp. v3.	$p^e_{i,t} = 0.25p_{-i,t-1} + 0.75p^e_{i,t-1}$
6	Fitted adaptive exp.	$p^e_{i,t} = 0.89p_{-i,t-1} + 0.11p^e_{i,t-1}$
7	Strong trend-following exp.	$p^e_{i,t} = p_{-i,t-1} + 0.75(p_{-i,t-1} - p_{-i,t-2})$
8	Medium trend-following exp.	$p^e_{i,t} = p_{-i,t-1} + 0.50(p_{-i,t-1} - p_{-i,t-2})$
9	Weak trend-following exp.	$p^e_{i,t} = p_{-i,t-1} + 0.25(p_{-i,t-1} - p_{-i,t-2})$
10	Strong contrarian exp.	$p^e_{i,t} = p_{-i,t-1} - 0.75(p_{-i,t-1} - p_{-i,t-2})$
11	Medium contrarian exp.	$p^e_{i,t} = p_{-i,t-1} - 0.50(p_{-i,t-1} - p_{-i,t-2})$
12	Weak contrarian exp.	$p^e_{i,t} = p_{-i,t-1} - 0.25(p_{-i,t-1} - p_{-i,t-2})$
13	Fitted extrapolative exp.	$p^e_{i,t} = p_{-i,t-1} + 0.08(p_{-i,t-1} - p_{-i,t-2})$
14	Fundamentalist	$p^e_{i,t} = p^{NE}_t$
15	SBNE	$p^e_{i,t} = \begin{cases} p_{-i,t-1}, & \text{if } \xi_t = \xi_{t-1} \\ p_{-i,t-m}, & \text{if } \xi_t \neq \xi_{t-1} \end{cases}$
16	SBAE	$p^e_{i,t} = \begin{cases} 0.75p_{-i,t-1} + 0.25p^e_{i,t-1}, & \text{if } \xi_t = \xi_{t-1} \\ 0.75p_{-i,t-m} + 0.25p^e_{i,t-m}, & \text{if } \xi_t \neq \xi_{t-1} \end{cases}$
17	SBTF	$p^e_{i,t} = \begin{cases} p_{-i,t-1} + 0.25\Delta p_{-i,t-1}, & \text{if } \xi_t = \xi_{t-1} \\ p_{-i,t-m} + 0.25\Delta p_{-i,t-m}, & \text{if } \xi_t \neq \xi_{t-1} \end{cases}$

Note: Each description provides a rule for how players expect the average guess of other players one-period-ahead.

The goodness of fit of a given rule to the experimental data is based on the aggregate one-period-ahead forecast error that is computed as the root mean squared error (RMSE):

$$RMSE(p^h_t) = \sqrt{\frac{\sum_{t=3}^{64}\sum_{g=1}^{G}(p^h_{g,t} - p^e_{g,t})^2}{62 \times G}}, \tag{7}$$

where $p^h_{g,t}$ is the prediction of rule $h \in \{1, ..., 17\}$ for the average expectation of group g in period t and $p^e_{g,t}$ is the actual average expectation of group g in period t. Here, the superscript G is the scale of RMSE. A lower value of RMSE points to a better fit. We measure the RMSE in three different ways: at the rematching 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 two past observations.

Panel A of Table 6 reports the RMSE for each of the seventeen rules. For baseline as well as pooled data, SBNE (rule 15) achieves the best fit. For the treatment data, is it slightly outperformed by SBTF (rule 17). In line with Hypothesis 4, the rules that are augmented with similarity-based learning (rules 15 to 17) yield a better fit than the remaining ones, with RMSE two times smaller than under the worst performing fundamentalist rule. The last line in Panel C shows how much switching to similarity-based learning improves the fits. As can be seen, the improvements emerge in the baseline condition in which identical shocks reoccur in a periodic manner. In the treatment condition, only the first and the last shocks are identical which leaves less space for applying similarity-based reasoning.[31] At the cluster level, the rules that are augmented with similarity-based reasoning provide the best fit for most (10 out of 12) rematching clusters. The data from the baseline (treatment) rematching clusters are best organized by the rules that are derived from adaptive (extrapolative) learning. Lastly, in line with Hypothesis 3, fundamentalism (rule 14) is always the worst fitting model regardless of the data aggregation level.

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

Result 4a: Reformulating the rules with similarity-based reasoning improves their fit to the data, especially when shocks are identical.

Result 4b: Overall, the SBNE rule has the best fit among all homogeneous expectation rules.

