UNDERSTANDING BEHAVIOUR THROUGH THE LENS OF BOUNDED RATIONALITY: EXPERIMENTS WITH HUMAN AND ARTIFICIAL AGENTS

LEE-PENAGOS, ALEJANDRO

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

THE UNIVERSITY OF NOTTINGHAM

SCHOOL OF ECONOMICS

JULY 2017

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Pursuing a Ph.D. is a difficult path that many consider to be mostly a lonely journey. However, I believe no one can accomplish great achievements entirely on its own. Indeed, the culmination of this dissertation was a journey on which many people, directly and indirectly, made significant contributions, both at the academic as well as at a personal level. I was, taking a cue from some Metallica lyrics, “by myself but not alone”.

Firstly, acknowledgements go to my thesis supervisors, Chris Starmer and Alex Possajennikov. I greatly thank both of them for all the support that any student can expect from fantastic supervisors. To Chris, who was my supervisor from day one, I particularly thank for the almost unlimited intellectual and administrative freedom he gave me on pursuing any topic that satisfied my curiosity. I explored a variety of topics ranging from happiness to fear of crime, that even if didn’t end up being part of my research, they were part of a vital trial-and-error process to finally identify the areas on which this dissertation focuses. His encouragement when switching methodologies from lab experiments and venturing into the (for us) uncharted territory of agent based models, was vital in allowing me to find what I consider now to be my long-term research interests and true philosophical inclinations. To Alex, who became my supervisor once a computational methodology was decided around half-way of my doctoral studies, I greatly thank his approachability and willingness to discuss ideas at almost any time. His insights regarding even the smallest of details, both mathematical and computational, represented a constant challenge and motivation to keep my arguments and analysis as sharp as possible.

At the academic level, there’s one more person to whom I am deeply grateful. I met John H. Miller during a graduate workshop at the Santa Fe Institute during the summer of 2015, as he was the director of such program, environment that has been the most intellectually stimulating of my life. I had the fortune to keep working with him afterwards, and I feel as if our interactions are like a prolonged extension of the intellectual challenge and curiosity that I now associate with that summer in the mountains. He became almost like an unofficial supervisor for the third chapter of this thesis, and his support ranged from detailed discussions on the implementation of code and algorithms, to writing style and career advice. Thanks to John for all his support during the last stretch of this journey.

At a personal level, many friends had a positive impact during my stay in the UK. Without a healthy social life, I probably wouldn’t have had the required stamina to complete this dissertation, so I’m greatly indebted to them for their indirect contribution. Particularly, I specially thank Alessio Gaggero and Elkin Castro. Is not easy to build family-like bonds away from home, but I’m proud to have found two “brothers from another mother” in the UK. With them, the scientific mindset went beyond formal research, and our field experiments in social dynamics conducted in all the bars of Nottingham (as well as some in Italy, Edinburgh and Spain), represented the most fun I probably have had in my life. To them, along with Jose Andre and Fahad “GURU!” Mohammad, I will always be thankful for the “maratonas semanais” days, always fostering and encouraging those much needed breaks from the heavy hours of research. Finally, I thank Franzi Gehringer for the (unfortunately short) time we had together. Her open mind, willingness to try new things and honest care for the

ones around her, represent a light that will shine in my direction even if we remain in opposite extremes of the planet. Never cleaning and organizing a house had been so delightful!

Finally, many thanks to my family and friends in Colombia, who were a great support during my final year of writing up back home. To my mom, Deborah, for her kind support, unconditional love and that mostly needed "home" feeling back in Bogotá. If there's one person to whom I owe the most for the completion of my doctoral studies, it is to her. To you, mom, I'm sorry to say that even if I wish, I cannot promise to stop the causes of that induced neck strain, for which I know you will laugh forever. To my brother, Juan Camilo, thanks for the time we had once again in this final year, reminding me how cool it is to fight over philosophical differences. To my cousin Andrés Felipe and the "Lebron Brothers" group, thanks for all the mindless virtual fun, which has helped me keep sanity during difficult times (particularly Mario's multimedia outlets). To Diego Novoa, thanks for allowing me the pleasure of preventing you from scoring a single goal, week after week. And to Andrea, Margarita and Alicia, thanks for reviving the feelings of having true teenage friends. Never has neighbourhood friendship been so special as in this last year. To the three of you, I say this: "thanks to you and my family for making this dissertation happen, hoping it contributes greatly in eradicating world hunger and bringing world peace" (there you go, I said it!).

PROLOGUE

This thesis consists of three separate papers, all of them independent and self-contained, making contributions in three different topics: money illusion, public goods games with altruistic punishment and coordination games including exogenous signals. The objective when starting this thesis was not to pursue research in one single particular topic, as sometimes is customary. With each paper I explored research questions according to what sparked my intellectual curiosity at the moment. However, if one is to put all three chapters under one single umbrella, in hindsight the overarching topic is the empirical and theoretical study of bounded-rationality. This short prologue is not intended in any way as a survey of the literature in bounded-rationality or a formal introduction to each individual chapter. Instead, it is a short, informal essay pointing towards currents of thought that have greatly influenced me and my work.

A famous Leo Tolstoy's quote says "All happy families are alike; every unhappy family is unhappy in its own way". Without the emotional connotation, rational behavior is, in a way, similar to Tolstoy's happy families. Rational behavior's axioms are well stablished and, some discussions aside, it seems straightforward to identify classical rationality. However, deviations from it are varied, and there's no single agreement on what type of behavior is referred to when the term bounded-rationality is used. Anecdotally, this is reflected in the Wikipedia entry for "list of cognitive biases", reporting over 170 ways in which behavior can be irrational. More formally, Ariel Rubinstein presents in his book "Modelling Bounded Rationality" (Rubinstein, 1998) an extensive comment from Herbert Simon that highlights how the two of them, great scholars in the field, have contrasting views regarding how bounded-rationality should be modelled and understood. In this prologue I will mention three specific lines of research on bounded-rationality. I believe they give some structure to the many possible ways in which behavior deviates from rational decision making. The implications of the differences between these currents of thought have been a key factor influencing the approaches I explore in this thesis.

Following the typology of Gigerenzer et al. (2011), there are three main influential lines of research related to bounded rationality: optimization under constraint, the heuristics-and-biases program and the fast-and-frugal heuristics approach.

The first of them, optimization under constraint, is the one that has influenced my work the least. Under this line of research, traditional optimization methods are still the main tool, but psychological plausibility is introduced by way of additional constraints. This approach can sometimes lead to a puzzling logic: the more constraints are added, the mathematics for optimization can become more difficult. So the more plausible a model tries to be by making subjects face more constraints, at the same time it can require agents to be capable of doing more complicated calculations. As an example, in a Quantal Response Equilibrium model (for example, McKelvey and Palfrey (1998)) agents are thought to be boundedly-rational by allowing them to make mistakes (include errors) in their choices of strategies. However, they are still assumed to make equilibrium calculations based on this errors, which are more difficult than if no mistakes were made. It is worth mentioning that one way to scape this puzzle is to think of the models as 'as-if' approaches, in line with Friedman (1953). In any case, the topics explored in this thesis did not lead to implementations of this approach, although they are an important line of research in bounded-rationality.

The second program, the heuristics-and-biases (Kahneman et al., 1982), is perhaps the most influential and widely recognized across all of social sciences. This approach brought to the spotlight decision making by way of heuristics, or simple rules of thumb that are to be used when optimization is out of reach. They can be thought of in some cases as akin to 'cognitive illusions', or quick decisions made by our intuitive reasoning that can lead us to be 'predictably irrational', borrowing the term from Dan Ariely. In this approach, it is normally assumed that heuristics either lead to biased decisions relative to traditional rationality, or, as in Payne et al. (1993), that they imply a trade-off between the accuracy of a decision and the effort (e.g., time) required to make it.

The heuristics-and-biases program has influenced my research through what I believe is a key epistemological insight from this line of thought. Sometimes, ground-breaking progress in knowledge can be made by understanding how people do *not* make decisions. As explained by Thaler (2015), documenting deviations from traditional rationality and showing that the latter could not explain several empirical facts was a key factor in the emergence of behavioural economics as a field. Indirectly, similar was the motivation for the first chapter of this thesis on Money Illusion. There, I test if subjects deviate from behaviour predicted by the notion of Money Illusion. I found that such theory cannot account for all the empirical observations from my experiments, suggesting that subjects are using different rules of decision depending on the environment they are facing. However, one more general conclusion that can be drawn from that chapter is that even if understanding heuristics as general deviations from rationality has allowed us to make incredible progress, it also has limitations. Is not clear how a heuristic works under different environments. To explore deviations from a given type of behaviour, it is less of an issue, but if one is to further explore in which situations, under which conditions and exactly how those deviations take place, specifying better what those rules of thumb exactly are becomes the next natural step.

And this is where the last approach, the fast-and-frugal heuristics program (Gigerenzer et al., 2002), comes into play. I will not focus on the critiques this approach presents towards the heuristics-and-biases program (discussed in Gigerenzer et al. (2011)), but on how it pushes it forward. The key point is that is focuses on taking heuristics beyond general deviations from rationality, endorsing the definition of rules of thumb as computational algorithms. By translating rules of decision into algorithms, the researcher necessarily has to clearly state the information available, how it is used, when to stop looking for more information and how the decision comes about. This algorithmic approach makes the decision process inherently quantifiable, fostering quantitative tests of its performance across different environments. For this thesis, even though no work in the fast-and-frugal tradition is presented, the notion of defining behaviour as concrete algorithms impacted the methodologies used starting from chapter two. The latter introduces a model to explain experimental data in a particular environment: public good games with punishment. The focus there is to model individual decision making by explicitly specifying an agent's learning and decision algorithms.

The fast-and-frugal heuristics approach has been perhaps the one with the most influence on my personal views, highly influencing the research presented in this thesis, when it comes to individual decision making. But its influence is also related to how it opened the door to the topics explored in the third and final chapter. By introducing decision making as well defined computational algorithms, it allowed this thesis to progress naturally towards the exploration of how those individual decisions interact with each other in order to create aggregate patterns. In environments where the interactions between agents is crucial to understand particular phenomena, especially when there is heterogeneity of behaviour, closed form mathematical solutions can become intractable, and aggregate simplifications (such as a representative agent) can leave the most interesting issues unexplained. A computational approach supports the study of complexity and emergent behaviour by way of modelling artificial, interacting agents via simulations. The literature in complexity is scattered given how novel its study has been compared to other approaches, but a great example of its philosophy and methodology applied to economic environments is found in Kirman (2010). The motto of complexity is that aggregate patterns cannot be inferred by analysing individuals separately and on their own: emergent behaviour cannot be grasped without the study of specific interactions among individuals, for which computer simulation complements traditional methodologies. The third and final chapter of this thesis studies the emergence of behaviour in evolutionary games (particularly in coordination games). It focuses on what types of individual behaviour can be learned when there is a dynamic feedback between system and individuals, both influencing each other. Its study in this case is possible thanks to the bridge that the computational tools create between micro and macro behaviour.

In order to close this prologue, it seems customary to make general conclusions derived from all three chapters. As mentioned above, given that each chapter tackles a different topic, particular conclusions can be found in each one of them, so I won't repeat them here. However, two main lessons, I believe, can be drawn from this thesis as a whole. The first is related to methodology, as hinted throughout all of these preliminary pages. Computational experiments and simulations are an essential aspect of the study of human behaviour, complementing traditional methodologies in environments where closed form solutions or big scale experiments are out of reach. The research conducted for the completion of this thesis has led me to the firm belief that their influence in social sciences will only grow over the decades, as it has done in other fields (such as physics or meteorology). But also, the three chapters have pointed towards the necessity for social sciences to foster their trans-disciplinary nature. This thesis started with the study of problems particularly relevant for economics, but its development quickly required gathering insights from psychology, computer science, philosophy and biology. And I am convinced that given the scope of modern times social problems such as global warming, ecological sustainability or disease control in a highly interconnected planet, no single field can encompass the necessary tools to tackle them thoroughly. So both main conclusions would support researchers in developing multi-disciplinary scientific toolkits that are not necessarily constrained by narrow definitions of individual fields. The nuts-and-bolts of social science cannot be grasped by one single discipline.

REFERENCES

FRIEDMAN, M., (1953). "The Methodology of Positive Economics," in: Essays In Positive Economics. University of Chicago Press, pp. 3–16, 30–43.

GIGERENZER, G., HERTWIG, R., PACHUR, T., (2011). Heuristics: The foundations of adaptive behavior. Oxford University Press, New York.

GIGERENZER, G., TODD, P., GROUP, A.B.C.R., (2002). Simple Heuristics That Make Us Smart. Oxford University Press.

KAHNEMAN, D., SLOVIC, P., TVERSKY, A., (1982). Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press.

KIRMAN, A., (2010). Complex Economics: Individual and Collective Rationality. Routledge.

MCKELVEY, R.D., PALFREY, T.R., (1998). "Quantal response equilibria for extensive form games." *Exp. Econ.*, Vol. 1, pp. 9–41.

PAYNE, J.W., BETTMAN, J.R., JOHNSON, E.J., (1993). The Adaptive Decision Maker. Cambridge University Press.

RUBINSTEIN, A., (1998). Modelling Bounded Rationality. MIT press.

THALER, R.H., (2015). Misbehaving: The Making of Behavioral Economics. W.W. Nortona & Company.

CHAPTER 1 .

LEE-PENAGOS, ALEJANDRO[⊥Ψ]

TESTING THE NOMINAL PAYOFFS DOMINANCE PRINCIPLE: AN EXPERIMENT ON MONEY ILLUSION

ABSTRACT

Fehr and Tyran (2001) have put forward the relevance of Money Illusion for economic theory and practice. Recently however, it has also raised controversy about the validity of its experimental design and results (Petersen and Winn, 2014). This paper puts to the test the predicted effects of money illusion in equilibrium selection. We find that framing effects can greatly affect subjects' behavior, but that the direction of such effects are not always as expected. Money illusion does matter, but is not clear yet exactly how.

KEY WORDS: Money Illusion, Experimental Economics, Bounded Rationality, Coordination Games, Payoffs Dominance.

JEL CLASSIFICATION: C70, C92

[⊥] University of Nottingham, School of Economics. Email: alejandro.lee@nottingham.ac.uk

[Ψ] Acknowledgements: The author thanks Chris Starmer for suggestions on the experimental design. To him, and also to Alex Possajennikov, acknowledgements are given for (several) discussions on the own framing of the paper, as well as for their helpful comments on this and all previous versions. To my friend Alessio Gaggero, for his kind help while running the experiments.

1.1 INTRODUCTION

There has been a long history of different approaches to study Money Illusion, term traditionally used to denote any failure in distinguishing monetary from real magnitudes (it dates back as far as to David Hume (Howitt, 2008), including the early work of Leontief (1936) and more recently Shafir et al. (1997), Cohen et al. (2005), Dzokoto et al. (2010), Brunnermeier and Julliard (2008), Kooreman et al. (2004), Cannon and Cipriani (2006), Mussweiler and Englich (2003) and Raghubir et al. (2012)). In the experimental economics literature, the work of Fehr and Tyran (2001) has been highly influential. The interest of this paper is in the latter approach.

Fehr and Tyran (2001) developed a novel experimental setup in order to investigate speed of convergence after monetary shocks, and their results have often been cited as evidence that nominal representations are relevant for economic theory and practice (Petersen and Winn, 2014). Their methodology and theoretical foundations have been further explained in Fehr and Tyran (2005) and Fehr and Tyran (2014), as well as used to explore the effects of the strategic environment on speed of convergence (Fehr and Tyran, 2008) and the effects of nominal prices as a coordination device for equilibrium selection (Fehr and Tyran, 2007). In the above papers by Ernst Fehr and Jean-Robert Tyran (FT from now on) [1], money illusion is interpreted as a behavioral rule of thumb which involves subjects *'taking nominal payoffs as a proxy for real payoffs'*[2], which can lead subjects to maximize nominal instead of real payoffs under some circumstances.

This concept of money illusion is closely related to the *payoff dominance principle,* a criterion for choosing between equilibria in games. This principle states that if one equilibrium payoff-dominates all others, then rational players will play their parts in that equilibrium (Harsanyi and Selten, 1988). The concept of money illusion, as put forward by FT, is the first to explicitly differentiate between *real* payoff dominance and *nominal* payoff dominance (Fehr and Tyran, 2007), suggesting the relevance of the latter for discussions about equilibrium selection. Hence, the nominal dominance principle is relevant not only as a particular instance of money illusion, but also as a hypothesis about subjects' behavior in games with multiple equilibria.

The objective of this paper is to test the predictive power of the nominal dominance principle. The approach for this test is to observe if the hypothesis that people behave as nominal maximizers when multiple equilibria are available, holds under different scenarios. The interest here is not just to find one particular environment where people don't behave as nominal maximizers (that would not be of much interest). The goal is to create an experimental environment where players are affected by how payoffs are presented (i.e. in real versus nominal terms) and do behave as nominal maximizers. Then, while keeping constant all real variables, change strategically irrelevant factors and observe if players still behave in the same way. The objective is to create a tension between nominal dominance and other salient features of the payoffs, trying to induce players to choose strategies contrary to what is hypothesized by the nominal dominance principle (i.e. to play actions that do *not* give the highest nominal payoff).

---

[1] This refers to the papers of these authors cited so far.

[2] As defined in Fehr and Tyran (2001), pp. 1239-1240, and Fehr and Tyran (2007), pp. 247.

Testing the nominal dominance principle is relevant not only because it is the underlying principle behind the concept of money illusion and its potential effects as an equilibrium selection mechanism, but also because there is an unsettled debate in the experimental literature regarding the effects of nominal maximization in the money illusion experiments (Petersen and Winn (2014), PW from now on). The questions raised by PW are both related to whether subjects indeed behave as nominal maximizers, as well as about the potential causes and theoretical accounts for this to happen. Our design, besides testing the predictive power of the nominal dominance principle, will also contribute towards this debate (explained in more detail section 1.2).

1.1.1 Methodology

For our test we conduct an experiment in a symmetric *n*-player pricing game with two stable Pareto-ranked equilibria. An important feature of our experiment is that we put forward, a-priori, a hypothesis regarding the effects of the nominal dominance principle. We test the predictive power of the nominal dominance principle as an equilibrium selection mechanism (as suggested by Fehr and Tyran (2007)). The hypothesis to be tested is that *players converge in the long-run to the equilibrium with the highest nominal payoffs*. Such convergence is an explicit consequence of money illusion, stated as a behavioral hypothesis[3]. Our approach is primarily empirical, testing whether the observed effects are in line with the nominal dominance principle or not. Theoretical accounts of observed behavior or disentangling its causes is not the objective of this paper.

The conducted experiment relies mainly on treatments changing the way payoffs are framed and presented to subjects, without altering the underlying incentives structure. This alters only the nominal payoffs or face-values observed by subjects, while keeping constant the real ones. These changes prove relevant for the nominal dominance principle, since according to it, players choose their actions based on nominal payoffs, not entirely on the real ones. By implementing different types of payoffs framing (including the ones used in previous money illusion literature), we test conditions under which the principle might hold or not.

Is worth clarifying a concept before explaining our treatments. A focal point is a solution that people will tend to use not because it is better than others *per se,* but because it seems natural or particularly attractive. It can aid coordination specially when there's a lack of communication by making people select a solution that they consider others will also consider natural[4]. In our experiment, when a particular equilibrium is, by design, intended to lure

---

[3] Of course, there might be other consequences or hypothesis to be derived from the theory of money illusion, but we are interested in this one given its binary nature: we can evaluate whether players converge or not to a particular equilibrium. As explained in section 1.2, we argue that difficulties in interpreting money illusion for a particular hypothesis to be tested, is one reason that has led previous tests of the effects of money illusion to be contentious.

[4] Schelling (1960) introduced this concept in game theory. In his classical example a group of students were presented with the following problem: "tomorrow you have to meet a stranger in New York City. Where and when do you meet them?" In this coordination game, any place and time in the city could be an equilibrium solution, but the most common answer was "noon at the Grand Central Station". Selecting one of the most common stations in the city was a natural expectation of what others would do, allowing coordination without the possibility of communication. Mehta et al. (1994) explores and finds evidence consistent with Schelling's arguments in an experimental setting.

players into playing it by changing the framing of payoffs, we will refer to it as a "focal equilibrium". For example, a treatment that, as in previous literature, alters framing to change which equilibrium has the highest nominal payoffs, would make such equilibrium focal under this terminology (since the design is intended to lure players into playing based on nominal payoffs). Here, a focal equilibrium is not intended to be a formal measure or definition, but only a way to refer to which equilibrium the design intends players to coordinate on (regardless of which underlying psychological effects might be taking place). Notice that actual convergence by subjects might or might not go in the same direction.

Given the above, our design implements three tests.

Test 1: our first goal is simple. Can we replicate the qualitative results of money illusion found in previous literature? In a similar way as in previous literature, we compare two treatments for subjects' convergence. A baseline treatment presents real payoffs without any particular intended framing, and according to traditional rationality, convergence is expected in the equilibrium with highest real payoffs. The second treatment alters the framing and makes the inefficient equilibrium the one with the highest nominal payoffs. By making the latter the focal equilibrium, we test if we can, as in FT's experiments, make players deviate from efficiency.

Our results will show that we can qualitatively replicate previous results. Our design can make players deviate from efficiency in a way consistent with the nominal dominance principle. This is important because it shows that whatever the results of following tests, they cannot be attributed to the real payoffs structure. Under our implemented game and incentives structure, the nominal dominance principle *can* hold.

Test 2: given that players did behave as nominal maximizers in the long run under Test 1, our second goal is to check if such behavior holds under similar scenarios.

For this test, is important to notice that high nominal payoffs are only one potential effect that could make an equilibrium focal. There can be, interconnected, other psychological effects taking place. For example, one such possible effect could be "salience", or which action seems more natural for players to play. We will refer to "other salient features of the payoffs" as any framing effects, *different from the nominal ordering of payoffs*, that can affect players' decisions[5]. Put differently, an equilibrium can be made focal *without* making it the one with highest nominal payoffs.

With this clear, Test 2 creates a tension between payoff dominance and other salient features of the payoffs. The implemented treatment makes the inefficient equilibrium the focal one, this time, without altering the ordering of nominal payoffs. To make it focal, two main framing changes are used: a constant is added to all payoffs and the order of the labels on the payoffs matrix are inverted. How these changes are intended to induce players into playing the inefficient equilibrium is detailed in section 1.4.2, but its effects can be in a way related to the Webber-Fechner Law (Robinson, 2010). The latter proposes that the just-noticeable difference between two stimuli is proportional to the

[5] The range of effects potentially related might be complex, and we do not intend or claim to disentangle them. Our interest is to observe potential deviations from the nominal dominance principle, without focusing on which particular effect is the cause.

magnitude of the stimuli. So by framing payoffs in higher levels, we intend to make subjects less prone to notice nominal payoffs differences between both equilibria (while being the same in real terms). Given the payoffs structure, this could lead subjects to perceive the inefficient equilibrium to be closer in payoffs terms to the efficient one.

In this treatment, payoffs dominance points towards efficiency, while focality towards inefficiency. This is a test to observe if the nominal dominance principle can hold under a more stringent scenario. If it holds, convergence should go towards efficiency.

Data will show that players in Test 2 do *not* converge to the equilibrium with highest nominal payoffs. This result has two implications. Primarily it shows, by way of counterexample, that even in such a controlled environment, the prediction of the nominal dominance principle, where subjects maximize nominal payoffs in the long-run, does not hold. But it also shows that other salient features of the payoffs can, on their own, drive behavior without any influence of money illusion. Put differently, since players converge into the Pareto-inefficient equilibrium *without* altering the payoff dominance of both equilibria, it means that money illusion is not a *necessary* condition for this behavior.

Test 3: After observing the above results, our final test puts the nominal dominance principle in a scenario where it is, arguably, more likely to hold. A potential argument is that players would behave as nominal maximizers only if the environment is difficult enough, or if the "veil of money" makes finding the optimal equilibrium more difficult (Fehr and Tyran, 2014). Simplifying rules of thumb might be more likely used by subjects when the optimal choices are not readily identifiable.

To address this, Test 3 removes the tension between the intended focal equilibrium and nominal payoffs dominance (introduced in Test 2), aligning both of them towards the Pareto-efficient equilibrium. Payoffs framing is changed in a way similar as in Test 1 (altering which equilibrium has higher nominal dominance), but here subjects face nominal payoffs that are more difficult to convert to real (i.e. their cognitive load is higher). High cognitive load is an effect suggested in the literature that can potentially make subjects more prone to suffer from money illusion (Fehr and Tyran, 2014).

Contrary to our a priori expectation, we find that players do not converge to the predicted equilibrium. The hypothesis of nominal dominance as an equilibrium selection mechanism, again, does not hold. Since conditions that are thought to make players more likely to behave as nominal maximizers are introduced by design, this result shows that nominal dominance is not a *sufficient* condition to drive behavior to the predicted equilibrium.

Finally, with the data collected from the three tests, we explore the notion of "lock-in" effects (Fehr and Tyran, 2007), which are considered important for having subjects coordinating in an inefficient equilibrium. The argument is that given a lack of communication, as well as incentives to avoid deviating from the group's average, initial actions can cause players to stick with those decisions (even if they imply long-term welfare losses). By comparing actions in the initial periods with convergence at the end of each treatment, our results show that such lock-in effects can happen sometimes, but that they are not present in all treatments.

1.2 MOTIVATION

This section will explain the core design used in this line of literature of money illusion [6], followed by an overview of the arguments and counterarguments that have been published regarding the challenge to this framework (i.e. experimental design and interpretation of money illusion results). This exchange between FT and PW will be referred to as 'the debate'.

1.2.1 The debate: the challenge to and the defense of the theory of money illusion

1.2.1.1 Core of the experimental design in money illusion

The main feature of the experiments on money illusion conducted by FT is that they present subjects with two versions of a big payoff matrix (30 x 30)[7]. In the experiments subjects have to choose a number between 1 and 30, framed as firms selecting a price, and their monetary gains depend on both their own price and the average of the other subjects in the group. The payoff matrix presents the experimental points that players receive for each possible combination of own and average price (leading to 900 possible outcomes). The first version of the matrix, called the '*real representation'*, is simply the payoff matrix showing without alteration the number of experimental points that can be obtained for each period of play (although at the end those points have to be converted into monetary units as is standard practice). The second version, called the '*nominal representation'*, is based on the real matrix but multiplies each payoff in it by the column label (average price of the group). Subjects are clearly instructed that the amount of points they obtain that can actually be exchanged for money has to be calculated (basically dividing each possible payoff by the average price).

Figure (taken from Fehr and Tyran (2005)) is an example of part of one of the nominal payoff matrices. Is worth noting that players couldn't see the highlighted slots (best responses) or the equilibrium (circled); those are shown just as illustration for the reader. In the example, for subjects to know the real payoffs of selecting a price of 27 (row label) when the group average is 13 (column label), they would have to divide, using pen and paper, 519 by 13 (which is 39.92).

[6] This refers to the papers that use the core features of the design of Fehr and Tyran (2001), including Fehr and Tyran (2014), (2008), (2007), (2005) (i.e. what we refer to as FT), but also including Petersen and Winn (2014), which started the debate.

[7] Technically. this experimental design, as well as the following debate, specifically refers to Fehr and Tyran (2001). However, since the features are at the core of all the other papers using the same methodology, they are related to FT and to the whole line of literature.

		Average price of others ($\bar{P}$)											
Selling Price P_i		9	10	11	12	13	14	15	16	17	18	19	20
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	....	272	274	274	274	276	285	303	321	327	322	311	296
20	....	300	306	308	309	312	323	343	363	370	364	351	334
21	....	326	337	343	346	351	364	387	409	418	412	397	378
22	....	346	365	377	385	393	408	434	459	470	465	449	427
23	....	357	387	407	421	434	453	482	511	524	521	506	483
24	....	358	398	429	452	471	495	527	559	578	579	566	544
25	....	349	397	439	472	500	530	565	601	625	633	627	608
26	....	330	386	436	480	517	552	591	629	660	679	682	671
27	....	305	364	421	473	519	560	600	640	678	709	727	727
28	....	277	335	395	452	505	550	591	631	676	720	753	771
29	....	248	304	363	422	478	526	565	605	654	708	758	795
30	....	221	272	328	386	442	490	527	565	616	677	741	796

Figure 1.1: Example of part of a nominal payoff matrix. Taken from Fehr and Tyran (2005). Subjects' tables were not marked with the best responses (highlighted slots) or the equilibrium (circled). Those are shown for illustration.

Notice that, formally, this nominal payoff matrix gives players the exact same monetary incentives as a real version of it, just that subjects "see" different numbers and are required to make additional arithmetic calculations. The latter, under standard rationality assumptions should not affect behavior: with proper incentives subjects would make the appropriate calculations and base their decisions only on the real values, whatever their nominal representation (framing) is. Hence, differences in behavior between groups treated with the two different kinds of matrices are considered to be the effect of money illusion: since the underlying real payoffs are the same, change of behavior under the nominal matrix is considered to be a form of bounded rationality.

Using this methodology, FT's results show that under a nominal representation agents converge more slowly towards equilibrium after a monetary shock (Fehr and Tyran, 2001). Nominal representations can also create coordination failures in pricing games, with agents being 'locked' in a Pareto-dominated equilibrium (Fehr and Tyran, 2007). In both cases, nominal representations led to relevant inefficiencies attributed to money illusion.

1.2.1.2 The challenge of Petersen and Winn (2014)

The design above has been challenged by PW. PW's main critique is that there are confounding factors in the design, challenging the interpretation of FT's results as being the effect of money illusion. The following are two confounding effects that are considered to affect the interpretation of the results:

1. Switching to a nominal representation increases the **cognitive load** faced by subjects. This means that the key variable at play is not necessarily money illusion per se.

   This addresses the issue that it is difficult for players, given the big amount of information they receive, to be able to consider all the options. Cognitive load refers to the amount of mental effort players have to make to process the information in the payoff matrix: with 900 potential options and making three digit divisions with pen and paper, the argument is that players could be overwhelmed by the difficulty of the task.

2. The nominal representation changes the **focal points** in the payoffs space.

   Changing where the higher nominal payoffs are located in the payoffs matrix might introduce new focal points not present under the real representation. For PW, the effects of focal points should be distinguished from the effects of money illusion.

PW consider these confounding effects of cognitive load and focal points to be a major flaw in the design. If it is either or both of them driving the main results of FT, then the conclusion that money illusion is causing them is misleading: is not about players using a rule of thumb, but about how they process the information in the payoff matrix. To address these issues, PW run the same experiment as in Fehr and Tyran (2001) but allow players to use a calculator and tools onscreen that make the arithmetic calculations easier (to diminish the effect of cognitive load), and also highlight in a different color the maximum real payoffs in the matrix in order to keep a constant focal point (addressing the second confound).

The results of PW show that when introducing these controls, the effects registered by FT are greatly diminished. Their main conclusion is that players do not maximize nominal payoffs as implied by FT's definition of "using nominal values as a proxy for real values", challenging the validity of the nominal dominance principle. However, to support these results and to explain how they challenge those of FT, the authors argue that players still suffer from *some* influence of the nominal representation. For this they rely on a reinterpretation of money illusion: a 'first order' money illusion is defined as players maximizing the nominal payoffs, which they find no evidence of. A 'second order' money illusion is defined as players *primarily* relying on real payoffs but taking into account nominal payoffs as well.

1.2.1.3 The reply by Fehr and Tyran (2014)

FT respond to this critique in Fehr and Tyran (2014). Their counter-argument and disagreement with PW's results is supported by a clarification of their initial definition of money illusion. They say that factors such as cognitive load and alteration of focal points are precisely part of their definition. They argue as follows:

"If people have difficulty piercing the veil of money, and are thus uncertain about their best choice, a rule of thumb of treating (changes in) nominal payoffs as a proxy for (changes in) real payoffs may affect their behaviour [...]. The very notion of a "proxy" means that the proxy is only used if the perfect solution is not available to an individual. In our context, this means that subjects are unlikely to simply be nominal income maximizers under the nominal frame. Instead, they will probably only use the above rule of thumb if they cannot pierce the veil of money – an inability that may result from, e.g., cognitive load or biased attention. Thus, the proxy hypothesis always presupposes some ***other sources of bounded rationality****. "*[8]

According to this clarification, the points addressed by PW shouldn't be a concern to their theory, for both cognitive load and change in focal points should be interpreted as "other sources of bounded rationality", factors that should be present for FT's interpretation of money illusion to hold. In FT's

---

[8] Fehr and Tyran (2014), p. 3. Italics and bolds our own.

view, PW's results only confirm their original results, validating the interpretation of money illusion as a principle of nominal dominance.

1.2.2 Key points of the debate

Let us summarize what we consider are the key takeaways of this exchange:

1. **In both FT's and PW's definitions, there are particular difficulties when interpreting empirically the concept of money illusion.** The experiments conducted by FT are mainly intended to study if nominal representations can lead subjects to deviate from traditional rationality[9]. For this goal, one can stablish the behavior predicted by traditional maximization, and if (any) deviations from it are observed, one can say that money illusion matters. With this in mind, money illusion defined as "a rule of thumb using nominal values as a proxy for real" is perfectly fit for such purpose. However, if one is to take a next step and predict behavior beyond deviations from real payoffs maximization, or explore which variables influence subjects suffering from money illusion (as PW did), interpreting such definition becomes more difficult. Does it lead to maximization of nominal payoffs in the long-run, or in a period by period basis? In which environments is such rule of thumb more likely to be observed? PW's concerns are interesting questioning of the design, but when trying to disentangle the effects of variables such as cognitive load and focal points, such questions become relevant. Even more, PW's conclusions led to further interpretation difficulties. They relied on new concepts like "second order" money illusion, defining it as players *primarily* relying on real payoffs, but also taking into account the nominal. This reinterpretation raises similar questions: How much is *primarily*? Does the theory hint under which circumstances would this 'less severe' instance of money illusion be observed? FT precisely argue that their initial interpretation of money illusion was closer to PW's second order concept than to the strict first order definition, hence dismissing PW's conclusion as a valid challenge to their results.
2. **In their test, PW made the subjects' environment easier.** If one agrees with FT's extension, made in their reply, of money illusion requiring the presence of "other sources of bounded rationality", then in that case PW's test is not conclusive in challenging the experimental design: PW's experiments give players tools to precisely remove such sources. An adequate challenge would require finding no evidence of money illusion without helping players with the cognitive load or explicitly giving them a focal point (i.e. without making their task easier). We interpret this requirement of having "other sources of bounded rationality" as a hint from FT for conditions in the environment that would make players more likely to behave as nominal payoffs maximizers.

[9] The title of Fehr and Tyran (2001) is "Does money illusion matter?", referring to whether money illusion matters in terms of behaviour relative to standard rationality predictions.

Taking into account the the points above, this paper will test the effects of money illusion in equilibrium selection. Fehr and Tyran (2007) argue that their *"results suggest that nominal payoff dominance is an equilibrium selection principle which drives behaviour in strategic settings"*[10]. We take such suggestion and test it explicitly. One advantage of testing money illusion as an equilibrium selection device is that it reduces the difficulties when interpreting results: one can observe whether players converge to the predicted equilibrium (i.e. the one with highest nominal payoffs) or not. This implication of money illusion is what we refer to as the 'nominal dominance principle'. By testing the nominal dominance principle as long-run convergence in the equilibrium with highest nominal payoffs, we focus on a quantifiable qualitative difference (i.e. convergence or not), reducing potential issues of *degree* or *how much* money illusion was observed. This relates to the first key point in the debate.

Also, our tests are conducted without making the subjects' task easier (e.g. allowing calculators) or subtracting what FT refer to as "other sources of bounded rationality". Even more, in one of our treatments the cognitive load of players is *increased*, which is arguably a way of making those 'other sources' more prominent. This relates to the second key point of the debate, or the environment in which money illusion is more likely to be observed.

1.3 EXPERIMENTAL DESIGN OVERVIEW

This section describes the general game used across all three tests. An overview of the treatment structure is given as well, leaving more specific details of their implementation for the next section, where results will also be presented.

1.3.1 Game structure

Players have to choose simultaneously, for $T = 30$ periods, a price $P_i \in \{1,2,\dots,35\}$. Players are randomly matched in groups of $N = 10$, and they know that the group remains the same throughout all T periods. Each treatment has three groups. After all players choose their price for the period, they are all informed of the median price of the rest of the group, $\tilde{P}_{-i}$, and their respective payoffs earned that round[11]. The payoffs depend only on P_i and $\tilde{P}_{-i}$, which are presented to the players via a 35 x 35 payoff matrix showing each possible combination, with own price in the rows and group median in the columns. The game is symmetric in the sense that all players receive the same matrix, which is common knowledge since the beginning. Deviating from the group's median price is costly, making this a coordination game. In order to avoid making the task easier (second key point in the debate), players are not allowed calculators, having only pen and paper as in the original experiments of FT[12].

[10] Fehr and Tyran (2007), p. 263. Italics our own.

[11] The median was used instead of the mean of the group in order to reduce the effects of strategic teaching. Since the mean is more sensitive to extreme outcomes, it would've been easier for some few players to send strong signals to the rest of the group, as happened in our experimental pilots.

[12] Players also had a maximum of ten minutes to get used to the matrices and understand the task after the instructions were read out loud by the experimenter. Both instructions and payoff

In the payoff structure there is a unique best reply for player i for every given level of $\tilde{P}_{-i}$. The equilibria of this game are located at the intersection of the best reply function with the 45-degree line (due to the game being symmetric). The two stable equilibria will each be called the *efficient* equilibrium (which arises for player i when she chooses $P_{ef} = P_i = \tilde{P}_{-i}$) and the *inefficient* equilibrium (arising when $P_{inf} = P_i = \tilde{P}_{-i}$). Notice that sub-index *'ef'* denotes the efficient one, and sub-index *'inf'* the inefficient.

1.3.2 Treatments overview

As outlined in the introduction, three tests are conducted. Figure 1.2 shows the connection between the four implemented treatments. The difference between them is only the framing of the payoffs matrices given to subjects and the arithmetic calculations players would have to do to estimate the real payoffs. Starting with the baseline, each treatment is constructed by implementing the corresponding change in framing shown in parenthesis, indicated by the direction of the arrows. Arrows also relate pairs of treatments compared for each test[13].

A key distinction is to be made between *Real* and *Nominal* treatments (the final letter in the name of each treatment is either R or N, respectively). A nominal treatment will always be implemented from its real counterpart. This nominal transformation is done by multiplying each possible payoff inside the corresponding *real* matrix by $\tilde{P}_{-i}$ (i.e. by the corresponding column label), as in previous literature. Players are accordingly instructed that they should make the corresponding calculations to know the real payoffs. By *nominal treatment* or *nominal implementation,* we refer exclusively to this particular conversion of the matrix.

It will also be useful to clarify here the concepts of payoffs dominance. We will refer to a **real dominant** equilibrium when such equilibrium is the Pareto-efficient one (i.e. has the highest *real* payoffs). A **nominal dominant** equilibrium is the one with the highest *nominal* payoffs. Notice that in a real treatment the real dominant equilibrium is always the nominal dominant one. However, in a nominal treatment this is not necessarily the case: the nominal dominant equilibrium can be Pareto-inefficient in real terms.

---

matrices were given printed to subjects, but inputs and feedback on payoffs was done onscreen. The game wouldn't begin until every player in the session would decide to move on. Never in any treatment was this time completely used. Players also had two minutes in each period for choosing price: when this time was over they received an onscreen reminder to choose, but the game wouldn't move on until they did. Finally, after each period they had one extra minute to check the history of all periods' selected price, group median price and previously obtained nominal payoffs. The screenshots of how the experiment was presented to subjects can be found in Appendix 1.7.9.

[13] Each treatment was run in one separate session, with all 30 subjects (three groups of ten players) at the lab at the same time. All treatments were conducted in the CEDEX lab at the University of Nottingham during February of 2014, and all subjects were students recruited using the ORSEE system (Greiner, 2003). Subjects earned on average 9.2 GBP pounds and treatments lasted no more than one hour, except InvN, which lasted about an hour and twenty minutes. The experiment was programmed and conducted with the software z-Tree (Fischbacher (2007).

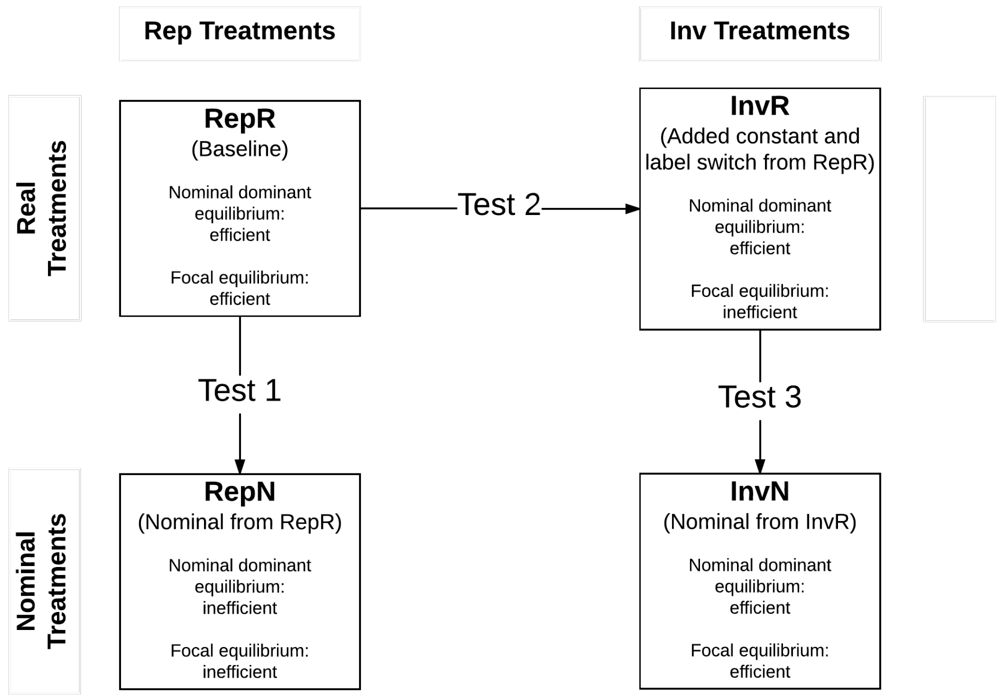


Figure 1.2: Treatment Summary

Test 1 has a simple goal: to replicate qualitatively the effects found in previous money illusion literature, where a nominal implementation causes players to deviate from real payoffs maximization. Treatments *RepR* and *RepN* are compared for this, the *Rep* name being a reminder of the goal of replication.