Note that these comparisons rely on the central 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: a given rule may at first perform poorly, but later on become more relevant due to the experience. For this

[31] This is most starkly observable when we compare the improvements in RMSE from refining rules with similarity-based learning for period 57, i.e. the first post-shock period in round 4. More details are provided in Table B2 in Appendix B that reports the changes in fits in the initial post-shock periods.

reason, we first refer to an evolutionary model of expectations: the HSM of Anufriev and Hommes (2012).[32] 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

$$p_{t+1}^{HSM} = \sum_{h=1}^{H} n_{h,t} p_{t+1}^{h}, \tag{8}$$

where $n_{h,t}$ is the impact factor of heuristic h at period t. This impact factor can be interpreted 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

$$U_{h,t} = -(\bar{p}_{-i,t} - p_t^h)^2 + \eta U_{h,t-1}, \tag{9}$$

where $\eta \in [0, 1]$ is a free parameter representing the weight assigned to the past performance compared to current. $\eta = 0$ implies that 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

$$n_{h,t} = \delta n_{h,t-1} + (1 - \delta) \frac{exp(\beta U_{h,t-1})}{\sum_{h=1}^{H} exp(\beta U_{h,t-1})}, \tag{10}$$

where the impact of the expectation heuristic h at period t depends on its accumulated impact and its relative performance normalized with the sum of all competing heuristics. There are two free parameters in (10). 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.[33]

To compute expectations with HSM and compare fits, one must first determine which

[32] Their model inherits its main features from Brock and Hommes (1997).

[33] $\beta = 0$ would imply equal impacts regardless of the differences in performances.

expectation heuristics to include, and then set their initial impacts as well as assign the values to free parameters η, δ, β. Following a common practice in the literature, we consider four different heuristics. We consider three classes of expectation rules – adaptive, extrapolative and similarity-based – and, for each of them, choose the best-fitting specification (rules 3, 8, 15, respectively).[34] We also include an equilibrium-based rule: fundamentalism (rule 14). We assign the initial impact factors equal to $n_{h,3} = 0.25$ for all h and set the free parameters to $\eta = 0.1, \delta = 0.4, \beta = 0.1$. By trial and error, we discover that this combination of free parameters fits the data best.[35]

The second heterogeneous expectation model we consider – MIXEXP – follows Haltiwanger and Waldman (1985, 1989) and is based on the assumption that each group is composed of n rational players and $5-n$ naïve players. The main reason for including it in our comparison is that it provides good descriptive fit to the experimental data in Cooper et al. (2017). Like them, we consider a single rational player per group ($n = 1$). The remaining players make naïve forecasts according to the SBNE. In each round, the subject whose prediction error in the first post-shock period is the smallest in a given group is considered as the rational player in that group.[36] The rational player forecasts consistently with RE:

$$p_{i,t}^{RE} = \bar{p}_{-i,t} + \epsilon_{i,t}, \tag{11}$$

where $\epsilon_{i,t} \sim \mathcal{N}(0,1)$.

Panel B in Table 6 reports RMSE for each of the heterogeneous expectation models.[37] These results show that the HSM fits better than all the other rules, including MIXEXP, at every data aggregation level. The improvement in fit compared to the best homogeneous expectation rule ranges between 15% and 53% depending on the scale and equals 32% at

[34]Note that for the class of adaptive expectations, the best fitting rule is naïve expectations (rule 1). However, since SBNE (rule 15) is already a refined version of that rule, we include rule 3 instead.

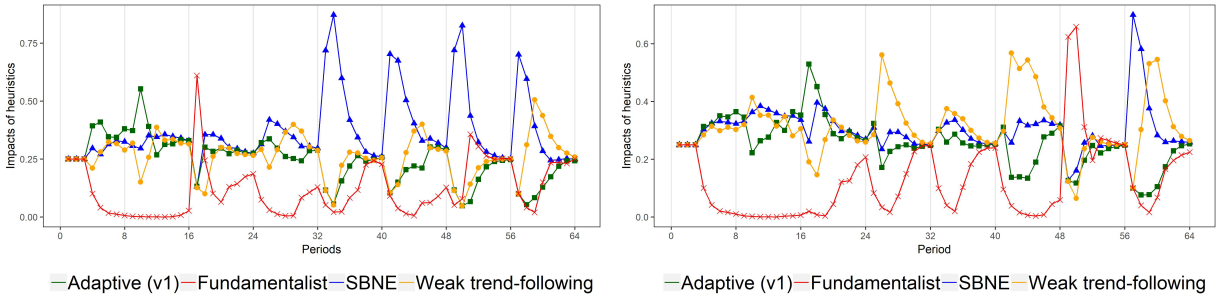
[35]As a robustness check, in Appendix C.1 we provide RMSE obtained under various combinations of free parameters. While the outcomes do vary in absolute terms, the relative standing of different rules for a given period remains fairly stable.

[36]We are indebted to Michael Waldman for suggesting us this strategy.

[37]For the HSM, RMSE is computed for periods 4-64 where expectations are endogeneously determined.

pooled level. Albeit not as well performing as the HSM, MIXEXP performs better than all the homogeneous expectation rules which makes it the second-best expectation model in our exercise.

Result 5: Models with heterogenous expectations better fit the data than those with homogeneous expectations.



(a) Baseline (b) Treatment

Figure 3: 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.

Finally, Figures 3a and 3b show the evolution of the impact factors of different heuristics over time in the baseline and treatment conditions. Notwithstanding the claim of "experience eliminates naivety", for both experimental conditions we find that SBNE attains the highest average impact factor in the final round of the game.[38] On the opposite extreme, fundamentalist expectations attract very low weights during the first two rounds, but their role increases toward the final round.

5 Discussion

The results outlined in Section 4.2 reveal that the refinement of expectation rules with similarity-based learning improves their fit considerably when shocks are identical. Notably, the SBNE rule performs well under both homogeneous and heterogeneous expectations.

[38]Table B3 in Appendix B reports the average of impact factors separately for pre-shock and post-shock phases.

Table 6 - 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	**7.12**	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	6.53	6.58	5.73	7.25	5.89	8.11	6.75	6.30	6.61	6.56	7.15	6.59	6.73	6.66	6.70
7	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
8	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
9	6.25	6.71	5.83	7.52	5.62	8.24	6.72	6.23	6.04	**6.31**	7.58	6.44	6.76	6.57	6.67
10	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
11	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
12	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
13	6.64	6.67	5.79	7.25	5.99	8.28	6.84	6.35	6.80	6.69	7.19	6.72	6.82	6.77	6.79
14	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
15	**4.61**	4.58	4.00	6.62	**4.08**	6.13	6.14	**6.23**	6.47	6.58	7.26	6.47	**5.10**	6.54	**5.86**
16	4.82	**4.55**	**3.99**	**6.61**	4.18	**5.99**	6.29	6.40	6.66	6.72	7.16	6.57	5.11	6.64	5.93
17	4.91	5.09	4.56	7.13	4.31	6.42	**6.11**	6.27	**6.00**	6.48	7.64	**6.36**	5.50	**6.50**	6.02
Panel B - Heterogeneous expectation models															
HSM	2.45	3.45	3.37	5.46	2.56	5.08	3.02	3.47	4.56	2.95	5.51	4.34	3.90	4.08	3.99
MIXEXP	4.13	3.94	3.44	5.63	3.38	5.15	4.71	5.62	5.09	5.12	6.31	5.34	4.36	5.39	4.90
Panel C - Changes in fit between models															
Δ HSM	47%	24%	16%	17%	37%	15%	51%	44%	24%	53%	23%	32%	24%	37%	32%
Δ SB	28%	31%	31%	10%	30%	26%	9%	0%	1%	-3%	0%	1%	24%	1%	12%

Note: RMSE calculated at different data aggregation levels. The first column indicates the number assigned to the rule in Table 5 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 and MIXEXP in panel B refer respectively to the RMSE of the HSM and of the model with one rational player. Panel C reports the changes in fits for three comparison groups. Δ HSM reports the improvement in fit by using the HSM compared to the best fitting homogeneous expectation rule. Δ SB reports the changes in fit between the best upgraded rule with similarity-based learning compared to the nonupgraded version of the same rule. A negative value of RMSE implies a loss in fit.

Under this rule, a player expects the last period's 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 characterized by the same change. This rule is consistent with the Gilboa and Schmeidler (1995) case-based decision theory and the similarity-based learning model of Plonsky et al. (2015), 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 for 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 so that there is less necessity for extrapolation. The naïve and weak trend-following rules therefore tend to perform better. For instance, this may explain why a trend model with extrapolation factor 1 performs poorly in the AMEs of Marquardt et al. (2019): weak trend-following and naïve heuristics create slow convergence toward NE, analogously to what we observe in our experiment.

Lastly, the results in the second part of Section 4.2 indicate that allowing heterogeneity in expectations through the HSM or MIXEXP improves the fits substantially in line with the evidence from the LtF literature. The evolution of impact factors computed through HSM implies that there is more heterogeneity in nonidentical shocks condition. This pattern

– if proves to be robust – suggests that the presence of real-world complications such as nonidentical shocks may make coordination over an expectation rule more difficult.