Test 2 tests if the effects previously found in the literature (and tested in Test 1) still hold under a more stringent scenario. It creates a tension between which equilibrium is focal (the inefficient) and which one presents higher nominal payoffs (the efficient)[14]. This is done by inverting focality (hence the *Inv* name), *without* altering nominal dominance as done in Test 1.

Notice that both treatments compared in Test 2, *RepR* and *InvR,* are real treatments. This is because the change in framing from RepR to InvR is not done via a nominal implementation (further details of this implementation are given in next section). The objective is to test if other salient features of the payoffs can have similar effects than what has been previously attributed to money illusion.

Finally, Test 3 is considered as a final, less stringent test than the one implemented in Test 2, with the objective of making the nominal dominance principle more likely to hold. It compares *InvR* with its nominal counterpart, *InvN*. This nominal implementation defuses the tension created under Test 2, also increasing the cognitive load of players. The latter is intended to address what FT referred to as "other sources of bounded rationality", or to implement

[14] The reader is reminded that we refer to a "focal equilibrium" as the one which the design intends players to focus their actions on. For example, nominal treatments (both here and in previous literature) usually change which equilibrium is nominally dominant with the intention of luring players into playing that equilibrium. In our terminology, this would make that equilibrium the focal one.

conditions in the environment that are considered to increase the probabilities of players relaying on behavioral rules of thumb such as money illusion. The test is considered less stringent because the efficient equilibrium is both focal and nominal dominant, and cognitive load is at the highest across treatments[15].

1.4 THREE TESTS: DESIGN AND RESULTS

The design details of each of the three conducted tests are presented sequentially. After each design, results are addressed for that test.

1.4.1 Test 1: replication of money illusion effects

1.4.1.1 Test 1: design

Our first test's objective is to replicate qualitatively whether under a real treatment players converge into P_{ef} (*efficient* equilibrium), but under the nominal they converge on P_{inf} (*inefficient* equilibrium), as done in previous literature.

The Rep treatments create a tension between the principle of *real* payoff dominance and the principle of *nominal* payoff dominance as equilibrium selection devices. If we can replicate, the following tests' results could not be attributed to characteristics of the payoffs structure (since it is always kept constant).

RepR is our base treatment. On it, the highest nominal payoff is the same as the real (which by definition is true in the *real* treatments): the real payoff for player i if $P_i = \tilde{P}_{-i} = P_{ef}$ is $\pi_{ef} = 100$ and the real payoff if $P_i = \tilde{P}_{-i} = P_{inf}$ is $\pi_{inf} = 61$. Under any *real* treatment, there is no tension between the highest real and nominal values. This means that by definition $\pi_{ef} > \pi_{inf}$, but also that $\pi_{ef}^n > \pi_{\text{inf}}^n$ in the real treatments (supra index n denotes nominal payoffs).

The game payoff matrix for RepR is presented in Appendix 1.7.1. The reader is encouraged to check both equilibriums in it: $P_{ef} = 1$ and $P_{inf} = 31$. Under RepN (Appendix 1.7.2), the same ordering of *real* payoffs is maintained. However, for RepN the *nominal* payoffs of converging in P_{inf} are higher than those of converging in P_{ef} ($\pi_{inf}^n > \pi_{ef}^n$). This is the main objective of RepN: to change which equilibrium is nominal dominant. Test 1 compares the convergence under RepR and RepN.

1.4.1.2 Test 1: results

Result 1 (Comparing RepR and RepN): *Under RepR, most subjects converge to the efficient equilibrium, but when a nominal representation is implemented, convergence is to the inefficient equilibrium, which is the*

---

[15] A high enough cognitive load as a potential condition for players to suffer from money illusion is suggested by FT in the debate: "[players] *will probably only use the above rule of thumb* [using nominal payoffs as a proxy for real] *if they cannot pierce the veil of money – an inability that may result from, e.g.,* ***cognitive load*** *or biased attention"* (Fehr and Tyran (2014), p. 3. Italics and bolds our own). Treatment InvN implements a high cognitive load with the expectation that this will increase the chances of players focusing on the nominal dominant and focal equilibrium.

nominal dominant. So the design and payoffs structure can qualitatively replicate previous results in the literature, which are consistent with interpreting nominal dominance as an equilibrium selection device.

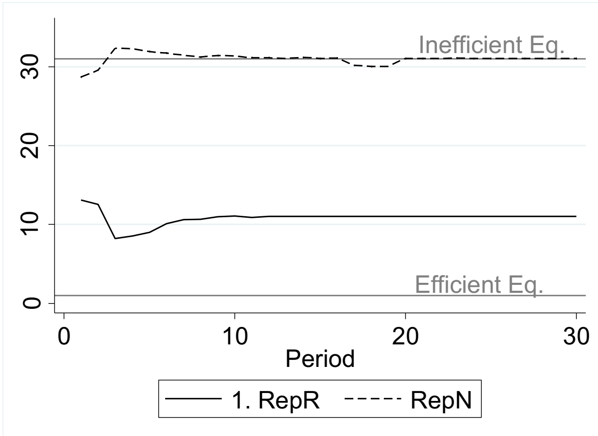


Figure 1.3: Average prices per treatment. RepR and RepN.

Figure 1.3 shows the average chosen price for all 30 subjects in each treatment for each period. The difference is clear: under RepN players converge exactly into the *inefficient* equilibrium ($P_{inf} = 31$) and under RepR players converge much closer to the *efficient* equilibrium ($P_{ef} = 1$). The effects of the nominal representation are clear: as in FT, nominal dominance can act as an equilibrium selection device.

However, one could ask why the convergence for RepR is not complete (the average price is $11 > P_{ef} = 1$). For this, Figure 1.4 is relevant: it shows the average price for each of the three groups in the treatments (remember that each treatment had 30 subjects organized in groups of 10). Panel (a) shows that two out of the three groups very quickly (before period 5) converge exactly into $P_{ef} = 1$. Only one of the groups ends up converging into $P_{inf} = 31$; it is the behaviour of this group the one pulling the average up. In panel (b) the three groups converge into $P_{inf} = 31$, showing the stark change of convergence under the nominal implementation. This is considered a qualitative replication of FT's results.

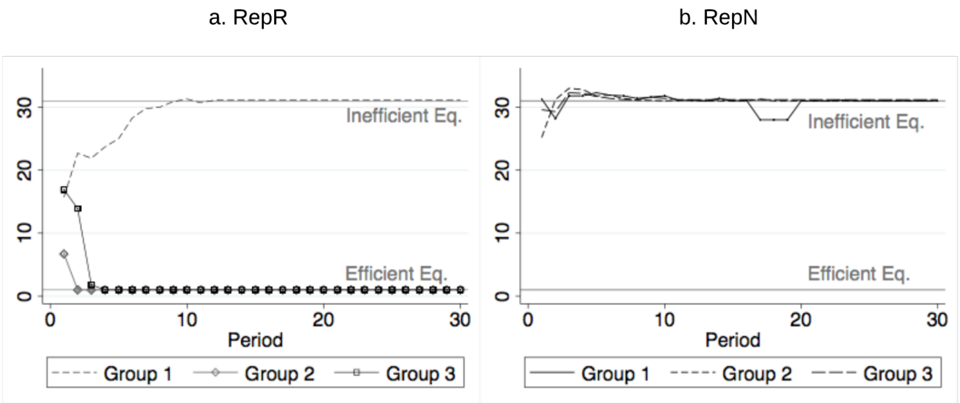


Figure 1.4: Average prices per group. RepR and RepN.

1.4.2 Test 2: tension between nominal dominance and the focal equilibrium

1.4.2.1 Test 2: design

Test 1 confirms FT's results: we can lure subjects to converge in the nominally dominant but inefficient equilibrium. Now suppose we have a setup where both nominal and real dominance point towards the same equilibrium (done by comparing two *real* treatments). Can we lure subjects to an inefficient equilibrium by altering the other salient features of payoffs? This is the objective of Test 2[16].

Notice that in both Rep treatments the nominal dominant is also the focal equilibrium. Test 2 alters the payoff matrix framing, from RepR to InvR, in order to make P_{inf} the intended focal equilibrium *without* altering nominal dominance. By creating this tension between the focal and nominal dominant equilibrium, the main goal it to test if the nominal dominance principle holds under a more stringent situation. If it does, players should converge into the nominal dominant equilibrium (efficient). By design, the objective is to try to lure players into the inefficient equilibrium, testing if the nominal dominance principle is still replicated under different environments[17].

Two changes in the matrix for InvR take place compared to RepR: first, all payoffs are 'scaled up', meaning that a constant is added to all of them. Second,

[16] Levitt and List (2009) identify experimental replication of results at three levels. The first and simplest one refers to using an experiments' own data and check the validity of statistical results. The second notion refers to running an experiment following a similar protocol to the one intended to be replicated. The third (and most general) consists in testing the hypotheses of the original study and pursue replication using a new research design. Our Test 1 can be related to the second type of replication, while our Test 2 can be related to the third.

[17] We emphasise again that we will avoid discussions about what formally constitutes a more or less salient point in the payoffs space. Although our design is intended to alter salient features of the payoffs and the psychological perception of players regarding an equilibrium, we will not claim "disentangling" or isolating any particular psychological effect as causing our results. Whether the principle holds or not will depend only on the aggregate results of convergence, irrespective of whether it was caused by salience or by another unintended psychological effect.

we implement a price 'label switch': we invert the price labels in the matrix. Figure 1.5 shows graphically both changes.

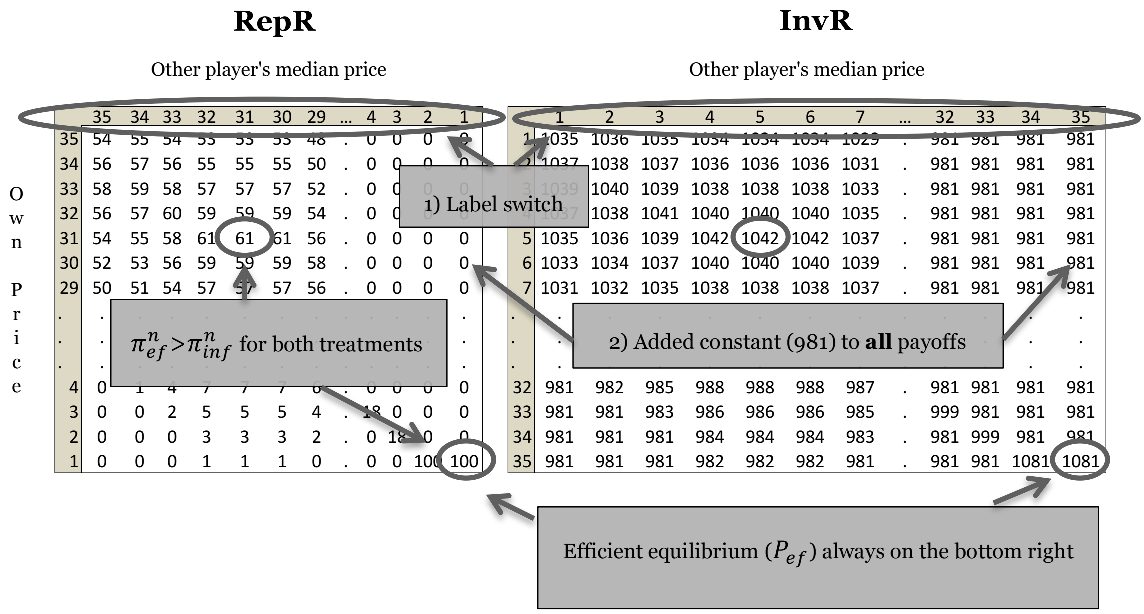


Figure 1.5: Creating InvR from RepR. 'Scaling' adds a constant to all payoffs. 'Label switch' inverts the price column labels, but leaves unaltered the rest of the payoffs.

When scaling payoffs by adding a constant to all of the payments in the matrix, players are accordingly instructed that they should subtract the constant in order to know their exact payment. This implies that the *nominal* values inside the payoff matrix are changed, but this **does not** create a tension between nominal and real dominance. In Figure 1.5, the circled payoffs correspond to π^n_{inf} (upper left in each matrix) and π^n_{ef} (down right in each matrix): notice that adding a constant simply scales the payoffs but keeps the same ordering. Whichever payoff is the highest in RepR, it remains the highest in InvR. Hence, in this treatment the important fact is that all payoffs keep the same nominal ordering.

The latter is precisely the reason why InvR is still considered a *real* treatment in relation to RepR: because the alteration of the nominal values by scaling up the payoffs doesn't alter the order of **any** of the payoffs in the matrix, even if subjects have to make one extra calculation (subtracting a constant)[18]. Since there is no tension between nominal and real dominance in the payoffs of InvR (since there's no such tension, by definition, in any real treatment), if the nominal dominance principle holds, convergence should go into the efficient equilibrium, in the same way it did under RepR.

[18] Technically, if experimental points are used, all payoffs in an experiment are nominal values different to their real monetary counterparts. The exchange rate used in any experiment to convert the points is one additional cognitive step (usually neglected), but since the ranking of payoffs is unaltered it is supposed, under rationality assumptions, to be irrelevant. Here, scaling up the payoffs has the same effect. Hence, given the use of experimental points, if a treatment is considered real or nominal depends on which other treatment it is being compared to. InvR is a real treatment compared to the baseline RepR.

The added constant is 981 as seen in Figure 1.5. The number itself is to some extent arbitrary, but it was chosen with the objective that π_{inf}^{n} becomes a stronger focal point. Scaling up the payoffs has three potential main effects that are also present when introducing a nominal treatment: first, having higher nominal values increases the cognitive load required to browse through the information. The reader can have a feeling of this by just browsing the payoff matrices for RepR and InvR (Appendix 1.7.1 and 1.7.3, respectively): a matrix with two digit numbers is less taxing than a matrix with mainly three and four digit numbers. Second, InvR creates a four digit numbers "zone" in the upper left part of the matrix where π_{inf}^{n} is located; experimental pilots and players' annotations after the experiments give anecdotal evidence suggesting that they can engage in some kind of local search that ignores other potentially relevant information. Finally, noticing a fixed change in a variable is usually more difficult when the level of that variable is higher. This can cause players to *perceive* the difference between π_{ef} and π_{inf} in InvR as smaller than it is, preventing them from playing P_{ef} (the cost of deviating from $\tilde{P}_{-i}$ when playing P_{ef} is higher than when playing P_{inf}. The *efficient* equilibrium is riskier by design).

The above effects can be in a way related to the Webber-Fechner law (Robinson, 2010), which states that the just-noticeable difference between two stimuli is proportional to the magnitude to the stimuli. Although the law has been widely studied for physical variables such as light, sound or temperature, it has also been studied for numerical perception. As an example, Dehaene and Marques (2002) find evidence of this law in people evaluating prices under different currencies: the standard deviation of estimated prices is proportional to their mean. So in our experiment, when the constant is added, subjects might perceive the difference in payoffs between both equilibria as smaller than it actually is.

The other change in the matrix, the label switch, can also have psychological effects. The names of each action, prices in this case, can have similar effects on which equilibrium is perceived as a more natural coordination point[19]. If a framing of higher or lower prices can have a psychological impact in how players make their decisions, the label switch can potentially affect which prices are chosen[20].

The above are the possible effects that might occur in order to induce a change of behaviour leading to convergence on P_{inf}, or put differently, to make it the focal equilibrium (compared to the expected and observed convergence in P_{ef} under RepR). However, we will not claim that any or all of such effects are formally happening, as this was not the the objective of the design (although the design was implemented under the intuition that they might take place). The approach here is interested in whether nominal dominance is an equilibrium selection device as in the interpretation of money illusion tested,

[19] The payoffs of an action can induce focal points, as intended with the scaling up. This can be referred to as "payoffs salience". But the labelling of the action can have also have similar effects, which can be thought of as as "label salience". These terms, "payoffs salience" and "label salience", are taken from Anbarci et al. (2015). The concepts however, although not specifically named that way, can be traced back to Schelling (1960).

[20] The label switch can have such effects, although its main purpose in the design is to allow comparison of InvR with InvN in Test 3 below. This will be clearer when details for it are given below.

not on the underlying mechanisms of such process. The test's objective is to refute or not the nominal dominance principle, not to explain or extend it.

1.4.2.2 Test 2: results

Result 2 (Comparing RepR and InvR): *Under RepR convergence goes mostly into the efficient equilibrium (reported above in Result 1). This is hypothesised by the payoffs dominance principle (either the nominal or real version). If the nominal dominance principle holds, convergence under InvR should also go into the efficient equilibrium. However, convergence under InvR goes into the inefficient equilibrium for all three groups. In this test, the nominal (and also, real) dominance principle does not hold*[21].

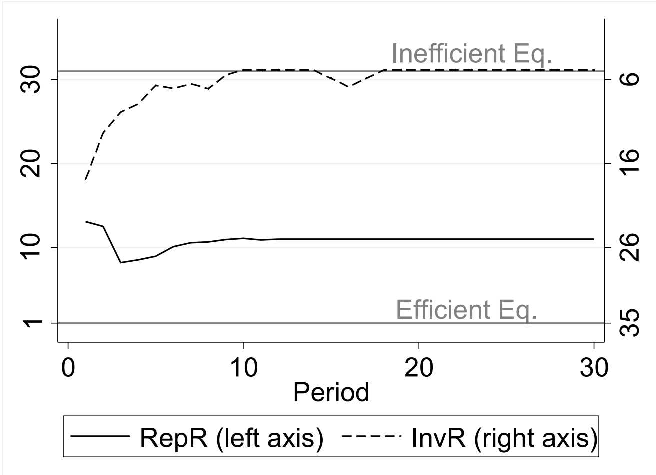


Figure 1.6: Average price per treatment. InvR and RepR.

Figure 1.6 and Figure 1.7 are the evidence for this result. Figure 1.6 shows a comparison of the average price for InvR and RepR. As shown before, the average price for RepR is very close to the *efficient* equilibrium, but under InvR convergence goes completely into the *inefficient* equilibrium. When analysing both figures, is worth noting that the vertical axis labels are different for both treatments; this is due to the label switch in InvR.

Here, once again is worth observing the convergence of each group, shown in Figure 1.7. Convergence for each group under RepR in panel (b) has been

[21] Result 2 was hypothesised a priori to be the effect of only adding the constant to InvR (i.e. to come only from changes in "payoffs salience"). The label switch was implemented for permitting further tests, when comparing InvR with its nominal counterpart, as detailed for Test 3. The label switch (or the effects of "label salience") was not hypothesised to have significant effects on behaviour. As part of the design from the beginning, another treatment (not reported) was run which added the constant to RepR without switching the labels. Surprisingly, only adding the constant did not have the same effects as when combined with the label switch. In hindsight, one could argue that the framing effect of high-low prices might be affecting players' decisions, or refer to possible mechanisms in which labels can change saliency and induce coordination in particular outcomes (e.g. Crawford and Iriberri (2007) on Hide-and-Seek games). Such discussion is avoided because it wasn't part of the initial hypotheses before running the experiment. However, the important part is that there are changes with a clear effect on behavior under InvR. Fortunately, where these effects come from is irrelevant for our conclusions regarding whether the nominal dominance principle holds or not. The objective from the beginning was not to separate each particular psychological effect.

shown before for test one, but is shown again for convenience of comparison for the reader. That way is easy to capture the fact that under InvR all three groups converge into the *inefficient* equilibrium (panel (a)), compared to the majority converging into the *efficient* under RepR. The nominal dominance principle, which would predict convergence into the nominal dominant equilibrium P_{ef}, does not hold. This result is also evidence against the payoff dominance principle in its real version: since InvR is a real treatment, real dominance does not hold either.

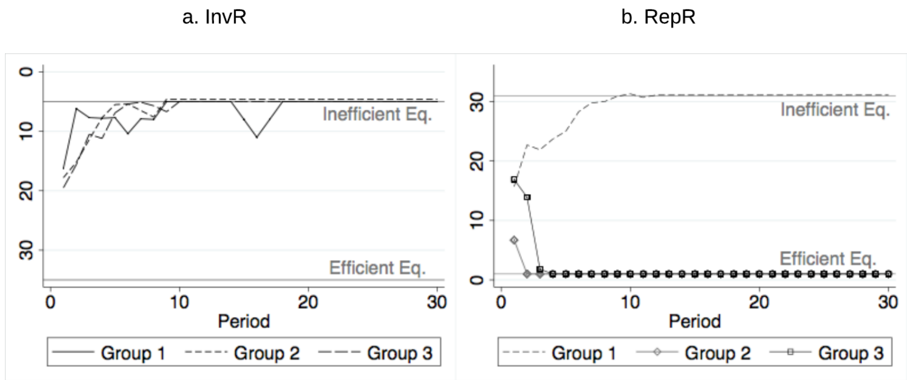


Figure 1.7: Average price per group. InvR and RepR.

This result's main implication is that the nominal dominance principle does not hold. However, another implication can be drawn that addresses the debate explained in the motivation section. In their experiments PW controlled for factors that they considered as confounds. Their hypothesis was that without them, nominal dominance would not affect behavior. FT answer was that those confounds are actually factors that should be present for subjects to suffer from money illusion, implying a broader psychological definition of the concept. Our approach in this test, instead of controlling for such factors, is to induce them directly *without* any implication from money illusion (since no nominal representation was implemented). So regardless of whether focal points can be considered as a confound or as part of the effects of money illusion, Result 2 shows that money illusion is not *necessary* to induce convergence into the inefficient equilibrium. Those other factors can, on their own, drive convergence away from efficiency.

1.4.3 *Test 3: can the nominal dominance principle "get back on its feet"?*

1.4.3.1 *Test 3: design*

Result 2 above constitutes evidence against the nominal dominance principle. The objective of this third, final test, is to create an environment where, arguably, the nominal dominance principle is more likely to hold. It is a test of whether the principle can hold under less stringent conditions than those implemented for Test 2. For Test 3, InvR is compared with its nominal counterpart, InvN (i.e. payoffs in InvR are multiplied by the corresponding column labels).

There are two main effects that are taking place under InvN. They are intended to address potential concerns in previous literature (hinted by FT in the debate) regarding the conditions under which money illusion is more likely to hold. Those concerns consider factors such as cognitive load and focal points as "other sources of bounded rationality", or effects that should go in tandem with money illusion in order to observe the expected behavioral effects. The design for InvN implements them in order to test if the nominal principle holds when they are accounted for. We interpret these effects as conditions in the environment that make money illusion more likely to make accurate predictions (i.e. convergence into the nominal dominant equilibrium).

The first such effect implemented under InvN, relative to InvR, is to eliminate the tension between the payoff dominant and the focal equilibrium. Such tension was by design introduced for InvR in Test 2. The nominal implementation for InvN aligns back salience and payoff dominance (both real and nominal) into the efficient equilibrium. For the nominal dominance principle to "get back on its feet" after the the results of Test 2, convergence should go into the efficient equilibrium, which is by design also intended as the focal point[22].

The second effect is related to the cognitive load faced by subjects when analyzing their payoff matrices. The argument is that a high cognitive load, or a difficult environment, can increase the probabilities for players to rely in potentially inefficient behavioral rules of thumb such as money illusion.

To account for this, InvN has higher nominal values than any other treatments. Almost all of the payoffs in InvN have four and five-digit numbers with a maximum face value of almost 38,000 (the reader can observe this in Appendix 1.7.4). This is due to the combined effects of scaling the payoffs and the nominal implementation. Compare this with, for example, the maximum nominal payoff in the other nominal treatment, RepN: there, the maximum is 2,000 and most payoffs have between one and three digits. This effect on itself makes the amount of information in the matrix more difficult to organise for the subjects. It also implies that the arithmetic operations required for a player who wants to calculate the real payoffs are more complicated. Besides a four or five-digit division, a subsequent subtraction (of the constant 981) is required for each payoff calculated. Remember that players, as in all other treatments, are not allowed calculators.

In Test 2, convergence under InvR was to the inefficient equilibrium. If the nominal principle holds in this environment, when comparing InvR with InvN, convergence should be inverted. In this case, money illusion should reverse the inefficiencies introduced under InvR and lead players to converge into the efficient equilibrium, which is real and nominal dominant, as well as the intended focal equilibrium by design.

1.4.3.2 Test 3: results

Result 3 (comparing InvR and InvN): *Convergence under InvR went to the inefficient equilibrium (as reported above for Test 2). If the nominal dominance principle holds under InvN, convergence should go into the*

[22] Which equilibrium becomes nominally dominant depends on which one is being multiplied by the highest group prices (column label). Here, the label switch introduced for InvR is key: it allows nominal dominance to go into the efficient equilibrium once payoffs are multiplied. This was the main design objective of the label switch.

efficient equilibrium (since both nominal and real dominance are aligned towards it, and cognitive load is considered to be higher). However, none of the three groups converged into it. This test is also evidence contrary to the nominal (as well as the real) dominance principle.

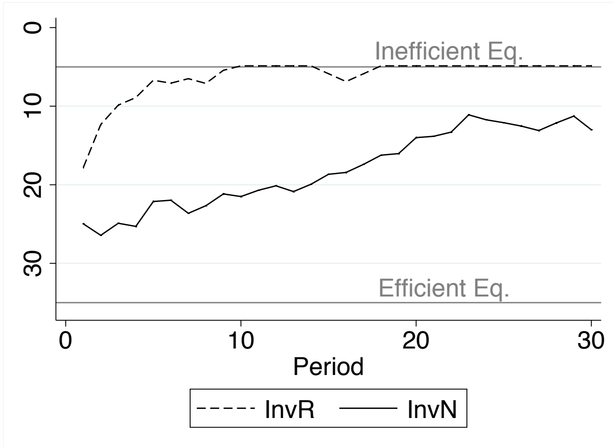


Figure 1.8: Average price per treatment. InvR and InvN

Evidence for Result 3 comes from Figure 1.8 and Figure 1.9. Figure 1.8 shows a stark difference between InvR and InvN. As shown in Result 2, convergence into the *inefficient* equilibrium is clear under InvR. If the nominal dominance principle holds, we should expect convergence towards the *efficient* equilibrium for InvN, since nominal and real dominance are aligned. Surprisingly, and contrary to our a priori hypothesis, this is not the case. Even if initial decisions under InvN seem on average to be closer to P_{ef} , the trend is leading prices *away* from it. Important is the fact that even if convergence is not achieved in the time span of the experiment (T=30), the trend of InvN is positive, meaning that even under longer time spans convergence would not go towards P_{ef} (assuming no structural changes)[23].

[23] Technically, the trend is negative due to the inversion of the price label. Notice that in Figure 1.8 the vertical axis is indeed inverted. This way of showing the results is done to make the graphical comparison easier: for all of our results, the *efficient* equilibrium is always down and the *inefficient* is always up, regardless of what treatment is being observed.

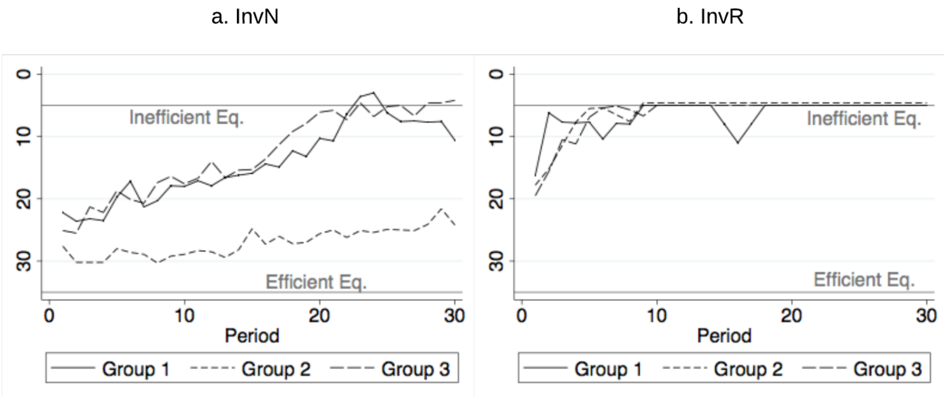


Figure 1.9: Average price per group. InvN and InvR

Figure 1.9 shows the same patterns disaggregated for each group in both InvN and InvR. Panel (a) shows that none of the three groups converge into the efficient equilibrium, and they all have a trend moving away from it. The average pattern is not an artifact of aggregation.

This final result is particularly striking because conditions under which money illusion is more likely to hold are incorporated: not only cognitive load is at the highest ("piercing the veil of money" is the most difficult)[24], but also the maximization of *nominal* payoffs is aligned with the maximization of *real* payoff, making the efficient equilibrium the focal one. Our a priori hypothesis was that the nominal dominance principle would hold under InvN. However, players did not maximize nominal (or even real) payoffs in the long-run (potential reasons for this are given in the next section).

Hence, Result 3, besides showing that the nominal principle does not hold, can also be interpreted in a way that addresses some concerns raised in the debate, regarding conditions in the environment considered to make money illusion more likely to hold. If our implementation of cognitive load and focal points are to be considered what FT call "other sources of bounded rationality", they are shown not to be *sufficient* in order to induce the predictions of the nominal dominance principle[25].

---

[24] An indication for this is that players spent on average more than twice the time in each period to make their price decision in InvN than in the other treatments.

[25] A reader familiar with this literature could argue that in this case money illusion is still causing sluggishness of adjustment and inefficiencies compared with the real treatment. This is true in the data (players in InvN were the ones earning less money), and such effects are mentioned previously in the literature as an important effect of money illusion. However, this is precisely the reason for choosing a coordination game for equilibrium selection. The test defines specific effects of money illusion on equilibrium convergence. Our results show that a nominal framing *does* cause behavioural effects; we do not claim that it doesn't affect behaviour at all. On that aspect, this result is evidence consistent with nominal representations causing behavioural changes, just not the ones tested.

1.5 "LOCK-IN" EFFECTS AND INITIAL DECISIONS

Our data naturally allows us to explore further another concept that was an important motivation for FT for considering money illusion as a coordination device. That is the notion of "lock-in" effects. The argument is that even if players individually learn after a few periods the real payoffs in the matrix, a coordination failure can cause the initial (and perhaps inefficient) decisions to have effects in the long-run. With a lack of communication between players and if deviating from the average group decision is costly for the individual, initial decisions can "lock" players in such initial decisions. Are such lock-in effects observed under our implemented treatments?

One way to answer this question is to observe the initial decisions of players and compare them with their decisions on the last periods. If players choose in the first period the same prices on which they convergence in the long-run, one could argue that this behaviour is consistent with FT's lock-in effects. Figure 1.10 shows the frequency of first period prices for each treatment.

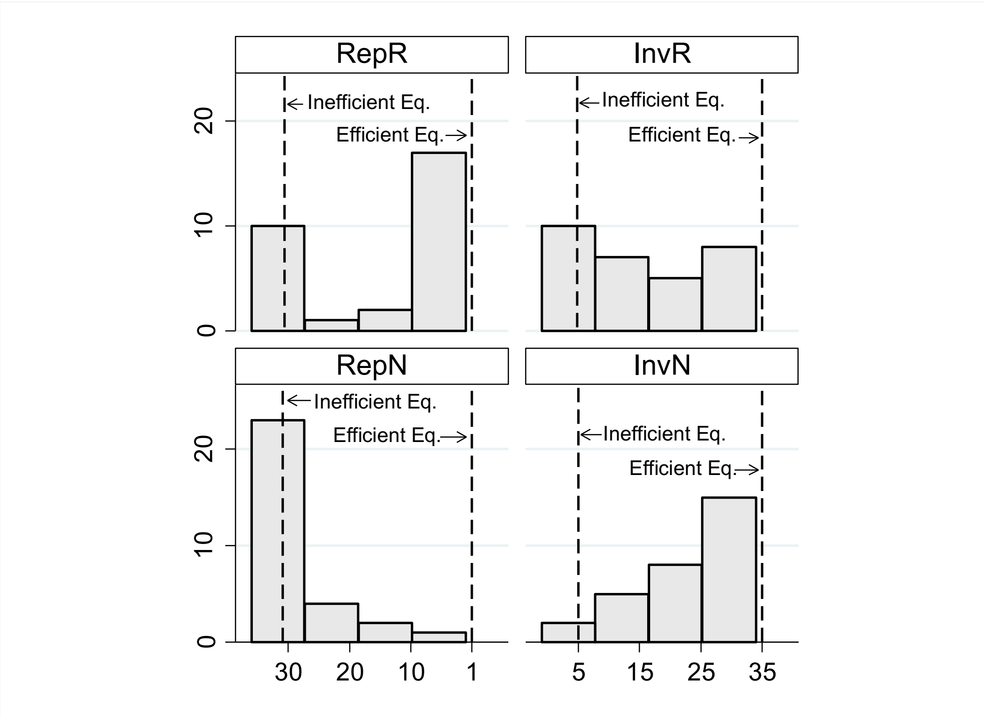


Figure 1.10: Frequency of prices chosen in the first period. All treatments.

First observe the two panels on the left for the Rep treatments. From Result 1 it is known that under RepR convergence went to the *efficient* equilibrium, while under RepN it went to the *inefficient* equilibrium. Was the same pattern observed since the beginning of the game? Yes, it was. The distribution of prices in the first period is clearly shifted when introducing the nominal treatment, with players selecting price in a way consistent with money illusion. In the replication treatments, it can be argued that there are indeed lock-in effects.

Creating a tension between the nominal dominant and the focal equilibrium from RepR to InvR (Test 2), clearly flattens the distribution, preventing the peak at P_{ef} and leading to convergence into the *inefficient* equilibrium. Since the distribution is flatter and doesn't have a clear peak in InvR, it would be difficult to argue whether there are or there aren't any lock-in effects in this treatment.

The interesting pattern, however, is observed under InvN. The treatment, as was expected, induced players into selecting prices in the first period much closer to the *efficient* equilibrium (the distribution is clearly skewed to the right), since nominal and real dominance are aligned. But Result 3 showed that players in InvN did **not** converge into the *efficient* equilibrium, and the trend of the all groups was driven away from it. This means that even if players at the beginning of the game could be choosing prices based on nominal dominance, such decision didn't stick in the long run.

What can one make of such results? They show that in the replication treatments, the data is consistent with both the implications of money illusion and with having initial group decisions that "lock" players in inefficient outcomes in the long run. This is in line with the original results found in Fehr and Tyran (2007). But once we modify the environment in further tests, such effects disappear. Even if players initially choose closer to the nominal and real dominant equilibrium, they are slowly driven *away* from it. Lock-in effects don't happen under InvN.

What could explain players moving away from their initial decisions? What could explain the slow drifting towards P_{inf} in InvN (Figure 1.9)? The objective of the tests conducted are to observe if the hypothesis put forward holds, not to give behavioral explanations of the observed outcomes. However, we can conjecture about the results. The observed trend towards the inefficient equilibrium could be consistent with a period by period best response on the part of the players. Perhaps the increased cognitive load triggered in players a behavioral rule of thumb according to which, instead of choosing the highest nominal payoffs as a potential coordination point, they engage in a short-term, period by period maximization of payoffs.

This conjecture could also be consistent with the way in which the payoff structure was designed. The whole experiment was designed with the intention of increasing the likelihood of players converging into P_{inf} under InvR, so the dynamics were set that way: best replies would on most parts of the matrix lead towards P_{inf}[26]. A period by period best response behaviour would slowly drift the prices towards the inefficient equilibrium, which would be consistent with the observed pattern under InvN.

Why would players under some treatments directly maximise long-run payoffs, and under others go for a period by period behaviour? The different framings for the payoffs can change the environment faced by the players, and under different environments they might use different rules of thumb (this is related to the concept of ecological rationality, presented for example in Gigerenzer et al. (2011)). But exploring further what rules of thumb subjects are actually using is not the objective of this paper. Its objective was to conduct tests of particular implications of the theory of money illusion, not to find possible alternatives to it.

[26] Put differently, the likelihood of falling into the "basin of attraction" of the inefficient equilibrium is higher than that for the efficient equilibrium.

1.6 CLOSING REMARKS

1.6.1 Summary

This paper explored the predictive power of the nominal dominance principle, defined as subjects converging on the long-run in the nominal dominant equilibrium. Such principle is closely related to the concept of money illusion, which involves subjects taking nominal payoffs as a proxy for real payoffs. In order to put this principle to the test, we conducted an experiment in a pricing, coordination game with two Pareto-ranked equilibria. In all three tests the implemented treatments consisted in changing the framing of payoffs, while keeping constant the real structure of incentives across all of them. The objective was to test changes in convergence towards either equilibrium when different framings were introduced.

Test 1 implemented two treatments that created a tension between real payoffs and nominal payoffs. The objective of this test was to try to replicate qualitatively results found in previous literature, where agents can converge into the equilibrium with highest nominal payoffs, as stated by the nominal dominance principle. Under our base treatment, two out of three groups converged to the Pareto-efficient equilibrium, as traditional rationality would suggest. But under a nominal representation, making the inefficient equilibrium the nominal dominant, led to full convergence to it (three out of three groups). Previous effects attributed to money illusion were replicated under our experimental design.

Test 2 put such result under a more stringent scenario. We created a treatment where both nominal and real dominance pointed towards the same efficient equilibrium. Could we lure players into the opposite direction? By adding a constant to all payoffs and inverting the order in the payoffs matrix labels, we tried to make players perceive the inefficient equilibrium as a more natural coordination point. A tension was created between focality and payoffs dominance. All three groups were effectively lured into coordinating in the inefficient equilibrium. Here, contrary to what was observed in Test 1, the nominal (and real) dominance principle did not hold. Even more, if as in previous literature, one considers focal points to be a confounding factor of Money Illusion, this result shows that such factors can, on their own, have similar behavioral effects.

In order to give the nominal dominance principle a chance to "get back on its feet", we defused the tension created in Test 2. Test 3 aligned back both focality and payoff dominance. Even more, conditions considered to make subjects more prone to converge into the nominal dominant equilibrium, such as cognitive load, were reinforced in the final treatment. Contrary to our expectations, none of the groups converged according to the nominal dominance principle. Although the framing of payoffs did have behavioral effects such as making convergence slower, they were not what would be predicted by our tested hypothesis on equilibrium convergence.

Finally, we tested if "lock-in" effects, an important concept that theoretically supports money illusion as a coordination device, took place in our treatments. Given a lack of communication between subjects, it might be that even if they learn individually how to choose the optimal solution, a coordination failure might prevent them from reaching that solution. They might be "locked-in" their initial decisions because it is costly to deviate from the group's median behavior. We compared decisions in the first and final rounds of the game. If such effects are taking place, one could expect those decisions to point towards the same equilibrium. We showed that such effects indeed took place in Test 1, but that they did not happen in Test 3. As with the predictions of the nominal dominance principle, lock-in effects were found in some of our treatments, but not in others.

1.6.2 Discussion

Our results can be related to the ongoing debate in the literature, regarding the effects of money illusion, in several ways. On the one hand, our experiment supports the idea that payoffs framing is relevant in terms of behavior, and that it can divert subjects from maximization of real payoffs. We believe our evidence, as well as previous literature, shows that subjects can indeed alter their choices when nominal values are introduced. For evidence against perfect real payoffs maximization and to highlight the relevance of bounded rationality in explaining subjects' behavior, we believe FT have put forward a solid experimental framework to study money illusion, with results that can be replicated. We have no doubt about the relevance of payoffs framing in affecting subjects' actions in these experiments. The difficulty lies in understanding more precisely how and when such effects take place.

We also believe that the concerns of PW, highlighting that there could be confounding effects such as focality or cognitive load on this line of research, are equally important. Leaving aside discussions on whether such effects are considered as part of a broader concept of money illusion or not, our results show that they can, on their own, drive behavior in similar ways as to what is attributed to money illusion. The relevant issue about those confounds is that they also have empirical effects even when isolated from money illusion, as shown in our experiment.

In order to conclude, a key message from our data is that there are framing effects that can alter subjects' behavior in several different ways. Sometimes, such effects drive behavior in directions consistent with nominal dominance, but some other times they don't. It seems that in order to clarify better such effects, we need to understand more about the underlying cognitive processes taking place. FT's experiments have presented an innovative experimental framework to study payoffs framings, and have put forward a theoretical basis with their interpretation of money illusion. They have laid the ground for further exploration. But if several years after FT's initial results, the concept of money illusion is to take the next step into better explaining behavior or making predictions beyond falsifying perfect rationality, it needs to be reformulated or expanded.