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 thanks to repetition, yet only slightly: despite four repetitions, convergence 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 the backward-looking rules organize the data well. In particular, rules refined with similarity-based learning approach outperform the others in terms of predictive power.

Our experiment successfully documents the robustness of 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 its acceleration over repetition. This evidence strengthens the empirical validity of the strategic environment effect, in line with an early conjecture by Haltiwanger and Waldman (1985). Furthermore, these patterns of learning to play equilibrium under strategic complementarity persist in a more complex environment with time-varying shocks. We fit a large set of expectation rules to provide an individual-based explanation to the observed aggregate dynamics. The SBNE rule, a simple learning rule first proposed by Cooper et al. (2017) and not yet studied in a comparative analysis, outperforms other rules in terms of descriptive accuracy.

The main implication of these findings is that inertia in adjustment may rather persist over time. The fact that the type of shock does not affect behavioral dynamics suggests that sluggishness is an inherent feature of strategic complementarity. Importantly, 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 nonrational expectations suffice to create stickiness, and potentially opens the door 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.[39]

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, we reckon that four rounds should be sufficient for observing accelerated adjustment in a relatively simple environment like ours. Another design choice 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.[40] 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. We believe that varying the form of the heterogeneity of shocks constitutes a possible agenda for future studies.

[39]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.

[40]Cooper et al. (2017) test this argument in an auxiliary treatment and find that the main results are qualitatively unchanged. In the light of this result, the question of matching scheme should be less of a concern.

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 × (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

Nombre cible = $0.75 \times$ (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 x 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 = $0.75 \times$ (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 x 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

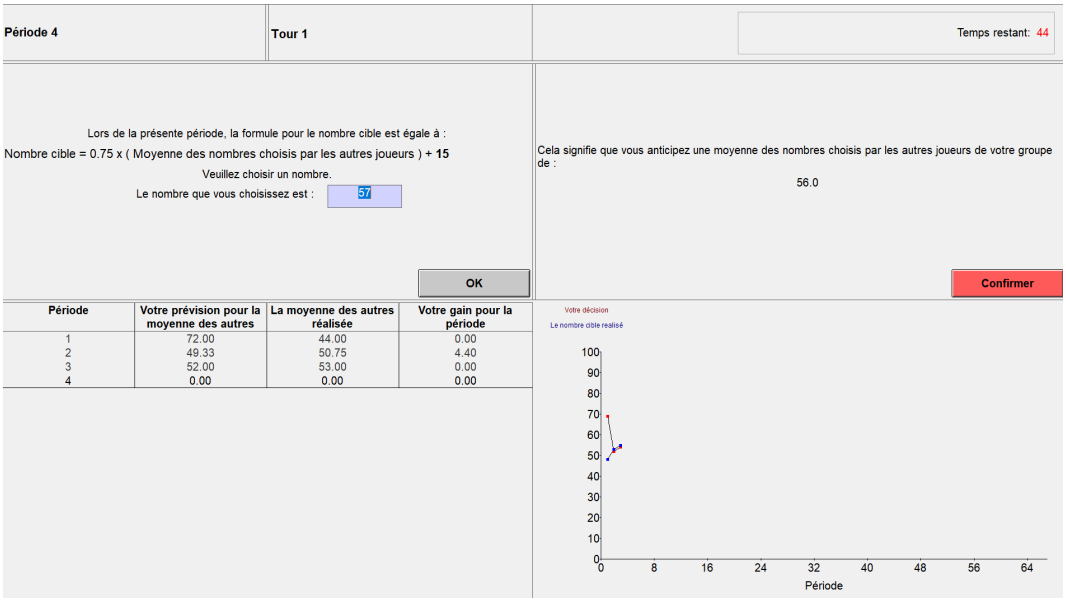


Figure 4: An (arbitrary) example for the decision screen that subjects used to enter their guess

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.