1.7 APPENDIX

1.7.1 RepR payoffs matrix

Your Selected Price	Other Players' Median Price																																		
	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
35	54	55	54	53	53	53	48	43	40	37	32	27	23	16	13	10	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	56	57	56	55	55	55	50	45	42	39	34	29	25	18	15	12	7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	58	59	58	57	57	57	52	47	44	41	36	31	27	20	17	14	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	56	57	60	59	59	59	54	49	46	43	38	33	29	22	19	16	11	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	54	55	58	61	61	61	56	51	48	45	40	35	31	24	21	18	13	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	52	53	56	59	59	59	58	53	50	47	42	37	33	26	23	20	15	9	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	50	51	54	57	57	57	56	55	52	49	44	39	35	28	25	22	17	11	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	48	49	52	55	55	55	54	53	54	51	46	41	37	30	27	24	19	13	9	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	46	47	50	53	53	53	52	51	52	53	48	43	39	32	29	26	21	15	11	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	44	45	48	51	51	51	50	49	50	51	50	45	41	34	31	28	23	17	13	9	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	42	43	46	49	49	49	48	47	48	49	48	47	43	36	33	30	25	19	15	11	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	40	41	44	47	47	47	46	45	46	47	46	45	45	38	35	32	27	21	17	13	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0
23	38	39	42	45	45	45	44	43	44	45	44	43	43	40	37	34	29	23	19	15	11	3	0	0	0	0	0	0	0	0	0	0	0	0	0
22	36	37	40	43	43	43	42	41	42	43	42	41	41	38	39	36	31	25	21	17	13	5	2	0	0	0	0	0	0	0	0	0	0	0	0
21	34	35	38	41	41	41	40	39	40	41	40	39	39	36	37	38	33	27	23	19	15	7	4	1	0	0	0	0	0	0	0	0	0	0	0
20	32	33	36	39	39	39	38	37	38	39	38	37	37	34	35	36	35	29	25	21	17	9	6	3	0	0	0	0	0	0	0	0	0	0	0
19	30	31	34	37	37	37	36	35	36	37	36	35	35	32	33	34	33	31	27	23	19	11	8	5	2	0	0	0	0	0	0	0	0	0	0
18	28	29	32	35	35	35	34	33	34	35	34	33	33	30	31	32	31	29	29	25	21	13	10	7	4	1	0	0	0	0	0	0	0	0	0
17	26	27	30	33	33	33	32	31	32	33	32	31	31	28	29	30	29	27	27	27	23	15	12	9	6	3	0	0	0	0	0	0	0	0	0
16	24	25	28	31	31	31	30	29	30	31	30	29	29	26	27	28	27	25	25	25	25	17	14	11	8	5	2	0	0	0	0	0	0	0	0
15	22	23	26	29	29	29	28	27	28	29	28	27	27	24	25	26	25	23	23	23	23	19	16	13	10	7	4	1	0	0	0	0	0	0	0
14	20	21	24	27	27	27	26	25	26	27	26	25	25	22	23	24	23	21	21	21	21	17	18	15	12	9	6	3	0	0	0	0	0	0	0
13	18	19	22	25	25	25	24	23	24	25	24	23	23	20	21	22	21	19	19	19	19	15	16	17	14	11	8	5	2	0	0	0	0	0	0
12	16	17	20	23	23	23	22	21	22	23	22	21	21	18	19	20	19	17	17	17	17	13	14	15	16	13	10	7	4	1	1	0	0	0	0
11	14	15	18	21	21	21	20	19	20	21	20	19	19	16	17	18	17	15	15	15	15	11	12	13	14	15	12	9	6	3	3	0	0	0	0
10	12	13	16	19	19	19	18	17	18	19	18	17	17	14	15	16	15	13	13	13	13	9	10	11	12	13	14	11	8	5	5	0	0	0	0
9	10	11	14	17	17	17	16	15	16	17	16	15	15	12	13	14	13	11	11	11	11	7	8	9	10	11	12	13	10	7	7	0	0	0	0
8	8	9	12	15	15	15	14	13	14	15	14	13	13	10	11	12	11	9	9	9	9	5	6	7	8	9	10	11	12	9	9	0	0	0	0
7	6	7	10	13	13	13	12	11	12	13	12	11	11	8	9	10	9	7	7	7	7	3	4	5	6	7	8	9	10	11	11	0	0	0	0
6	4	5	8	11	11	11	10	9	10	11	10	9	9	6	7	8	7	5	5	5	5	1	2	3	4	5	6	7	8	9	13	0	0	0	0
5	2	3	6	9	9	9	8	7	8	9	8	7	7	4	5	6	5	3	3	3	3	0	0	1	2	3	4	5	6	7	15	0	0	0	0
4	0	1	4	7	7	7	6	5	6	7	6	5	5	2	3	4	3	1	1	1	1	0	0	0	0	1	2	3	4	5	13	0	0	0	0
3	0	0	2	5	5	5	4	3	4	5	4	3	3	0	1	2	1	0	0	0	0	0	0	0	0	0	0	1	2	3	11	18	0	0	0
2	0	0	0	3	3	3	2	1	2	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	0	18	0	0
1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	100	100

1.7.2 RepN payoffs matrix

Your Selected Price	Other Players' Median Price																																		
	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
35	1890	1870	1782	1696	1643	1590	1392	1204	1080	962	800	648	529	352	273	200	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	1960	1938	1848	1760	1705	1650	1450	1260	1134	1014	850	696	575	396	315	240	133	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	2030	2006	1914	1824	1767	1710	1508	1316	1188	1066	900	744	621	440	357	280	171	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	1960	1938	1980	1888	1829	1770	1566	1372	1242	1118	950	792	667	484	399	320	209	90	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	1890	1870	1914	1952	1891	1830	1624	1428	1296	1170	1000	840	713	528	441	360	247	126	51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	1820	1802	1848	1888	1829	1770	1682	1484	1350	1222	1050	888	759	572	483	400	285	162	85	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	1750	1734	1782	1824	1767	1710	1624	1540	1404	1274	1100	936	805	616	525	440	323	198	119	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	1680	1666	1716	1760	1705	1650	1566	1484	1458	1326	1150	984	851	660	567	480	361	234	153	80	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	1610	1598	1650	1696	1643	1590	1508	1428	1404	1378	1200	1032	897	704	609	520	399	270	187	112	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	1540	1530	1584	1632	1581	1530	1450	1372	1350	1326	1250	1080	943	748	651	560	437	306	221	144	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	1470	1462	1518	1568	1519	1470	1392	1316	1296	1274	1200	1128	989	792	693	600	475	342	255	176	105	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	1400	1394	1452	1504	1457	1410	1334	1260	1242	1222	1150	1080	1035	836	735	640	513	378	289	208	135	14	0	0	0	0	0	0	0	0	0	0	0	0	0
23	1330	1326	1386	1440	1395	1350	1276	1204	1188	1170	1100	1032	989	880	777	680	551	414	323	240	165	42	0	0	0	0	0	0	0	0	0	0	0	0	0
22	1260	1258	1320	1376	1333	1290	1218	1148	1134	1118	1050	984	943	836	819	720	589	450	357	272	195	70	26	0	0	0	0	0	0	0	0	0	0	0	0
21	1190	1190	1254	1312	1271	1230	1160	1092	1080	1066	1000	936	897	792	777	760	627	486	391	304	225	98	52	12	0	0	0	0	0	0	0	0	0	0	0
20	1120	1122	1188	1248	1209	1170	1102	1036	1026	1014	950	888	851	748	735	720	665	522	425	336	255	126	78	36	0	0	0	0	0	0	0	0	0	0	0
19	1050	1054	1122	1184	1147	1110	1044	980	972	962	900	840	805	704	693	680	627	558	459	368	285	154	104	60	22	0	0	0	0	0	0	0	0	0	0
18	980	986	1056	1120	1085	1050	986	924	918	910	850	792	759	660	651	640	589	522	493	400	315	182	130	84	44	10	0	0	0	0	0	0	0	0	0
17	910	918	990	1056	1023	990	928	868	864	858	800	744	713	616	609	600	551	486	459	432	345	210	156	108	66	30	0	0	0	0	0	0	0	0	0
16	840	850	924	992	961	930	870	812	810	806	750	696	667	572	567	560	513	450	425	400	375	238	182	132	88	50	18	0	0	0	0	0	0	0	0
15	770	782	858	928	899	870	812	756	756	754	700	648	621	528	525	520	475	414	391	368	345	266	208	156	110	70	36	8	0	0	0	0	0	0	0
14	700	714	792	864	837	810	754	700	702	702	650	600	575	484	483	480	437	378	357	336	315	238	234	180	132	90	54	24	0	0	0	0	0	0	0
13	630	646	726	800	775	750	696	644	648	650	600	552	529	440	441	440	399	342	323	304	285	210	208	204	154	110	72	40	14	0	0	0	0	0	0
12	560	578	660	736	713	690	638	588	594	598	550	504	483	396	399	400	361	306	289	272	255	182	182	180	176	130	90	56	28	6	5	0	0	0	0
11	490	510	594	672	651	630	580	532	540	546	500	456	437	352	357	360	323	270	255	240	225	154	156	156	154	150	108	72	42	18	15	0	0	0	0
10	420	442	528	608	589	570	522	476	486	494	450	408	391	308	315	320	285	234	221	208	195	126	130	132	132	130	126	88	56	30	25	0	0	0	0
9	350	374	462	544	527	510	464	420	432	442	400	360	345	264	273	280	247	198	187	176	165	98	104	108	110	110	108	104	70	42	35	0	0	0	0
8	280	306	396	480	465	450	406	364	378	390	350	312	299	220	231	240	209	162	153	144	135	70	78	84	88	90	90	88	84	54	45	0	0	0	0
7	210	238	330	416	403	390	348	308	324	338	300	264	253	176	189	200	171	126	119	112	105	42	52	60	66	70	72	72	70	66	55	0	0	0	0
6	140	170	264	352	341	330	290	252	270	286	250	216	207	132	147	160	133	90	85	80	75	14	26	36	44	50	54	56	56	54	65	0	0	0	0
5	70	102	198	288	279	270	232	196	216	234	200	168	161	88	105	120	95	54	51	48	45	0	0	12	22	30	36	40	42	42	75	0	0	0	0
4	0	34	132	224	217	210	174	140	162	182	150	120	115	44	63	80	57	18	17	16	15	0	0	0	0	10	18	24	28	30	65	0	0	0	0
3	0	0	66	160	155	150	116	84	108	130	100	72	69	0	21	40	19	0	0	0	0	0	0	0	0	0	0	8	14	18	55	72	0	0	0
2	0	0	0	96	93	90	58	28	54	78	50	24	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	45	0	54	0	0
1	0	0	0	32	31	30	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	200	100

1.7.3 InvR payoffs matrix

Your Selected Price \ Other Players' Median Price	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	1035	1036	1035	1034	1034	1034	1029	1024	1021	1018	1013	1008	1004	997	994	991	986	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
2	1037	1038	1037	1036	1036	1036	1031	1026	1023	1020	1015	1010	1006	999	996	993	988	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
3	1039	1040	1039	1038	1038	1038	1033	1028	1025	1022	1017	1012	1008	1001	998	995	990	984	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
4	1037	1038	1041	1040	1040	1040	1035	1030	1027	1024	1019	1014	1010	1003	1000	997	992	986	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
5	1035	1036	1039	1042	1042	1042	1037	1032	1029	1026	1021	1016	1012	1005	1002	999	994	988	984	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
6	1033	1034	1037	1040	1040	1040	1039	1034	1031	1028	1023	1018	1014	1007	1004	1001	996	990	986	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
7	1031	1032	1035	1038	1038	1038	1037	1036	1033	1030	1025	1020	1016	1009	1006	1003	998	992	988	984	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981
8	1029	1030	1033	1036	1036	1036	1035	1034	1035	1032	1027	1022	1018	1011	1008	1005	1000	994	990	986	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981
9	1027	1028	1031	1034	1034	1034	1033	1032	1033	1034	1029	1024	1020	1013	1010	1007	1002	996	992	988	984	981	981	981	981	981	981	981	981	981	981	981	981	981	981
10	1025	1026	1029	1032	1032	1032	1031	1030	1031	1032	1031	1026	1022	1015	1012	1009	1004	998	994	990	986	981	981	981	981	981	981	981	981	981	981	981	981	981	981
11	1023	1024	1027	1030	1030	1030	1029	1028	1029	1030	1029	1028	1024	1017	1014	1011	1006	1000	996	992	988	981	981	981	981	981	981	981	981	981	981	981	981	981	981
12	1021	1022	1025	1028	1028	1028	1027	1026	1027	1028	1027	1026	1026	1019	1016	1013	1008	1002	998	994	990	982	981	981	981	981	981	981	981	981	981	981	981	981	981
13	1019	1020	1023	1026	1026	1026	1025	1024	1025	1026	1025	1024	1024	1021	1018	1015	1010	1004	1000	996	992	984	981	981	981	981	981	981	981	981	981	981	981	981	981
14	1017	1018	1021	1024	1024	1024	1023	1022	1023	1024	1023	1022	1022	1019	1020	1017	1012	1006	1002	998	994	986	983	981	981	981	981	981	981	981	981	981	981	981	981
15	1015	1016	1019	1022	1022	1022	1021	1020	1021	1022	1021	1020	1020	1017	1018	1019	1014	1008	1004	1000	996	988	985	982	981	981	981	981	981	981	981	981	981	981	981
16	1013	1014	1017	1020	1020	1020	1019	1018	1019	1020	1019	1018	1018	1015	1016	1017	1016	1010	1006	1002	998	990	987	984	981	981	981	981	981	981	981	981	981	981	981
17	1011	1012	1015	1018	1018	1018	1017	1016	1017	1018	1017	1016	1016	1013	1014	1015	1014	1012	1008	1004	1000	992	989	986	983	981	981	981	981	981	981	981	981	981	981
18	1009	1010	1013	1016	1016	1016	1015	1014	1015	1016	1015	1014	1014	1011	1012	1013	1012	1010	1010	1006	1002	994	991	988	985	982	981	981	981	981	981	981	981	981	981
19	1007	1008	1011	1014	1014	1014	1013	1012	1013	1014	1013	1012	1012	1009	1010	1011	1010	1008	1008	1008	1004	996	993	990	987	984	981	981	981	981	981	981	981	981	981
20	1005	1006	1009	1012	1012	1012	1011	1010	1011	1012	1011	1010	1010	1007	1008	1009	1008	1006	1006	1006	1006	998	995	992	989	986	983	981	981	981	981	981	981	981	981
21	1003	1004	1007	1010	1010	1010	1009	1008	1009	1010	1009	1008	1008	1005	1006	1007	1006	1004	1004	1004	1004	1000	997	994	991	988	985	982	981	981	981	981	981	981	981
22	1001	1002	1005	1008	1008	1008	1007	1006	1007	1008	1007	1006	1006	1003	1004	1005	1004	1002	1002	1002	1002	998	999	996	993	990	987	984	981	981	981	981	981	981	981
23	999	1000	1003	1006	1006	1006	1005	1004	1005	1006	1005	1004	1004	1001	1002	1003	1002	1000	1000	1000	1000	996	997	998	995	992	989	986	983	981	981	981	981	981	981
24	997	998	1001	1004	1004	1004	1003	1002	1003	1004	1003	1002	1002	999	1000	1001	1000	998	998	998	998	994	995	996	997	994	991	988	985	982	982	981	981	981	981
25	995	996	999	1002	1002	1002	1001	1000	1001	1002	1001	1000	1000	997	998	999	998	996	996	996	996	992	993	994	995	996	993	990	987	984	984	981	981	981	981
26	993	994	997	1000	1000	1000	999	998	999	1000	999	998	998	995	996	997	996	994	994	994	994	990	991	992	993	994	995	992	989	986	986	981	981	981	981
27	991	992	995	998	998	998	997	996	997	998	997	996	996	993	994	995	994	992	992	992	992	988	989	990	991	992	993	994	991	988	988	981	981	981	981
28	989	990	993	996	996	996	995	994	995	996	995	994	994	991	992	993	992	990	990	990	990	986	987	988	989	990	991	992	993	990	990	981	981	981	981
29	987	988	991	994	994	994	993	992	993	994	993	992	992	989	990	991	990	988	988	988	988	984	985	986	987	988	989	990	991	992	992	981	981	981	981
30	985	986	989	992	992	992	991	990	991	992	991	990	990	987	988	989	988	986	986	986	986	982	983	984	985	986	987	988	989	990	994	981	981	981	981
31	983	984	987	990	990	990	989	988	989	990	989	988	988	985	986	987	986	984	984	984	984	981	981	982	983	984	985	986	987	988	996	981	981	981	981
32	981	982	985	988	988	988	987	986	987	988	987	986	986	983	984	985	984	982	982	982	982	981	981	981	981	982	983	984	985	986	994	981	981	981	981
33	981	981	983	986	986	986	985	984	985	986	985	984	984	981	982	983	982	981	981	981	981	981	981	981	981	981	981	982	983	984	992	999	981	981	981
34	981	981	981	984	984	984	983	982	983	984	983	982	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	982	990	981	999	981	981
35	981	981	981	982	982	982	981	981	981	982	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	981	988	981	981	1081	1081

1.7.4 *InvN payoffs matrix*

	Other Players' Median Price																																		
Your Selected Price	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	1,035	2,072	3,105	4,136	5,170	6,204	7,203	8,192	9,189	10,180	11,143	12,096	13,052	13,958	14,910	15,856	16,762	17,658	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
2	1,037	2,076	3,111	4,144	5,180	6,216	7,217	8,208	9,207	10,200	11,165	12,120	13,078	13,986	14,940	15,888	16,796	17,676	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
3	1,039	2,080	3,117	4,152	5,190	6,228	7,231	8,224	9,225	10,220	11,187	12,144	13,104	14,014	14,970	15,920	16,830	17,712	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
4	1,037	2,076	3,123	4,160	5,200	6,240	7,245	8,240	9,243	10,240	11,209	12,168	13,130	14,042	15,000	15,952	16,864	17,748	18,658	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
5	1,035	2,072	3,117	4,168	5,210	6,252	7,259	8,256	9,261	10,260	11,231	12,192	13,156	14,070	15,030	15,984	16,898	17,784	18,696	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
6	1,033	2,068	3,111	4,160	5,200	6,240	7,273	8,272	9,279	10,280	11,253	12,216	13,182	14,098	15,060	16,016	16,932	17,820	18,734	19,640	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
7	1,031	2,064	3,105	4,152	5,190	6,228	7,259	8,288	9,297	10,300	11,275	12,240	13,208	14,126	15,090	16,048	16,966	17,856	18,772	19,680	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
8	1,029	2,060	3,099	4,144	5,180	6,216	7,245	8,272	9,315	10,320	11,297	12,264	13,234	14,154	15,120	16,080	17,000	17,892	18,810	19,720	20,622	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
9	1,027	2,056	3,093	4,136	5,170	6,204	7,231	8,256	9,297	10,340	11,319	12,288	13,260	14,182	15,150	16,112	17,034	17,928	18,848	19,760	20,664	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
10	1,025	2,052	3,087	4,128	5,160	6,192	7,217	8,240	9,279	10,320	11,341	12,312	13,286	14,210	15,180	16,144	17,068	17,964	18,886	19,800	20,706	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
11	1,023	2,048	3,081	4,120	5,150	6,180	7,203	8,224	9,261	10,300	11,319	12,336	13,312	14,238	15,210	16,176	17,102	18,000	18,924	19,840	20,748	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
12	1,021	2,044	3,075	4,112	5,140	6,168	7,189	8,208	9,243	10,280	11,297	12,312	13,338	14,266	15,240	16,208	17,136	18,036	18,962	19,880	20,790	21,604	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
13	1,019	2,040	3,069	4,104	5,130	6,156	7,175	8,192	9,225	10,260	11,275	12,288	13,312	14,294	15,270	16,240	17,170	18,072	19,000	19,920	20,832	21,648	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
14	1,017	2,036	3,063	4,096	5,120	6,144	7,161	8,176	9,207	10,240	11,253	12,264	13,286	14,266	15,300	16,272	17,204	18,108	19,038	19,960	20,874	21,692	22,609	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
15	1,015	2,032	3,057	4,088	5,110	6,132	7,147	8,160	9,189	10,220	11,231	12,240	13,260	14,238	15,270	16,304	17,238	18,144	19,076	20,000	20,916	21,736	22,655	23,568	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
16	1,013	2,028	3,051	4,080	5,100	6,120	7,133	8,144	9,171	10,200	11,209	12,216	13,234	14,210	15,240	16,272	17,272	18,180	19,114	20,040	20,958	21,780	22,701	23,616	24,525	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
17	1,011	2,024	3,045	4,072	5,090	6,108	7,119	8,128	9,153	10,180	11,187	12,192	13,208	14,182	15,210	16,240	17,238	18,216	19,152	20,080	21,000	21,824	22,747	23,664	24,575	25,506	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
18	1,009	2,020	3,039	4,064	5,080	6,096	7,105	8,112	9,135	10,160	11,165	12,168	13,182	14,154	15,180	16,208	17,204	18,180	19,190	20,120	21,042	21,868	22,793	23,712	24,625	25,532	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
19	1,007	2,016	3,033	4,056	5,070	6,084	7,091	8,096	9,117	10,140	11,143	12,144	13,156	14,126	15,150	16,176	17,170	18,144	19,152	20,160	21,084	21,912	22,839	23,760	24,675	25,584	26,487	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
20	1,005	2,012	3,027	4,048	5,060	6,072	7,077	8,080	9,099	10,120	11,121	12,120	13,130	14,098	15,120	16,144	17,136	18,108	19,114	20,120	21,126	21,956	22,885	23,808	24,725	25,636	26,541	27,468	28,449	29,430	30,411	31,392	32,373	33,354	34,335
21	1,003	2,008	3,021	4,040	5,050	6,060	7,063	8,064	9,081	10,100	11,099	12,096	13,104	14,070	15,090	16,112	17,102	18,072	19,076	20,080	21,084	22,000	22,931	23,856	24,775	25,688	26,595	27,496	28,449	29,430	30,411	31,392	32,373	33,354	34,335
22	1,001	2,004	3,015	4,032	5,040	6,048	7,049	8,048	9,063	10,080	11,077	12,072	13,078	14,042	15,060	16,080	17,068	18,036	19,038	20,040	21,042	21,956	22,977	23,904	24,825	25,740	26,649	27,552	28,449	29,430	30,411	31,392	32,373	33,354	34,335
23	999	2,000	3,009	4,024	5,030	6,036	7,035	8,032	9,045	10,060	11,055	12,048	13,052	14,014	15,030	16,048	17,034	18,000	19,000	20,000	21,000	21,912	22,931	23,952	24,875	25,792	26,703	27,608	28,507	29,430	30,411	31,392	32,373	33,354	34,335
24	997	1,996	3,003	4,016	5,020	6,024	7,021	8,016	9,027	10,040	11,033	12,024	13,026	13,986	15,000	16,016	17,000	17,964	18,962	19,960	20,958	21,868	22,885	23,904	24,925	25,844	26,757	27,664	28,565	29,460	30,442	31,392	32,373	33,354	34,335
25	995	1,992	2,997	4,008	5,010	6,012	7,007	8,000	9,009	10,020	11,011	12,000	13,000	13,958	14,970	15,984	16,966	17,928	18,924	19,920	20,916	21,824	22,839	23,856	24,875	25,896	26,811	27,720	28,623	29,520	30,504	31,392	32,373	33,354	34,335
26	993	1,988	2,991	4,000	5,000	6,000	6,993	7,984	8,991	10,000	10,989	11,976	12,974	13,930	14,940	15,952	16,932	17,892	18,886	19,880	20,874	21,780	22,793	23,808	24,825	25,844	26,865	27,776	28,681	29,580	30,566	31,392	32,373	33,354	34,335
27	991	1,984	2,985	3,992	4,990	5,988	6,979	7,968	8,973	9,980	10,967	11,952	12,948	13,902	14,910	15,920	16,898	17,856	18,848	19,840	20,832	21,736	22,747	23,760	24,775	25,792	26,811	27,832	28,739	29,640	30,628	31,392	32,373	33,354	34,335
28	989	1,980	2,979	3,984	4,980	5,976	6,965	7,952	8,955	9,960	10,945	11,928	12,922	13,874	14,880	15,888	16,864	17,820	18,810	19,800	20,790	21,692	22,701	23,712	24,725	25,740	26,757	27,776	28,797	29,700	30,690	31,392	32,373	33,354	34,335
29	987	1,976	2,973	3,976	4,970	5,964	6,951	7,936	8,937	9,940	10,923	11,904	12,896	13,846	14,850	15,856	16,830	17,784	18,772	19,760	20,748	21,648	22,655	23,664	24,675	25,688	26,703	27,720	28,739	29,760	30,752	31,392	32,373	33,354	34,335
30	985	1,972	2,967	3,968	4,960	5,952	6,937	7,920	8,919	9,920	10,901	11,880	12,870	13,818	14,820	15,824	16,796	17,748	18,734	19,720	20,706	21,604	22,609	23,616	24,625	25,636	26,649	27,664	28,681	29,700	30,814	31,392	32,373	33,354	34,335
31	983	1,968	2,961	3,960	4,950	5,940	6,923	7,904	8,901	9,900	10,879	11,856	12,844	13,790	14,790	15,792	16,762	17,712	18,696	19,680	20,664	21,582	22,563	23,568	24,575	25,584	26,595	27,608	28,623	29,640	30,876	31,392	32,373	33,354	34,335
32	981	1,964	2,955	3,952	4,940	5,928	6,909	7,888	8,883	9,880	10,857	11,832	12,818	13,762	14,760	15,760	16,728	17,676	18,658	19,640	20,622	21,582	22,563	23,544	24,525	25,532	26,541	27,552	28,565	29,580	30,814	31,392	32,373	33,354	34,335
33	981	1,962	2,949	3,944	4,930	5,916	6,895	7,872	8,865	9,860	10,835	11,808	12,792	13,734	14,730	15,728	16,694	17,658	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,496	28,507	29,520	30,752	31,968	32,373	33,354	34,335
34	981	1,962	2,943	3,936	4,920	5,904	6,881	7,856	8,847	9,840	10,813	11,784	12,766	13,734	14,715	15,696	16,677	17,658	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,460	30,690	31,392	32,967	33,354	34,335
35	981	1,962	2,943	3,928	4,910	5,892	6,867	7,848	8,829	9,820	10,791	11,772	12,753	13,734	14,715	15,696	16,677	17,658	18,639	19,620	20,601	21,582	22,563	23,544	24,525	25,506	26,487	27,468	28,449	29,430	30,628	31,392	32,373	36,754	37,835

1.7.5 RepR Instructions

Welcome to the experiment.

These instructions will be read out loud by the experimenter. Please follow when he starts.

You can earn money in this experiment (up to 15 Pounds). Your income will be calculated in points. Those points you earn will be converted into British Pounds to be paid at the end of the experiment.

Please **do not communicate** with other participants during the experiment. Please raise your hand at any time if you have any questions regarding the instructions.

This experiment has 30 periods. All participants are members of a group consisting of ten (10) people. None of you know who is in each group, but the composition of the group remains the same throughout the experiment. Only the decisions in your own group are relevant for your earnings. Decisions by other groups are irrelevant for you.

All group members are in the role of firms. In each period, all firms must **simultaneously** set a price from 1 to 35 (1 and 35 included). How much a firm earns depends on the price it chooses and on the median price all **other** firms in the group choose.

The income table (find a printed copy on your desk) shows your **point income.** All firms have the same tables. *Example:* Suppose you choose a price of 30 and the other firms choose prices of 20 on median. In this case your point income is 20 points for that round. Those points you earn will be converted into British Pounds according to the following exchange rate: 200 points = 1 Pound.

Here is how the experiment proceeds: at the beginning of each period, you choose a selling price (a number from 1 to 35). At the end of each period you are informed about the actual median price of the other firms and about your actual point income.

Do you have any questions?

1.7.6 RepN instructions

Welcome to the experiment.

These instructions will be read out loud by the experimenter. Please follow when he starts.

You can earn money in this experiment (up to 15 Pounds). Your income will be calculated in points. Those points you earn will be converted into British Pounds to be paid at the end of the experiment.

Please **do not communicate** with other participants during the experiment. Please raise your hand at any time if you have any questions regarding the instructions.

This experiment has 30 periods. All participants are members of a group consisting of ten (10) people. None of you know who is in each group, but the composition of the group remains the same throughout the experiment. Only the decisions in your own group are relevant for your earnings. Decisions by other groups are irrelevant for you.

All group members are in the role of firms. In each period, all firms must **simultaneously** set a price from 1 to 35 (1 and 35 included). How much a firm earns depends on the price it chooses and on the median price all **other** firms in the group choose.

The income table (find a printed copy on your desk) shows your **nominal point income.** All firms have the same tables. *Example:* Suppose you choose a price of 30 and the other firms choose prices with a median of 20. In this case your *nominal* point income is 400 points for that round.

For the determination of your earnings at the end of the experiment, only the real point income is relevant. This holds for all firms. To calculate your real point income from your nominal point income, you have to divide the nominal point income by the median price of other firms (the column label in the income table). Therefore, the nominal and the real point income are related as follows:

Real point income = Nominal point income / Median price of other firms

In the example above, your nominal point income is 400 points, but your **real** point income is 20 points (= 400 points / 20). Those real points you earn will be converted into British Pounds according to the following exchange rate: 200 points = 1 Pound.

Here is how the experiment proceeds: at the beginning of each period, you choose a selling price (a number from 1 to 35). At the end of each period you are informed about the actual median price of the other firms and about your actual point income.

Do you have any questions?

1.7.7 InvR instructions

Welcome to the experiment.

These instructions will be read out loud by the experimenter. Please follow when he starts.

You can earn money in this experiment (up to 15 Pounds). Your income will be calculated in points. Those points you earn will be converted into British Pounds to be paid at the end of the experiment.

Please **do not communicate** with other participants during the experiment. Please raise your hand at any time if you have any questions regarding the instructions.

This experiment has 30 periods. All participants are members of a group consisting of ten (10) people. None of you know who is in each group, but the composition of the group remains the same throughout the experiment. Only the decisions in your own group are relevant for your earnings. Decisions by other groups are irrelevant for you.

All group members are in the role of firms. In each period, all firms must **simultaneously** set a price from 1 to 35 (1 and 35 included). How much a firm earns depends on the price it chooses and on the median price all **other** firms in the group choose.

The income table (find a printed copy on your desk) shows your **point income.** All firms have the same tables. *Example:* Suppose you choose a price of 30 and the other firms choose prices with a median of 20. In this case your point income is 986 points for that round.

The earned points will be converted into British Pounds at the end of the experiment according to the following rule: in each round you are guaranteed a minimum of 981 points, which is the lowest point income given by the income table. Only points you make above that number (in each period) will be converted according to the following exchange rate: 200 points = 1 Pound. Another way to put it is that at the end of the experiment 29,430 points (30 periods X 981 points) will be subtracted from your point income; then the amount of British Pounds will be calculated.

The experiment proceeds in the following way: at the beginning of each period, you choose a selling price (a number from 1 to 35). At the end of each period you are informed about the actual median price of the other firms and about your point income.

Do you have any questions?

1.7.8 InvN instructions

Welcome to the experiment.

These instructions will be read out loud by the experimenter. Please follow when he starts.

You can earn money in this experiment (up to 15 Pounds). Your income will be calculated in points. Those points you earn will be converted into British Pounds to be paid at the end of the experiment.

Please **do not communicate** with other participants during the experiment. Please raise your hand at any time if you have any questions regarding the instructions.

This experiment has 30 periods. All participants are members of a group consisting of ten (10) people. None of you know who is in each group, but the composition of the group remains the same throughout the experiment. Only the decisions in your own group are relevant for your earnings. Decisions by other groups are irrelevant for you.

All group members are in the role of firms. In each period, all firms must **simultaneously** set a price from 1 to 35 (1 and 35 included). How much a firm earns depends on the price it chooses and on the median price all **other** firms in the group choose.

The income table (find a printed copy on your desk) shows your **nominal point income.** All firms have the same tables. *Example:* Suppose you choose a price of 30 and the other firms choose prices with a median of 20. In this case your *nominal* point income is 19,720 points for that round.

For the determination of your earnings at the end of the experiment, only the real point income is relevant. This holds for all firms. To calculate your real point income from your nominal point income, you have to divide the nominal point income by the median price of other firms (the column label in the income table). Therefore, the nominal and the real point income are related as follows:

Real point income = Nominal point income / Median price of other firms

In the example above, your nominal point income is 19,720 points, but your **real** point income is 986 points (= 19,720 points / 20).

The earned real points will be converted into British Pounds at the end of the experiment according to the following rule: in each round you are guaranteed a minimum of 981 real points, which is the lowest real point income given by the income table. Only real points you make above that number (in each period) will be converted according to the following exchange rate: 200 points = 1 Pound. Another way to put it is that at the end of the experiment 29,430 real points (30 periods X 981 real points) will be subtracted from your real point income; then your earned British Pounds will be calculated.

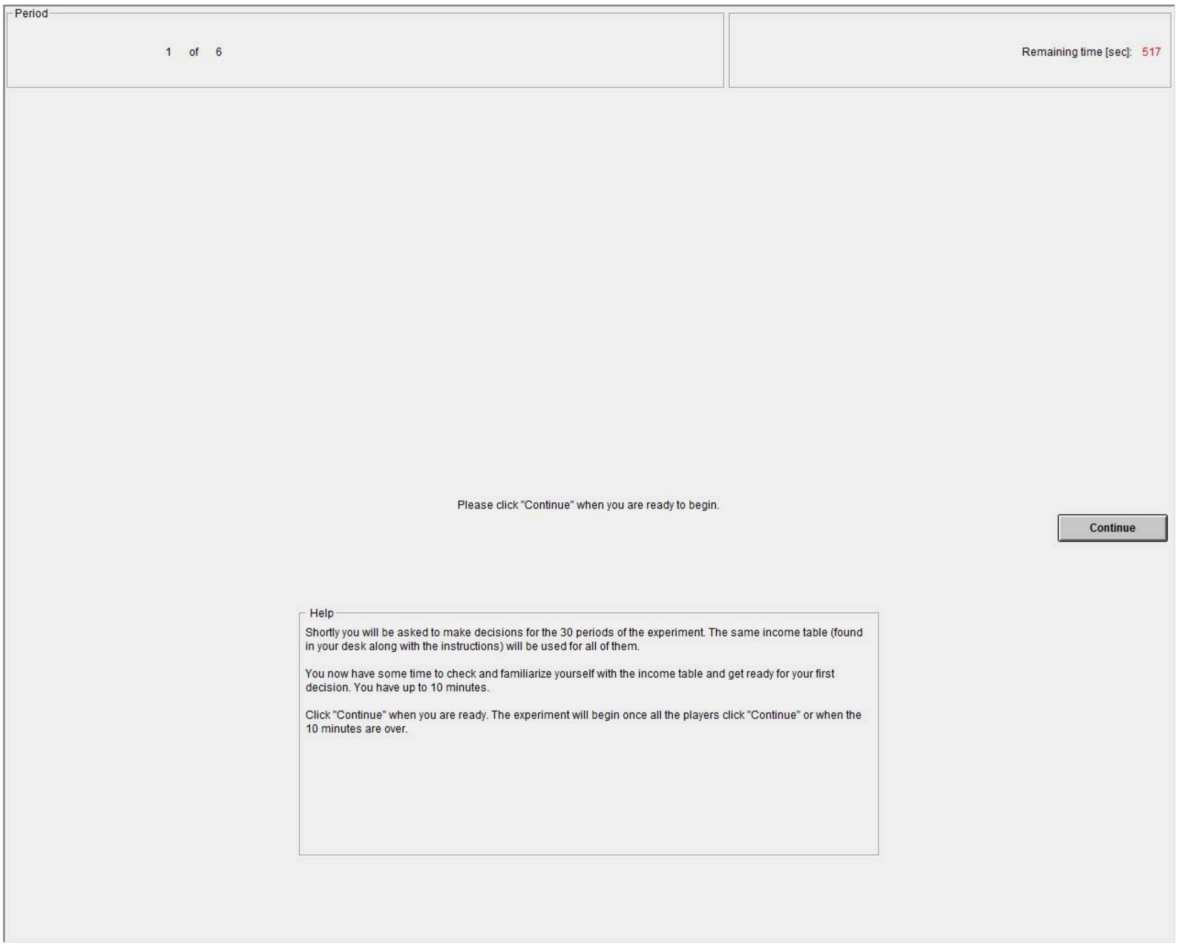
The experiment proceeds in the following way: at the beginning of each period, you choose a selling price (a number from 1 to 35). At the end of each period you are informed about the actual median price of the other firms and about your point income.

Do you have any questions?

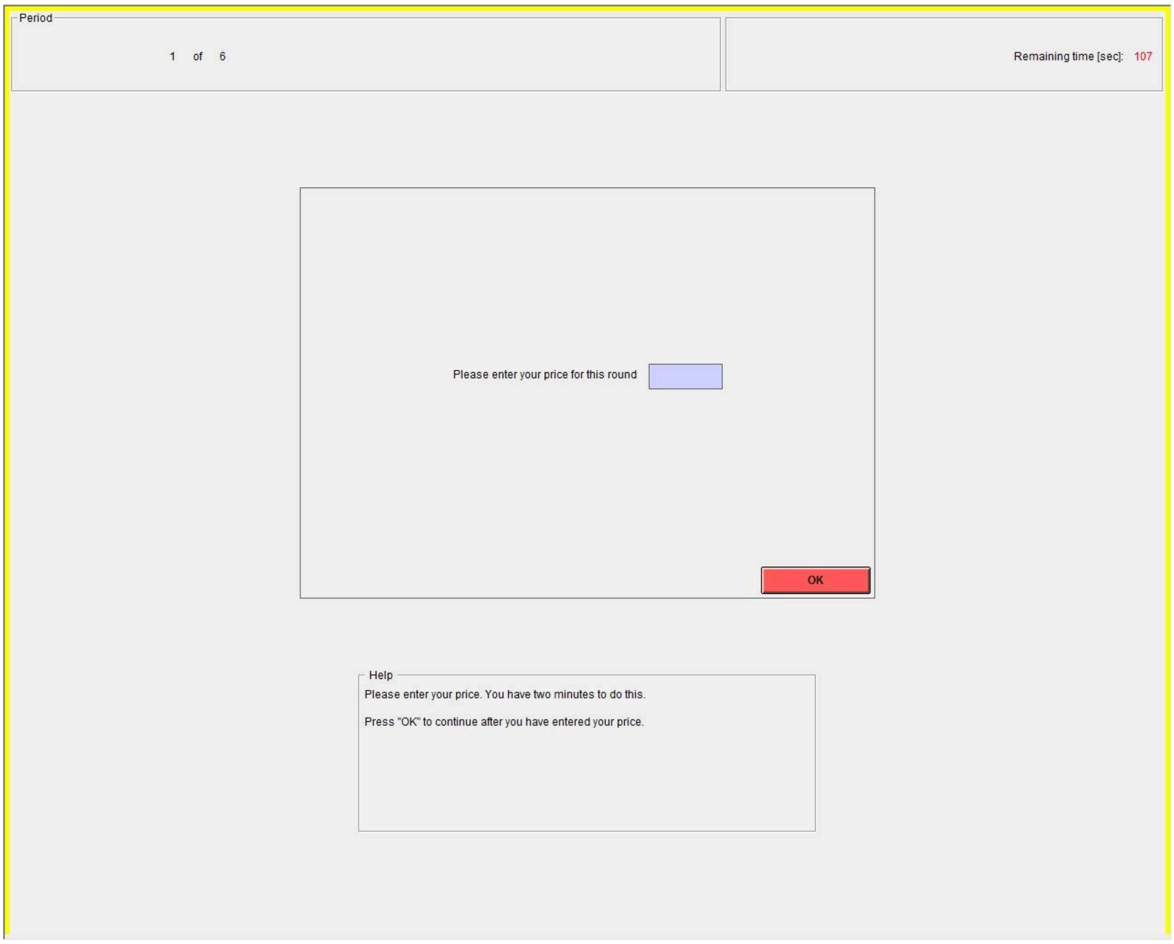
1.7.9 ***Experiment screenshots***

Total periods (top left corner) in the experiments was actually 35, as pointed in the main text. Screenshots show a different value due to them being taken during software test runs.

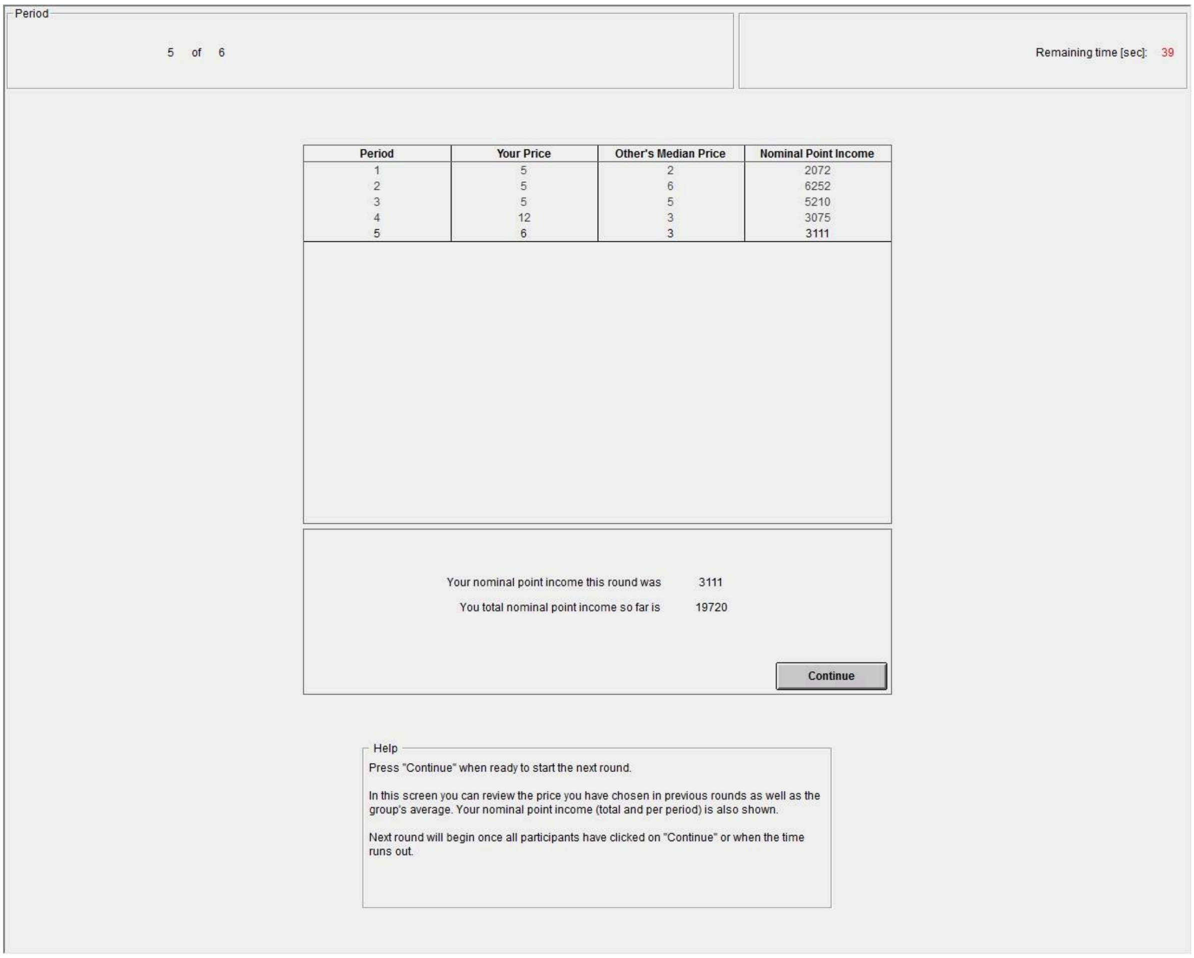
1.7.9.1 Beginning of the experiment



1.7.9.2 Input screen



1.7.9.3 *Review of previous rounds (nominal payoffs). From InvN treatment*



REFERENCES

ANBARCI, N., FELTOVICH, N., GÜRDAL, M.Y., (2015). "Payoff Inequity Reduces the Effectiveness of Correlated-Equilibrium Recommendations." *Work. Pap. Monash Univ.*,.

BRUNNERMEIER, M.K., JULLIARD, C., (2008). "Money Illusion and Housing Frenzies." *Rev. Financ. Stud.*, Vol. 21, pp. 135–180.

CANNON, E., CIPRIANI, G. PIETRO, (2006). "Euro-Illusion: A Natural Experiment." *J. Money, Credit Bank.*, Vol. 38, pp. 1391–1403.

COHEN, R.B., POLK, C., VUOLTEENAHO, T., (2005). "Money Illusion in the Stock Market: The Modigliani-Cohn Hypothesis." *Q. J. Econ.*, Vol. 120, pp. 639–668.

CRAWFORD, V.P., IRIBERRI, N., (2007). "Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental 'Hide-and-Seek' Games." *Am. Econ. Rev.*, Vol. 97, pp. 1731–1750.

DEHAENE, S., MARQUES, J.F., (2002). "Cognitive euroscience: scalar variability in price estimation and the cognitive consequences of switching to the euro." *Q J Exp Psychol A*, Vol. 55, pp. 705–731.