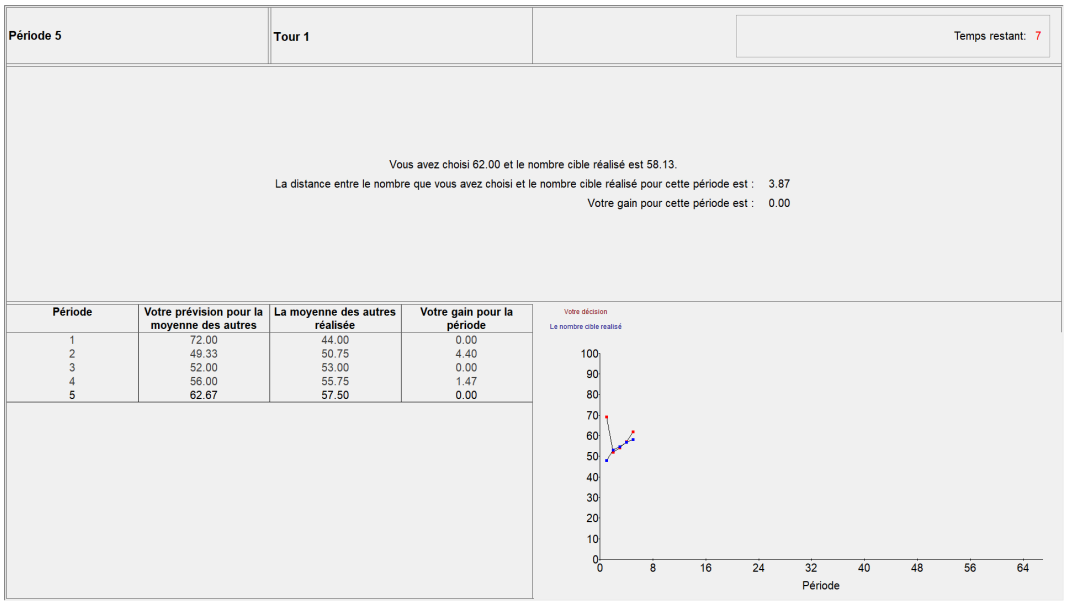


Figure 5: An (arbitrary) example for the feedback screen that subjects used to see the results of the period

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

Bienvenue!

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 devrons 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 chauqe 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

Veuillez 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.

Veuillez 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 nonbinding 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 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.

B Additional Figures and Tables

Table B1 - Descriptive statistics of the sample

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

Note: Standard deviations are reported in parentheses.

Table B2 - Improvements from similarity Refinement in the initial post-shock period

	Baseline			Treatment
Periods	25	41	57	57
$\Delta SBNE$	-0.117	0.383	0.682	0.286
$\Delta SBTF$	-0.363	0.152	0.451	0.210
$\Delta SBAE$	0.000	0.502	0.723	0.335

Note: Each row reports the percent change in RMSE when the SBNE, SBTF or SBAE rules are taken instead of their non-refined equivalents. The columns correspond to the data from the intial post-shock periods (one per round)

Table B3 - Average of impact factors in the HSM

Panel A - Average impact factors of pre-shock phases								
	Baseline				Treatment			
Rounds	Rule 15	Rule 3	Rule 8	Rule 14	Rule 15	Rule 3	Rule 8	Rule 14
1	0.31	0.37	0.28	0.03	0.32	0.34	0.30	0.04
2	0.29	0.27	0.24	0.21	0.31	0.34	0.26	0.09
3	0.47	0.19	0.21	0.12	0.29	0.26	0.31	0.14
4	0.42	0.17	0.19	0.23	0.23	0.20	0.21	0.35
Panel B - Average impact factors of post-shock phases								
	Baseline				Treatment			
Rounds	Rule 15	Rule 3	Rule 8	Rule 14	Rule 15	Rule 3	Rule 8	Rule 14
1	0.33	0.36	0.30	0.00	0.36	0.31	0.34	0.00
2	0.34	0.29	0.31	0.06	0.27	0.24	0.35	0.13
3	0.45	0.22	0.27	0.06	0.31	0.22	0.43	0.03
4	0.37	0.15	0.32	0.16	0.37	0.16	0.34	0.13

Note: Each value in panel A and B represents the average impact factors computed for the HSM described in Section 4.2 for pre-shock or post-shock phases. The pre-shock phase of first round includes periods 4 to 8.