DZOKOTO, V., MENSAH, E., TWUM-ASANTE, M., OPARE-HENAKU, A., (2010). "Deceiving Our Minds: A Qualitative Exploration of the Money Illusion in Post-redenomination Ghana." *J. Consum. Policy*, Vol. 33, pp. 339–353.

FEHR, E., TYRAN, J.-R., (2001). "Does Money Illusion Matter?" *Am. Econ. Rev.*, Vol. 91, pp. 1239–1262.

FEHR, E., TYRAN, J.-R., (2005). "Individual Irrationality and Aggregate Outcomes." *J. Econ. Perspect.*, Vol. 19, pp. 43–66.

FEHR, E., TYRAN, J.-R., (2007). "Money illusion and coordination failure." *Games Econ. Behav.*, Vol. 58, pp. 246–268.

FEHR, E., TYRAN, J.-R., (2008). "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia." *Econometrica*, Vol. 76, pp. 353–394.

FEHR, E., TYRAN, J.-R., (2014). "Does Money Illusion Matter? Reply." *Am. Econ. Rev.*, Vol. 104, pp. 1063–1071.

FISCHBACHER, U., (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Exp. Econ.*, Vol. 10, pp. 171–178.

GIGERENZER, G., HERTWIG, R., PACHUR, T., (2011). Heuristics: The foundations of adaptive behavior. Oxford University Press, New York.

GREINER, B., (2003). An Online Recruitment System for Economic Experiments, Forschung und wissenschaftliches Rechnen. Goettingen: Ges. fuer Wiss. Datenverarbeitung, 79-93, 2004..

HARSANYI, J.C., SELTEN, R., (1988). "A general theory of equilibrium selection in games." *MIT Press Books*, Vol. 1.

HOWITT, P., (2008). "money illusion," in: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke.

KOOREMAN, P., FABER, R., HELEEN, H., (2004). "Charity Donations and the Euro Introduction: Some Quasi Experimental Evidence on Money Illusion." *J. Money, Credit Bank.*, pp. 1121–1124.

LEONTIEF, W.W., (1936). "The Fundamental Assumption of Mr. Keynes' Monetary Theory of Unemployment." *Q. J. Econ.*, Vol. 51, pp. 192–197.

LEVITT, S.D., LIST, J.A., (2009). "Field experiments in economics: The past, the present, and the future." *Eur. Econ. Rev.*, Vol. 53, pp. 1–18.

MEHTA, J., STARMER, C., SUGDEN, R., (1994). "The Nature of Salience: An Experimental Investigation of Pure Coordination Games." *Am. Econ. Rev.*, Vol. 84, pp. 658–673.

MUSSWEILER, T., ENGLICH, B., (2003). "Adapting to the Euro: Evidence from bias reduction." *J. Econ. Psychol.*, Vol. 24, pp. 285–292.

PETERSEN, L., WINN, A., (2014). "Does Money Illusion Matter? Comment." *Am. Econ. Rev.*, Vol. 104, pp. 1047–1062.

RAGHUBIR, P., MORWITZ, V.G., SANTANA, S., (2012). "Europoly Money: How Do Tourists Convert Foreign Currencies to Make Spending Decisions?" *J. Retail.*, Vol. 88, pp. 7–19.

ROBINSON, G.H., (2010). "Fechner's Law," in: The Corsini Encyclopedia of Psychology. John Wiley & Sons, Inc.

SCHELLING, T., (1960). The strategy of conflict. Harvard University Press.

SHAFIR, E., DIAMOND, P., TVERSKY, A., (1997). "Money Illusion." *Q. J. Econ.*, Vol. 112, pp. 341–374.

CHAPTER 2 .

LEE-PENAGOS, ALEJANDRO[⊥Ψ]

MODELLING CONTRIBUTIONS IN PUBLIC GOOD GAMES WITH PUNISHMENT

ABSTRACT

Theoretical models have had difficulties to account, at the same time, for the most important stylized facts observed in experiments of the Voluntary Contribution Mechanism. A recent approach tackling that gap is Arifovic and Ledyard (2012), which implements social preferences in tandem with an evolutionary learning algorithm. However, the stylized facts have evolved. The model was not built to explain some of the most important findings in the public good games recent literature: that altruistic punishment can sustain cooperation. This paper extends their model in order to explain such recent findings. It focuses on fear of punishment, not punishment itself, as the key mechanism to sustain contributions to the public good. Results show that our model can replicate both qualitatively and quantitatively the main facts. Data generated by our model differs, on average, in less than 5% compared to relevant experiments with punishment in the lab.

KEY WORDS: Public Good Games, Punishment, Agent Based Modelling, Learning Algorithms, Other Regarding Preferences, Bounded Rationality.

JEL CLASSIFICATION: C63, C70, C73, C92

[⊥] School of Economics, University of Nottingham. Email: alejandro.lee@nottingham.ac.uk

[Ψ] Acknowledgements: Special thanks to Chris Starmer and Alex Possajennikov for comments on this and previous versions of the paper. To Simon Gächter and Robin Cubitt for general commentaries, as well as to the participants of the WEHIA 2015 Conference in Nice, France, and of the Brown Bag Seminar (march, 2015) at the Nottingham School of Economics.

2.1 INTRODUCTION

Experiments with the Voluntary Contribution Mechanism (VCM)[27] have been a workhorse of social sciences in order to foster our understanding of human cooperation. Data from these 'n-players prisoner's dilemma' experiments (when played repeatedly for several rounds) have shown consistent patterns. Some authors (e.g. Holt and Laury (2008)) consider that there are five main stylized facts to focus on: i) that average contributions start around 50%, declining with time but not reaching zero, ii) that individuals vary considerably in their contributions (heterogeneity), iii) that higher values of the marginal productivity of the public good lead to increases in average contributions, iv) that increases in the size of the group lead to an increase in the average rate of contribution, and v) that there's a 'restart effect', so that when subjects are told that the game will restart, contributions increase and are similar as in first rounds.

Several authors have developed alternatives to the traditional Nash equilibrium approach in order to explain these main stylized facts, since traditional profit maximization would predict contributions of exactly zero in all periods of the experiments. Such alternatives have included decision errors (Anderson et al., 1998)), decision errors with altruism (Goeree et al., 2002), evolutionary dynamics (Miller and Andreoni, 1991), cooperative gain seeking (Brandts and Schram, 1996) and forward-looking signalling (Isaac et al., 1994), among others. Recently, important advances have been made by Fischbacher and Gächter (2010), highlighting the role of social preferences, beliefs and behavioural heterogeneity in order to explain the decline of contributions. However, none of these approaches could explain the main experimental findings *at the same time* (Holt and Laury, 2008). As put forward by Fischbacher and Gächter (2010) themselves, "the facts are clear, but the explanations are not"[28].

In order to close this gap, Arifovic and Ledyard (2012) (AL from now on) have developed IELORP[29], a model focused on explaining simultaneously the above mentioned patterns. Their model claims to do so by using primarily two building blocks. First, agents are endowed with Other Regarding Preferences (ORP), so that an agent's utility depends also on the payoffs of others, accounting for social motives. Second, agents learn their equilibrium, long-run strategies over time based on an Individual Learning Algorithm (IEL), which sets the dynamics towards convergence. AL claim that this model is robust to parameter changes, and importantly, that it has been tested in different environments and experiments, successfully explaining the data. This makes IELORP a strong contender among the many models to explain the stylized facts.

However, even if IELORP captures many of the most interesting earlier facts for repeated VCM experiments, it wasn't built to explain some of the most

[27] Arranging players in small groups, experimenters endow each individual with a resource, usually tokens representing real money, and each one of them can individually decide whether to contribute to a public good or to keep its own endowment. If everyone contributes, the group is better off, but if everyone else contributes, an individual can increase its own payoffs by not doing so, creating a tension between social and individual motives.

[28] pp. 541

[29] Although the acronym is not the easiest to remember, it will be kept the same in order to retain AL's original convention

important recent observations. After the seminal papers of Fehr and Gächter (2000) and Fehr and Gächter (2002), substantial work has been dedicated to understanding the effects of altruistic punishment on maintaining cooperation. Arguably, the sustainability of contribution levels when different types of punishment are available is one of the most important facts on the public goods game literature, not only for the economics literature, but for social sciences in general: Bowles and Gintis (2013) have most of their main stylized facts on human cooperation closely linked to altruistic punishment, on which they base several of their evolutionary models. Also, Guala (2012) discusses the importance of the experimental evidence on punishment for theories of strong reciprocity and its external validity, and Chaudhuri (2011) presents a recent literature survey on public goods games experiments which emphasizes punishment as one of the key mechanisms to sustain cooperation. The stylized facts have evolved.

This paper's objective is to model behavior in public goods games and account for the sustainability of cooperation when punishment is allowed. We test the usefulness of AL's modelling approach by extending IELORP. Can the model be extended to also explain some of the most relevant stylized facts found in experiments with punishment? Can it be done while maintaining its main assumptions and core building blocks (i.e. learning and other regarding preferences)? Our results will give a positive answer to these questions.

Our model includes punishment as a simple rule of thumb (Gigerenzer et al., 2002), based on empirical observations of how subjects assign punishment across several experiments[30]. We will show what we consider to be key stylized facts of punishment, including the possibility of it sustaining cooperation, but only when the costs of punishing are low enough relative to the impact it has on the punished player (Nikiforakis and Normann (2008), Egas and Riedl (2008)). To explain those facts, our model focuses on "fear of punishment", not punishment itself, as the main mechanism to sustain contributions (Fudenberg and Pathak, 2010). Intuitively, what the model does is to penalise strategies (in terms of utility) that are expected to be punished, based on the difference of contributions between agents: contributions sufficiently below the group's average, are expected to be punished. This allows the learning algorithm (IEL) to reinforce higher contributions, hence sustaining cooperation.

Our methodology is as follows. After introducing formally the linear VCM and the IELORP model (section 2.2), we attempt to replicate AL's model and test its previously reported results (section 2.3). Given the computational nature of IELORP, this replication is vital before extending the model. For this, we independently code the model and test if we can replicate the main findings by AL regarding how closely it tracks previous experimental data. At this point, we do not fit the model or calibrate any of its parameters, but rather test if we find the same results with the same parameter estimated by AL. Then, after introducing the punishment facts (section 2.4), we formally present our model and defend its methodology (section 2.5). Here is worth mentioning our calibration strategy (section 2.6.1). One of the main reasons for using IELORP as our starting point for modelling behaviour, is the previously reported

[30] A work also using heuristics to explore punishment is Pahl-Wostl and Ebenhöh (2004). Although similar in terms of the relevance it gives to empirically based heuristics for modelling behaviour, their approach is completely different from ours, theirs not including any kind of preferences or learning.

stability and out of sample robustness of its parameters. Our model calibrates (i.e. fits) to relevant experimental data only new parameters introduced in the punishment extension, but keeps the exact same values of AL for all the original IELORP parameters. This is a more stringent test than fitting again all the parameters, testing further the robustness and out of sample capabilities of the model.

Our results (section 2.6.2) will show that our extended model can replicate the main stylized facts of the punishment literature. The quantitative test of the model is done by running Monte Carlo simulations. The good fit of the model to the experimental data reflects that simulated contribution levels differ, on average, less than 5% compared to experiments on the lab. Overall, these results show not only that learning and other-regarding preferences reflect general behavioural insights that can explain the data on repeated VCM experiments, but also that they are compatible with more environment-specific, simple rules of behaviour, explaining how punishment can prevent contributions decline. Our model suggests that boundedly-rational behaviour that ignores information and relies on fast and frugal heuristics, can account for the most relevant facts observed in the repeated public goods games experimental data.

2.2 IELORP MODEL

We will start by presenting the IELORP model, introducing first the notation used for the VCM, followed by the explanation of the original AL model's two components: social preferences and individual evolutionary learning. These two components describe, respectively, the characteristics and behaviour of the agents. Their characteristics are given by the assumptions of what players care about in their utility functions, in this case, Other Regarding Preferences (ORP). The behavioural component is a non-strategic, Individual Evolutionary Learning algorithm (IEL). The latter explains how agents, given their characteristics and information about the environment, decide their contributions. This presentation follows closely that of AL and does not include punishment.

2.2.1 Linear Voluntary Contribution Mechanism

The VCM' structure is now widely known in the literature, so the following description is only intended as a way to introduce the notation. The core setup is as follows.

N agents (indexed $i = 1,2,\dots,N$) have a linear payoff function $\pi^i = p^i(w^i - c^i) + y$, where w^i is the initial endowment of a private good, c^i is their contribution to the production of the public good with $c^i \in [0, w^i]$ and y is the amount of public good produced. $1/p^i$ is the agents' willingness to pay in the private good for a unit of the public good. The production function of the public good is considered to be linear as $y = M\sum_{j=1}^{N} c^j$ with M being the marginal product of the public good. The game is given by the N players, their payoffs π^i and their possible contribution levels $c^i \in [0, w^i]$. The focus will be on symmetric games where all players have the same $p^i = 1$ and the same endowment $w^i = w$. In this case, if $M < 1$, notice that each agent i has a dominant strategy in contributing zero (choosing $c^i = 0$). If $M > (1/N)$, aggregate payoff is maximized when all agents choose $c^i = w$. Thus, the traditional commons dilemma is the tension between the individual (private) and public interest that arises when $(1/N) < M < 1$.

2.2.2 Other Regarding Preferences (ORP)

Since the influential work of Fehr and Schmidt (1999), there has been an extensive literature exploring utility functions that take into account not only own payoffs, but also those of other agents. Substantial empirical evidence shows that people indeed present this kind of social preferences, and that disregarding them by relying only on traditional selfish motivations prevents adequate understanding of relevant economics issues such as laws governing cooperation and collective action, effects and determinants of material incentives, which contracts and property rights arrangements are optimal, and important forces shaping social norms and market failures (Fehr and Fischbacher, 2002). Following this route, IELORP introduces other regarding preferences by endowing some agents (but not all) with components of social preference and envy.

For each player payoffs are given by $\pi^i(c) = w - c^i + M\sum c^j$ with an average group payoff of $\bar{\pi} = \sum \pi^i/N = w - \bar{c} + MN\bar{c}$, with $\bar{c} = \sum c^i/N$. The utility function for player i is given by

$$u^i(c) = \pi^i(c) + \beta^i\bar{\pi}(c) - \gamma^i \max\{0, \bar{\pi}(c) - \pi^i(c)\} \qquad (1)$$

with $\beta^i \geq 0$ and $\gamma^i \geq 0$.

In equation (1) the first term of the right hand side accounts for the interest for personal payoffs, with the second term being the interest for a social component (i.e. utility for the group's average payoffs with a weight of β^i). The third one represents the agents receiving disutility for being taken advantage of (i.e. receiving a payoff below the group average, that happens when $\bar{\pi} > \pi^i$).

Notice that heterogeneity is introduced by allowing parameter values (β, γ) to be different for each i. In IELORP these parameters are assumed exogenous (i.e. subjects come to the lab endowed with given preferences that don't change during the experiments). To model this, agents are given particular values (β^i, γ^i) from a population distribution $F(\beta, \gamma)$. $F(\beta, \gamma)$ is such that for each simulated agent, $(\beta, \gamma) = (0,0)$ with probability P. With probability $(1 - P)$, β^i and γ^i are drawn independently from $U([0, B])$ and $U([0, G])$ respectively, where $U(D)$ is the uniform density on the interval D[31]. The specific values of the parameter triplet (P, B, G) are discussed in section 2.6.1.

It is relevant to know what the possible one-shot Nash equilibrium levels of contribution can be. Given the utility function with other regarding preferences (equation (1)), the experimental parameters (N,M) and heterogeneity across (β, γ), only three types of Nash equilibrium behaviour are possible: free riding ($c^i = 0$), fully contributing ($c^i = w^i$), and conditionally cooperating ($c^i = \bar{c} = (\sum_i c^i)/N$). Is worth noting that in IELORP, free riding, altruism or conditional cooperation are considered as 'behavioural' types, not

[31] Another way to put this is that under the distribution $F(\beta, \gamma)$, with probability P, $(\beta, \gamma) =$ (0,0). Otherwise (with probability $(1 - P)$), $F(\beta, \gamma) = U([0, B]) \times U([0, G])$.

inner traits of the agents. An agent with the same "inner" parameters (i.e. β^i and γ^i) can show different equilibrium behaviour for different values of N and M. Put differently, the equilibrium strategy of the agent can vary depending on the environment (i.e. experimental setup)[32].

The above description of other regarding preferences accounts for the characteristics of agents. But we haven't defined exactly how they make their decisions. The model does not assume that decisions are made through traditional deductive reasoning. However, the one-shot Nash types of behavior are relevant because most learning algorithms would find such solutions given enough time. It is not a problem for most algorithms to find the dominant strategy of an agent. Technically, free-riding and fully-contributing can be defined as dominant strategies, but since conditional cooperation entails a strategy that is contingent on others' contributions, it cannot be defined as dominant. However, the same logic applies. So IELORP models agents learning such equilibrium strategies inductively. Let us turn now turn to this second aspect of the model, specifying their behavior.

2.2.3 Individual Evolutionary Learning (IEL)

The next step is to model how the agents choose their strategy c^i in each period. In many applications of evolutionary algorithms to economics (e.g. Andreoni and Miller (1995)), each agent is considered to be one strategy and the whole population of strategies jointly implements a behavioral algorithm (social learning). However, in other applications, individual learning is modelled with each agent having a set of strategies; evolution takes place not on the entire population of strategies but on the set belonging to one individual (Arifovic and Ledyard, 2011). As explained next, the latter is the approach followed by IEL. Let us first explain the learning algorithm in a general form for repeated games, and then use it specifically for a public goods game environment.

2.2.3.1 General form of the learning algorithm

The idea is that the repeated game has a stage game G that is played for T rounds. In $G = \{N, X, V, I\}$, N is the number of agents (indexed $i = 1,2,\dots,N$), X^i is the action space of i, $v^i(x^1,\dots,x^N)$ is the payoff of i if the joint strategy choice is x, and $I^i(x_t)$ is the information reported to i at the end of each round. In the lab, the experimenter controls all of these. In round t each i chooses $x_t^i \in X^i$ and is told information $I^i(x_t)$ about what happened. Then the next round is played. A behavioral model must explain how the sequence of choices for i, $(x_1^i, x_2^i, \dots, x_T^i)$ is made, given what i knows at each round t.

IEL has two primary variables: first, a finite set of potential actions for each agent i at each round t, $A_t^i \subset X^i$. Second, a probability measure ψ_t^i on A_t^i. A_t^i consists of J alternatives: this free parameter J can be thought (loosely) as a measure of the agent's processing capacity. In each round t the agent chooses randomly an alternative from A_t^i using the probability density ψ_t^i on A_t^i, and then chooses the action $x_t^i = a_t^i$. One way to see it is that a mixed strategy on X^i at t is induced by (A_t^i, ψ_t^i). At the end of each period t the agent is informed of $I^i(x_t)$. The heart of the behavioral model is that at the beginning of next

[32] The exact conditions for each type are presented in Appendix 2.8.2.

round $t+1$ the agent computes a new A^i_{t+1} and a new ψ^i_{t+1}. The three key components of IEL are as follows, starting at the end of round t knowing A^i_t, ψ^i_t and $I^i(x_t)$:

1. *Experimentation*[33]*:* this allows agents to try new strategies that perhaps might never be tried otherwise. With probability ρ and for each $j = 1,2,\dots,J$, action $a^i_{j,t}$ is replaced by a new contribution strategy selected at random from X^i. The distribution used for this replacement is normal $\sim N(a^i_{j,t},\sigma)$. So not only J but also ρ and σ, constitute the free parameters of the learning model.
2. *Replication:* a key component of the model is the concept of *foregone utility,* which refers to the payoffs that an action that was not played could've given to the agent. For example, in a public goods game, say an agent contributed 10 tokens to the public good in a particular round. At the end of that round, knowing his own contributions and those of the group, he can calculate his own payoffs. Those payoffs are observed based on the actual decision he made (his actual utility). But having observed a particular contribution of the group, he can make a similar counterfactual calculation. He may ask "How much would've been my payoffs, if instead of having contributed 10 tokens, I would've contributed, say, 15 tokens? What about 20 tokens?". The utility that he would have received for playing those 15 or 20 tokens, represent the foregone utilities for those potential contributions (taking as a given the group's contribution). The 'replication' part of the algorithm allows strategies in the set of potential actions to increase their probability of being chosen (by replicating, or replacing other actions with poorer performance), based on such foregone utility.

   Formally, let $v^i(a^i_{jt}|I^i(x_t))$ be the *foregone* utility of alternative j at time t given the information $I^i(x_t)$. The key assumption here is that the foregone utility $v^i(a^i_{jt}|I^i(x_t))$ is a counterfactual valuation function that must be specified for each application of the IEL learning model (specified for public goods games below). So given v^i, replication takes place as follows: For $j = 1,\dots,J$, $a^i_{j,t+1}$ is chosen as follows. From a uniform distribution, pick randomly (with replacement) two members of A^i_t. Let such two members be $a^i_{k,t}$ and $a^i_{l,t}$. Then

$$a^i_{j,t+1} = \begin{cases} a^i_{k,t}, & if \quad v^i(a^i_{kt}|I^i(x_t)) \geq v^i(a^i_{lt}|I^i(x_t)) \\ a^i_{l,t}, & if \quad v^i(a^i_{kt}|I^i(x_t)) < v^i(a^i_{lt}|I^i(x_t)) \end{cases}$$

   Replication in period $t+1$ favors alternatives with many replicates in A^i_t as well as those that, if would've been used in t, would've paid well. Actions that would've provided favorable situations given the actual contributions of others, will replicate in A^i_t. A^i_t will become more homogeneous as most alternatives become replicates of the best performing ones.
3. *Selection:* after experimentation and replication have taken place, selection occurs. Simply put, the probability of an agent choosing a

[33] This experimentation is similar in spirit to mutation in some biological models which randomly introduce changes.

particular action to play, depends on the foregone utility of that action relative to the foregone utilities of other potential actions.

Formally, each action $a^i_{k,t+1}$ has the following probability of being chosen:

$$\psi^i_{k,t+1} = \frac{v^i\left(a^i_{k,t+1}\middle|I^i(x_t)\right) - \varepsilon^i_{t+1}}{\sum_{j=1}^{J}(v^i\left(a^i_{j,t+1}\middle|I^i(x_t)\right) - \varepsilon^i_{t+1})}$$

for all $i = \{1,2,\dots,N\}$ and $k = \{1,2,\dots,J\}$, where

$$\varepsilon^i_{t+1} = \min_{a\epsilon A^i_{t+1}} \{0, v^i(a|I^i(x_t))\}.$$

If there are negative foregone utilities, what the latter does is to normalize all payoffs by adding a constant equal to the lowest payoff in the set (in absolute value).

All that is left to specify after describing how the agent calculates A^i_{t+1} and ψ^i_{t+1} starting from A^i_t and ψ^i_t, is to specify how the model is initialized. The assumption is a very naïve behavior: things begin randomly. A^i_1 is randomly populated with J draws from a uniform distribution from X^i. Also $\psi^i_{k,1} = 1/J$ for every k.

2.2.3.2 Application to VCM

Now the behavioral model is complete by having the two key elements of IEL, A and $v(a|I(x))$. In order to apply it to a VCM environment, one has to specify both of them, which is very straightforward. Let $A = [0, w]$. Since players receive an endowment w in the traditional VCM, their action space is the interval between zero and such endowment. Their decision is how much contribution they give out of w to the public good, so $c^i\epsilon[0, w]$. For specifying the value function, one requires to specify the information players receive, $I^i(x_t)$. Without punishment, in a public goods game players are informed the sum of the group's contributions, $\hat{c}_t = \sum_j c^j_t$. Since players know c^i_t (own contribution), they could calculate $\mu^i = \frac{\hat{c}_t - c^i}{N-1}$, which is the average of the contribution of the other players in the group. So let $I^i(c_t) = \mu^i_t$.

The functional form of the foregone utility v^i is based on the utility function in equation (1). Knowing the profits function $\pi^i = w^i - c^i + M\sum_{j=1}^{N} c^j$, v^i can be expressed as a function of c^i and μ^i_t as follows:

$$v^i(c^i|\mu^i) = c^i\left[(M-1) + \beta^i\left(M - \frac{1}{N}\right) - \gamma^{*i}\left(\frac{N-1}{N}\right)\right] + (N-1)\mu^i\left[M + \beta^i\left(M - \frac{1}{N}\right) + \gamma^{*i}\right] + w(1 + \beta^i) \qquad (2)$$

where $\gamma^{*i} = \begin{cases} \gamma^i, & \bar{\pi} \geq \pi^i \\ 0, & otherwise \end{cases}$

So it is this function $v^i(a|I^i(c_t) = \mu^i)$ the one used for the replication and selection procedures.

2.2.4 IELORP and previous literature

How is IELORP different from previous models? One can describe IELORP as endowing agents with an equilibrium behavior (free-riding, altruism or conditional cooperation) given by the Other Regarding Preferences. Such traits will reflect an agent's behavior in the long-run, but it is the learning mechanism what will determine the dynamics for such behavior to be reached. Neither of these ideas, however, are novel.

Other Regarding Preferences are now quite common in the literature (Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002)). For key VCM stylized facts (introduced in section 2.1), notice that no model that assumes completely selfish behavior could account accurately for contribution levels that decline over time but that remain positive. If free-riding is the only dominant strategy, eventually agents will converge into contributing exactly zero to the public good, which is not what experimental data shows (e.g. Isaac and Walker (1988)). IELORP underpins the same behavioral principles as such previous work. In fact, the implemented utility function (equation (1)) can be expressed as linear transformations of the ones used by Fehr and Schmidt (1999) or Charness and Rabin (2002). However, the differences in the specific functional form, as claimed by AL, are in order to explain that behavior changes when the group size is changed (tested in section 2.3). This is one aspect that differentiates IELORP with respect to previous literature.

Learning mechanisms are not novel either (Roth and Erev (1995), Camerer and Ho (1999)). Even more, they have also been used in tandem with Other Regarding Preferences in order to explain public goods games (Anderson et al. (2004), Cooper and Stockman (2002) and Janssen and Ahn (2006)). An important reason to model learning is that the stylized facts for VCM show that agents don't start playing right from the beginning of the game their long-run strategies (such as free-riding). Learning presents an explanation on why it takes time for people to reach equilibrium, hence making models more consistent with the empirical evidence. The claim by AL, however, is that their implemented learning algorithm is better at capturing speed of convergence towards equilibrium behavior, since the algorithms of the above models are not ideal for repeated games with strategy spaces that are a continuum. Also, AL claim that IEL's free parameters don't need to be recalibrated when tested in different games (Arifovic and Ledyard (2011), (2007), (2004)). The latter is key, because too many degrees of freedom is unlikely to be desirable for most models; if their values need to be calibrated only once, then IEL's usefulness can go beyond fitting data and be tested out of sample. The latter is a strong motivation for this work and to test further the usefulness of AL's modelling approach.

Let us now turn to testing whether IELORP can be replicated, and to check if we can independently reproduce its main characteristics.

2.3 TESTING IELORP PREVIOUS EVIDENCE

AL's claim is that IELORP can track several stylized facts in VCM experimental data (introduced in section 2.1). For testing the model, in their main results they compared their simulated data with experiments conducted by Isaac and Walker (1988) (IW from now on). Under the belief that replicability is a critical component of the scientific method, particularly in computational models (Wilensky and Rand, 2007), we use the same IW dataset and our own independent implementation of IELORP to verify AL's results[34]. This implementation consists on coding the model independently based on the information given in AL's paper. We also use their same parameter values in order to compare our simulations with IW's data and check if the same qualitative and quantitative results found by AL can be replicated.

2.3.1 Qualitative test

The experiments conducted by IW had subjects in the lab playing a repeated public goods game experiment (for ten rounds) under a partners setting (i.e. group composition was not changed). Their main results, which are tied to the stylized facts on which AL focused, are related to how average contributions to the public good change when group size (N) and marginal productivity (M) are altered. Plotting average contributions across groups for each period of the game, one should observe (as in the stylized facts in section 2.1) that they start around 50% percent of the endowment, and start declining with time without reaching zero. And although such negative trend in contributions should be observed for different values of N and M, contribution levels should be different: group size and marginal productivity affect how much players contribute. Figure 2.1 presents these empirical facts in IW data with the solid lines (ignore the dashed-lines for now), each data point representing the average contribution across subjects in six groups for each period. The design is 2x2 (four treatments), group size taking values of N= (4,10), and marginal productivity of the public good values M= (0.3,0.75). Endowment is normalized to w=10. Figure 2.1 shows that higher M leads to higher contributions. For example, with group size equal to four players (left panel), contributions across all periods are higher when M=0.75 compared to M=0.3, even if for both treatment contributions decline over time. The same holds in the right panel for group size equal to ten.

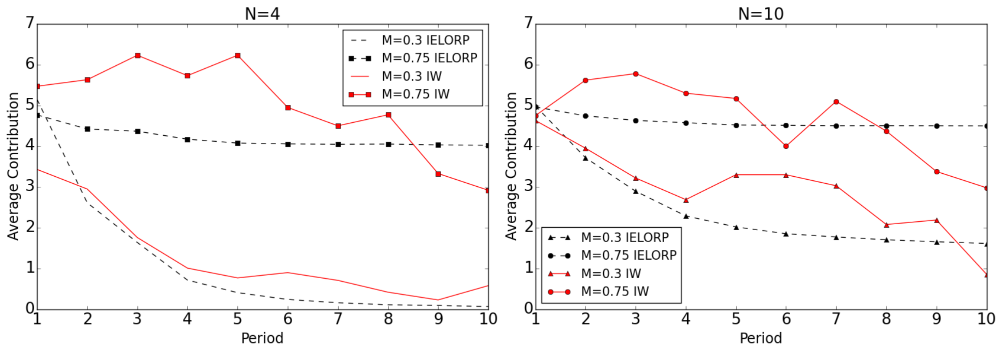


Figure 2.1: Comparison of our independent replication of IELORP (simulations) versus experimental data of Isaac and Walker (1988) *(IW). Includes four treatments: group size taking values N=4 (left panel) and N=10(right panel), and marginal productivity of the public good taking values M= (0.3,0.75).*

So Figure 2.1 shows the above empirical evidence from subjects in the lab, but the main goal of this section is to observe if our implementation of IELORP

[34] Our implementation of IELORP, as well as all simulations in this paper, were conducted using the agent-based-modelling software NetLogo, version 5.2 (Wilensky, 1999).

can replicate such findings. For this, simulations were run with our independent implementation of the model[35]. The experimental parameters were kept analogous to those in IW (e.g. group size, marginal productivity). The model parameters were taken directly from AL's estimations[36]. Notice that we do not fit the model here, but rather test if we can replicate the qualitative patterns of IW's data by using the parameter values previously estimated by AL. These simulations are represented by the dashed lines in Figure 2.1. Each data point represents the average contribution per period across 100 simulated groups of artificial agents, with each run of the model being analogous to one of the 6 groups in IW data (this makes simulated data "smoother", since it presents more observations). Qualitatively, it can be observed that for each treatment the simulated data is very similar to the experimental, presenting similar trends as well as having similar changes in contributions for the different values of N and M. This is considered as evidence that our implementation of IELORP replicates qualitatively the main features in IW's data, in a similar fashion as presented by AL. But what about quantitatively?

2.3.2 Quantitative test

One of the main measures AL use to test IELORP quantitatively is the squared error of how much the simulated differs from the experimental data. They estimate that on average, such difference is 3.4%. Let us explain how that measure is calculated, showing if our replication presents similar results.

Let $\bar{c}_{sim}^{10}(r)$ denote the average contribution for all simulated agents with IELORP across all ten periods on treatment r, for the particular parameter combination used (100 simulations). Let $\bar{c}_{sim}^{3}(r)$ be the analogous but only for the average of the last three periods, and $\bar{c}_{IW}^{10}(r)$ and $\bar{c}_{IW}^{3}(r)$ be such averages from IW data (across the six group observations for each treatment). The squared deviations between the simulated data and the experimental data were computed. That way the SE (Squared Error) was calculated as

$$SE = \sum_{r=1}^{R} [\bar{c}_{IW}^{10}(r) - \bar{c}_{sim}^{10}(r)]^2 + [\bar{c}_{IW}^{3}(r) - \bar{c}_{sim}^{3}(r)]^2 \qquad (3)$$

where R is the total number of treatments. For the present case of IW, $R = 4$. In order to have results that can be compared with experiments having different values for R, the SE is normalized. The reported value for the NSE (Normalized Squared Error) is

$$NSE = \sqrt{\frac{SE}{2R}}$$

---

[35] N artificial agents are created for each simulation, endowing them with Other Regarding Preferences parameters as explained in section 2.2.2, and playing for 10 periods. A new draw of parameters is done for each different run.

[36] $J = 100, \rho = 0.033, \sigma = \frac{w}{10} = 1, P = 0.48, B = 22, G = 8$. Is worth noting that AL calibrated parameters P,B and G to best fit IW data. The others, however, corresponding to the learning algorithm, were taken directly from previous work (Arifovic and Ledyard, 2011, 2007, 2004), appealing to its transferability.

It can be seen that the NSE is a standard measure of the difference between the simulated and experimental data. It also takes into account the average of the last three rounds in order to take into account the model's convergence, not just the average across all periods.

In our IELORP replication, NSE=0.43. The values to calculate it such as $\bar{c}_{IW}^{10}(r)$, belong to the interval [0,w]. Since we normalized to w=10, they are the average contribution for such an endowment. That value of NSE then represents an average error between our simulated data and IW's of 4.3%. The small difference of this value with that reported by AL (less than one percentage point), can reasonably be attributed to the inherent randomness of the simulations. With this, we consider that our implementation replicates IELORP's main features at the qualitative as well as the quantitative level[37].

2.4 PUNISHMENT STYLIZED FACTS

Since Fehr and Gächter (2000) and Fehr and Gächter (2002), the public goods game literature has highlighted the relevance of punishment as a fundamental mechanism to sustain cooperation. Our model is intended to capture relevant features of punishment experimental data beyond what was initially modeled by AL: the stylized facts have evolved to include punishment. This section's objective is to present four main stylized facts on punishment found in lab experiments. This presentation is not intended to survey the punishment literature, since other authors have already done so elsewhere (see, for example, Chaudhuri (2011)). The stylized facts presented below were chosen given what we considered, a priori, were the most relevant ones[38]. Some other important experimental results will be referred to indirectly, but such discussion if left for section 2.5.3.

Before presenting the facts, let us briefly present the traditional punishment setup and notation.

2.4.1 The punishment experimental setup

A traditional public goods game experiment with punishment works in the following way[39]. After players have decided on their contributions as they would if punishment is not allowed (i.e. in the VCM environment presented in section 2.2.1), a second stage is added. In this stage they are informed about how much the other individuals in the group contributed. Then, if they want, they can decide to buy punishment points (reducing their own income) to reduce the income of one or more of the other players. Let p_{ij} denote the amount of punishment points that player i assigns to reduce the income of player j (where $i, j = 1, \dots, N \; for \; j \neq i$), and e denote the effectiveness of each

---

[37] AL highlight other features of the model as well as other experimental setups where it was tested. Although we didn't formally explore those, they are worth mentioning since they further motivate our interest in IELORP as the base for our punishment model. They are summarised in Appendix 2.8.1.

[38] Thanks to Simon Gächter for discussions on this regard.

[39] This notation follows that of Nikiforakis and Normann (2008), since their data set is the one used later for testing the model (section 2.6).

punishment point: that is, how much the income of j is reduced for each p_{ij} assigned to him. Then the payoffs for i are described by

$$\pi^i = w - c^i + M\sum_{j=1}^{N} c^j - \sum_{j\neq i} p_{ij} - e\sum_{j\neq i} p_{ji}$$

The last two terms on the equation reflect how an agent's payoffs are affected when punishment is introduced: the agent takes the cost of punishing others in the group, as well as the cost of being punished by others, the latter multiplied by the effectiveness level set by the experimenter.

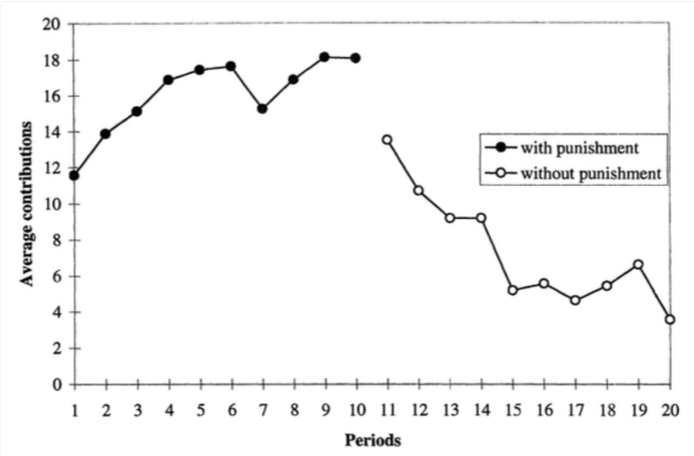
2.4.2 Stylized facts

2.4.2.1 Punishment can sustain cooperation

This is the most important of the facts. It has been shown that the contribution levels in public good games increase significantly compared to setups without punishment. Punishment can reverse the decline of cooperation.

The experiment pioneering this fact was Fehr and Gächter (2000). Their main treatments consisted on having both partners (group composition never changes) and strangers (groups are randomly reshuffled after each round) setups, as well as allowing and not allowing punishment. Some groups played 10 rounds of the public goods game with punishment followed by 10 round of no punishment (sequence 1), and others vice-versa (sequence 2). Figure 2.1 presents average period contributions in their data for the partners setup (data is similar for strangers, not presented). The stylized fact of punishment being able to sustain cooperation is captured by the slope of contributions: without punishment, the slope is negative, meaning that contributions decline over time. When punishment is allowed, the slope is positive, meaning that contributions don't decline and cooperation is sustained. Such results have been replicated several times in different labs across the world (Chaudhuri, 2011).

Panel A: Sequence 1 (with punishment followed by no punishment)



Panel B: Sequence 1 (no punishment followed by with punishment)

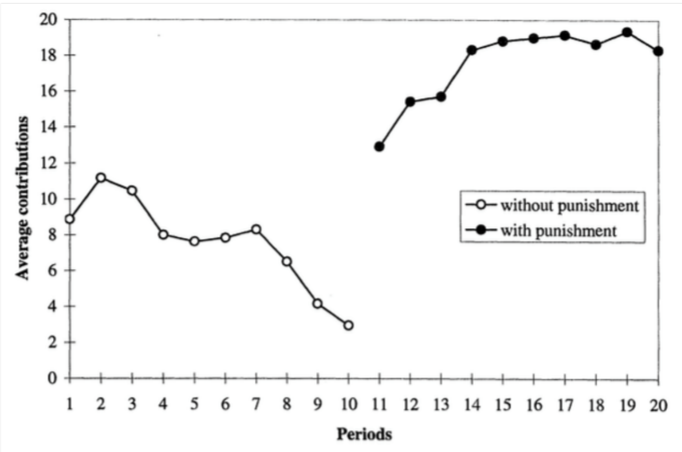


Figure 2.2: Average contributions over time (partners setup). Sequence 1 had subjects play 10 rounds with punishment allowed, followed by 10 rounds without punishment. Sequence 2 reversed that order. Source: Fehr and Gächter (2000)

2.4.2.2 The levels of contribution depend on the "effectiveness" of punishment

"Effectiveness" is defined as the experimental parameter that determines how many tokens (experimental points) are deducted from a punished player for each punishment point allocated by a punishing player. For example, if player A spends one point punishing player B in a given round, and the experimenter deducts two points from player B's payoffs due to such punishment, then the effectiveness is equal to two (ratio two to one).

Bowles and Gintis (2013) (p. 32) refer to this fact as "social preferences are not irrational". It means that even if people have preferences for social outcomes and care about others, as with any other good how much of it is consumed is affected by its "price" (i.e. effectiveness). When punishing is cheaper, levels of cooperation increase. Egas and Riedl (2008) and Nikiforakis and Normann (2008) are two studies showing clearly this fact in public good games. Figure 2.3 shows data from the former, under a strangers setup. Treatments here consisted of changing both punishment cost (how many tokens is the punisher player deducted in order to assign one token of punishment to the punished player) and impact (how much are payoffs of the punished player deducted for each token of punishment assigned to him). In this case, effectiveness can be defined as the ratio of cost to impact. The data shows that contributions declined in all treatments except in the one with highest effectiveness, where it actually increased with time. However, notice that in all the treatments with a negative slope, higher effectiveness was still associated with higher contributions, even if they declined over time.

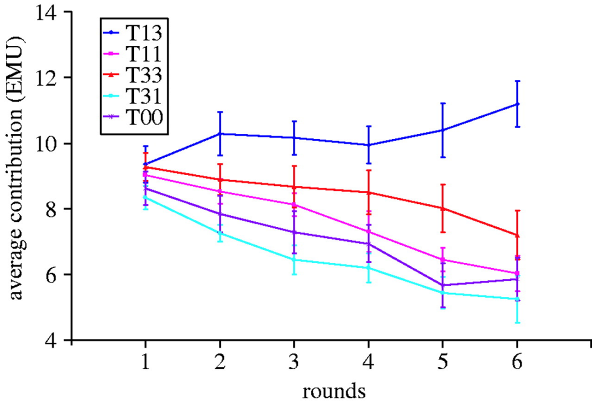


Figure 2.3: Average contributions in a public goods game. Each treatment T changes how much it costs one player to buy one token of punishment as well as how much that token deducts the payoffs of the punished player. For example, T31 means that one has to pay three tokens (cost) in order to deduct one point (impact) from another player. In this case, "effectiveness" would be the ratio of 1 to 3 (cost to impact ratio). Source: Egas and Riedl (2008)

Similar patterns can be observed in the data of Nikiforakis and Normann (2008), presented in Figure 2.4. Under a partners setup, their treatments changed the effectiveness level. The main difference with Egas and Riedl (2008) is that punishment cost is always constant at one token. Each treatment is labelled from "0" to "4". For example, in treatment "3", a player can deduct his own payoffs by one token in order to deduct three tokens from another player. Data shows clearly that the higher the levels of effectiveness, the higher the contribution levels to the public good.