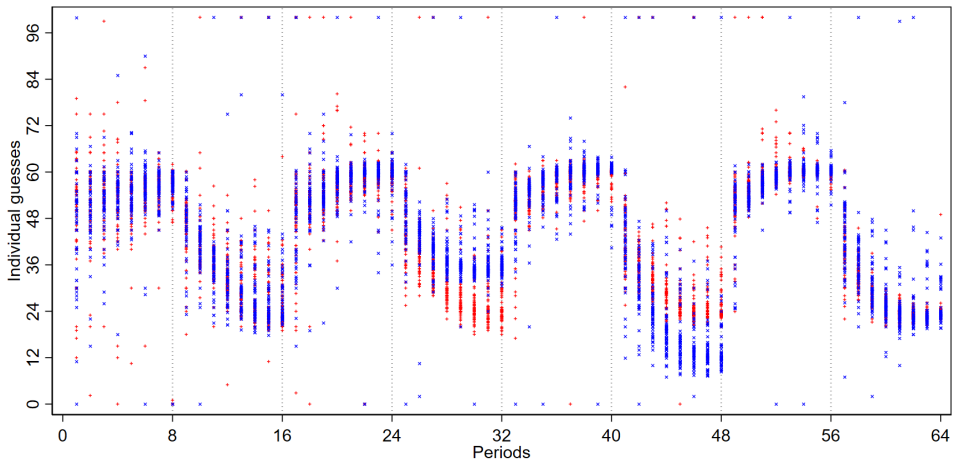


Figure 6: Distribution of individual guesses across periods. 120 subjects per period. Plus signs (cross signs) represent baseline (treatment) conditions.

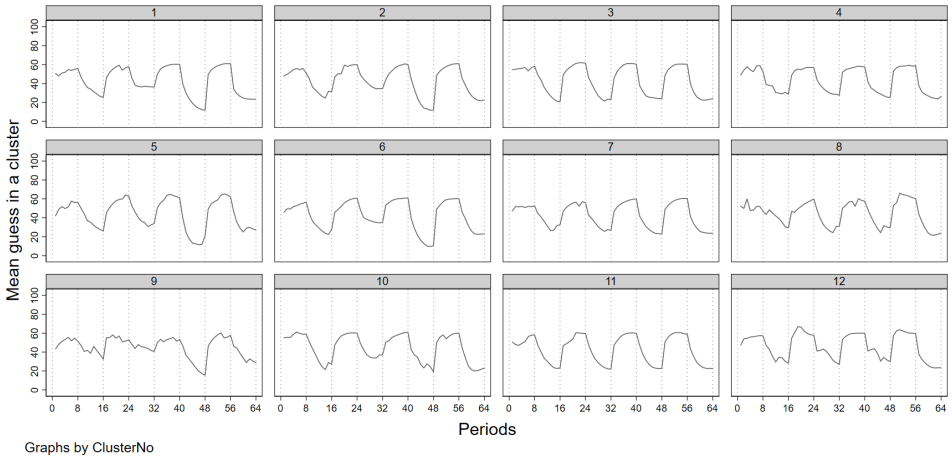


Figure 7: Average guess of clusters by periods. The identification number of clusters are denoted above each figure. Clusters 3, 4, 7, 8, 11, 12 belong to baseline sessions and clusters 1, 2, 5, 6, 9, 10 belong to treatment sessions.

C Robustness Analyses

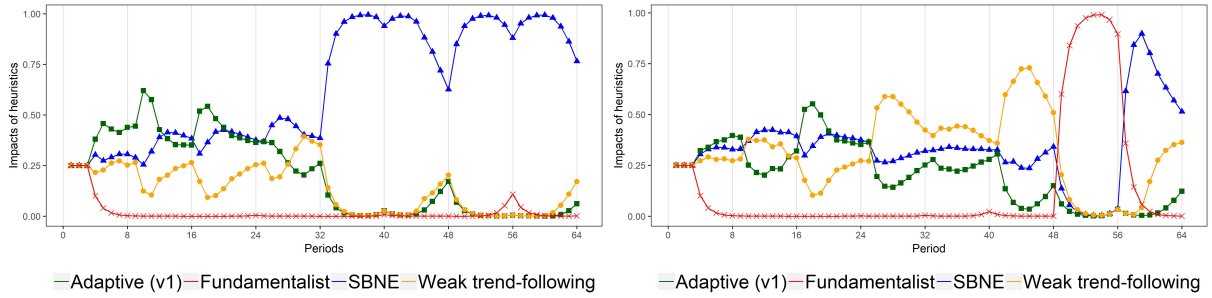
C.1 Calibration of Free Parameters in HSM

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 versions of HSM where in each step one free paramater gets closer to the one in Bao et al. (2012). Table C1 reports the results and Figure 8 projects the impact factors as time series. Results show that the fit worsens when any of the parameter increases *ceteris paribus* and the RMSE of HSM 4 is almost equal to the best homogeneous expectation rule. Nonetheless, the fits of HSM 2 and 3 are 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 η and β but it is somehow sensitive to the abrupt changes in δ, at least for the 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 C1 - RMSE of HSM under different values of free parameters

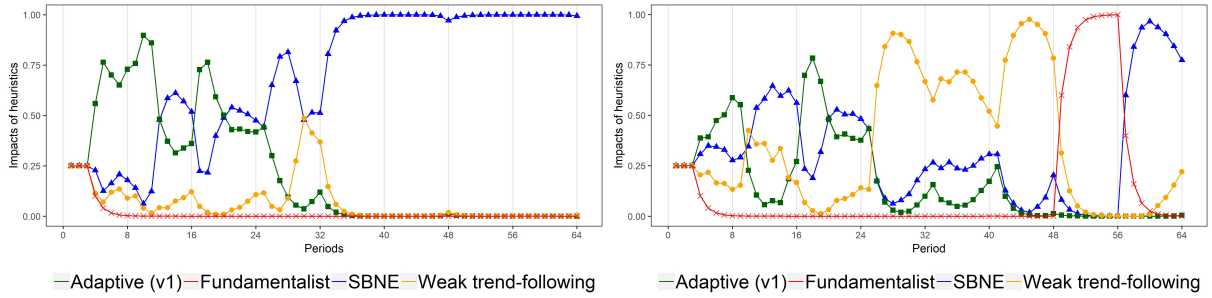
Comparison Level	HSM 1	HSM 2	HSM 3	HSM 4
Baseline	3.90	4.43	4.54	5.10
Treatment	4.08	5.68	5.68	6.46
Pool	3.99	5.09	5.14	5.82
η	0.1	0.7	0.7	0.7
δ	0.4	0.4	0.4	0.9
β	0.1	0.1	0.4	0.4

Note: HSM 1 is equivalent to our benchmark HSM reported in Table 6. HSM 4 has the benchmark parameters combination from Bao et al. (2012) and HSM 2 and 3 provide an intermediate change in parameters.



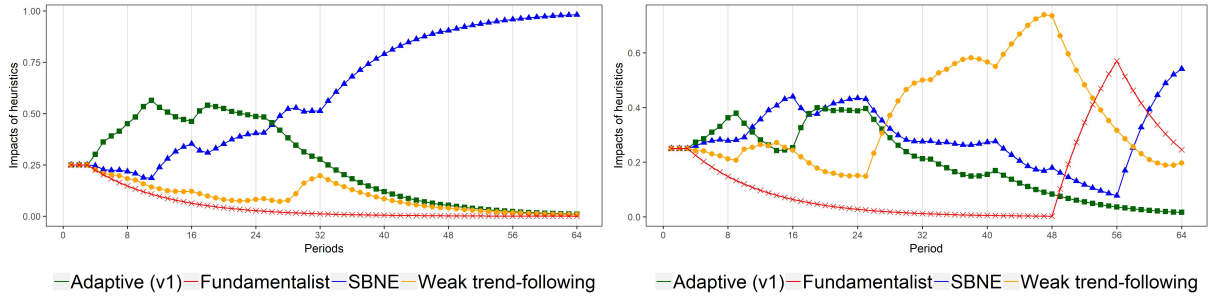
(a) Baseline - HSM 2

(b) Treatment - HSM 2



(c) Baseline - HSM 3

(d) Treatment - HSM 3



(e) Baseline - HSM 4

(f) Treatment - HSM 4

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

C.2 Cognitive Skills and Individual Expectations

In this Appendix 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

$$\frac{1}{16}\sum_{t=1}^{T=16}\frac{|p_{i,t}^{e}-\bar{p}_{-i,t}|}{\bar{p}_{-i,t}}, \tag{12}$$

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. These test scores are designed to measure some of the cognitive skills of subjects. Baccalaureate grade is the score that subjects obtained upon completion of their secondary education. The CRT score is computed through three items and is designed to measure one's ability to reflect on a question and override reporting the gut response. The RMET score is computed through thirty-six items and is designed to measure one's capacity to infer the internal emotional states of others. Memory task is designed to measure one's short-term memory capacity through the number of items that one can remember with the correct order after seeing a sequence of numbers. The representativeness (availability) score is computed through two (five) items and designed to measure one's propensity for reasoning according to the representativeness (availability) heuristic.

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. We use a random effects specification ($N = 480$ per regression).

Tables C2 and C3 report the corresponding estimates. Overall, the included test scores only weakly explain the variation in either measure of interest. A higher RMET score predicts a lower forecast error and a better fit for the fundamentalist rule. By contrast, in groups where subjects are more prone to representativeness heuristic, Nash play is less likely to occur.

The round dummies are systematically found to explain the dependent variables. Compared to round 1, in round 4 the values of ARPE drop and the fits of all the compared rules improve. Note that even with this improvement, 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 the rules. So, the improvement in the goodness of fit over time is attenuated under nonindentical shocks. The coefficient of this interaction term is also positive for ARPE (in line with Hypothesis 2), but not statistically significant.