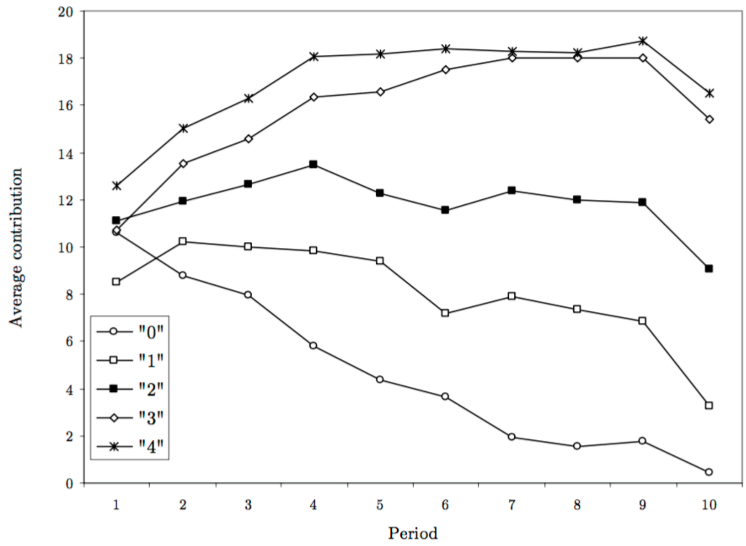


Figure 2.4: Average contributions in a public goods game with punishment. Each line presents a treatment with different effectiveness levels. For example, treatment "3" means that for each token of punishment assigned, the punished player payoffs are reduced by 3. Source: Nikiforakis and Normann (2008).

Summarising, both studies show that when effectiveness is too low, contributions decline on average. As effectiveness increases, contributions monotonically increase. With high enough effectiveness, cooperation can be sustained.

2.4.2.3 First period contributions remain the same with or without punishment and for different levels of effectiveness

This fact can be observed again in the experiments of Egas and Riedl (2008) and Nikiforakis and Normann (2008) (Figure 2.3 and Figure 2.4). The latter's main result is that average contributions are monotonic on all tested effectiveness levels; however, such condition holds for every period except the first one. In both studies initial average contributions are around half of the endowment, and are not statistically different for any effectiveness level, including zero (i.e. no punishment). This fact seems to highlight the non-strategic nature of punishment. The data of Fehr and Gächter (2000) and Fehr and Gächter (2002) show the same: with and without punishment, first round contributions are not statistically different.

2.4.2.4 When punishment sustains cooperation, group welfare is increased after sufficient rounds

One important caveat is to be made regarding the studies referenced in this section showing that punishment can sustain cooperation. Even if contributions levels are higher with punishment, this does not mean that group welfare is also necessarily higher. Punishment can lead to inefficiencies: the costs of punishing can outweigh the benefits of higher contributions in terms of group's payoffs. This is the case in Fehr and Gächter (2000) and Fehr and Gächter (2002), where punishment led to lower average net earnings. In both Egas and Riedl (2008) and Nikiforakis and Normann (2008), net earnings were higher only in the one treatment with the highest effectiveness. In all the other treatments, the social costs of punishment exceeded the benefits of higher contributions to the public good. Is then punishment mostly inefficient for group welfare?

To address this, Gächter et al. (2008) ran treatments with and without punishment, but allowing the experiments to last both 10 and 50 rounds under a partners setup. Their objective was to test if group welfare would increase after more periods of play were allowed. Figure 2.5 shows the net earnings of their experiment. Observing both treatments that allowed 50 rounds with and without punishment (P50 and N50), it can be seen that punishment allowed higher earnings. Without punishment, earnings were higher during the first periods of play (as was observed in previous studies), but once cooperation was stablished, punishment was unambiguously beneficial. The explanation is that once high levels of contributions are stablished, punishment is rarely needed: the credible threat of punishment, not punishment itself, is what sustains cooperation. So that the more rounds are played, the more the benefits of cooperation will outweigh the costs of punishment.

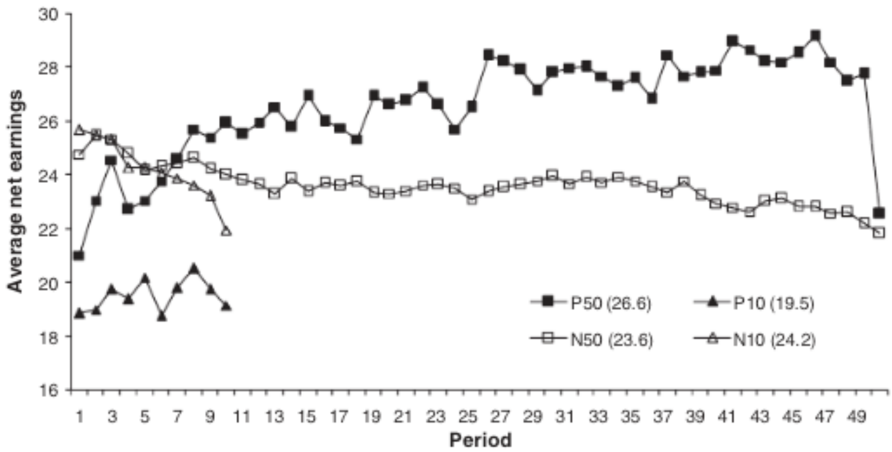


Figure 2.5: Average net earnings in a public goods game. Treatments included both no punishment and punishment setups ('N' and 'P' respectively), as well as time spans of both 10 and 50 rounds. For example, treatment P50 means that punishment was allowed and that the game lasted 50 rounds. Numbers in parenthesis show average earnings across all periods for each treatment. Source: Gächter et al. (2008)

Given the above facts on punishment, let us now turn to our presentation of the model.

2.5 MODELLING PUNISHMENT

In this section we present our modelling approach. Our model extends IELORP by introducing expectations of punishment: agents include such expectation in their counterfactual evaluation of potential actions (foregone utility) so that the learning algorithm favors strategies expected not to be punished. Our extended model will be referred to as 'Punishment Heuristics' (PH) from now on.

Punishment expectations are modeled as a simple rule of thumb, or as a 'fast and frugal heuristic' (Gigerenzer et al., 2011). Before formally introducing the model, we show that our punishment expectations are inspired in data showing that players in the lab use similar rules of thumb to assign punishment. This is both a motivation for our approach (showing it is plausible given the data), but will also give the reader some intuition before formalizing the model. After presenting PH, we discuss both how our heuristics approach fits IELORP theoretically, as well as other relevant aspects of the model. For example, is worth mentioning that the expectation of whether an action would be punished or not is the only necessary ingredient in the model to explain the stylised facts: this means that such expectation is considered exogenous, since it can't change through experience in the model. This makes it independent from actual punishment (i.e. allocation of punishment points): modelling punishment decisions is not required for obtaining the presented main results. Such independence between the expectations and actual experienced punishment might seem strange from a strategic game theoretical point of view, but there are empirical reasons that make this a valid approach. Section 2.5.3 will discuss this and other points, after PH is presented.

2.5.1 Motivation for punishment as a simple rule of thumb

How do subjects across experiments decide on which other players to punish? Figure 2.6, taken from Hetzer and Sornette (2013) can shed some light. The authors used the data from three different experiments (Fehr and Gachter (2002), (2000) and Fudenberg and Pathak (2010)) and calculated how much, on average, a player in a given group spends in punishing other players in relationship with pairwise deviations of contribution levels. Such pairwise deviations are defined as the difference between the contribution level of the punisher player with the contribution level of the punished one. Figure 2.6 shows that the more negative the deviations are, the more punishment is assigned. This data hints at players assigning punishment when other players contribute less than themselves, increasing punishment linearly when such contributions are lower.

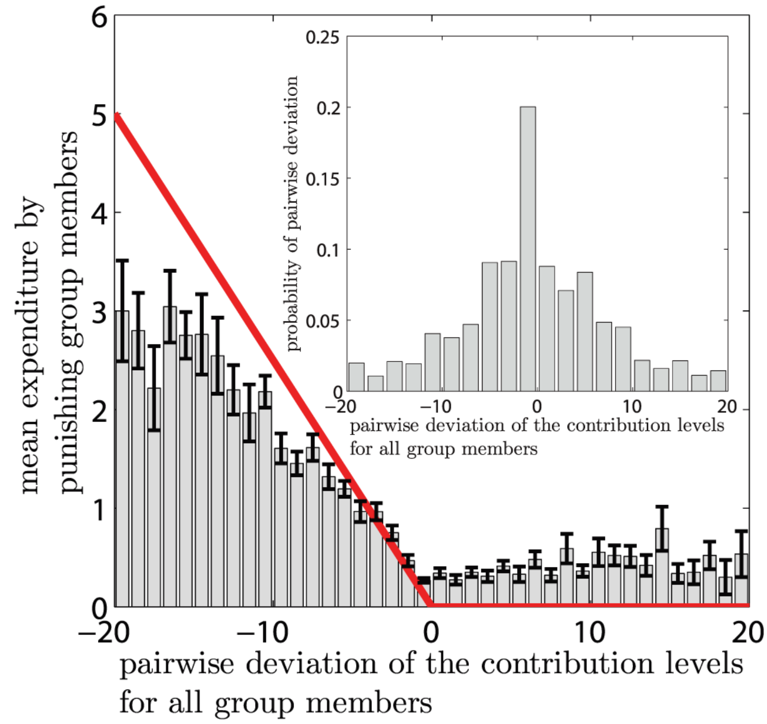


Figure 2.6: Mean expenditure of a given punishing member as a function of the deviation between that member's contribution and that of the punished member (for all pairs of subjects within a group). The straight line crossing zero shows the average decision rule for punishment: the more negative the deviations, punishment increases in a linear way. Error bars indicate the standard error around the mean. Data from the experiments of Fehr and Gachter (2002), (2000) and Fudenberg and Pathak (2010). *Source:* Hetzer and Sornette (2013)

PH will assume that players expect to receive punishment in a similar fashion. Our artificial agents will expect to be punished when they contribute less than other players in the group. However, instead of using pairwise comparisons, they compare their contribution with the group average, which simplifies their calculations. Agents will expect that only contributions below the group average will be punished, and that the higher the difference with respect to that average, the more punishment they will receive[40].

The above rule of thumb is closely related to how different effectiveness levels affect punishment and the sustainability of cooperation (stylized fact in section 2.4.2.2), vital in our modelling approach. Conclusions from Egas and Riedl (2008) can help in understanding the connection, since they find similar pairwise deviations for explaining punishment. As a reminder, their experiments changed both cost and impact of punishment, implying different levels of effectiveness. Their main results (Figure 2.3) showed that for higher effectiveness levels, higher contributions were observed. But they have other conclusions relevant for motivating PH. First, their results show that "*surprisingly, the marginal propensity to increase punishment with increasing deviations in contribution is the same for all four punishment*

[40] Figure 2.6 shows that positive deviations (i.e. contributing more than others) can also be punished, although such effect is smaller than for negative deviations. This "anti-social" punishment towards co-operators (Herrmann et al., 2008) is neglected in our model.

treatments"[41]. This means that for all effectiveness levels, one less token of contribution is associated with the same amount of punishment, meaning that a linear effect similar to that in Figure 2.6 (slope for negative differences) can be expected for different effectiveness.

Second, and perhaps more importantly, Egas and Riedl (2008) conclude that "*the cost and the impact of punishment have a significant effect on the* ***threshold*** *of deviation in contribution at which participants start to punish free-riders […] One surprising upshot of these results is that the force of punishment effectiveness can be pinned down to one single variable: the threshold level of free-riding that goes unpunished*"[42]. A key feature of PH will be related to the latter results: effectiveness levels affect the threshold at which agents start expecting to be punished. In other words, the lower the effectiveness levels are, agents will expect to be able to "get away" unpunished with lower contributions.

The above results give us an intuition about how we implement expectations of punishment. First, agents expect to be punished when they contribute less than their peers (below the group's average contributions), expecting more punishment the lower their contribution. Second, they expect a threshold, below the group's average contribution, for which they will start to be punished (i.e. small deviations from the group would not be punished). The key factor is that the higher the effectiveness levels, the smaller that threshold is, meaning that they expect to be punished easier. Let us now introduce PH formally.

2.5.2 The model

Agents' mechanism to evaluate if a particular action a_t^i would be punished depends on an estimated reference point R_t^i. If the action is lower than the reference point, the agent assumes that it would be punished. The reference point depends on two components: the last period's average group contribution $\bar{c_t} = \sum c_t^i / N$ and a tolerance value T, such that $R_{t+1}^i = \bar{c}_t - T$. How the tolerance value is estimated depends on the effectiveness parameter of the experiment, e. The latter is controlled by the experimenter, representing how many tokens a punished player is deducted from her profits when another player has assigned her one punishment token. So T is calculated as

$$T = \frac{w}{L^e} \qquad (4)$$

where $L > 1$ is the main free parameter of PH. Notice that T is the same for every period and every agent[43]. L represents how the tolerance T of players changes in response to different values of e.

---

[41] pp.875, italics added.

[42] pp.875, their own bolds, italics added.

[43] Here, since we simplify that the reference point is based on $\bar{c}_t$ (i.e. the whole group's average contribution, instead of the average of the other players in the group), then $R_{t+1}^i = R_{t+1}$ for all i, giving all agents the same reference point in a given period. The notation is kept as R_{t+1}^i because it is more general, and the model can easily be changed to allow for heterogeneity. In this case, we move forward with the homogeneity assumption as a particular case, testing how far can we go with this simpler assumption.

A behavioural interpretation of R_t^i and equation (4) is straightforward. T represents an agent's belief of how much it can get away with without being punished. Higher T means that agents believe lower contributions with respect to the group's average would still go unpunished. How is such tolerance estimated? It is entirely based on e. Lower e implies that punishment is less effective (more costly), so agents expect that others will punish less[44]. The key feature of Equation (4) is to impose an inverse relationship between e and T.

Then, if an action is expected to be punished, how much punishment is expected? This amount will be denoted by z_t^i. The basic idea is that the farther the potential contribution a_t^i is with respect to R_t^i, the more punishment is expected. The latter is modelled by allowing agents to calculate $z_t^i = (R_i^t - a_t^i)K$, where K is the second free parameter of PH. Notice that here an agent expecting punishment from all the other agents in the group doesn't care about where does the punishment comes from (i.e. from which player). The agent simply assumes that for each token of contribution below R_t^i it will receive a certain amount of punishment K. Then the amount of tokens expected as punishment z_t^i when evaluating action a_t^i is

$$z_t^i = \begin{cases} (R_t^i - a_t^i)K & if \quad a_t^i < R_t^i \\ 0 & otherwise \end{cases} \tag{5}$$

Equations (4) and (5) determine z_t^i for a_t^i depending on two free parameters, L and K (calibration procedures for these will be explained in section 2.6.1.1). The next step is to connect this PH expectation mechanism with the IELORP model.

In IELORP, the foregone utility $v_t^i(a|I^i(c^t) = \mu^i)$ is calculated for every $a_{j,t}^i$ with $j = \{1, \dots, J\}$. A similar calculation will apply with punishment, with the difference that v_t^i will be modified to include z_t^i, as implied in the following equation:

$$\begin{aligned} v^i(c^i, z_t^i|\mu^i) = c^i \left[(M-1) + \beta^i\left(M - \frac{1}{N}\right) - \gamma^{*i}\left(\frac{N-1}{N}\right)\right] \\ + (N-1)\mu^i\left[M + \beta^i\left(M - \tfrac{1}{N}\right) + \gamma^{*i}\right] + \\ w(1+\beta^i) - ez_t^i \end{aligned} \tag{6}$$

The reader will notice that equation (6) is almost identical to equation (2), with the only difference that it subtracts at the end the term ez_t^i. This indicates that the actions a_t^i that are expected to be punished will be penalized according to the calculated z_t^i multiplied by e. Two points are important to mention here: first, notice that in experiments that allow punishment the information revealed to subject i also includes punishment received in last rounds and individual contributions of other players. In this case, $I^i(c^t)$ would include

[44] Some authors (e.g. Anderson and Putterman (2006)) have referred to this effect as a demand for punishment that is decreasing on its price. We reserve from referring to it this way because a demand curve at the aggregate level does not imply that the law of demand holds at the individual level. Also, a demand interpretation is based on a rational choice model, different to the approach here.

more information besides μ^i. However, the model assumes that agents ignore such information[45]. Second, the model assumes that the agents don't have the computational capabilities to evaluate how the payoffs including punishment decisions would affect the social or inequality component of the utility function (equation (1)). One way to think about it is that players evaluate their payoffs as if there would be no punishment. Then, after that calculation is made, they take any expectation of punishment received as a personal cost, ignoring its social impact. This assumption implies that players care about the social outcomes when deciding about contributions, but not when evaluating received punishment [46]. Notice that from the payoff equation when punishment is included, $\pi^i = w - c^i + M\sum_{j=1}^{N} c^j - \sum_{j\neq i} p_{ij} - e\sum_{j\neq i} p_{ji}$, ez_t^i is equivalent to an expectation of $e\sum_{j\neq i} p_{ji}$ which enters the utility only in the "selfish" component of equation (1).

With a value function that includes their punishment expectations (equation (6)) and a way to calculate how much punishment each potential action would receive (equation (5)), given a reference point (equation (4)), PH is complete. It includes a way for agents to make counterfactual assumptions of received punishment which are integrated into the learning mechanism IEL, using the same computations (i.e. experimentation, replication and selection) for its reinforcement mechanism.

Before showing results, the next subsection addresses some points worth discussing on why this particular modelling approach for extending PH was chosen.

2.5.3 Discussion of modelling strategy

A first point that is important to address is how an exogenous expectation of punishment fits AL modelling strategy. Initial intuitions regarding its extension pointed towards allowing agents to update the foregone utility of the evaluated contributions based on observed punishment. In IELORP, v_t^i is evaluated, for each action a_t^i, based on observed past contributions (particularly on μ_t^i). That way, by observing last round's group contributions and assuming that other agents wouldn't change their strategy, it is possible to calculate exactly how much profits π^i each a_i^t would represent. Why not do the same with punishment? The reason is that in experiments without punishment, the payoffs are given directly by the experimenter and depend only on the known functional form of π^i (taking group contribution as a given). With punishment, such functional form doesn't allow a direct calculation of profits. Observing that action a_1^i has been punished at time t, says nothing about whether action a_2^i would be punished or not. The agent observes that one action is punished (or not punished), but extrapolating that information to other actions would require some additional belief on other players' punishment behaviour. PH endows agents with such belief. In IELORP, the

45 See for example Gigerenzer et al. (2011). In Gigerenzer's line of research, a fast and frugal heuristic is a rule of thumb that allows agents to make smart decisions by ignoring information. PH could be thought of in a similar line.

46 This is consistent with players increasing their contribution levels when only symbolic punishment points are allowed (Masclet et al., 2003). This could be interpreted as players taking punishment received only as a personal cost, since symbolic punishment doesn't affect social payoffs.

naïve expectation that other players play the same contributions as last round is enough to allow counterfactual evaluations. For the more complex environment with punishment, an additional (and perhaps still naïve) expectation of what would be punished is now implemented.

Let us also address two potential criticisms about PH and the approach presented above.

The first one is that intuitively, it seems unnatural to have an expectation of punishment that is independent from actual punishment. This means that agents still expect punishment even if none is being allocated at all (i.e. is not being modelled). To answer to this point, the experiments conducted by Fudenberg and Pathak (2010) are illuminating. Their design has subjects playing a traditional repeated public goods game with punishment, with the twist that it included treatments for not allowing players to observe punishment decisions by others until the end of all periods. That means that throughout the game players do not know if their punishment is affecting the behaviour of others or if they themselves are being punished. In the words of the authors: "Our experiment shows that subjects will engage in costly punishment even when it will not be observed until the end of the session, which supports the view that agents enjoy punishment. Moreover, players continue to cooperate when punishment is unobserved, perhaps because they (correctly) anticipate that shirkers will be punished: *Fear of punishment can be as effective at promoting contributions as punishment itself*" (pp. 78, italics added). This is consistent with PH: agents expect punishment without any requirement for observing it.[47]

But suppose now that punishment is indeed modelled. Even if there's still independence in the expectations and allocation of punishment, one could argue intuitively that players should learn from observing the punishment received. The response to this critique is intuitive as well. Such argument would definitely be true in more realistic time spans beyond what is allowed in the lab, where enough learning opportunities are given: with time, people will learn if their free-riding goes unpunished. However, the model assumes that in the time span of the examined public goods games, the player doesn't have a way to reach this result without prior assumptions. For example, imagine a player that contributes 10 tokens to the public good when the average was 11 in the last round. If the player is not punished this round, what inferences should be made? Would a contribution of 10 tokens never be punished? What would've happened if the contribution would've been 9 tokens? PH assumes that players answer these questions based on given beliefs, derived perhaps from the institutional framework of the experiment. At least in the time span of the experiments analysed, it is assumed that players stick to them.

Finally, a simpler point to motivate the modelling strategy without including actual punishment (only expectations of it), is that the model is simpler that way (an Occam's razor argument). If adding punishment

[47] Of course the claim is not that PH is a general overarching model of behaviour (assuming one actually exists). It is instead intended to represent particular rules of thumb under a specific environment. In this case, the environment is a repeated public goods game where punishment is regarded as "legitimate" according to culturally accepted norms (see for example Ertan et al. (2009)), which includes a real threat of punishment. An interpretation of PH could be that players expect punishment due to the belief on the institutions implemented (i.e. the experimental norms in the lab) instead of the observed punishment.

decisions, or allowing agents to use more information in their calculations, or allow them to take into account the social costs of punishment, can help to explain better the stylized facts, then perhaps the additional complexity might be worth it. If not, then until other stylized facts are intended to be explained by the researcher, the simplicity argument is relevant. Whether PH can successfully account for the stylized facts or not is answered below in section 2.6. For now, we examine how far we can go in explaining the data using this simple model.

To summarise, IELORP assumes behaviour that is adaptive but also includes a simple forward looking component through naïve expectations: the assumption that other agents in the group will maintain the same contributions observed last period. PH's core assumption is that agents, when punishment is included, add another (perhaps still naïve) expectation component related to the reference point. The counterfactual nature of IEL makes the inclusion of another naïve expectation strategy fit appropriately into AL's modelling strategy.

2.6 MODEL CALIBRATION AND MAIN RESULTS

This section presents the main results from PH. An important objective of this paper is to test AL's modelling approach by extending IELORP to include punishment. The test is not a "horse race" comparing different models, but rather an attempt to extend IELORP and check the robustness of its parameters along with the plausibility to include another sort of information into the model (i.e. heuristics). Would the extended model produce data that is quantitatively similar to that from lab experiments? Foreshadowing the results, they will show that the model can reproduce quite accurately the main punishment stylized facts described above, keeping the same parameter estimations used by AL (calibrating only the two new parameters included in PH). We believe this out-of-sample parameter stability is a strong robustness test for the modelling approach, and shows that it is flexible for researchers without needing to recalibrate in every data set. This section will first describe the calibration of the newly introduced PH's parameters, followed by the main results. We close this chapter discussing stylized facts of section 2.4.2 that are not addressed with the main data.

2.6.1 Parameters and calibration procedure

PH's free parameters are eight in total: IEL has as free parameters $\{J, \rho, \sigma\}$, the ORP distribution of types is determined by $\{P, B, G\}$ and finally the expectation of punishment has $\{L, K\}$. Under the belief that eight free parameters in a model can give too much degrees of freedom to the researcher, the robustness of IELORP parameters across experiments is a way to address this issue: its parameters have been tested by AL across domains and data sets *without* recalibration[48]. So here the approach followed is to test PH using exactly the same parameter numbers estimated by AL. Two goals are achieved by doing this: the amount of degrees of freedom is reduced to only two (PH's

[48] As mentioned in Appendix 2.8.1.

free parameters) and the robustness of those parameter values is tested with an extended model out-of-sample.[49]

Table 2.1 summarises the values used for the parameters of the model. The values of J, ρ, σ, P, B and G are taken exactly from the estimations of AL. The values of PH were calculated as explained next. When using this set of parameter values, the model will be referred to as PH*.

IEL	**ORP**	**PH**
$J = 100$	$P = 0.48$	$L = 3.3$
$\rho = 0.033$	$B = 22$	$K = 14$
$\sigma = w/10 = 2$	$G = 8$	

Table 2.1: Parameter values for PH. IEL and ORP values taken from Arifovic and Ledyard (2012). *PH values estimated*

2.6.1.1 Estimation of PH parameters (L and K)

In PH, the parameter L is a measure of how sensitive is the expectation of tolerance (T) with respect to changes in the effectiveness of punishment, e. K represents the amount of punishment players expect for each point their own contribution is lower than the group's average. For example, in a group of N=4, $K = 6$ represents that an agent expects an average of 2 punishment points from each player if its own contribution is one point below his reference point R_t^i. K and L are estimated by generating simulated experiments with PH, keeping the values for IEL and ORP parameters as given by Table 2.1. Unless specified differently, all simulations where conducted with such corresponding values.

The estimation was conducted using the data of Nikiforakis and Normann (2008) (NN from now on)[50]. In their experiments, the main treatment is the variation of e. Each treatment, under a partners setting, takes values of $e = \{0,1,2,3,4\}$. Each one of the five treatments is named according to the value of e: for example, when $e = 0$, (i.e. no punishment), the treatment is called Treatment 0. NN experiments kept the values of N=4 and M=0.4 constant across all treatments, with $w = 20$. Such values are the same used in all the simulations reported here unless specified differently.

Each run (or trial) simulated N agents for ten periods by drawing β^i and γ^i according to the distribution and parameter values explained in section 2.2.2.

---

[49] The learning parameters have been tested in different environments, so they seem quite robust according to AL. For the distribution of types (ORP parameters), AL try both a single estimation as well as a recalibration when using different data sets (they claim that a lot of data is necessary for having one single distribution of types across games). However, their conclusion is that the recalibration keeps almost the same results, which is why we test the PH extension while keeping those same parameter values.

[50] As discussed in section 2.4.2, other dataset that captures the stylized fact that cooperation changes with effectiveness is that of Egas and Riedl (2008), which uses an stranger setup (contrary to NN, which used partners). We chose the experiments of NN mainly for two reasons: first, although AL tested IELORP under both partners and strangers setup, they mention that further research is still required to conclude about the model under strangers. Second, NN alters effectiveness levels while always keeping constant the cost of one punishing token (equal to one). On the contrary, Egas and Riedl (2008) change both cost as well as impact, which can introduce framing effects not intended to be captured by our adaptive agents.

A grid search was conducted for the duplet (K, L) selecting the values with better fit to NN data[51]. An initial wide search was conducted. K was given values from 0 to 15 in steps of one, and L was given values in the interval from 1 to 5 in steps of 1. For each treatment 1, 2, 3 and 4 and each combination of (K, L), 100 trials were conducted. Then a second narrow search was conducted with K taking values from 12 to 15 in steps of one, and L with values from 2 to 4 in steps of 0.1. For the latter search, 100 trials were also run for each parameter combination. The best fit was chosen according to a standard approach as follows.

As in section 2.3.2, the SE (Squared Error) was estimated, this time for each treatment and parameter combination of K and L. Using similar notation, let $\bar{c}_{PH}^{10}(r)$ denote the average contribution for all simulated agents with PH* across all ten periods on treatment r for a particular combination of (K, L). Let $\bar{c}_{PH}^{3}(r)$ be the analogous for the average of the last three periods, and $\bar{c}_{NN}^{10}(r)$ and $\bar{c}_{NN}^{3}(r)$ be the same but for NN experimental data (across the six group observations for each treatment in their experiment). The squared deviations between the simulated data and the experimental data were computed with the objective of finding the minimum of their sum. That way, MSE, the Minimum Squared Deviation was calculated as[52]

$$MSE = Min \sum_{r=1}^{R} [\bar{c}_{NN}^{10}(r) - \bar{c}_{PH}^{10}(r)]^2 + [\bar{c}_{NN}^{3}(r) - \bar{c}_{PH}^{3}(r)]^2 \qquad (7)$$

where R is the total number of treatments. For the present case of NN, $R = 4$ (treatments 1,2,3 and 4)[53]. As before, the error is normalized, so that the reported value of the NMSE (normalized mean squared error) is

$$NMSE = \sqrt{\frac{MSE}{2R}}$$

The lowest value of NMSE was generated by the parameter values shown in Table 2.1. Interpretations of the values for L and K are as follows. A value of L=3.3, given equation (4), implies that with an effectiveness of 1, the tolerance level is about 6 tokens, so agents expect that only contributions more than 6 tokens below last round's average would be punished. For effectiveness equal to 2 and 3 the tolerance is, respectively, about 2 and 0.5, showing that the higher the effectiveness, agents expect to be punished more easily. On the other hand, K=14 represents that even contributions one point below the reference point R_t^i are expected to be highly punished.

2.6.2 Simulations and statistical tests

[51] This procedure follows AL calibration for IELORP.

[52] This is almost the same as equation (3). However, besides being useful to remind the reader of the procedure, here it specifies that this time we focus on finding the minimum SE in order to select the parameter values.

[53] Since the results of the model under Treatment 0 (i.e. no punishment) are not affected by the parameters L and K, it was not included in the calibration.

2.6.2.1 Main results

The fit of PH* to NN data can be grasped quickly with the value of NMSE=0.840. Since $\bar{c}_{PH}^{10}(r)$ and the other averages for its calculation belong to the interval [0,w], they are the average contribution for an endowment of $w = 20$. That way, it corresponds to less than a 5% error in the fit across all the four treatments. The model generates data that is on average very close to the experimental dataset on the lab (an error of 4.2%).

But to observe closer the dynamics and patterns of the data and have a better idea of how good the performance of the model is, Figure 2.7 shows both PH* simulations and NN data across the ten periods of the experiments for different levels of effectiveness. Each point of the simulated data is the average group contribution across 100 trials (analogous to 100 groups or 100 observations for each treatment) for each period. NN data consists of six observations per period, each one a different group. Due to having fewer observations for NN data, the simulations present a "smoother" pattern than the experimental data.

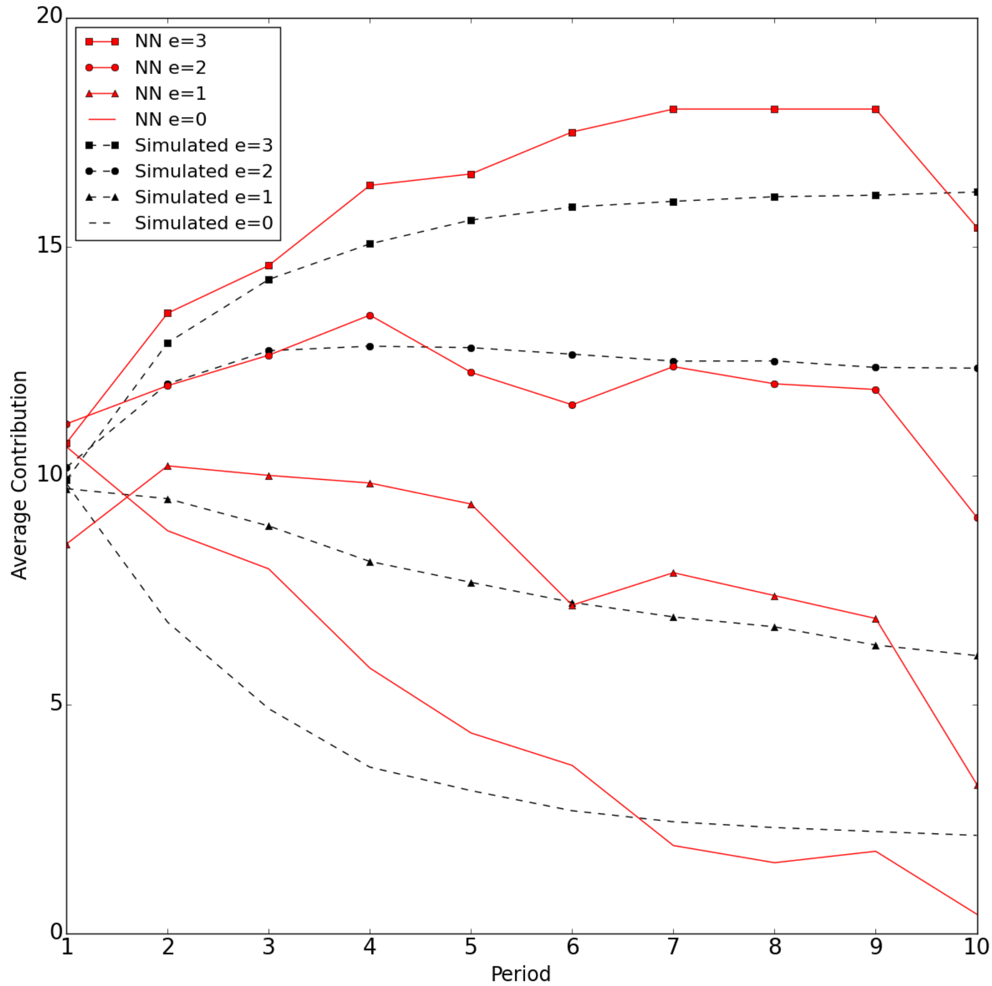


Figure 2.7: Experimental vs. simulated data for treatments e=0, e=1, e=2 and e=3. Simulations generated with PH (blacked dashed lines). Experimental data source is* Nikiforakis and Normann (2008) *(NN, red lines).*

Figure 2.7 shows how close each treatment is replicated by the model[54]. The simulations capture closely both the levels of contribution as well as its dynamics, and the figure can easily be related to the stylized facts in section 2.4.2. The first fact, that punishment can sustain cooperation (or put differently, that it can prevent its decline), can be observed for effectiveness equal to two (e=2) and higher, since the trend for such treatments is not negative. The second fact, that such contributions can be sustained only for high enough levels of punishment, is also easily observed and constitutes the main empirical focus of NN. As in their experimental data, in PH average contributions decrease over time with e=0 and e=1, they are constant with e=2 and increase with e=3. In the model, this effect comes from the higher tolerance levels associated with lower values of the effectiveness parameter. The third fact is that first period contributions are not statistically different across treatments (i.e. effectiveness levels). Figure 2.7 shows that average contributions (for both simulated and experimental data) increase monotonically in the effectiveness of punishment in every period, except for the first. Graphically this is observed by noticing that all first period data is clustered close to ten (half the endowment). In the model this is explained by the fact that in the first period agents don't have experience of a previous average contribution, hence they can't have any reference point. This leaves agents expecting the same punishment (or lack thereof) for each potential contribution until a reference point is formed in the second period. Analysis of the fourth and final stylized fact regarding welfare, is addressed below in section 2.6.3.

2.6.2.2 Statistical tests

To further test the fit of the model, two-sample Kolmogorov Smirnov (KS) tests were conducted. 1,000 runs of PH* were simulated for each treatment, calculating for each run the values of $\bar{c}_{PH}^{10}(r)$ and $\bar{c}_{PH}^{3}(r)$. Two tests were conducted for each treatment, one for the average contribution of all periods and the other for the average of the last three. These results are reported in Figure 2.8. As can be seen almost none of the tests can reject the null hypothesis that the simulated and experimental data come from the same distributions. This is further evidence of the good fit of the model to NN data.

	Combined Kolmogorov Smirnov test (two-sample). Corrected p-value				
Data used	Treatment 0	Treatment 1	Treatment 2	Treatment 3	Treatment 4
All periods	0.669	0.547	0.180	0.383	0.010***
Last 3 periods	0.291	0.320	0.132	0.126	0.200

*Figure 2.8: Kolmogorov-Smirnov tests, using average contributions for all ten periods and for the last three. Reported is the corrected p-value of the two sample test, under the null hypothesis that both the simulated and experimental data come from the same distribution. *** for significance at 1% level.*

[54] Reference to treatment e=4 (not included in the figure) is done below.

Is worth noting that Treatment 4, for the average of all ten periods, is the only one for which the KS rejects the null hypothesis of the test at more than 5% (or 10%) significance level. What is happening in the data? Observing Figure 2.9 can help to understand better what the model can replicate (and where it has a limitation). Comparing the simulated data for Treatments 3 and 4 shows that qualitative patterns from NN are still captured by the model. Besides the random initialization (i.e. first round behaviour), which is in line with the empirical facts, the figure shows that the model replicates higher contribution levels, in the final periods, for e=4 compared to e=3. This can be observed by comparing both simulated data of PH* in the figure. However, one can notice that for e=4 the experimental data is higher than the simulated in every period (except the last one). In PH*, the main variable determining the contribution levels is the reference point R_t^i , which depends on the tolerance level (T). If the effectiveness level is high enough, the value of T will be lower than one, which happens with $e = 3$. Higher effectiveness still lowers the value of T, but since the minimum contribution is one token[55], with $e \geq 4$ the effect of T on which actions are penalized (in terms of foregone utility) becomes imperceptible. Hence, with $e = 4$, contribution levels slightly increase due to the effect of e in the value function (i.e. by multiplying the amount z_t^i in equation (6)) but not because the reference point is changing. That is why the model has difficulty replicating the higher contribution levels in the data, reflected in the p-value of the KS test.

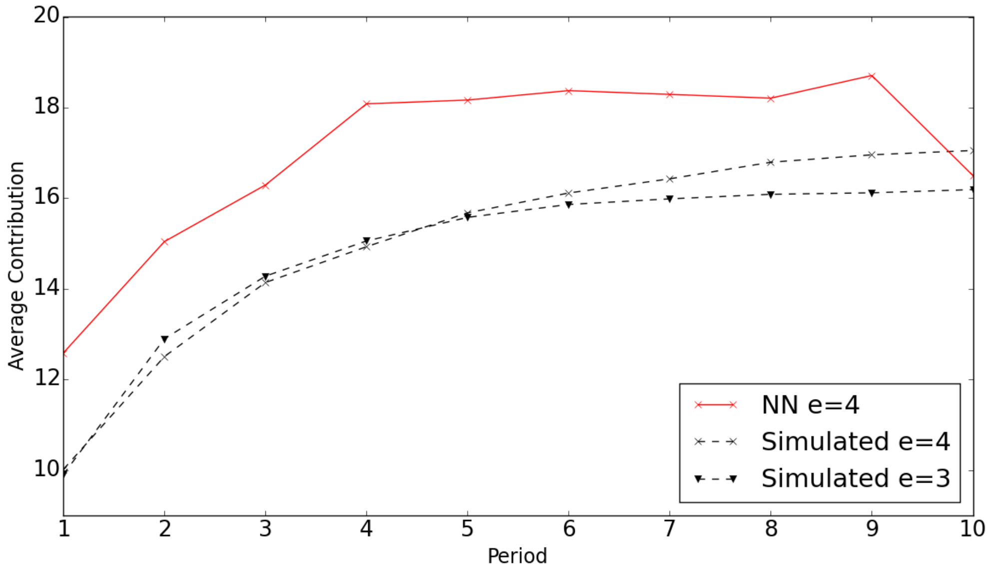


Figure 2.9: Experimental vs. simulated data for Treatment e=3 (simulations) and e=4 (simulations and experimental data). Simulations are generated with PH (dashed, black lines). Experimental data source is* Nikiforakis and Normann (2008) *(NN, red line).*

Given the above, another calibration for only treatments 1,2 and 3 (excluding 4) was also conducted as a robustness check. The obtained average fit was virtually identical (NMSE=0.844, compared to 0.840), but more

[55] This is to replicate the fact that most experiments allow only discrete changes in contribution (e.g. one token). However, technically in PH the experimentation component of the learning mechanism can introduce contributions that are not integers.

importantly, the parameter values remained exactly the same as those reported in Table 2.1 (K=14 and L=3.3).

To summarise, the main results presented show that PH* generates data that has an average error of less than 5% compared to NN experimental data. Graphically one can observe the fit for the different levels of effectiveness: the model closely tracks both the dynamics and the convergence levels of average contributions observed in the data. Such patterns are confirmed for all treatments (except for e=4) by the corresponding statistical tests. This fit was obtained *without* recalibrating the six original parameters of IELORP. Such values were taken directly from previous calibrations in the literature, conducted for different datasets that did not include punishment. The parameters that were estimated here were the ones included in our extension to allow expectations of punishment. The close fit of the model as well as the out of sample robustness of the previously estimated parameters, are evidence of both the good performance of the model as well as of the flexibility of AL's modelling approach.

2.6.3 Further analysis of stylized facts

As seen above (Figure 2.7), NN data captures our main stylized facts related to contribution levels, and they are closely replicated by the model: cooperation can be sustained by punishment, but only with high enough effectiveness levels. And the higher the effectiveness, the higher the contributions levels in all periods, except for the first one. However, there is one fact that has not been analyzed so far: that eventually, given enough time, punishment will increase group welfare (our fourth stylized fact). Is PH consistent with this fact?

Even without comparing directly simulated net earnings with experimental data, one can realize how the model actually does account for this fact. The learning algorithm will eventually find those potential contributions not expected to be punished, since they represent higher payoffs. Such contributions will replicate so that the set of potential actions of each agent becomes homogenous. With time, agents will not choose actions expected to be punished, except for random mutations. So under any effectiveness level that sustains cooperation, given that agents will learn to avoid expected punishment, the social benefits of cooperation will unambiguously outweigh the costs of punishment.

To illustrate better this decline in potential punishment, we ask the following question: "how many players would be punished every period, if every agent punishes others based on the same rule on which they expect to be punished?" That is, if agents punish any contributions below the last round's group average minus the tolerance level T. We ran additional simulations including such a simple punishment behavior. To keep it simple, we included punishment just as a binary decision of either to punish or not. To notice that punishment will go to zero, is not necessary to worry about how much punishment is given[56].

[56] Earlier versions of the model included more complicated punishment rules, calibrating parameters of when and how much to punish based on the experimental data. However, the final model didn't include those (keeping only fear of punishment) because they required more free

Figure 2.10 presents the average number of players that would receive (any) punishment under the above rule, across 100 simulations of PH* for each treatment. As can be observed, eventually in all treatments the number of punished agents converges virtually to zero. The small positive number of punished agents is due to some of them randomly trying new alternatives (mutation). As expected, the higher the effectiveness level, the longer it takes for agents to avoid punishment completely. Intuitively, this is due to agents being 'more tolerant' towards punishment (higher T) when effectiveness levels are lower. Technically, this is because a lower T makes the set of potential actions not expected to be punished smaller, hence making the learning algorithm to take longer to find them. So the model is consistent with the main conclusions of Gächter et al. (2008): once cooperation is established (through expectations of punishment, not punishment itself in the model), punishment is rarely needed, so with time its costs become negligible.

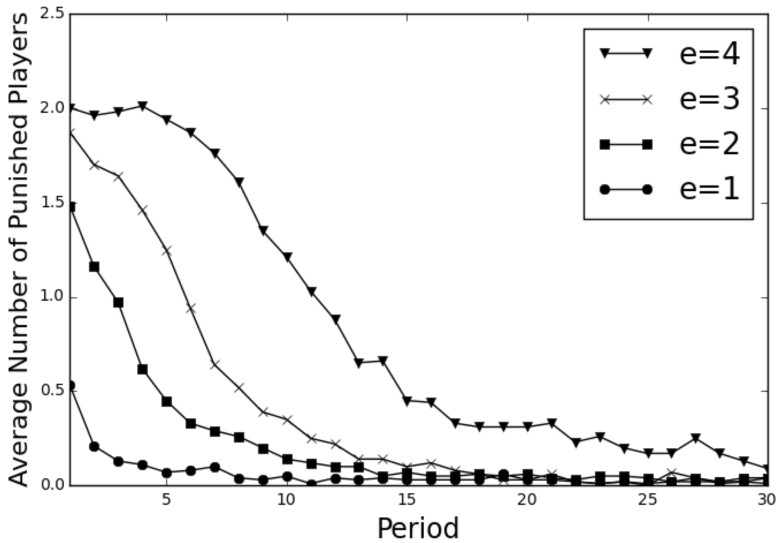


Figure 2.10: Average number of agents punished for different effectiveness levels (across 100 simulations per treatment). Punishment is modeled: an agent punishes any other in the group that contributes below the group average minus the tolerance level T.

2.7 Closing Remarks

Experimental research has spanned a wide range of literature on public goods games containing many facts on human behaviour that are at odds with traditional game theoretical approaches. Recent models with boundedly rational agents have emerged trying to close this gap, but few can claim to have done it thoroughly. The model developed by Arifovic and Ledyard (2012) in the context of voluntary contribution mechanisms, IELORP, has been claimed to predict many of the main stylized facts in this literature, with remarkable transferability to other experiments and out-of-sample robustness of its parameter values. This makes IELORP a contender in closing that gap.

parameters and complicated the model without actually adding much to the explanation of the stylized facts.

However, the model wasn't designed to explain one of the most important facts on public good games: that punishment mechanisms can sustain cooperation and prevent the tragedy of the commons. We have presented here an extension to IELORP that relies on the expectations of punishment, not punishment itself, as a way to sustain cooperation in public good games. This extension can be seen as contributing to the literature in two significant ways: first, it is considered a test to IELORP's modelling approach. This is not a traditional test in the sense of comparing it with different models, but as a mean to answer the questions, Is the model flexible for researchers to be extended, maintaining its core components of learning and other regarding preferences? Would the extension retain the same parameter calibrations used in the literature before? The answers to these questions are positive. Results presented here show that while keeping exactly the same parameter values (calibrating only the new ones included), the model can replicate four main stylized facts on the experimental punishment literature, producing data quantitatively similar to that of human subjects in the lab.

Second, the model presents an interesting modelling approach on its own: it combines core behavioural principles, such as learning, with ad-hoc rules of thumb that are tailored specifically to the environment under study. This is presented under the belief that in the toolkit of the social scientist, all approaches can be used as long as they give useful insights.

Is the model presented here useful? One way to think of whether a model is useful of not, is to ask whether it presents new questions that could be explored empirically[57]. In this regard, one clear topic addressed by the model is the learning time spans of subjects. Empirical evidence has shown that even unobserved punishment can deter free-riding. In the model, agents have a given expectation of punishment that can't change, assuming a short time span similar to that of short lab experiments. But how long would it take subjects in the lab to learn? Once cooperation has been stablished, how long would it take players to adapt to new institutional frameworks, such as removing punishment (without being explicitly told about the change)? In PH, even after removing the punishment expectation, once cooperation is sustained, contributions remain high for several periods: the strategy set of agents is populated by the equilibrium strategy, and is not until experimentation takes place that new strategies can be tried. This can take several rounds depending on the model parameters. Would human players sustain cooperation for long, or would they quickly revert to free-riding? What factors (e.g. social norms) could affect this behaviour? These are considered empirical questions that the model hints to be explored as future research.

Finally, there are theoretically relevant approaches that come to mind after working closely with PH. The model highlights both a learning mechanism as well as the use of (exogenously given) rules of thumb in order to explain the patterns in the data. An interesting question is how these rules of thumb come to be used, or which others can be learned adaptively. Making the process of trying and developing new rules of thumb explicit, is a modelling approach that would make such adaptation process endogenous. Work on such mechanisms of inductive learning points in that direction (e.g. Holland et al., 1986). Some implementations of similar approaches in economic environments have been done in the literature: see for example Kirman (2010) for examples in financial and fish markets, Miller (1988) and Zhang (2015) for prisoner's dilemma, or

[57] Here again, thanks to Simon Gächter for pointing this out.

Lee-Penagos (2016) for coordination games. Allowing agents to learn and adapt different rules of behaviour is suggested as a next step to understand better how agents adapt their behaviour in a complex environment as a public goods game with punishment in the lab.

2.8 APPENDIX

2.8.1 Additional features of IELORP in public good games

Besides the main results in the main text, AL tested the model in different datasets and treatments from different experimenters in public goods games without punishment (not formally explored in our replication). Their overall conclusion is that IELORP fits the data very well. The following are other environments (besides IW) in which AL claim that the model has been successful at predicting out of sample[58].

1. **Partners vs strangers:** to explore differences between these two setups, AL tested IELORP in an strangers setting with the data of Andreoni (1995) (remember that IW's used a partners setup). Andreoni also had different experimental parameter values (N=5 and M=0.5). Without recalibrating any of the parameters, the model generated data that differed on average with the experimental one in 4.9%. As a caveat, AL present a discussion on IELORP's explanation for the difference between partners and strangers setups. Is worth mentioning that although their model fitted the data accurately, more data is required to conclude strongly about the partners vs. strangers explanation given by the model.
2. **Rank based payoffs:** Andreoni (1995) also presented treatments where subjects are not paid according to their profits, but on how their profits ranked compared to the rest of the group. By modifying the value function v^i accordingly and again keeping the same parameter values, IELORP differed on average on 4.1% with the experimental data.
3. **Experience vs inexperienced:** the treatments mentioned already, which varied the values of N and M in the data of Isaac and Walker (1988), provided previous experience to the players (i.e. played some practice rounds). However, some groups didn't receive such experience. Sessions with M=0.3 and N=4 were compared with IELORP's generated data. Average difference was 6.6% without recalibration of any parameter. AL conclude that experience of the subjects is not something that needs to be controlled for in the model.
4. **Restart effect:** Introduced by Andreoni (1988). The effect consists in that after subjects finish the initially announced periods of the experiment, they are informed that they will play additional ones. After the announcement, contributions to the public good raise and start declining again. IELORP captures this effects by randomly populating again the set of available strategies A_t^i (agents "rethink" the problem). Croson (1996) replicated Andreoni's experiments. AL used both data sets to test their explanation of the restart effect. Although more data is required to confirm it (due to small sample

[58] We mention "other" environments because even if the whole IELORP model was firstly implemented for the Isaac and Walker (1988) data, the learning component IEL had been designed and tested before for different experiments, but not jointly with the ORP component. To that degree, the IEL behavioural model was tested out of sample with IELORP.

size) IELORP presents similarities with the data that don't discard it as a potential explanation.

Two additional points are worth mentioning regarding the stability of the parameters and their transferability to other domains.

First, robustness of the above results to changes in the parameter values is tested by AL. Their conclusion is that the model doesn't require re-calibration when transferring it to different experiments and conditions for public goods games without punishment. Also, when re-calibration was indeed conducted (for the ORP parameters) there was only a marginal benefit in the fit to the data, a strong point in favor of the model robustness to parameter changes. For AL's main results several ranges for the parameters of both IEL and ORP were tested. Their conclusion is that all of the model parameters are robust and changes within "reasonable ranges" affect very little the model's performance.

Second, and perhaps more interesting, is that the learning model (IEL) was initially designed for other kind of repeated games. It was implemented first to study Groves-Ledyard mechanisms for public good allocations (Arifovic and Ledyard, 2011, 2004) as well as for call markets (Arifovic and Ledyard, 2007). Remarkably, AL claim that the IEL model not only has replicated data across such domains accurately, but that it has done so using exactly the same parameter values (the triplet (J, ρ, σ)). The fact that IEL kept those same values when extended with ORP is a strong test of the model transferability, and a motivation to use it and test it further with our implementation of punishment.

Finally, is worth referring the reader to AL's final discussion on the model's shortcomings. An example of those is not including reputation concerns, which makes the model not well suited to strategic coordination games that require more sophistication. The latter, for example, would require agents that can learn strategies beyond one single period of history.

2.8.2 Conditions for each equilibrium behavior

In IELORP, in equilibrium each agent will have one of three equilibrium behaviors: free riding ($c^i = 0$), fully contributing ($c^i = w$) or conditional cooperation ($c^i = \bar{c}$). On which strategy an agent converges will depend both on its ORP parameters (β^i *and* γ^i) and the experimental parameters (groups size (N) and public good marginal productivity (M)). Hence, altruism or conditional cooperation are behaviors that arise from other regarding preferences, but only when the environment provides the setting for it. The conditions, as presented by AL, are as follows:

$$c^i = \begin{Bmatrix} 0 \\ \bar{c} \\ w \end{Bmatrix} if \begin{Bmatrix} 0 \geq \left[\left(M - \frac{1}{N}\right)\beta^i + M - 1\right] \\ \gamma^i\left(\frac{N-1}{N}\right) \geq \left[\left(M - \frac{1}{N}\right)\beta^i + M - 1\right] \geq 0 \\ \gamma^i\left(\frac{N-1}{N}\right) \leq \left[\left(M - \frac{1}{N}\right)\beta^i + M - 1\right] \end{Bmatrix}$$

REFERENCES

ANDERSON, C.M., PUTTERMAN, L., (2006). "Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism." *Games Econ. Behav.*, Vol. 54, pp. 1–24.

ANDERSON, S.P., GOEREE, J.K., HOLT, C.A., (1998). "A theoretical analysis of altruism and decision error in public goods games." *J. Public Econ.*, Vol. 70, pp. 297–323.

ANDERSON, S.P., GOEREE, J.K., HOLT, C.A., (2004). "Noisy Directional Learning and the Logit Equilibrium." *Scand. J. Econ.*, Vol. 106, pp. 581–602.

ANDREONI, J., (1988). "Why free ride?: Strategies and learning in public goods experiments." *J. Public Econ.*, Vol. 37, pp. 291–304.

ANDREONI, J., (1995). "Cooperation in Public-Goods Experiments: Kindness or Confusion?" *Am. Econ. Rev.*, Vol. 85, pp. 891–904.

ANDREONI, J., MILLER, J.H., (1995). "Auctions with Artificial Adaptive Agents." *Games Econ. Behav.*, Vol. 10, pp. 39–64.

ARIFOVIC, J., LEDYARD, J., (2004). "Scaling Up Learning Models in Public Good Games." *J. Public Econ. Theory*, Vol. 6, pp. 203–238.

ARIFOVIC, J., LEDYARD, J., (2007). "Call market book information and efficiency." *J. Econ. Dyn. Control*, Vol. 31, pp. 1971–2000.

ARIFOVIC, J., LEDYARD, J., (2011). "A behavioral model for mechanism design: Individual evolutionary learning." *J. Econ. Behav. Organ.*, Vol. 78, pp. 374–395.

ARIFOVIC, J., LEDYARD, J., (2012). "Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism." *J. Public Econ.*, Vol. 96, pp. 808–823.

BOLTON, G.E., OCKENFELS, A., (2000). "ERC: A Theory of Equity, Reciprocity, and Competition." *Am. Econ. Rev.*, Vol. 90, pp. 166–193.

BOWLES, S., GINTIS, H., (2013). A Cooperative Species. Princeton University Press.

BRANDTS, J., SCHRAM, A., (1996). "Cooperation and Noise in Public Good Experiments: Applying the Contribution Function Approach." *Work. Pap. CREED*,.

CAMERER, C., HUA HO, T., (1999). "Experience-weighted Attraction Learning in Normal Form Games." *Econometrica*, Vol. 67, pp. 827–874.

CHARNESS, G., RABIN, M., (2002). "Understanding Social Preferences with Simple Tests." *Q. J. Econ.*, Vol. 117, pp. 817–869.

CHAUDHURI, A., (2011). "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature." *Exp. Econ.*, Vol. 14, pp. 47–83.

COOPER, D.J., STOCKMAN, C.K., (2002). "Fairness and learning: an experimental examination." *Games Econ. Behav.*, Vol. 41, pp. 26–45.

CROSON, R.T.A., (1996). "Partners and strangers revisited." *Econ. Lett.*, Vol. 53, pp. 25–32.

EGAS, M., RIEDL, A., (2008). "The economics of altruistic punishment and the maintenance of cooperation." *Proc. R. Soc.*, Vol. 275, pp. 871–878.

FEHR, E., FISCHBACHER, U., (2002). "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives." *Econ. J.*, Vol. 112, pp. C1–C33.

FEHR, E., GÄCHTER, S., (2000). "Cooperation and Punishment in Public Goods Experiments." *Am. Econ. Rev.*, Vol. 90, pp. 980–994.

FEHR, E., GÄCHTER, S., (2002). "Altruistic punishment in humans." *Nature*, Vol. 415, pp. 137–140.

FEHR, E., SCHMIDT, K.M., (1999). "A Theory of Fairness, Competition, and Cooperation." *Q. J. Econ.*, Vol. 114, pp. 817–868.

FISCHBACHER, U., GÄCHTER, S., (2010). "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *Am. Econ. Rev.*, Vol. 100, pp. 541–556.

FUDENBERG, D., PATHAK, P.A., (2010). "Unobserved punishment supports cooperation." *J. Public Econ.*, Vol. 94, pp. 78–86.

GÄCHTER, S., RENNER, E., SEFTON, M., (2008). "The Long-Run Benefits of Punishment." *Science (80-.).*, Vol. 322, pp. 1510.

GIGERENZER, G., HERTWIG, R., PACHUR, T., (2011). Heuristics: The foundations of adaptive behavior. Oxford University Press, New York.

GIGERENZER, G., TODD, P., GROUP, A.B.C.R., (2002). Simple Heuristics That Make Us Smart. Oxford University Press.

GOEREE, J.K., HOLT, C.A., LAURY, S.K., (2002). "Private costs and public benefits: unraveling the effects of altruism and noisy behavior." *J. Public Econ.*, Vol. 83, pp. 255–276.

GUALA, F., (2012). "Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate." *Behav Brain Sci*, Vol. 35, pp. 1–15.

HERRMANN, B., THÖNI, C., GÄCHTER, S., (2008). "Antisocial Punishment Across Societies." *Science (80-.).*, Vol. 319, pp. 1362–1367.

HETZER, M., SORNETTE, D., (2013). "An Evolutionary Model of Cooperation, Fairness and Altruistic Punishment in Public Good Games." *PLoS One*, Vol. 8, pp. e77041.

HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E., THAGARD, P.R., (1986). Induction: processes of inference, learning, and discovery. MIT press, Cambridge, MA.

HOLT, C.A., LAURY, S.K., (2008). "Chapter 90 Theoretical Explanations of Treatment Effects in Voluntary Contributions Experiments," in: Charles, R.P., Vernon, L.S. (Eds.), Handbook of Experimental Economics Results. Elsevier, pp. 846–855.

ISAAC, R.M., WALKER, J.M., (1988). "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." *Q. J. Econ.*, Vol.

103, pp. 179–199.

ISAAC, R.M., WALKER, J.M., WILLIAMS, A.W., (1994). “Group size and the voluntary provision of public goods.” *J. Public Econ.*, Vol. 54, pp. 1–36.

JANSSEN, M.A., AHN, T.K., (2006). “Learning, Signaling, and Social Preferences in Public-Good Games.” *Ecol. Soc.*, Vol. 11.

LEE-PENAGOS, A., (2016). “Learning to Coordinate: Co-Evolution and Correlated Equilibrium.” *CEDEX Discuss. Pap. No. 2016-11*,.

MILLER, J.H., (1988). “The Evolution of Automata in the Repeated Prisoner’s Dilemma.” *Two Essays Econ. Imperfect Information, Ph.D.University Michigan*,.

MILLER, J.H., ANDREONI, J., (1991). “Can evolutionary dynamics explain free riding in experiments?” *Econ. Lett.*, Vol. 36, pp. 9–15.

NIKIFORAKIS, N., NORMANN, H.-T., (2008). “A comparative statics analysis of punishment in public-good experiments.” *Exp. Econ.*, Vol. 11, pp. 358–369.

PAHL-WOSTL, C., EBENHÖH, E., (2004). “An adaptive toolbox model: a pluralistic modelling approach for human behaviour based on observation.” *J. Artif. Soc. Soc. Simul.*, Vol. 7.

ROTH, A.E., EREV, I., (1995). “Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term.” *Games Econ. Behav.*, Vol. 8, pp. 164–212.

WILENSKY, U., (1999). “Netlogo. http://ccl.northwestern.edu/netlogo/.” *Cent. Connect. Learn. Comput. Model. Northwest. Universtiy, Evanston*,.

WILENSKY, U., RAND, W., (2007). “Making Models Match: Replicating an Agent-Based Model.” *J. Artif. Soc. Soc. Simul.*, Vol. 10, pp. 2.

ZHANG, W., (2015). “Can Errors Make People More Cooperative? Cooperation in an Uncertain World.” *Work. Pap. Purdue Univ.*,.

CHAPTER 3 .

LEE-PENAGOS, ALEJANDRO[⊥Ψ]

LEARNING TO COORDINATE: CO-EVOLUTION AND CORRELATED EQUILIBRIUM

ABSTRACT

In a coordination game such as the Battle of the Sexes, agents can condition their plays on external signals that can, in theory, lead to a Correlated Equilibrium that can improve the overall payoffs of the agents. Here we explore whether boundedly rational, adaptive agents can learn to coordinate in such an environment. We find that such agents are able to coordinate, often in complex ways, even without an external signal. Furthermore, when a signal is present, Correlated Equilibrium are rare. Thus, even in a world of simple learning agents, coordination behavior can take on some surprising forms.

KEY WORDS: Battle of the Sexes, Correlated Equilibrium, Evolutionary Game Theory, Learning Algorithms, Complex Adaptive Systems, Coordination Games.

JEL CLASSIFICATION: C63, C73, C72, D83

[⊥] School of Economics, University of Nottingham. Email: alejandro.lee@nottingham.ac.uk

[Ψ] Acknowledgements: Special thanks to John H. Miller for insightful discussions and comments, as well as for sharing some of his own computational routines. To Chris Starmer and Alex Possajennikov for comments on this and previous versions of the paper. Also, to participants of the Graduate Workshop in Computational Social Sciences (2015) at the Santa Fe Institute, where this project started. Financial support from the ESRC funded Network for Integrated Behavioural Sciences (NIBS) is acknowledged (ES/K002201/1).

3.1 INTRODUCTION

"If there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium"

-Roger Myerson, winner of the Nobel Memorial Prize in Economic Sciences[59]

Aumann (1974) introduced the concept of Correlated Equilibrium (CE), which is a generalization of the traditional Nash Equilibrium (NE). Under a mixed strategy interpretation of Nash, players randomize their strategies independently of each other. In a Correlated Equilibrium such independence is not necessary: players have probability distributions based on an exogenous signal or randomization device whose distribution is common knowledge. Players map their decisions from the outcomes of such a signal to their potential actions, making their actions correlated with each other. Mutual best responses to the belief that the other players will condition their actions based on the signal is considered a correlated equilibrium.

Notice that the signal (or exogenous randomization device) has no direct influence on the payoff matrix of the game, but it can nonetheless affect the equilibrium payoffs of the players. This is not possible under NE. The definition of CE allows solutions where the signal can both affect or not the behaviour of agents. This makes it a more general concept that also includes NE, where the signals can play no role whatsoever. Perhaps this is why Myerson believes that aliens would have probably learned first to play the CE[60]. However, the presence of the external signal and its effect on equilibrium convergence is puzzling. It requires players to be endowed with incredible computational powers and to know the other players' payoffs. Players also have to know the signal's distribution and a specific mapping from signal to actions in order to interpret it as a recommendation of what to play. From a normative point of view, such assumptions might be adequate. But from a positive or descriptive one, it is not clear how (or if) players could actually learn this information under less straining rationality assumptions.

This paper's objective is to explore, under a canonical coordination game (Battle of the Sexes), the effects of an exogenous signal on equilibrium selection when perfect rationality assumptions are relaxed. It focuses on the behaviour of learning, adaptive, boundedly-rational agents, with an emphasis on understanding how they use the signal in order to coordinate. It takes Myerson's idea about the discovery of CE to be easier than NE as a hypothesis to be tested. Can boundedly-rational agents learn to use exogenous signals to coordinate? If so, how could this happen? Will such agents learn to condition their behaviour on the signal as implied under a CE solution, or do they converge to a different equilibrium? These are the key questions explored here.

To tackle this objective, we develop a computational model with artificial adaptive agents playing a repeated Battle of the Sexes game. Analyses are made

[59] Leyton-Brown and Shoham (2008) (p. 24) or Solan and Vohra (2002) (p. 92). Interestingly, this famous quote is often attributed to Myerson, but we couldn't find the direct source.

[60] In his quote, the "discovery" of the correlated equilibrium by the extra-terrestrial "players" is interpreted as them playing it in real life (i.e. to condition their actions on the exogenous signal), versus having their game theorists understand and describe the concept.

via Monte Carlo simulations. The model represents each agent as a strategy that observes inputs from the environment (such as a rival's action or an exogenous signal) and based on those observations, the agent outputs as an action in the game. We use 'finite automata', which is a mathematical model of discrete inputs and outputs that can represent boundedly rational behaviour. Such agents are allowed to adapt and change their behaviour via a learning algorithm (a 'Genetic Algorithm'). The latter simulates social learning at the population level by implementing selection and mutation processes that tend to reinforce better performing strategies and to eliminate poorer performing ones. This constitutes an evolutionary approach that explores what types of strategies emerge in the long-run.

In order to explore the impact of the exogenous signal in coordinating behaviour, computational experiments are conducted for two treatments: a baseline *No-Signal* model of the traditional game (without signal), along with a main *Signal* treatment. In the latter, agents are allowed to observe and potentially use an exogenous randomization device to coordinate.

This methodology presents several advantages for answering the above questions. First, given the interest of modelling bounded rationality, finite automata allow the representation of agents with limited memory and processing power. While they can observe the behaviour of the other agents they interact with as well as the exogenous signal, they don't have access to others' payoffs or the distribution of the signal. Hence they can only react to the observed inputs from the environment without assuming a priori complete information or infinite computational capabilities. Second, the learning algorithm implements a computational evolutionary process that allows strategies to evolve endogenously; the adaptive behaviour of the agent is given by the evolutionary dynamics of the model. This allows a wide range of strategies to potentially arise, with emerging behaviour that can potentially be difficult to predict beforehand. Such an algorithm can find strategies that were not directly specified by the researcher.

This paper contributes to the literature in its exploration of exogenous signals and correlated equilibrium by using adaptive agents. It studies the long-run effects of an exogenous randomization device on coordinating behaviour. Previous literature has also investigated coordination games by using adaptive agents, but this is the first one to allow the implementation of an exogenous signal and the exploration of its implications on equilibrium selection and evolution of individual strategies.

The model has the structure of an evolutionary tournament including two populations. In each time step, all agents in one population play a repeated Battle of the Sexes game against every other agent in the rival population. Overall scores are kept, and based on those, agents with better payoffs have a higher probability to replicate themselves and replace other agents in their own population. They undergo random mutations at the end of each time step, and the process is repeated for several thousands times simulating long term evolutionary processes.

Our results show that under both implemented treatments (with and without the exogenous signal) the system switches constantly between three different types of equilibrium or attractors, and contrary to what was expected a priori, it never stabilises on one of them. This type of behaviour is sometimes known as 'punctuated equilibria', where the system remains in equilibrium for long periods of time but then presents sudden transitions into a different equilibrium. These three equilibria are i) constantly coordinating in one of the pure Nash solutions of the game, ii) symmetric alternation between the two pure Nash solutions (i.e. taking turns between the two coordination points of the game) and iii) *biased* alternation, where agents also take turns between the two coordination points, but one of them is played more often than the other. To the best of our knowledge, this is the first time that this latter behaviour has been documented in coordination experiments, whether computational or in the lab. Unexpectedly, we found no treatment differences in terms of payoffs and efficiency: both with and without the signal agents learn to coordinate quite well.

A key finding is that agents can indeed learn to condition their actions by consistently following the exogenous signal. However, even if such behaviour can be learned, the probability of it happening is very low (around 5%). While agents sometimes condition their actions based on the signal, they can also learn to alternate and coordinate their behaviour by completely ignoring it.

Hence, consistent with recent experimental literature (discussed below), our results cast doubt about CE being an accurate description of common coordination behaviour. If our adaptive computational agents can be somehow analogous to intelligent life from another planet, they will not learn CE before NE.

Finally, our methodology allowed us to identify interesting behaviour that we couldn't predict a priori. Not only do some strategies learn to use the signal while others can coordinate by completely ignoring it, but the *same* strategy can ignore the signal, use it partially, or interpret it in different ways depending on the history of the game.

3.2 Battle of the Sexes (BOS) Game

Figure 3.1 shows the payoff matrix for the traditional Battle of the Sexes (BOS) game. This game has two pure Nash strategy equilibria, with both players playing *A* (action profile (A,A)) or both playing *B* (action profile (B,B)), corresponding to the upper-left and down-right corners of the matrix respectively. In either case, one player's expected payoff is 2 and the other's is 3. Include now the simplest possible randomization device: both players observe the same outcome of a fair coin toss before deciding their actions, with a 50% probability of observing H (*Heads)* and 50% T (*Tails)*.

		Column Player	
		A	B
Row Player	A	2,3	0,0
	B	0,0	3,2

Figure 3.1: Payoff Matrix in Battle of the Sexes Game

Traditionally, H or T is interpreted as an exogenous signal or a non-binding recommendation for players on what actions to choose. For example, with probability 0.5 both players are 'recommended' to play A (i.e. the recommended action pair is (A,A)) when, say, Heads shows up and (B,B) otherwise (when Tails). This is a *correlated strategy,* which is given by this joint distribution over the set of pure strategy pairs. Notice that in this case, the expected payoff for both players is 2.5 (since each outcome AA or BB would be played with 50% probability), which differs from the expected payoffs of any of the two NE[61].

This correlated strategy is also a CE because no player wishes to depart from following the recommendation. For example, when Heads shows up with recommendation (A,A) and given that player Column will follow it, player Row would decrease its payoffs by not playing what is recommended: if Row decides to play B, his payoffs would be zero instead of two. The same is true for player Column, whose payoffs would go from three to zero in the analogous situation.

The CE concept requires each player to assume that the rival will follow the recommendation given. It also requires common knowledge of the distribution of signal as well as every other agents' payoff. Here we will relax these assumptions. As explained in section 3.4.2, in our model the signal will be observed by the agents without any common knowledge assumption, and it is the dynamics of the model that will determine if they learn to use it consistently to coordinate or not. Also, there will not be any given function mapping the signal to particular actions (i.e. no recommendations): whether agents learn to give particular meanings to the signal or not will be determined endogenously by the evolutionary process of the model.

The payoffs of the game can be formalized graphically as in Figure 3.2. The line $\overline{ABC}$ is the boundary of the convex hull, so all payoffs combinations on the line or inside of the triangle are feasible with appropriate randomization. The maximum attainable payoff for a single player must occur at one of the vertices of the convex hull (i.e. when a pair of pure strategies is played). In this case those points are A = (2,3) and B = (3,2). In this BOS game, A and B are also the two pure Nash equilibria. The line $\overline{AB}$ forms the set of Pareto optimal solutions. Point D = (2.5, 2.5) is the CE discussed earlier. An interesting characteristic of this point is that it is not only Pareto efficient, but is also an *egalitarian* equilibrium: a priori, before the coin toss, both players have the same expected payoffs[62].

To avoid confusion in the analysis that follows, we need to carefully specify what we mean by a CE in our model, or by 'behavior consistent with CE'. Technically, many solution concepts, including pure Nash, are also a CE. However, the interest here is to focus on CE that requires agents conditioning their actions on the signal. So for a CE, we will require agents to condition on the fair coin toss without ignoring it. At the aggregate level this implies payoffs close to the egalitarian equilibrium; at the individual level, as in the correlated

[61] Although this paper will not allow the possibility of mixed strategies, is worth noting that these expected payoffs cannot be obtained by players randomizing on their own (i.e. without the signal). The coin toss in this case allows payoffs that cannot be obtained under a mixed Nash equilibrium concept.

[62] Aumann (1987) suggests the fair coin toss as one of the most simple randomization devices, making it a good candidate for studying the emergence of CE. Another equilibrium studied in the literature is for the Chicken game (Duffy and Feltovich, 2010), with the characteristic that the CE is outside of the convex hull of NE (i.e. the randomization allows higher Pareto efficient payoffs). However, as those authors argue, such CE can be more difficult to learn (at least for humans) because it requires three recommendation profiles, instead of the two implemented here. Cason and Sharma (2007) find that humans' difficulty in learning a CE comes from the uncertainty about their rival's actions, not a lack of incentives (i.e. higher payoffs). One objective here is to test a CE that could arguably be the easiest to learn.

strategy above, using the signal as if a recommendation profile is being followed.

Other behavior, even if under traditional theoretical assumptions (which are relaxed in our model) could also be labeled as CE, will be referred to independently in order to maintain focus on this particular form of signal use.

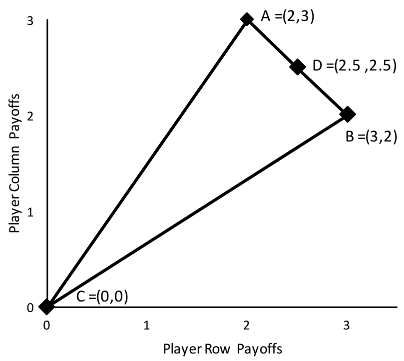


Figure 3.2: Set of attainable payoffs of BOS game under a correlated strategy pair. $\overline{ABC}$ is the boundary of the convex hull, hence any payoffs on or inside of this hull are attainable with the appropriate randomization. Point D represents the correlated equilibrium given by a fair coin toss as the randomization device.

For a more formal presentation of the one shot game and some equilibrium concepts, see Appendix 3.7.1.

3.3 RELATED LITERATURE

3.3.1 On Correlated Equilibrium and learning

Aumann (1974) introduced the concept of CE into the literature and refined it in Aumann (1987), showing that Bayesian rationality implies convergence to a CE. However, this required players to have the same prior beliefs regarding the distribution of the exogenous signal. Some following papers focused on giving conditions or learning rules for achieving convergence. In Foster and Vohra (1997), such convergence is based on players making "calibrated forecasts". This implies evaluating the complete past actions of all rivals, and using this to make perfect probabilistic forecasts that match beliefs with randomized strategies. Fudenberg and Levine (1999) presented an alternative mechanism requiring similar memory capabilities. Hart and Mas-Colell (2000) introduced convergence via "regret", with players making *better* choices instead of using *best* responses (i.e. they switch to actions that would have given higher payoffs than the ones used in the past). This latter approach relaxes some of the rationality assumptions in previous work, but still requires players to have a complete memory of all past actions and calculate the potential payoffs of all of the strategies that could have been played under all potential scenarios. These approaches require very sophisticated players, with unbounded memory and computational capabilities, playing indefinitely. In contrast, the approach here is to model agents with limited memory and no prior beliefs, and test whether evolutionary learning processes at the population level can lead them to learn the CE.

Recent experiments in the lab have focused on CE. These studies have been conducted by Cason and Sharma (2007), Duffy and Feltovich (2010), Bone et al. (2013), Duffy et al. (2014) and Anbarci et al. (2015). A key result arising in all of them is that while some subjects do follow the recommendations given, they do so inconsistently, casting some doubt on the descriptive power of the CE concept[63]. However, as conjectured by Cason and Sharma (2007) in their conclusions, perhaps in longer time spans subjects might learn to consistently follow the recommendations. The evolutionary approach with adaptive agents presented here addresses this issue by conducting long-run analyses that would be impossible to conduct in the lab. Also, it is worth noting that all the experiments above give subjects common knowledge about the distribution of the recommendations as well as what is being recommended to the rival. While such information is useful in helping subjects understand the experiment, how is it that agents come to know such information in a different environment?

3.3.2 On methodology

This paper uses artificial adaptive agents to study the learning and evolution of behavior consistent with CE. Modelling artificial adaptive agents serves as a great compliment to theoretical analysis in economic theory (Holland and Miller, 1991), and it has been used in a wide range of social science topics like market institutions (Gode and Sunder, 1993), pricing (Arifovic, 1994), auctions (Andreoni and Miller, 1995), the evolution of norms (Axelrod, 1986), elections (Kollman et al. (1992)), political institutions (Kollman et al., 1997), loyalty in fish markets (Kirman and Vriend, 2000) and the emergence of communication (Miller et al., 2002), among many others.

Agents presented in this work are boundedly rational with limited information and memory. They are embedded with a mechanism that promotes constant adaptation to their changing environment. Since all agents adapt to each other at the same time, they constitute a co-evolving complex adaptive system. Such adaptive behavior is modelled by means of a genetic algorithm (Holland, 1992), which captures the idea of social learning: strategies that are successful are more likely to be copied by other agents and hence spread in the population, but strategies that are unsuccessful are more likely to be distorted in the learning process. The algorithm strikes a balance between exploration and exploitation (i.e. looking for new solutions versus exploiting the ones that have already been found), which constitutes a classic conundrum in problem-solving (Holland, 1992; Holland et al., 1986).

Each agent is defined as a finite state automaton. Rubinstein (1986) was the first to introduce automata into game theory as representations of strategies.

---

[63] Our evolutionary methodology makes it impossible to make quantitative comparisons with the results obtained in the short time span possible in the lab. However, in section 3.5, we observe qualitative patterns that also emerge in these experiments, giving some external validity to the model presented. The experimental literature is also relevant because results in the lab can inspire new scenarios to explore computationally and vice versa. It is our belief that complementarities and mutual feedbacks exists in social sciences between studies conducted with humans and machines (Duffy (2006) or Poteete et al. (2010) present overviews of this methodological complementarity. Andreoni and Miller (1995) is an example of lab experiments working in tandem with computational simulations).

Miller (1988) introduced the idea of using evolutionary algorithms to model adaptive learning in games (Miller (1996), Ioannou (2013) and Zhang (2015)). Some recent studies have explored the use of automata in coordination games (such as Browning and Colman, 2004; Hanaki, 2006; Ioannou and Romero, 2014a; Ioannou and Romero, 2014b) but no one has studied exogenous signals or CE.

Here we explore with adaptive agents the long run emergence of CE behavior. Arifovic et al. (2015) used *individual* learning to see if adaptive agents can replicate quantitatively the short-term behavior of subjects in the laboratory, including exogenous recommendations. In contrast the approach here uses *social* learning at the population level to focus on the long-term evolution of signal conditioning.

3.4 THE COMPUTATIONAL MODEL

3.4.1 Overall structure

The game used in this paper is the repeated Battle of the Sexes (BOS) as presented in section 3.2. Each agent represents a strategy, and agents face each other in a computational tournament.

More specifically, agents are represented as finite automata (their formalization explained in detail in section 3.4.2). The model has two populations, COL and ROW, each one consisting of N agents. Each time step of the model is called a generation, denoted as t. At each t, each agent in population COL plays R rounds of the BOS game against each other agent in population ROW. The average score (payoffs) of each agent is recorded across all $R \times N$ rounds of play in one generation. Agents select their strategies by imitating the strategies used by other successful agents, with the average score being the (fitness) measure used of success. Hence, strategies with lower scores will tend to disappear from the population while those with higher scores will tend to spread. This is due to the learning algorithm (detailed in section 3.4.3) giving successful strategies higher probability of being copied by other agents. This learning happens at the end of each generation, with agents copying only strategies that are in their own population, thus the ROW and COL populations evolve independently of each other.[64]

The computational experiments conducted here consist of two main treatments: *No-Signal* and *Signal*. Under *No-Signal*, agents play without any randomization device or exogenous signal. In the main treatment, *Signal,* agents play under the same game structure, but are allowed to observe an exogenous signal (given by the fair coin toss) at the beginning of each round.

3.4.2 Artificial agents as finite automata

[64] The choice of the structure of the game, mainly repeated interactions (instead of one-shot) and having two populations instead of one, makes learning potentially easier and should give the emergence of the CE the best possible chance. Experiments conducted by Duffy and Feltovich (2010) show that humans in the lab learn more frequently to follow the exogenous signals in coordination games when they play repeatedly versus playing in one-shot interactions.

Each agent is defined as a class of finite automata using a Moore machine (Moore (1956)), which is a mathematical model with discrete inputs and outputs [65]. The system can be in any of a finite number of internal configurations, called "states". States summarize the past set of inputs and determine the automaton's behavior for subsequent outputs.

A finite automaton can be described as a four-tuple (Q, q_0, f, τ), where

- Q is a finite set of internal states,
- $q_0 \epsilon Q$ is specified to be the initial state,
- $f\colon Q \rightarrow A_i \epsilon \{A, B\}$ is an output function that maps each state into an action of the machine, and
- $\tau: Q \times W \rightarrow Q$ is a transition function assigning a state to every two-tuple of state and observed input.

Here, $W = A_{-i} \epsilon \{A, B\}$, where A_{-i} is the action implemented by the other agent. In this case the only input used by an agent to decide its next action is the action implemented by its rival. In the BOS such input can be A or B, giving agents two potential inputs to respond to. This is how the agents are implemented for the *No-Signal* treatment.

In the *Signal* treatment, each automaton is allowed to respond to four different inputs. Let $S \epsilon \{H, T\}$ be an exogenous random signal with a probability distribution $\left[\frac{1}{2}, \frac{1}{2}\right]$ (e.g., a fair coin toss showing either Heads (H) or Tails (T)), and having the same value H or T for any pair of interacting agents at a given round (i.e. both agents observe the same signal). Thus, in this treatment $W = A_{-i} \times S$ with $W \epsilon \{(A, H), (A, T), (B, H), (B, T)\}$ giving all four possible combinations of the other agent's action and observed signal.

An intuitive way to describe an automaton is by using a transition diagram. Figure 3.3 shows two examples of such diagrams. The nodes in the transition diagrams represent the internal states. The arrows originating from each node represent the transition function with the labels showing the input (rival's action and signal) required for a transition. The arrows point towards the state that the automaton transitions to after observing the corresponding input. The initial state of the machine is given by the "start" arrow.

The automaton in Figure 3.3 (a) for the *No-Signal* treatment shows a strategy that starts by playing A in the first round. Afterwards, it does the same as the rival did in the last round: whenever it observes A it transitions to the state playing A, and whenever it observes B it transitions to the state playing B. This is the famous Tit-for-Tat strategy (Axelrod (1980)). In the *Signal* treatment (Figure 3.3 (b)), transitions are coded using two letters, the first representing the rival's last action (A or B) and the second representing the observed signal (H or T). So a transition showing, say, AT, means that such a transition occurs when the agent observed the rival playing A in the last round and the signal is T for the current one. There are four possible transitions for each node in the diagram. The strategy here starts playing A and when it observes a signal of T, regardless of the rival's past action or the machine's current internal state, it will play A. This is easily noticed by observing that all

[65] There are other types of finite automata such as Mealy machines. Choosing Moore machines as the type of automata implemented is due to it being the standard in previous game theoretical literature. We see no evident reason to deviate from this convention.

the arrows that have a signal of T go into the initial state. Similarly, whenever it observes H, regardless of the rival's action, it plays B. This strategy gives a consistent interpretation of the signal: play A when T, B when H. This is one possible strategy that could be consistent with CE behaviour.

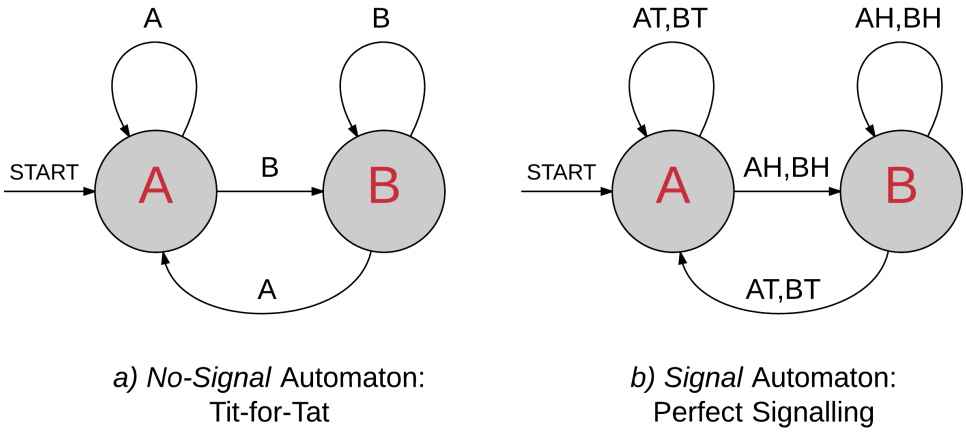


Figure 3.3: Examples of automata for both No-Signal and Signal treatments.

In order to use the learning routines (explained in section 3.4.3) the automata need to be coded as finite length strings. Figure 3.4 shows the coding for both treatments. The *No-Signal* automata is coded as a 25-length string, where the first element provides the initial state of the machine (Figure 3.4(a)). Then, there are eight three-element packets, each representing one of the eight internal states of the automaton[66]. In these packets, the first element gives the action the agent takes when it is in that particular internal state (i.e. to play either A or B). The other elements are the transitions to make when observing the different inputs (i.e. the rival's action): the second element is the transition when the rival is observed to play A, and the third element is the transition when observed to play B (Figure 3.4(b)). The coding for the *Signal* treatment is very similar, with the difference that it requires a longer string (41 elements instead of 25). This is because including the signal allows four possible inputs, requiring four transition per internal state (instead of two). Hence for each state, as in Figure 3.4(d), the first element is the action to be taken, and the following elements are the transitions for all four possible combinations of the rivals' action in the last round and the observed signal in the current.

[66] The number of states used in the machines is in line with previous literature. For example, Ioannou (2013) also uses eight internal states arguing that it allows for a variety of automata that can incorporate a diverse array of characteristics. It is worth noting that more complex machines (more states) do not necessarily mean better strategies. As pointed by Rubinstein (1986), more complex plans of actions are more likely to break down, are more difficult to learn, and can require more time to be executed. Gigerenzer et al. (2011) has several examples of simple rules of thumb that perform better than complex strategies.