Table C2 - RMSE of expectation rules and individual characteristics

Variables	SBNE	SBAE	SBTF	Fundamentalism
Intercept	10.52***	10.37***	11.42***	19.14***
	(3.36)	(3.13)	(3.39)	(2.83)
Treatment	-0.75	-0.58	-0.81	0.24
	(1.42)	(1.41)	(1.55)	(1.62)
Round 2	0.74	0.52	0.81	-2.26***
	(1.19)	(1.22)	(1.26)	(0.99)
Round 3	-1.69	-1.95*	-1.37	-4.38***
	(1.14)	(1.13)	(1.23)	(0.95)
Round 4	-3.67***	-3.81***	-3.43***	-8.05***
	(1.28)	(1.202)	(1.38)	(0.87)
Round 2 × Treatment	-0.18	-0.06	-0.16	-2.98**
	(1.31)	(1.34)	(1.38)	(1.16)
Round 3 × Treatment	2.28	2.51	1.74	4.38**
	(1.55)	(1.54)	(1.62)	(1.71)
Round 4 × Treatment	6.47***	6.43***	6.25***	3.24***
	(1.40)	(1.34)	(1.48)	(1.24)
Baccalaureate grade	2.99**	2.66**	2.66**	2.35*
	(1.22)	(1.24)	(1.11)	(1.32)
CRT score	-1.53	-1.54*	-1.65*	-0.81
	(0.95)	(0.88)	(0.97)	(0.74)
RMET score	-2.56	-1.88	-3.05	-5.41***
	(1.94)	(1.73)	(1.97)	(1.99)
Memory score	-2.95	-2.43	-3.13	-2.55
	(3.44)	(3.33)	(3.51)	(3.52)
Represent. score	0.81	0.61	1.16	2.89***
	(1.23)	(1.15)	(1.24)	(0.89)
Availability score	-0.75	-0.56	-0.93	-0.49
	(2.07)	(1.95)	(1.99)	(1.36)
R^2 overall	0.08	0.08	0.08	0.22

Note: Coefficients from random effect regression models. The dependent variables are the RMSE of four expectation rules per subject computed for each round ($N = 480$ per regression). The independent variables are standardized test scores, round and treatment dummy variables and their interactions. Robust standard errors clustered at the rematching cluster level (12 clusters per condition) are reported. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively.

Table C3 - Prediction errors and individual test scores

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.174***	0.176***	0.186***	0.219***	0.208***	0.154***	0.176***	0.224***
	(0.022)	(0.027)	(0.023)	(0.027)	(0.034)	(0.025)	(0.024)	(0.041)
Treatment	-0.001	-0.001	-0.000	-0.000	-0.001	-0.001	-0.001	0.001
	(0.024)	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)	(0.024)	(0.024)
Round 2	-0.043**	-0.043**	-0.043**	-0.043**	-0.043**	-0.043**	-0.043**	-0.043
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Round 3	-0.066***	-0.066***	-0.066***	-0.066***	-0.066***	-0.066***	-0.066***	-0.066***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Round 4	-0.107***	-0.107***	-0.107***	-0.107***	-0.107***	-0.107***	-0.107***	-0.107***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Round 2 × Treatment	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Round 3 × Treatment	0.065*	0.065*	0.065*	0.065*	0.065*	0.065*	0.065*	0.065*
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Round 4 × Treatment	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)
Baccalaureate grade		-0.002						0.016
		(0.021)						(0.019)
CRT score			-0.024*					-0.020
			(0.014)					(0.014)
RMET score				-0.069**				-0.070**
				(0.029)				(0.031)
Memory score					-0.053			-0.021
					(0.042)			(0.037)
Represent. score						0.030*		0.030
						(0.017)		(0.016)
Availability score							-0.003	-0.018
							(0.026)	(0.022)
R^2 overall	0.138	0.138	0.147	0.152	0.145	0.148	0.138	0.170

Note: Coefficients from random effect regression models. The dependent variable is the individual ARPE computed for each round ($N = 480$ per regression). The independent variables are standardized test scores, round and treatment dummy variables and their interactions. Robust standard errors clustered at rematching cluster level (12 clusters per condition) are reported. *, **, *** indicate statistical significance at the 10, 5, 1% level, respectively.

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.

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. *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.

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

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. *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. *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. *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.

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. *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. *Review of Economic Studies 76*(3), 1125–1147.

Prevost, M., M.-E. Carrier, G. Chowne, P. Zelkowitz, 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.