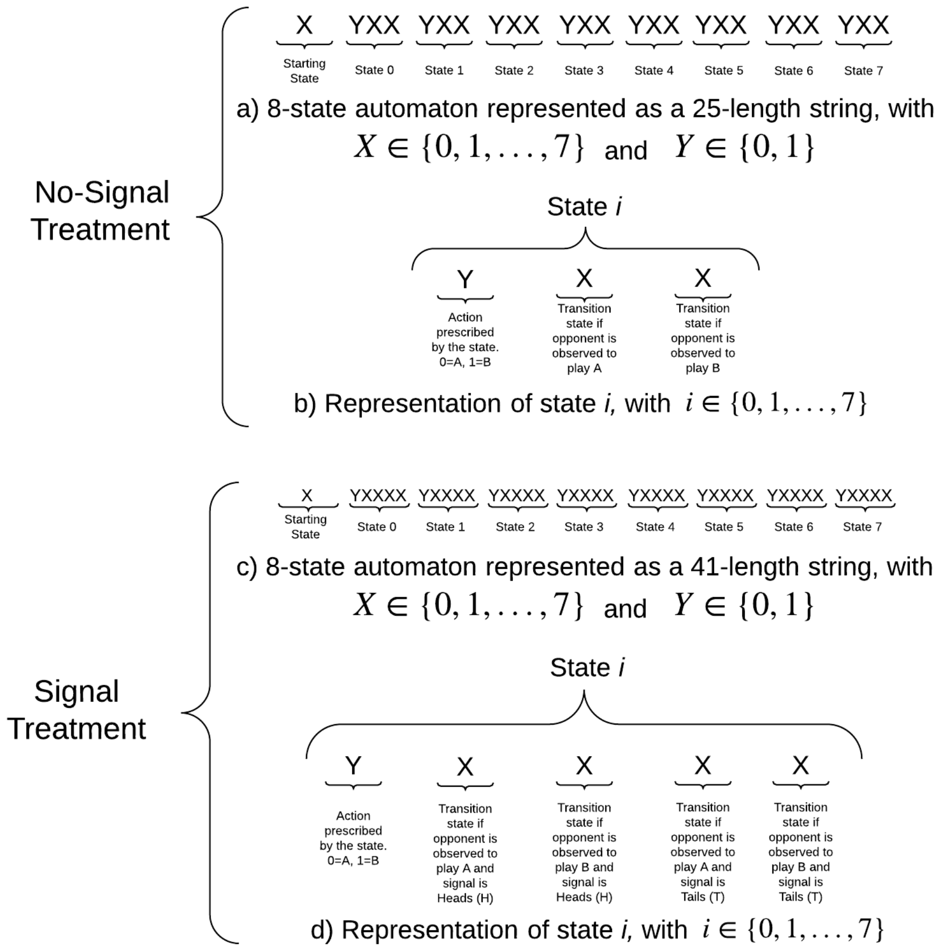


Figure 3.4: Coding of automata for both No-Signal and Signal treatments.

There are some important technical points inherent to the use of automata. Notice that the machines don't have any sort of "expectations" of what the rival will do and that their behavior is purely backwards looking, which is one way to represent simple, boundedly-rational strategies in evolutionary processes. Also, although no separate computational memory is implemented, the internal state of the machine contains the relevant history of the game. A strategy that is based on the past n moves of its opponent will require a maximum of 2^n internal states: for example, the Tit-for-Tat strategy requires the automata to remember only the last action of the opponent, hence it requires two states. Even if the automata here is modelled with eight internal states, only a subset of these states may be accessible to a machine given the starting state and transitions. The number of potential configurations of machines is rather large. In the *No-Signal* treatment, there are $8^{16} \times 2^8$ different arrangements of the strings. However, since many of the configurations lead to the same behavior, the number of unique strategies is lower. For example, two-state machines have $2^7 = 128$ possible arrangements (genotypes) but only 26 unique strategies (phenotypes)[67]. Finally and related

[67] For 3-state machines (with also two inputs and two outputs) the number of unique strategies is 5,832. Notice the exponential growth in the number of possible phenotypes. This

to the latter, automaton theory, as in Harrison (1965), proves that isomorphic automata that represent the same behavior can be mapped to a minimal state machine in the canonical form. This means that many different machines can lead to the same behavior, all of them being able to be represented by a single 'minimal' automaton. These are referred to as "behaviorally equivalent" or "minimized" machines.

3.4.3 Evolution of strategies

3.4.3.1 Motivation for the learning mechanism

The learning algorithm used in this paper is derived from a class of optimization routines from computer science called genetic algorithms (GA), introduced by Holland (1975). GAs are computer programs that mimic the processes of biological evolution in order to solve problems and to model evolutionary systems. We use GAs for two main reasons: its technical advantages and its analogy as a learning mechanism reflecting bounded rationality.

The algorithm has several advantages over other optimization methods. It is designed to work well in *difficult* domains, meaning domains that involve discontinuities, nonlinearities (many local optima), noise and high dimensionality (these issues arise in the strategy space in the tournament analyzed here). Contrary to calculus-based methods that require derivatives in order to perform an effective search for better structures, GAs require payoffs associated with the individual strings, making it ideal for game theoretical environments with their well-defined payoffs structure. All of the above makes GAs a more canonical optimization method than many other search schemes[68].

Evolutionary processes such as a GA explicitly model a dynamic process describing how agents adjust their choices over time by learning from experience; this makes the GA a useful tool for observing the learning (or lack thereof) of coordinating behavior with an exogenous signal. In the same line as Kandori et al. (1993), this evolutionary approach gives a concretely defined, step by step process of how an equilibrium can emerge based on trial and error mechanics. Even if biological interpretations are usually given to such processes, the algorithm's processes can be reinterpreted as bounded rationality, reflecting the limited ability on the player's part to receive, decode and act upon information they get in the course of the game. As in Kandori et al. (1993) three main hypotheses are relevant and related to this learning interpretation, reflecting its adequacy in order to model adaptive, boundedly rational agents. First, the *inertia* hypothesis holds since not all players react instantaneously to their environment. This is because given the imperfect observations agents have (for example, regarding payoffs and strategic choices of other agents), changing one's strategy can be costly. Second, the *myopia* hypothesis holds since there is substantial inertia in the system with only a small fraction of agents changing their strategies simultaneously, resulting in agents making only moderate changes. The myopia hypothesis also captures a key factor in *social* learning: imitation or emulation. Agents learn what are good strategies in a complex environment (where they cannot calculate best

---

makes calculations on the exact number of possible strategies for machines with more internal states increasingly costly in computational terms.

[68] For further discussions on genetic algorithms, see Mitchell (1998).

responses) by observing what works well for others. In such an environment strategies that remain effective in the present are likely to remain effective in the near future. Also, myopic agents do not take into account the long-run implications of their actions or strategies. Finally, the *mutation* hypothesis holds given that with some small probability agents will play an arbitrary strategy, capturing the exploration aspect of most learning processes.

3.4.3.2 Details of the Genetic Algorithm implementation

The mechanics of the implemented GA (for both *No-Signal* and *Signal* treatment) are as follows: two populations (ROW and COL) are randomly initialized with 40 agents each at *t*=1 (first generation). This initialization consists of generating for each agent a random finite-length string automaton as in Figure 3.4 (with uniform probability across the alternatives)[69]. Then each automaton is tested against the environment: this consists of each agent in population ROW playing 50 rounds of the repeated BOS game against each of the 40 agents in COL population. Scores are stored for all automata, with the score for each agent being the average payoffs earned across all games.

Two new *offspring populations*, each with 40 agents, are created based on the current *parent populations* (i.e. the populations existing at the beginning of the generation). Each population evolves independently, so the offspring of the COL population will be based only on the parent COL population (the same applies for ROW). Offspring populations are created based on two operators: *selection* and *mutation*. For selection, the top 20 scorers are chosen and given a copy in the new population. The other 20 needed to keep populations constant are chosen via pairwise tournaments by randomly picking two agents (with replacement), and keeping the one with the highest score. Such tournaments are repeated 20 times in order to keep population size constant.

Before moving on to the next generation, the 20 strategies picked via the pairwise tournament go through mutation process. Each automaton has a 0.5 probability of being randomly altered. If a strategy undergoes mutation, one of the internal states is randomly selected and with a 0.5 probability the action of that state is changed (thus, if the state had an action of A, it is changed to B and vice versa); otherwise, a randomly chosen transition (from the chosen state) is changed with uniform distribution for the alternatives[70].

Finally, once both ROW and COL offspring populations have been created, scores are reset to zero and a new generation of the algorithm is begun (i.e. agents are again tested against the environment, scores are assigned, and

---

[69] Randomly generated populations will favour minimized (behaviourally equivalent) machines that represent strategies with only one internal state (i.e. always play A or always play B). When the maximum internal states allowed is equal to two, the probability of generating a machine that always plays A is 31% (analogous for always playing B). When three internal states are allowed, this probability is 20%. Making such calculations for more internal states becomes increasingly costly; however, the dynamics of the GA will quickly start favouring strategies that perform better.

[70] There are other ways to implement selection and mutation. GAs are a broad class of algorithms with many variations, but fortunately they are fairly robust to different parametric and algorithmic choices. The mutation parameters and mechanism used are the same as in Miller et al. (2002) and Miller and Moser (2004). In general, within reasonable changes, results will be consistent. However, if taken to an extreme, too small mutation rates eliminate exploration and will lead the system to converge based only on the selection process. If mutation is too high, the system will always be exploring, unable to settle down and exploit information. The chosen mechanism tends to be in a reasonable "sweet spot" to balance this out.

populations undergo selection and mutation). An overview of the whole process is given in Figure 3.5.

1) Initialise two random populations (ROW and COL) with 40 agents each. Set *t*=1 (first generation)
2) Test each agent against the environment: play 50 rounds of BOS against each agent in the rival population, saving average scores.
3) For ROW population, form a new population of 40 agents in the following way:
    a) Copy top 20 scorers from old population (will also be potential parents)
    b) Pairwise tournament: choose randomly 2 potential parents from the population of 20 copied in (a), with replacement. The one with the highest score gets one child copy of itself
    c) With 50% probability, mutate the child:
        i) Randomly choose one internal state
        ii) With 50% probability, switch the action of that state
        iii) If didn't change action in step (ii) (50% prob.), randomly choose one transition of the state and change it with uniform probability across alternatives.
    d) Repeat steps (b) and (c) until the new ROW population has 40 agents.
4) Do step (3) for COL population
5) Increment *t* by 1 (next generation), reset scores to zero and iterate (go to step (2)).

Figure 3.5: Structure of the evolutionary process (works the same for both No-Signal and Signal treatment)

3.5 RESULTS

Given the model we can analyze its behavior. The following five questions address the overarching research goals presented in the introduction, serving as a roadmap for the evidence ahead. They will be answered in the order presented.

1) Will the system converge to an equilibrium?

   A priori, is not clear if an equilibrium will emerge. We hypothesize that without the signal agents will converge into one of the pure Nash equilibria. With it, our hypothesis is that they will converge in Turn-Taking (alternation), taking turns symmetrically in both coordination points of the game. For both treatments the hypothesis is that the system will stabilize in the corresponding equilibrium and remain there.

2) Will the presence of the signal allow agents to coordinate more easily? That is, will the system be more efficient when the signal is included?

We hypothesize that when the signal is included, agents will miscoordinate less often leading to higher payoffs.

3) Are there other treatment differences, if any, in terms of the aggregate behavior of the system?

   We have no other a priori hypotheses regarding treatments differences besides the ones addressed in questions 1 and 2, but we leave the possibility for unexpected results. With the power of hindsight, we know that there are indeed other differences that are worth exploring once answers to questions 1 and 2 above are known.

4) Conditioned on observing Turn-Taking (alternation) as hypothesized in question 1, will agents be actually conditioning on the signal in a way consistent with CE?

   This question might seem subtle, but its analysis is key to understanding the emergence of CE. Notice that agents might alternate or take turns in the two coordination points by either using the signal or by completely ignoring it. Both types of behavior would seem similar at the aggregate level, but only conditioning on the signal would be consistent with CE as defined here. We hypothesize that agents will learn to condition their actions based on the signal.

5) At the micro level, how are agents coordinating? That is, how do we characterize the strategies that evolve?

   Analysis of questions 1 to 4 are made at the aggregate level of the system (e.g. average payoffs, coordination rates). But one of the advantages of using automata and computational methods is that we can directly observe each and every strategy in the system at any point in time. Here we use a methodology based on *pairs of interacting* strategies to characterize them and understand their exact behavior. A priori, given the immensity of the possible strategy space, we don't have any particular expectation of the type of strategies that would evolve besides the ability to invoke both pure Nash and alternating behavior. However, as we will see, novel and interesting behavior evolved that we didn't predict beforehand.

3.5.1 Regimes and epochs

We start by focusing on what type or types of equilibrium are selected under the *No-Signal* treatment. Figure 3.6 shows the average payoffs obtained by each population across all rounds of play. Five panels are shown, each one of them corresponding to a different run of the model. Some key patterns can be observed and some characteristics inferred based only on the average payoffs.

Note that the system never fully stabilizes. Instead, it is characterized by punctuated equilibria: the system locks for several generations in a kind of stasis where average payoffs per population are quite stable, followed by a sudden transition into a different (and similarly stable) configuration[71].

---

[71] The assertion that the system "never" stabilises is based on longer runs. Some of the earliest literature on similar models ran simulations for around 50 generations. Recent work has used between 1,000 and 2,000 generations. Besides the five simulations, the model has been run several times up to 5,000 and 10,000 generations. One very long simulation that will be

Three kinds of equilibrium behavior are identified. Remembering that both game's pure Nash equilibria have payoffs of (3,2) and (2,3), the run in the top panel of Figure 3.6 shows consistent coordination on either (A,A) or (B,B). In this run, one population is consistently receiving average payoffs very close to three and the other very close to two. Thus, one population is 'dominating' the other in terms of payoffs. The transitions here only change which population is getting the higher payoffs.

The second equilibrium behavior observed, for example, on the third panel around the 1,000 generations mark, has both populations obtaining average payoffs close to 2.5. Given the structure of the model, without the exogenous signal this means that the agents have found a way to coordinate on some sort of turn-taking behavior. They are alternating symmetrically between the two coordination points, although it is not clear if they are alternating each turn. They could, for example, by playing three times in a row (A,A), then three times in a row (B,B), and so on.

The third equilibrium that arises in the model was not foreseen. It can be observed in the bottom panel, around generations 1,100 to 1,700. Here agents use '*biased* turn-taking': although they take turns, it is not symmetric. Agents are playing, for example, two rounds at (A,A), followed by one round of (B,B) and then back to (A,A). This gives both agents a chance to play to their preferred coordination point, but one of them having its way more often. This is the first time such behavior has been documented in a BOS game, either in simulated or experimental data. The micro analysis showing exactly what strategies emerged for all three equilibria will be done section 3.5.5.2, allowing us to understand how such coordination happens.

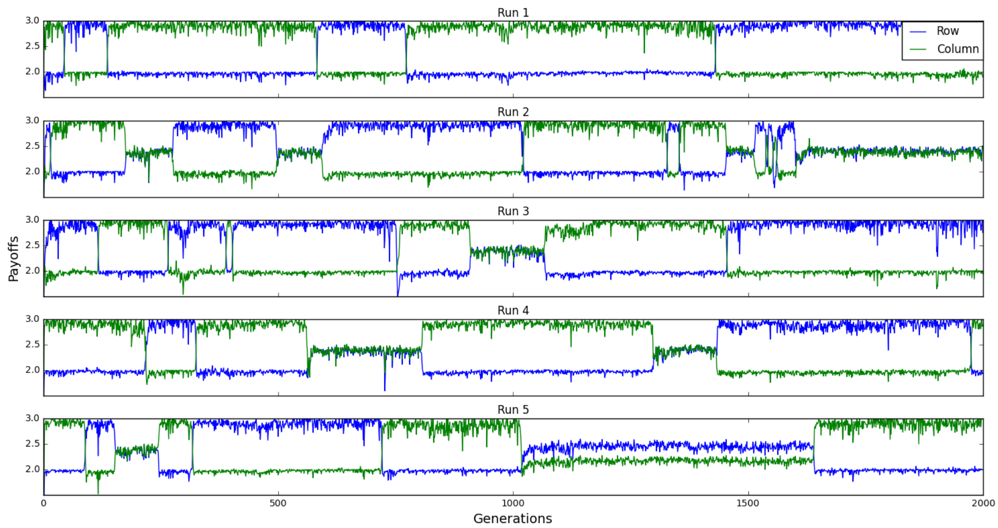


Figure 3.6: No-Signal treatment. Average payoffs per population. Each panel is one different run of the model, each consisting of 2,000 generations.

It is convenient to have a formal way to describe and name these equilibria: each generation, t, will be classified under one of the following *regimes* based on the a-posteriori probability of observed play. Let AA_p^t be the percentage of rounds for any pair of agents playing (A,A) during generation t, and BB_p^t the

reported below was run for 100,000 generations. In all runs the system displayed punctuated equilibria.

analogous for (B,B). Then each generation is classified into one of four regimes according to the following rules[72]:

- *Domination A (B)*: if $AA_p^t\ (BB_p^t) > \ 0.8$
- *Turn-Taking:* if $(0.4 > \ AA_p^t < 0.55)\ and\ (0.4 > \ BB_p^t < 0.55)$
- *Biased Turn-Taking A (B):* if

  $$\left(0.15 > \ AA_p^t(BB_p^t) < 0.4\right)\ and\ \left(0.55 > \ BB_p^t(AA_p^t) < 0.80\right)$$

- *Other:* if none of the above.

An *epoch* is defined as a streak of consecutive generations under the same regime. Technically, it is a window of at least ten generations with the same regime where no more than three are being classified under a different regime (hence allowing for some "mistakes"). For example, 500 generations in a row classified under the regime "Domination A" (allowing for a few mistakes) is considered as **one '**Domination A' epoch.

In order to have representative measures of the system's behavior, one very long simulation (with *t*=100,000) was run for each treatment[73]. Compared to the *t*=2,000 of initial simulations, the longer time span gives us a good measure of the system's statistical properties. All of the following data for each treatment is based on the corresponding long simulation[74].

Based on such long simulations, less than 1% of generations are classified under the 'Other' regime. So the system under the *No-Signal* treatment can be accurately described in terms of the three main regimes Domination, Turn-Taking and Biased Turn-Taking.

What is the equilibrium behavior of the system when the signal is included? Surprisingly, it is very similar to the *No-Signal* treatment. One can grasp an intuitive feeling for this by observing appendix 3.7.2, where figures for the 100,000 simulations and five short ones for the *Signal* treatment are presented. The reader will notice that the payoffs present very similar patterns compared to the *No-Signal* treatment. Formally, based on the corresponding 100,000 generations simulation, the system can also be classified in more than 99% of the time in one of the three main regimes, and constant transitions between them are also observed. This means that at the aggregate level, both with and without the signal the model presents similar behavior in terms of the regimes that emerge. Other treatment differences will be addressed below,

---

[72] The threshold values for each regime were chosen in order to allow a convenient classification, and the analysis is robust to reasonable changes.

[73] Having one very long simulation instead of aggregating several short ones for the main analysis was chosen for a reason: as will be seen below, some epochs can be rather long, characteristic that would be lost with short simulations.

[74] Although one might initially have concerns for the effects of the random initial conditions, given enough time and due to the switch between epochs (i.e. the phase transitions), the system will eventually forget its past. Each type of epoch (i.e. regime) can be seen as an attractor of the model, and by visiting them all the system is no longer dependent on the initial conditions. This would be different if the system would lock in one of the attractors forever, which would make initial conditions critical.

including the probability of finding the system in each regime (confirming this result).

The evidence so far can be summarized as follows:

Result 1*: The behavior of the system can be described in terms of three main regimes: Domination, Turn-Taking and Biased Turn-Taking. The system never stabilizes in one particular regime, but instead presents transitions switching from one long epoch to another in short time spans. This applies for both Signal and No-Signal treatments*[75].

3.5.2 Efficiency

The next question we consider is the efficiency of the system. Table 3.1 presents the average payoffs in the long run as well as the average coordination rates. The latter is measured as the percentage of rounds across all generations where any pair of agents play a coordination point (either (A,A) or (B,B)). In terms of payoffs both treatments have virtually the same value of 2.4, which is very close to the Pareto optimal of 2.5[76]. Coordination rates also show that the system is highly efficient. In both treatments agents play one of the pure Nash strategies (i.e. a coordination point) in more than 95% of rounds. Comparing this with the expected coordination rates for agents playing mixed strategies (48%) or even playing randomly (50%), it can be seen that the system is equally efficient with or without the use of the signal. Contrary to what was hypothesised a priori, the signal doesn't really help agents solve the coordination problem.

Result 2*: Under both treatments the system is quite efficient: the probability of agents coordinating in one of the two pure Nash equilibria is close to 95% with and without the signal. Payoffs are virtually the same and very close to the Pareto optimal of 2.5, so we conclude that there are no treatment effects in terms of payoffs or efficiency. Agents learn to coordinate equally well with or without the exogenous signal.*

---

[75] Simulations using an alternative selection mechanism also have been run. Instead of selecting 20 top scorers to go directly into the next generation and then using a pairwise tournament, the alternative was to conduct the tournaments directly for the whole population, without guaranteeing any strategy a direct copy. Simulations are robust to this result, namely the regimes observed and the constant transitions between them.

[76] Average payoffs by population are, for the *Signal* treatment 2.38 and 2.43, and for the *No-Signal* treatment, 2.26 and 2.48. Due to the large amount of observations, differences are statistically significant, although they seem relatively small in economic terms. Such small differences can occur due mainly to the presence of some very long epochs, particularly for the Biased Turn-Taking regime (as shown below).

Average Payoffs		Average Coordination Rate	
No-Signal	*Signal*	*No-Signal*	*Signal*
2.36	2.4	95%	96%

Table 3.1: Average Payoffs and Coordination Rates for both No-Signal and Signal treatments. Treatment differences are barely noticeable.

3.5.3 Probabilities of each regime

We turn now to the differences in regime frequencies. Figure 3.7 presents the probability of randomly choosing one generation and having it classified under each regime. The percentages presented are equivalent to the ratio of the number of generations classified under each regime to the total number of generations in the run (here *t*=100,000). This provides a measure of how much time the system spends in each regime. For easy of exposition, notice that 'Domination A' and 'Domination B' are aggregated simply as 'Domination' (the same applies to Biased Turn-Taking).

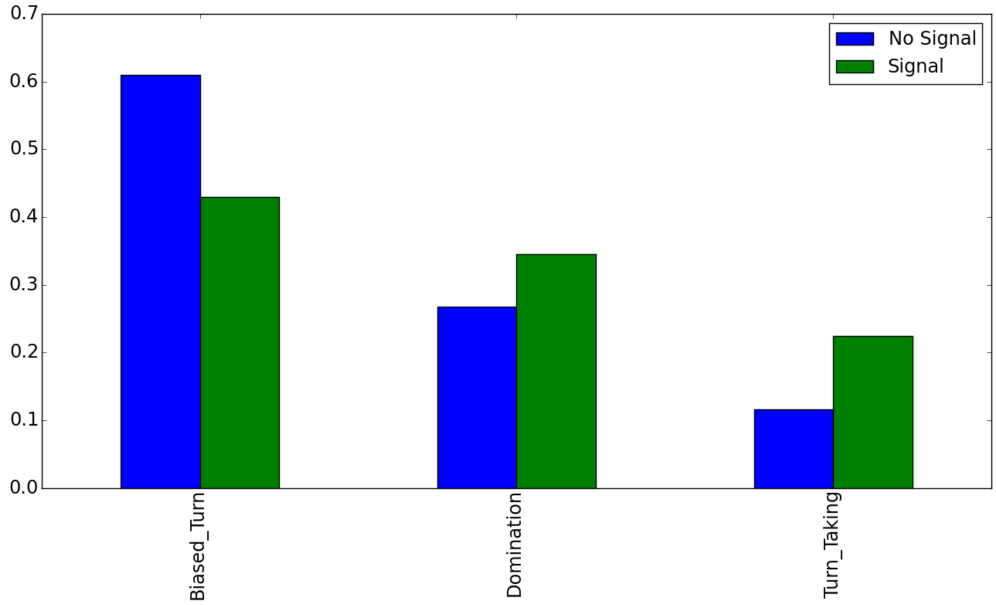


Figure 3.7: Percentage of the time that the system spends in each regime. Measured as the ratio of generations classified under each regime to the total generations in the run (t=100,000).

Figure 3.7 shows that Turn-Taking is the least frequent of the three regimes, both with and without the signal. This suggests that CE may not be a more likely equilibrium concept than pure Nash. There's also no evidence that Turn-Taking would be learned "first". All simulations ran started with a short period of learning (usually no more than ten generations) followed by a Domination epoch. This is due to the random generation of automata favoring strategies that always play A or always play B (as mentioned before). So in this model, behavior consistent with pure Nash equilibrium is both more frequent and happens before any kind of Turn-Taking. Figure 3.7 also shows that the system spends most of the time in a Biased Turn-Taking regime under both treatments. Why is this the case?

One potential explanation for the prevalence of Biased Turn-Taking is that the system transitions more often into these epochs than into the others. Figure 3.8 shows the total number of transitions the system underwent (a), and how are those distributed across the three regimes, i.e. the percentage of transitions into each regime (b). It can be seen that even if the system transitions more often under the *No-Signal* treatment, the distribution is the same under both treatments. For both treatments, Biased Turn-Taking is the regime to which the system transitions into *least* frequently. If Biased Turn-Taking is the more frequent regime, but also the one to which the system transitions into less frequently, the length of the epochs must be driving our results[77].

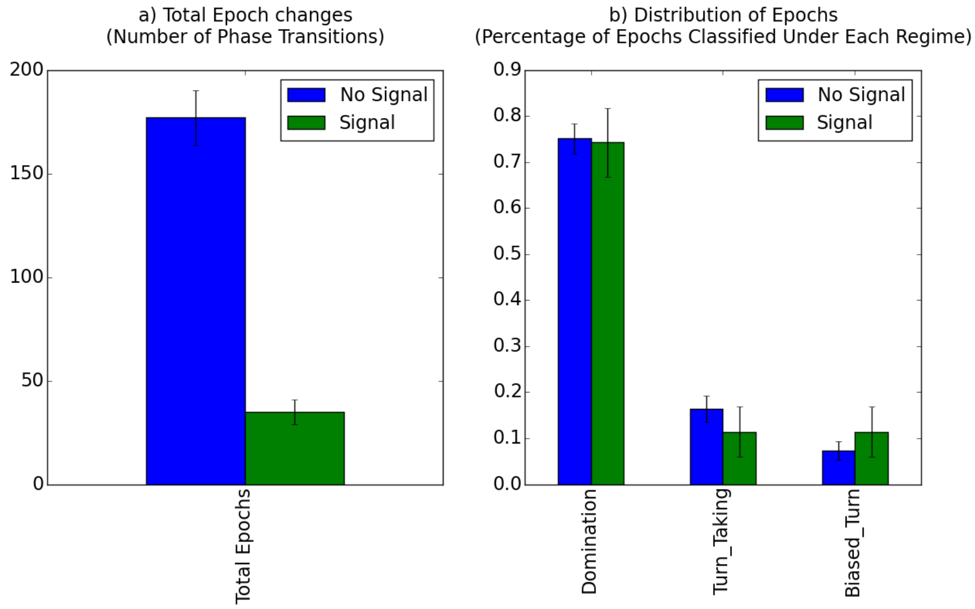


Figure 3.8: Panel a): Number of different epochs observed per treatment (i.e. time the system underwent a phase transition). Panel b): Distribution of epochs across regimes. Under the Signal treatment the system has less phase transitions (left panel), but the relative proportions across the regimes is the same for both treatments (right panel).

Figure 3.9 shows the average length of epochs per regime. As expected, the *Signal* treatment has longer epochs than *No-Signal*. But more importantly it also shows that Biased Turn-Taking has the longest epochs of all regimes for both treatments. So even if the number of Biased Turn-Taking epochs is low, their length makes it more frequent.

[77] Further tests on understanding better the difference in the frequency of transitions across treatments have been conducted (not reported), although preliminary results show that the causes might be quite complex. See section 3.6.3, on “future research”, for additional comments on this regard.

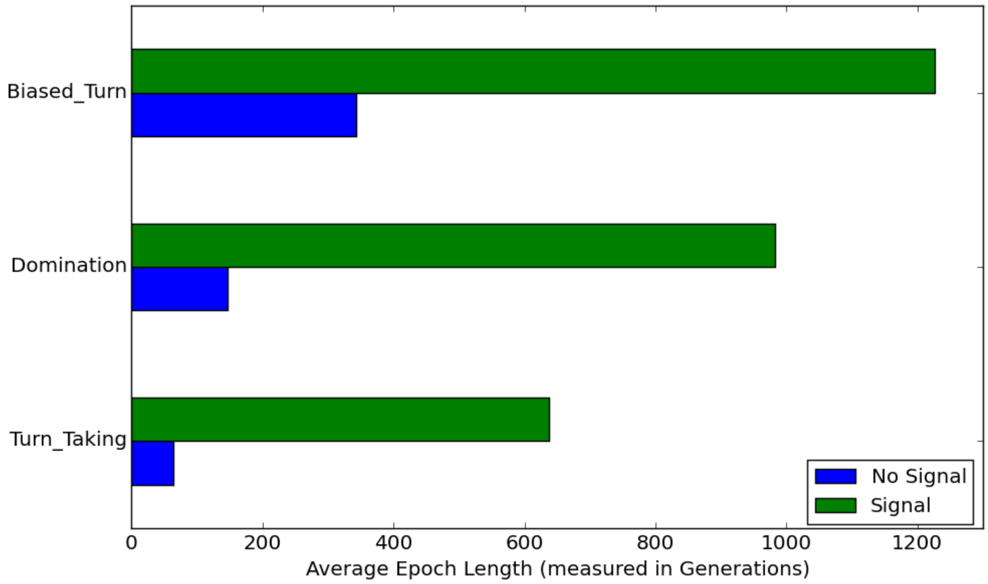


Figure 3.9: Average length of epochs per regime. Signal treatment presents the longest epochs.

Finally, it is worth emphasizing the difference in time spent under the Turn-Taking regime across treatments. In this case, the probability of a generation being classified as Turn-Taking goes from 11% without the signal to 22% when it is included (which can be seen in Figure 3.7) [78]. So, even if Turn-Taking is the least frequent regime, it is more likely to be found with the signal than without it.

Let us summarize these findings as follows:

***Result 3**: Two main treatment effects are identified: first, the system undergoes fewer epoch changes under the Signal treatment: a total of 35 compared to 177 for No-Signal. Second, the probability of finding the system under a Turn-Taking regime increases with the Signal from 11% to 22%. However, Turn-Taking is the least probable regime for the system. The system spends most of its time under Biased Turn-Taking regimes, with such epochs being longer, rather than more frequent.*

This result rules out CE being more frequent than other equilibrium concepts such as pure Nash, but it doesn't say anything about agents actually following the signal. For this, the strategies that are being used under Turn-Taking epochs need to be evaluated in a different way, both at the macro and micro level.

3.5.4 Searching for CE behavior at the aggregate level

In order to analyze behavior consistent with CE, the reader is reminded that, here, CE is being used only to refer to an equilibrium in which agents

[78] Differences here are statistically significant: since the sample is so large and the units of observation are generations, each one taken as an independent observation (with t=100,000), standard errors of the mean are on the order of 1×10^{-5} , resulting in very small confidence intervals.

condition their actions by using the signal. This avoids using CE to describe other types of behavior such as pure Nash.

The first way to explore if there are Turn-Taking epochs in which agents are following the signal, is to develop an aggregate measure based on the probabilities of agents playing each action conditioning on the signal. The intuition is that if agents are following the signal, one should observe, on average, that the probability of playing the same action (say A) should be high when the same signal is observed (say Heads). On the contrary, if the signal is being ignored, one should not expect the same action for each signal. Although this measure doesn't show exactly how agents are coordinating (such analysis is done in section 3.5.5), it will allow us to identify if there are epochs in which the signal is consistently being followed.

Let $p(row = A|S = Tails)$ be the observed probability in a particular generation for an agent from population ROW to play A, given that the observed signal for that round was Tails. Then, using analogous notation for a player from population COL, action B and signal Heads, we define our *Correlated Equilibrium Measure* in generation t (CEM_t) as follows:

$$CEM_t = \begin{array}{c} p(row = A|S = Tails)p(col = A|S = Tails) \times \\ p(row = B|S = Heads)p(col = B|S = Heads) \\ + \\ p(row = A|S = Heads)p(col = A|S = Heads) \times \\ p(row = B|S = Tails)p(col = B|S = Tails) \end{array}$$

Notice that $CEM_t \in [0,1]$. If agents are using the signal, $CEM_t \approx 1$. Under a Domination regime, $CEM_t \approx 0$. If agents are Turn Taking but ignoring the signal, CEM_t will be somewhere in-between.

We find that the behaviour of the values of CEM are very stable within single epochs. Agents use the signal in the same way within epochs, meaning that within a single one, agents tend to use the signal in the same way. Appendix 3.7.3 shows the CEM values vs. the average payoffs for the *Signal* treatment.

Average CEM	Regime	Number of Epochs
0.12	Turn-Taking	2
0.86	Turn-Taking	2
0.11	Biased Turn Taking	1
0.23	Biased Turn Taking	1
0.40	Biased Turn Taking	2
0.00	Domination	26

Table 3.2: Average Correlated Equilibrium Measure (CEM) for all observed epochs under the Signal treatment. Calculated as the average CEM_t of all generations within a single epoch. Different epochs under the same regime can have the same average CEM, which is reflected in the "Number of Epochs" column.

Table 3.2 presents the average values of CEM_t for all different epochs observed under the *Signal* treatment. The values in the left column are the average CEM_t across all generations within a single epoch. Different epochs can have the same *CEM* value, which is shown in the "Number of Epochs" column. The first two rows of the table indicate that out of a total of four observed Turn-Taking epochs, the average value of CEM_T is 0.12 for two of them and 0.86 for the other two. Thus, in two of the Turn-Taking epochs agents are following the signal. This is our first evidence showing that agents have indeed learned to play CE. Yet, despite agents being able to learn coordination by using the exogenous signal, they can also ignore it completely and alternate as they would do in the absence of a signal[79].

Unexpectedly, the *CEM* also shows that the behaviour under Biased Turn-Taking regimes can vary widely in its use of the signal. This behaviour will be explored below when analysing at the micro level the strategies that emerged, but it is worth mentioning that agents use the signal in different ways: this is what leads to the various observed intermediate values of the CEM in Table 3.2.

How important are the epochs where agents are learning to use the signal? The total time the system spends under a Turn-Taking regime in the *Signal* treatment is 22% (Figure 3.7), corresponding to four different epochs. However, the two epochs with a high average *CEM* constitute only 6.2% of the total time. So even if the evidence shows that agents can indeed learn to alternate their actions by following the signal, this happens rarely in the system[80].

---

79 Duffy et al. (2014) found similar results in their experiments. They document evidence in a BOS game where subjects exhibit both types of behaviour, alternating both by using the signal as well as by ignoring it.

80 Why are agents not learning CE more often? One potential answer is that the learning algorithm is having difficulties in finding complex solutions (strategies) that include processing the signal. If the latter is true, one could argue that the results are driven by an inefficient algorithm instead of some deeper property of the system's dynamics. Appendix 3.7.4

Result 4*: Agents can learn to play CE and alternate their actions tied to an external signal. However, the likelihood of finding such behaviour is small. Agents can also learn to alternate by completely ignoring the signal. No evidence is found of CE being learned faster, or more frequently, than other types of behaviour.*

Thus while agents can indeed learn to play by conditioning on the signal, such learning occurs very rarely, and CE may not be the best descriptive notion of actual behaviour.

The one remaining question is related to how exactly are agents coordinating. Regimes and epochs classification hint at what agents are playing and gives us a characterization of the system at the macro level, but several different strategies at the micro level can lead to the same aggregate patterns. For example, even without the signal, Turn-Taking behaviour could be happening by playing (A,A) four times in a row followed by (B,B) four times, or by alternating one time on each. Understanding precisely what strategies have evolved is also important for the Biased Turn-Taking regimes. Not only does the system spend most of its time under such epochs, but the different values observed for CEM suggest that coordination happens under a wide range of behaviours. Such heterogeneity is impossible to grasp based on the aggregate measures presented so far as exploring such findings requires a more fine-grained micro analysis of what strategies evolved under each regime.

3.5.5 Micro Analysis

3.5.5.1 Individual Machines

Here we observe the exact structure of the most successful strategies playing under each regime. How are strategies responding to both the signal and the rival's actions? One first approach to understand these micro characteristics of the agents is to observe the top evolved individual machines.

Figure 3.10 shows some of the most frequent machines for each regime, chosen by randomly picking one epoch and selecting the most frequent strategy in one population[81]. The most frequent machine for one Domination epoch (Column population) is shown in panel (a), showcasing a very simple kind of behavior: play A no matter what. Perhaps surprisingly, simple strategies can perform very well in complex environments (see for example Gigerenzer et al. (2002) or Gigerenzer et al. (2011)). Strategies for Turn-Taking and Biased Turn-Taking are a bit more complex, but still far away from using all eight states. Even so, it becomes difficult to gain a clear insight about the system by observing only individual strategies. For example, it is hard to infer directly from the Turn-Taking machine (panel (b)) if that strategy follows the signal. For some particular cases (such as the automaton in panel (b) of Figure 3.3) this can be easier, but in general it is not trivial.

---

implements a test that addresses this issue. Results show that without the strategic component of the game, agents can easily learn to alternate their actions by using the signal.

[81] The reader is reminded that the machines all have eight internal states, but that some of those states can be inaccessible or redundant (e.g. a machine with all eight states having an action of A has the same behaviour as a machine with one single state with action A). The shown machines are the minimal equivalents.

To be able to make such inferences one often needs to observe also the opponents' strategies. These are shown in Figure 3.11 for the Domination and Turn Taking regimes.

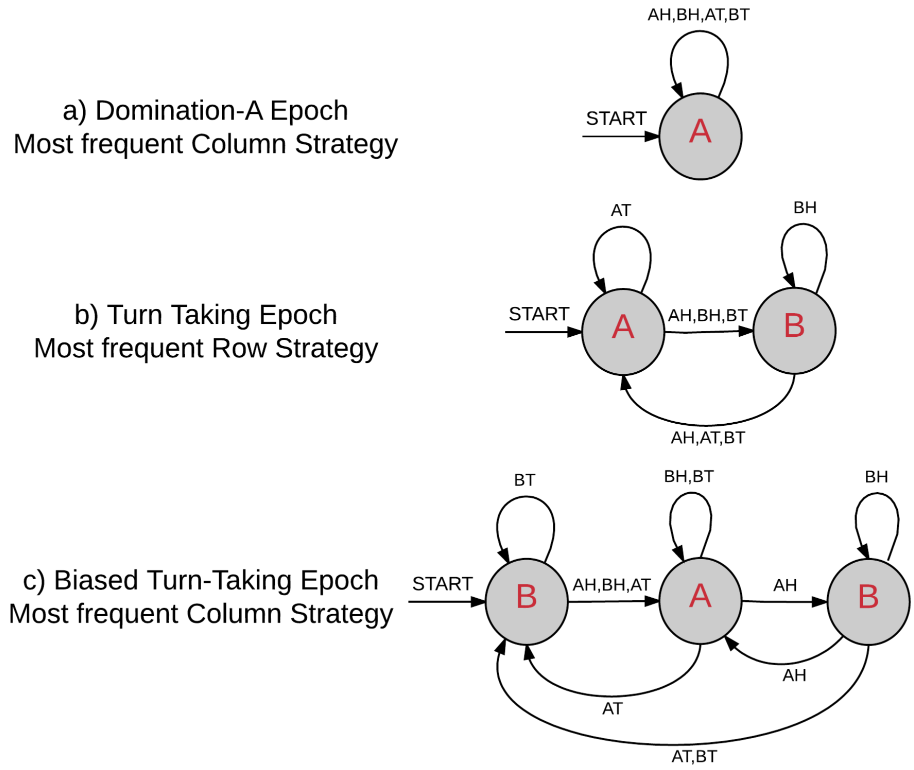


Figure 3.10: Some evolved strategies from Signal treatment. These were chosen by randomly selecting an epoch of the corresponding regime and choosing the most frequent strategy for one of the populations.

As can be seen, the top (most frequent) strategies in the opposing population are more complex. By observing the two interacting machines in panel (b) (of both Figure 3.10 and Figure 3.11), it is difficult to infer if they are following the signal or not. How exactly are they managing to coordinate?[82]. So directly observing the strategies may not be the best way to analyze the system at the micro level, unless one limits the strategies to a few internal states.

Another way to analyze the machines, previously used in the literature (e.g. Miller (1996) or Ioannou (2013)) is to generate average measures based on the accessible states of the machines. For example, checking how many of the accessible states in each machine have particular behavioral traits has been used to describe cooperation games (e.g. how many states punish defections, or how many forgive one).

Although this approach has proven very useful, it doesn't come without limitations. To illustrate this, observe that larger strategies may not necessarily

[82] These two strategies, when playing against each other, actually do follow the signal.

use all of their states even if they are accessible[83]. A machine could only visit a subset of the accessible states if no rival machine gives it the necessary input. So focusing the analysis on measures of the states of the individual machines can be misleading, because it could include behavior that is never actually used.

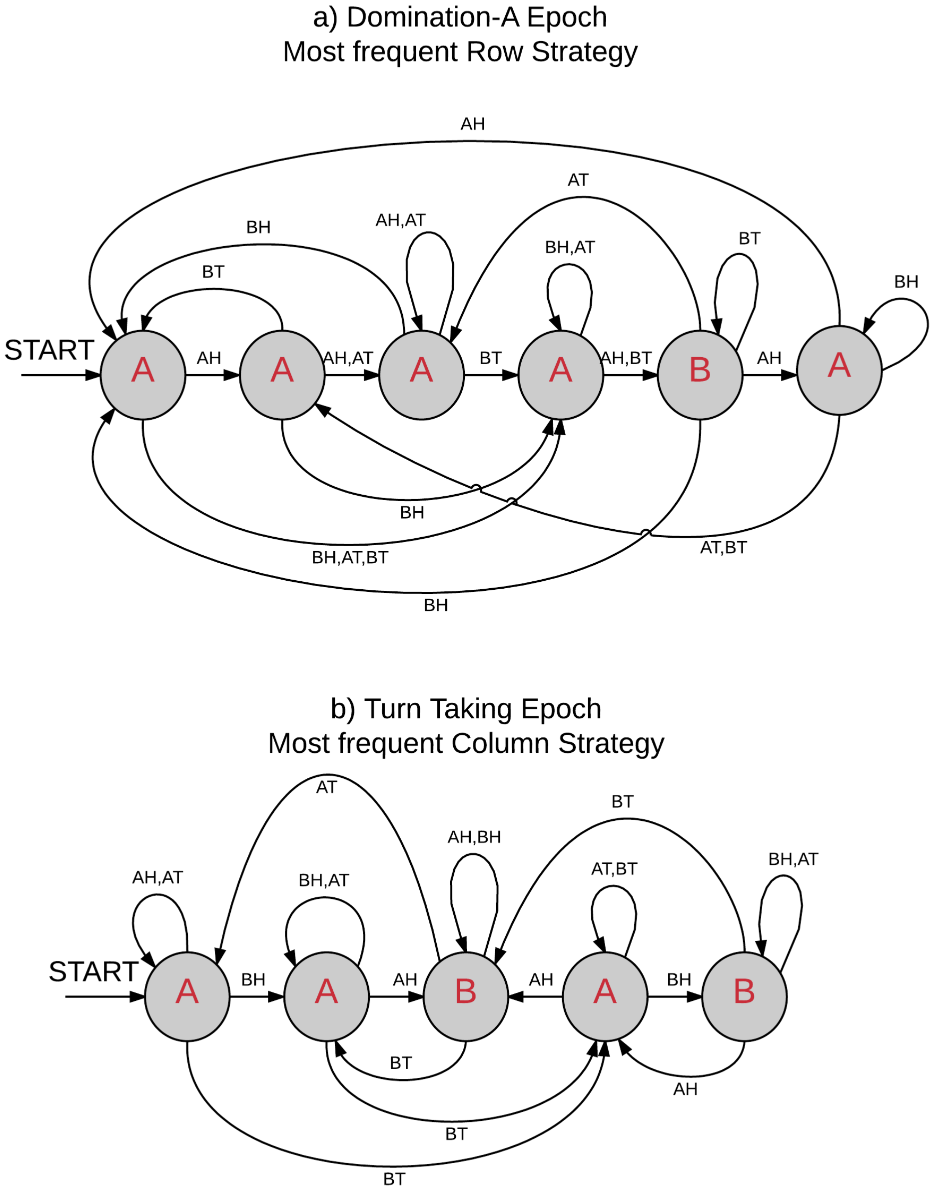


Figure 3.11: Evolved complex individual strategies. They were chosen by randomly selecting an epoch of the corresponding regime and choosing the most frequent strategy for one of the populations. The strategies presented are considered

[83] As a reminder, a state is accessible if there exists at least one combination of inputs (i.e. opponent's last action and exogenous signal) that can lead the machine to be in that state.

among the complex ones (i.e. having more internal states). Is very difficult to infer the behavior of the system by observing them.

In summary, focusing the micro behavior on the analysis of individual machines presents two potential difficulties: first, single machines don't capture the interaction between strategies. Second, average measures of the accessible states can be misleading, for not all of them are necessarily visited. So how can such analysis be done? In order to solve these issues, this paper uses 'Joint Machines' analysis[84].

3.5.5.2 Joint Machines

The interaction between any two automata can be modeled as a Joint Machine (JM). A JM is a 'meta' machine that represents, in a single automaton, the observed behavior of two automata playing each other. An example is appropriate to understand it.

Figure 3.12, in panels (a) and (b) shows automata for the *No-Signal* treatment. Is not straightforward to understand how are they coordinating by directly observing them, but panel (c) shows the corresponding JM. Both interacting machines start playing B in their initial state, which is represented by a starting state of the JM with action BB. The machine in (a), after observing B, transitions to its last state with action A, while the machine in (b) transitions to a state with action B (also its last). These actions are captured by the second state of the JM, with a joint action of AB. Following the same logic, using the input received by each machine and the state they transition into, the JM captures the actions in states that are visited. In Figure 3.12, by observing the JM in (c), it is easy to notice that after the two initial rounds, both machines will take turns, alternating their coordination point from AA to BB and back to AA, indefinitely. These machines correspond to a Turn-Taking regime.

Notice that JMs' actions are no longer the action of one particular strategy, but those of both interacting machines that are being represented. A state of the JM is given by corresponding states of the two interacting machines. So if the action of the JM is, for example AB, it means that in that particular state one agent plays A and the other B. This representation makes a JM a simpler representation of complex behavior.

---

[84] This approach is an original idea of, and has been developed by, professor John H. Miller. The implementations here are based on his own original algorithms via personal communication.

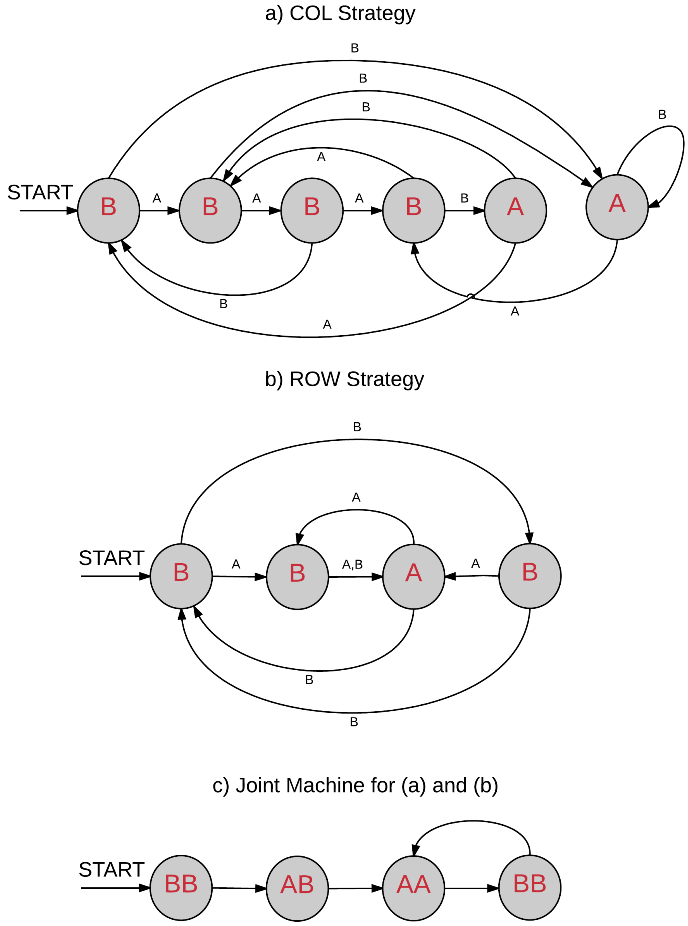


Figure 3.12: Example of Joint Machine for No-Signal treatment. When the machines in (a) and (b) play each other, their interaction can be represented as the Joint Machine in (c). These machines evolved under a Turn-Taking regime.

Without the signal there is no stochastic component, so the JM is completely deterministic (as the one in panel (c) of Figure 3.12). With the signal, transitions of the JM will depend only on the stochastic observed signal H or T. In any case, since the constituent automata are finite, the JM at some point will return to one state-pair that has already been visited, and from there cycle between a subset of states indefinitely[85]. The focus in what follows of this section is on the *Signal* treatment, but some intuition about JMs under *No-Signal* can be found in Appendix 3.7.5.1.

Turn-Taking Joint Machines:

[85] In the formal definition of automata in section 3.4.2, the following are the differences when the automata defined is a JM instead of a single strategy. For both *Signal* and *No-Signal* treatments, the JM actions are $A_i \epsilon \{AA, BB, AB, BA\}$. For *Signal*, now $W = S \in \{H, T\}$, meaning that the machine no longer depends on the input A_{-i} (opponent's action last round) since such information is already contained in the actions of each internal state. Under *No-Signal* the machines are simpler: W is no longer defined since the JM doesn't depend in any input or state of the world. The transitions are deterministic with $\tau: Q \rightarrow Q$, with each state Q having one single transition into another Q.

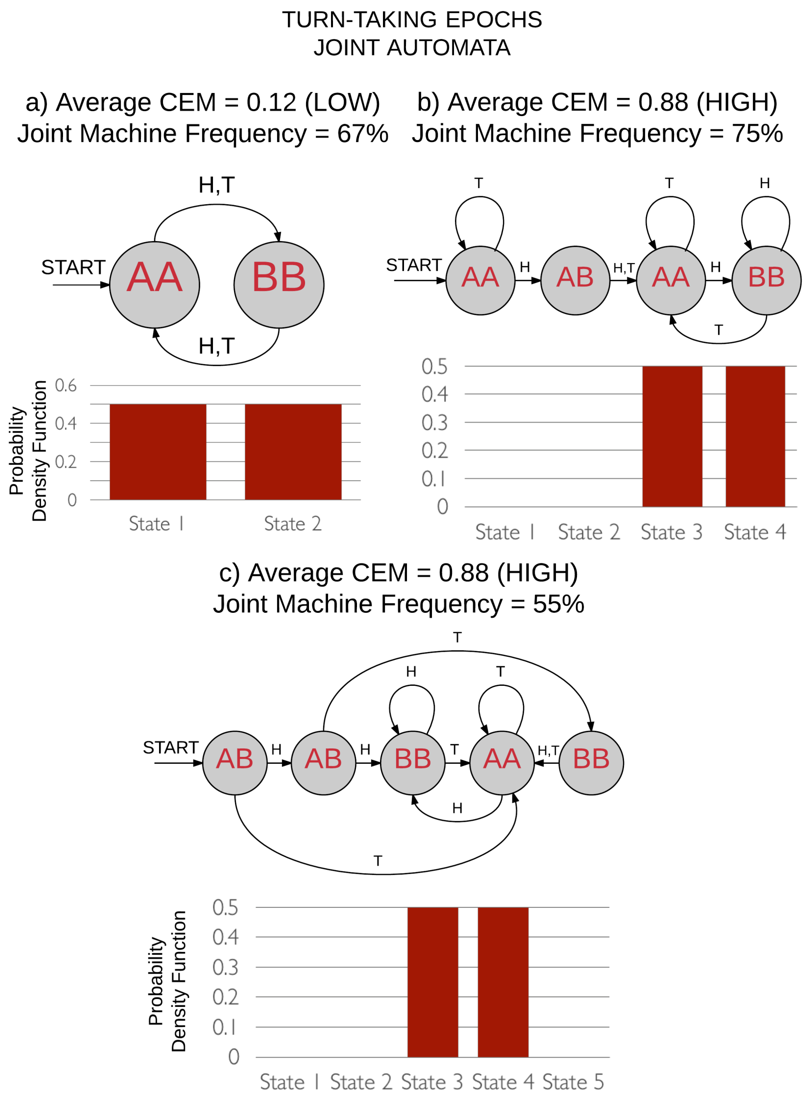


Figure 3.13: Joint Automata that evolved during the Turn-Taking epochs of the Signal treatment. Each machine was picked from the corresponding epoch (with low or high CEM value). One generation was randomly chosen from that epoch, and the most frequent Joint Machine is the one shown. Probability Density Functions show the long-run probability of finding the machine in each particular internal state.

Observe the JM presented in panel (a) in Figure 3.13, which is one of the three representative JMs shown for three different Turn-Taking epochs. The machine has only two states. In the starting one, both machines play A, and whatever the observed signal is (either H or T), it will always transition into the second state. In the second state, the action is BB, and again the transitions are the same regardless of the signal, returning into the initial state. This JM represents two strategies that when interacting will take turns playing (A,A), then (B,B), then (A,A) and so on. Notice that this machine completely ignores the signal, but still manages to perfectly alternate. This is precisely what the CEM captures at the aggregate level. Such machines belong to a Turn-Taking regime with $CEM_T = 0.12$: a low value reflecting that under such regime agents are not relying on the signal to coordinate their actions.

The JMs presented in Figure 3.13 are representative of the behavior observed during each epoch[86]. They were chosen by randomly picking a generation from an epoch with the corresponding CEM and then selecting the most frequent JM. For example, for the machine in panel (a), its frequency is 67%. This means that 67% of all the pairs of strategies playing each other in such generation are described by this automaton[87].

Associated with each machine, there is a Probability Density Function (PDF). It shows the probability of finding the machine in each state in the very long run. States that have zero probability would only be visited before the JM starts cycling, so in the long run their probability tends to zero. Those states with positive probabilities are the ones characterizing the core behavior of the system, and will be referred to as the *cycling states*. Finally, it is worth noting that the cycling states are also very stable across epochs. Even if the JMs don't represent 100% of the interactions, usually the states in the cycle do. Two JMs can have different states before reaching the cycle, but once there, their behavior is very similar. This is the case for JMs in panels (b) and (c), having different states with low probability, but the same cycle. JMs, and particularly the states with positive probabilities in the PDF, are an excellent tool for understanding the micro behavior of the system.

Let us also explain the behavior found in panels (b) and (c). Such JMs give us another formal way to understand the CE learned by the agents. Notice the cycling states (again, the ones with positive probability in the PDF). Even if both machines are from different epochs and have different states, their cycling behavior is identical. In both JMs the behavior alternates between AA and BB depending on the signal: in any of the two cycling states, whenever the signal is T, it will transition to the actions AA. Whenever it is H, it will transition to actions BB. This shows that the machines have learned to interpret the signal and coordinate based on it. As expected, on average, most JMs found under Turn-Taking epochs (the three panels) will play 50% of the times AA and 50% of the times BB. The difference —what is being captured by the CEM—is whether their transitions depend on the signal or not. This can be easily grasped in the JMs by observing the transitions in the cycling states.

Biased Turn-Taking Joint Machines:

[86] A total of four Turn-Taking epochs were identified for the *No-Signal* treatment. Only three machines are shown because the JM that doesn't follow the signal (panel (a)) was found to be representative under two of them. The other two epochs with high CEM values are shown in order to highlight that even if the machines are different, their core behaviour can be the same.

[87] With 40 agents in each population, the total number of possible Joint Machines in each generation is $40 \times 40 = 1{,}600$.

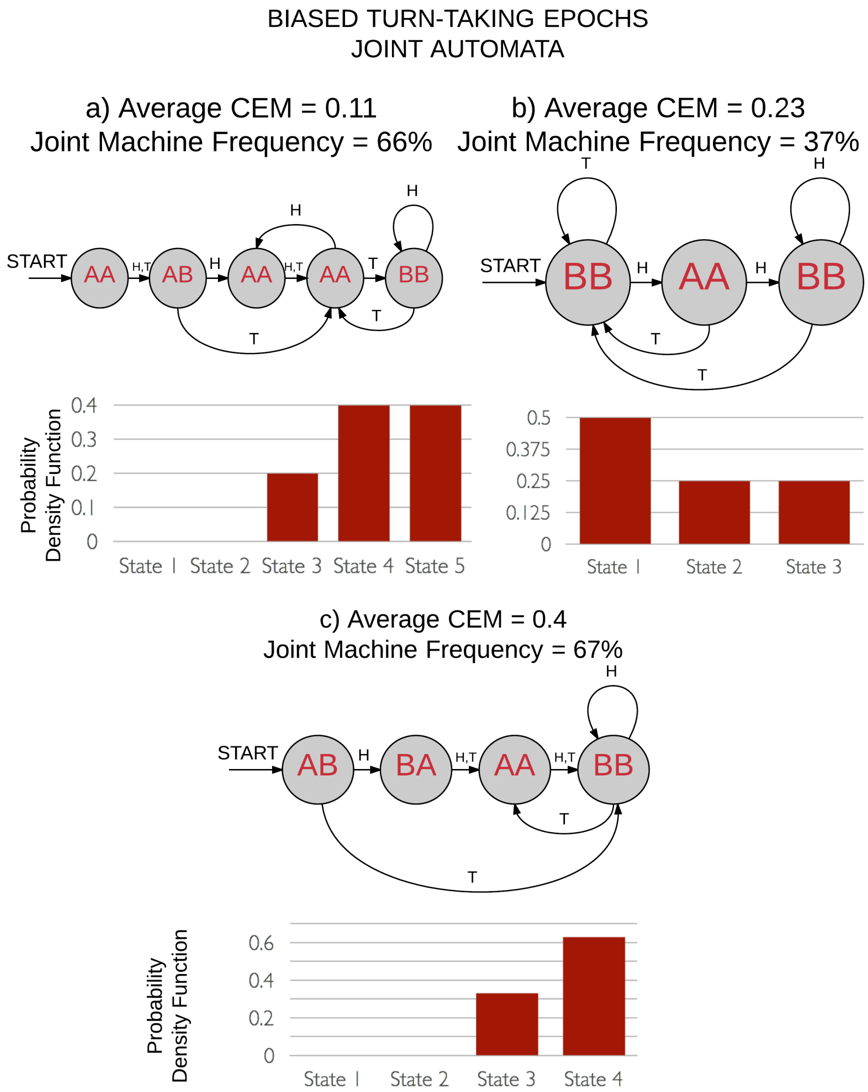


Figure 3.14: Joint Automata that evolved during a Biased Turn-Taking epoch of the Signal treatment. Each machine was picked from an epoch having a different CEM value. One generation was randomly chosen from that epoch, and the most frequent machine is the one shown. Probability Density Functions show the long-run probability of finding the machine in each particular internal state.

The corresponding analysis for the Biased Turn-Taking regime is also presented. Representative JMs for epochs with different CEM values are shown in Figure 3.14. Notice that the ratio of AA to BB actions varies across JMs (observe the probabilities of each machine being in AA or BB during the cycling states). This characteristic is impossible to grasp by observing only the aggregate classification based on the regime.

Each machine uses the signal differently in each state. For example, the JM in panel (a) interprets the signal consistently in state 4 and state 5, but not in its other states. In state 4, it transitions to AA given H and to BB given T. This means that the two constituent strategies found a way to coordinate by using the signal in that particular state. In state 5 an interpretation to the signal is also given, but it is the opposite of that in state 4: BB when H and AA when T. In state 3, the machine completely ignores the signal and always transitions to state 4. Here again, such behavior would have been impossible to observe based only on the aggregate CEM measure of 0.11 (and very difficult to grasp based on the individual machines). The JM's cycling states and the associated PDFs allows an understanding of how partial signal following is happening. Similar intuitions can be made for the other JMs in the Biased Turn-Taking epochs.

Perhaps the reader could have had an accurate a priori intuition of the kind of behavior observed under the Turn-Taking regime based on the values of CEM. Even so, the JMs present a much more intuitive and clear analysis of how machines coordinate. But for the Biased Turn-Taking epochs, such a priori expectations are more unlikely: the varied and perhaps less intuitive ways in which agents follow the signal were not hypothesized and were surprising. Potentially finding *some* strategies that follow the signal and others that don't was initially thought of. But observing such behavior under one single interaction (one single pair of agents) that represents strategies able to follow the signal or ignore it at the *same* time, was unexpected. This is a nice example of how adaptation can come up with marvelous and unexpected solutions that would be difficult to anticipate.

Result 5: Analysis based on Joint Machines (which summarizes any two interacting strategies) is more clear and robust than analyzing individual machines. For the Turn-Taking epochs, such analysis shows how some agents completely ignore the signal and how others perfectly condition on it. For the Biased Turn-Taking epochs, it shows that single machines can at the same time ignore, partially use, or perfectly follow the signal depending on the history of the game (i.e. their internal state).

3.6 Closing Remarks

3.6.1 Summary

This paper uses an explicit evolutionary process, simulated by a genetic algorithm, in order to analyze the effects of an exogenous signal in a repeated Battle of the Sexes coordination game. Its focus is on analyzing the strategies that emerge when coordinating with boundedly rational agents.

Contrary to what was expected, with and without the signal, coordination behavior was quite similar, presenting the same types of equilibria (such as pure Nash and alternation, both symmetrical and asymmetrical). Interestingly, the system doesn't settle down to a single equilibrium, but rather exhibits a constant transition from one to the other. Efficiency in terms of payoffs is the same with and without the signal, meaning that agents coordinate equally well under both setups. The main difference found was the frequency of transitions from equilibrium to equilibrium: with the signal, there is more stability (i.e. less transitions).

Our adaptive agents can indeed learn to coordinate consistently using the signal as a "recommendation" of what to play. However, such behavior is learned very rarely (around 5% of the time), making other strategies a more likely descriptive notion of observed behavior.

This is the first work using adaptive agents in the long run to study coordination games that include a signal. The above results constitute our main findings regarding Correlated Equilibrium. It also analyzes automata by focusing on Joint Machines: ‘meta’ automata that summarizes several interacting agents in a single representation. This analysis permitted us to analyze complex agents. Regarding Correlated Equilibrium, this analysis showed that when the signal is included *some* strategies can alternate their actions by using the signal while *others* can do so by completely ignoring the signal. It also allowed us to see that signal interpretation is not necessarily as intuitive as one might think, and that complex strategies learned to use signals in very different ways. For example, the *same* strategy can, depending on the history of the game, sometimes use the signal as different “recommendations” of play, partially use it, or completely ignore it. This complexity on how strategies use the signal would be difficult to observe without this methodology.

Previous studies of signal use in coordination game have mainly been done with experiments. Conclusions from such experiments show that even though some subjects can indeed learn to alternate their actions by following an exogenous signal, this rarely happens, as they can also alternate by completely ignoring signals (as in Duffy and Feltovich (2010)). One of the main advantages of evolutionary simulations is that they allow agents to learn over considerably longer time spans than in the lab. The limited time span of the lab has led some authors (e.g. Cason and Sharma (2007)) to speculate that signal conditioning would probably be learned much more often if agents were given more time. Our results, however, show that this is not necessarily the case, reinforcing previous results that cast doubt on the notion of Correlated Equilibrium as an accurate description of commonly observed behavior.

3.6.2 Discussion

The Battle of the Sexes game has both coordination and conflict elements (Camerer (2003), p.354; Lau and Mui (2008), p.154). This “mixed motive” social situation arises because both players want to coordinate and choose the same action (a social or shared motive) but also disagree on the activity they want to coordinate on (an individual motive). Our results show that the coordination dimension is solved most of the time, with or without the signal: the system is equally efficient most of the time. But the degree of conflict inherent in the solutions (equilibria) found by the agents can vary at different moments in time. When agents are taking turns symmetrically, they have found a solution without any conflict in terms of received payoffs, but when playing one of the Nash solutions consistently or under asymmetric turn taking, the conflict dimension is not solved. We can make a distinction in the behavior of the system in terms of the time span analyzed. In the short run, coordination seems to dominate over conflict. But since regimes are subject to change and transitions, in the (very) long run the conflict issues are averaged out. So in the long run the system has both coordination an absence of conflict (or efficiency and equality), but at any moment in time only coordination is found for sure.

Regarding the effects of the signal, in theory it could help agents solve both coordination and conflict. However, since agents learn to coordinate quite well without it, the signal is addressing a problem that doesn’t need help to be solved. The signal could also solve the conflict dimension, but in evolutionary terms, it only does that in occasion according to our model. At the heart of this distinction, is the game theoretic induction approach to solve these problems. Theoretically, agents could reason a priori and arrive to a common understanding about how to use the signal to solve both coordination and conflict, hence playing conditioning on the signal. But this would require a lot of reasoning and common knowledge. And notice that such outcome is only one possibility consistent with traditional rationality, since one agent being completely stubborn and only playing its preferred action, with the other complying, is a Nash Equilibrium.

In summary, it seems that the signal doesn't have the expected effect in behavior because agents don't really need it to coordinate. And even if the signal could solve the conflict dimension in the short run, the system can still operate under different degrees of conflict, since it doesn't lead to miscoordination or efficiency losses. In the long run, without the signal, both dimensions are solved, so the introduction of the signal seems redundant.

3.6.3 Future research

One of the main behavioral differences found between the *No-Signal* and *Signal* treatments was the difference in number of transitions. Tests on alternative treatments have been conducted, hinting that such results can be related to how the mutation rates interact with the number of transitions in the machines. However, results are not conclusive. The problem seems more complex than anticipated, requiring the development of better performing software than the one currently being used. Not only being able to run simulations for longer time spans could aid in this regard (which would reduce potential effects of very long epochs), but would also allow more efficient exploration of other potential variables that could also be related[88].

Answering the above is also related to more general questions, to be pursued in the mid and long-term. Recent efforts in evolutionary biology have focused on understanding similar phase transitions in natural systems, and other areas ranging from statistical physics, to artificial life to evolutionary robotics, have already made some contributions in understanding general principles of such changes across domains[89]. The computational nature of our model makes detailed analysis of all its components feasible, at least in principle. Understanding what mutations at the micro level are necessary for the system to transition, what aggregate measures show that the system is "ripe" for a sudden change and what precise evolutionary pathways are followed when this happens, will certainly shed some light not only in better understanding equilibrium behavior in systems with boundedly rational agents, but also into understanding phase transitions in evolutionary, artificial and social systems.

[88] Several of this tools have already been implemented. Some measures such as evolutionary "waste" or inefficiency in the construction of the machines, or unused behaviour related to unvisited states present in the machines (reflecting potential for change in the system) have already been explored. However, their examination is currently very expensive in computational terms, requiring further development on the implemented software.

[89] Solé (2016) presents a recent review of contributions across different fields. Sornette (2004) is an example of how understanding phase transitions is relevant for social sciences, in this case financial markets.

3.7 APPENDIX

3.7.1 *Formal presentation of correlated equilibrium and the one shot BOS game*[90]

3.7.1.1 *Correlated strategy pairs: relations with pure and mixed strategies*

In a game one shot game with two players having two possible actions the general form of a correlated strategy pair is

		Player 2	
		C	D
Player 1	A	p_1	p_2
	B	p_3	p_4

where $p_1 + p_2 + p_3 + p_4 = 1$. Such strategy can be represented as a 4-dimensional vector $\pi = (p_1, p_2, p_3, p_4)$, meaning that (A,C) is played with probability p_1, (A,D) is played with probability p_2, etc.. Under correlated strategy π the expected payoffs or rewards of player i are denoted as $R_i(\pi)$ and calculated with respect to the joint distribution of the actions to be taken. So such payoffs are given by a linear combination of the p_i:

$$R_i(\pi) = p_1 R_i(A, C) + p_2 R_i(A, D) + p_3 R_i(B, C) + p_4 R_i(B, D)$$

Notice the relationship between a correlated strategy pair and other strategy types. If $p_i = 1$ for some i, then the correlated strategy pair is a pair of pure strategies. If π is of the form $(qr, q[1-r], [1-q]r, [1-q][1-r])$ then it corresponds to a pair of mixed strategies. Here, Player 1 takes action A with probability q and Player 2 takes action C with probability r, with such probabilities being independent of the action of the rival. This makes the set of correlated strategy pairs an extension of the set of mixed strategy pairs.

In general, to attain a correlated strategy pair communication is required, with an agreement on it before the game is played. However, the agreement is not (and cannot be made) binding, so players are free to ignore any recommendation.

3.7.1.2 *Conditions for a CE in a 2x2 matrix game*

According to strategy pair $\pi = (p_1, p_2, p_3, p_4)$, Player 1 is recommended (by the randomization device or the external third party) to play A with probability $p_1 + p_2$. Given that Player 1 is recommended to play A, the probability of Player 2 being recommended to play C is $\frac{p_1}{p_1+p_2}$.

In a CE each player should maximise her expected payoffs $R_i(\pi)$ given the recommendation (signal) she receives. So if Player 1 is recommended to play A, her expected payoffs under such a correlated strategy pair are

[90] A textbook presentation on correlated equilibrium can be found in Myerson (1997). The one in this appendix was greatly benefited from the lecture notes of Dr. David Ramsey used at the University of Limerick, found online at http://www3.ul.ie/ramsey/Lectures/Operations_Research_2/gametheory4.pdf (last visited on April 25 of 2016).

$$R_1(\pi) = \frac{p_1 R_1(A,C)}{p_1 + p_2} + \frac{p_2 R_1(A,D)}{p_1 + p_2}$$

If Player 1 ignores her recommendation to play A and she plays B instead, her expected payoffs are

$$R_1(\pi) = \frac{p_1 R_1(B,C)}{p_1 + p_2} + \frac{p_2 R_1(B,D)}{p_1 + p_2}$$

For stability it is required that

$$\frac{p_1 R_1(A,C)}{p_1 + p_2} + \frac{p_2 R_1(A,D)}{p_1 + p_2} \geq \frac{p_1 R_1(B,C)}{p_1 + p_2} + \frac{p_2 R_1(B,D)}{p_1 + p_2}$$

which leads to

$$p_1 R_1(A,C) + p_2 R_1(A,D) \geq p_1 R_1(B,C) + p_2 R_1(B,D)$$

For the sake of completion, notice that the above expression is not defined in the case where $p_1 = p_2 = 0$ since we would be dividing by zero. However, in this case Player 1 is never recommended to play A and this condition might then be ignored.

The same line of argument given above can be used for the conditions corresponding to the following recommendations: i) Player 1 to play A, ii) Player 1 to play B, iii) Player 2 to play C and 4) Player 2 to play D.

Hence, the four condition for a correlated equilibrium, respectively for the above recommendations are:

$$p_1 R_1(A,C) + p_2 R_1(A,D) \geq p_1 R_1(B,C) + p_2 R_1(B,D)$$
$$p_3 R_1(B,C) + p_4 R_1(B,D) \geq p_3 R_1(A,C) + p_4 R_1(A,D)$$
$$p_1 R_2(A,C) + p_3 R_2(B,C) \geq p_1 R_2(A,D) + p_3 R_2(B,D)$$
$$p_2 R_2(A,D) + p_4 R_2(B,D) \geq p_2 R_2(A,C) + p_4 R_2(B,C)$$

There are some relationships between correlated equilibria and other types of equilibria that are worth mentioning. First, any Nash equilibrium pair of strategies is also a correlated equilibrium. Second, a pair of mixed strategies that is not a Nash equilibrium is not a correlated equilibrium. Third, any randomization over Nash equilibria is also a correlated equilibrium. Finally, any randomization over a set of strong Nash equilibria can be attained by joint observation of a public signal[91].

3.7.1.3 Battle of the sexes correlated equilibrium

The CE solution of interest in this paper for the BOS game, as indicated in the main text, is the one given by a fair coin toss as the exogenous signal. Formally, such CE is described as $\pi = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$. This equilibrium is both an *utilitarian* and an *egalitarian* equilibrium. Let us define such properties formally and then use the concrete payoffs examined in this paper in order to derive such solution.

[91] The two last conditions allow the easy graphical representation of the convex hull for correlated equilibria, as in the main text. In such case for the BOS, both (A,C) and (B,D) are strong Nash equilibrium, so any correlated strategy pair that picks (A,C) with probability p and picks (B,D) otherwise, is a correlated equilibrium.

1) **Utilitarian equilibrium**: an equilibrium which maximizes the sum of the expected payoffs of the players

2) **Egalitarian equilibrium:** an equilibrium which maximizes the minimum expected payoff of a player.

Since the expected payoff of players are linear combinations of p_i, the criteria above can be expressed as a maximization of a linear combination of p_i. So equilibria of such types can be derived by defining the problem as a linear programming one. For this, consider the following payoff matrix, with the same rewards of interest as in the main text:

		Player 2	
		A	B
Player 1	A	2,3	0,0
	B	0,0	3,2

The utilitarian equilibrium can be found by solving the following problem:

$$\max z = (2+3)p_1 + (0+0)p_2 + (0+0)p_3 + (3+2)p_4 = 5p_1 + 5p_4$$

subject to

$$p_i \geq 0 \; for \; i = 1,2,3,4$$

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$2p_1 + 0p_2 \geq 0p_1 + 3p_2 \Rightarrow p_1 \geq \frac{3p_2}{2}$$

$$0p_3 + 3p_4 \geq 2p_3 + 0p_4 \Rightarrow p_4 \geq \frac{2p_3}{5}$$

$$3p_1 + 0p_3 \geq 0p_1 + 2p_3 \Rightarrow p_1 \geq \frac{2p_3}{5}$$

$$0p_2 + 2p_4 \geq 3p_2 + 0p_4 \Rightarrow p_4 \geq \frac{3p_2}{2}$$

The first two restrictions represent the conditions for (p_1, p_2, p_3, p_4) to define a joint distribution. The final four conditions are the ones required for the solution to be a correlated equilibrium (as defined before).

One could solve this problem for all p_i, but there is a simpler way if one knows the pure Nash equilibria for this problem. Here, (A,A) and (B,B) are Nash equilibria that maximize the sum of the payoffs to the players over the set of pure strategy pairs. And any randomization over these two Nash equilibria is a correlated equilibrium that gives the same sum of payoffs. Hence, any π of the form $\pi = (p, 0, 0, 1-p)$ is a utilitarian equilibrium. So $\pi = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$ is a utilitarian equilibrium.

Let's turn now to the egalitarian equilibrium. For this, it is convenient to notice that the BOS game is not symmetric but still has a degree of symmetry. A 2x2 game where both players can choose either action A of action B will be called *quasi-symmetric* if the following conditions hold (which is indeed the case for BOS):

$$R_1(i,j) = R_2(j,i)$$

$$R_1(i,i) = R_2(j,j), where\ i \neq j, and\ i,j \in \{A,B\}$$

In words, this means that a payoff vector on the leading diagonal is the reverse of the other payoff vector on that diagonal.

As a result, at an egalitarian equilibrium of a quasi-symmetric game both players must obtain the same expected payoffs. So to find an egalitarian equilibrium of a quasi-symmetric game, the problem is to maximize the expected sum of the payoffs using the same constrains as before, but adding a new one: that both players should obtain the same payoffs. Hence the problem, is

$$\max z = 5p_1 + 5p_4$$

subject to (as before)

$$p_i \geq 0\ for\ i = 1,2,3,4$$

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$2p_1 + 0p_2 \geq 0p_1 + 3p_2 \Rightarrow p_1 \geq \frac{3p_2}{2}$$

$$0p_3 + 3p_4 \geq 2p_3 + 0p_4 \Rightarrow p_4 \geq \frac{2p_3}{5}$$

$$3p_1 + 0p_3 \geq 0p_1 + 2p_3 \Rightarrow p_1 \geq \frac{2p_3}{5}$$

$$0p_2 + 2p_4 \geq 3p_2 + 0p_4 \Rightarrow p_4 \geq \frac{3p_2}{2}$$

and adding the condition

$$2p_1 + 3p_4 = 3p_1 + 2p_4 \Rightarrow p_1 = p_4$$

As before, any correlated equilibrium of the form $(p, 0, 0, 1-p)$ maximises the sum of expected payoffs. And observing that setting $p = \frac{1}{2}$ holds for that new last condition, one can then define $\left(\frac{1}{2}, 0, 0, \frac{1}{2}\right)$ as the egalitarian equilibrium of interest.

3.7.2 Additional overview of average payoffs

Statistical analyses of the model are based on the simulations presented on Fig. A. One very long run with 100,000 generations is run for each treatment. Fig. B presents five shorter simulations for the *Signal* treatment, analogous to the figure presented on the main text for *No-Signal*.

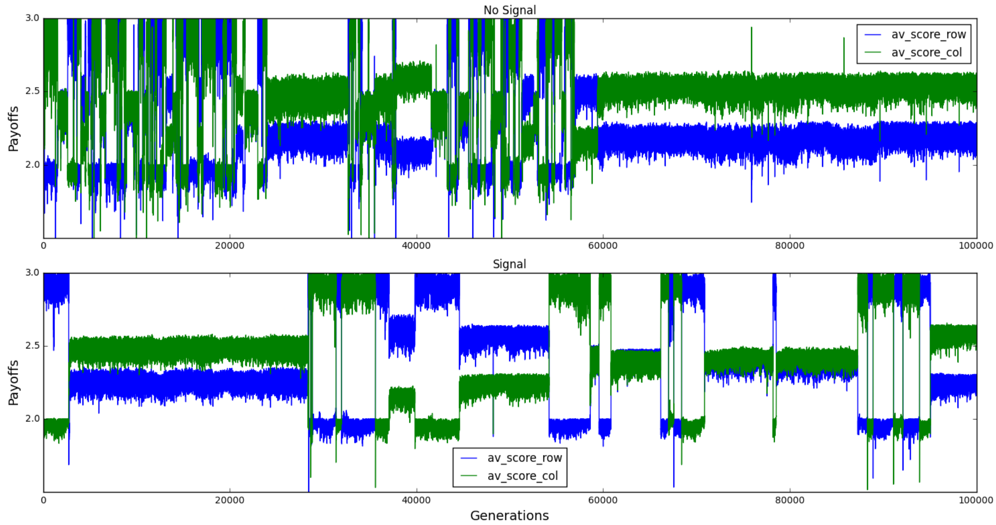


Fig. A: Average payoffs per population. Longest simulations for both Signal and No-Signal treatment with 100,000 generations. Statistical analyses in the main text are based on these runs of the model.

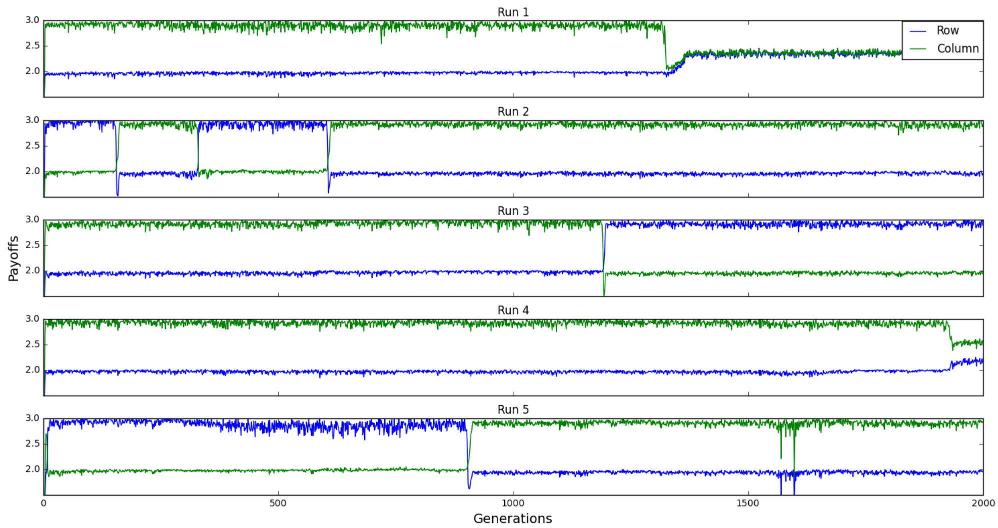


Fig. B: Signal treatment. Average payoffs per population. Each panel is one different run of the model, each consisting of 2,000 generations.

3.7.3 Correlated Equilibrium Measure (CEM)

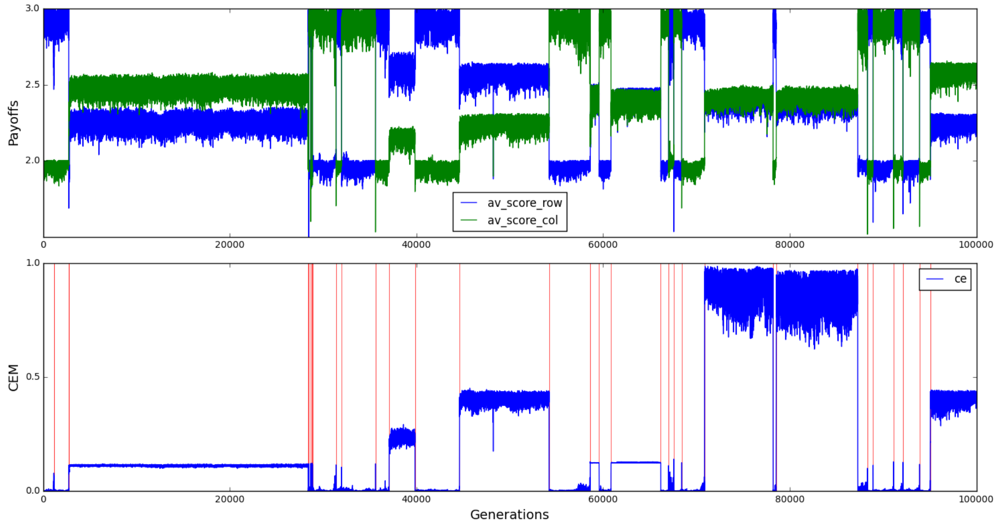


Fig. C: Average Payoffs on Signal treatment (top panel) vs. Correlated Equilibrium Measure CEM (bottom panel). Vertical lines in the bottom panel indicate the end of an epoch. It can be seen that within single epochs, the CEM is quite stable, meaning that agents use (or not) the signal in the same way consistently under each single epoch.

3.7.4 A learning test

Why agents don't learn to play CE more often? If the cause is that the algorithm finds it difficult to explore the larger strategy space when the signal is included, then not finding CE more often wouldn't be due to an interesting feature of the strategic interactions of the agents, but rather to having an inefficient (or perhaps wrongly 'tuned') learning mechanism. In order to address this concern, a test for the model in the *Signal* treatment was run by changing the payoffs of the game. The test is implemented by modifying the payoffs depending on the outcome of the signal in each round as follows:

Payoffs if Signal = Heads

		Player 2	
		A	B
Player 1	A	3,3	0,0
	B	0,0	0,0

Payoffs if Signal = Tails

		Player 2	
		A	B
Player 1	A	0,0	0,0
	B	0,0	3,3

Notice that with these payoffs there's no conflict of interests between the agents. If they are able to follow the signal in this environment, it means the algorithm is not having difficulties exploring the larger strategy space (compared to *No-Signal*). The model was run five different times up to 2,000 generations. Under this setup both generations will have the exact same payoffs and the Pareto optimal is now three for both populations.

In all of the simulations agents quickly learned to follow the signal and coordinate appropriately. After some generations (around 30) the system's behaviour becomes stable. No transitions are observed, average payoffs settle very close (2.85) to the Pareto optimal and the average CEM_T is very close to one (equals 0.9). This indicates that the learning mechanism has no problems finding CE strategies. If agents don't learn CE is due to the strategic environment, not due to something inherent to the implementation of the GA. Fig. D shows graphically this information (only 200 generations are reported due to the model becoming very stable and not presenting relevant changes).

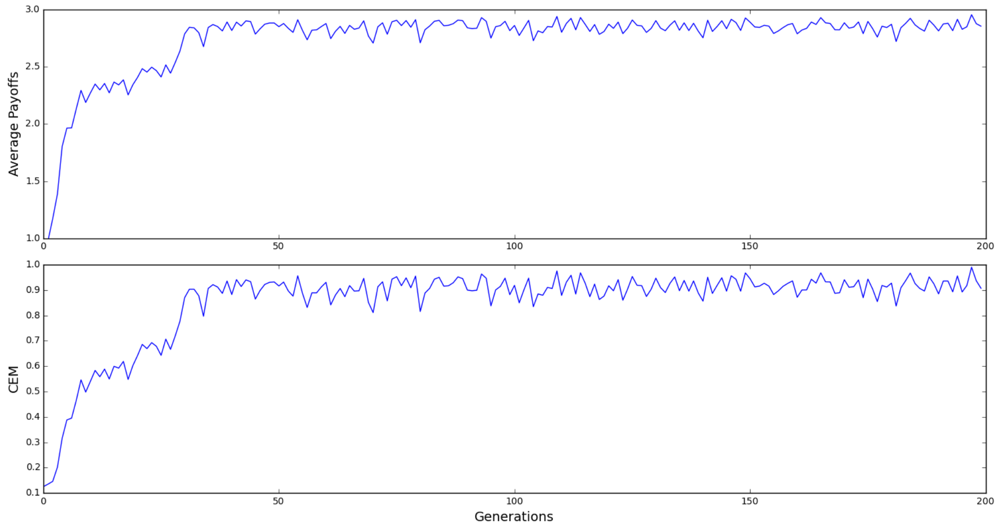


Fig. D: Average payoffs and CEM for the learning test. Around generation 30 agents have learned to follow the signal almost perfectly, shown by the high payoffs and high CEM.

3.7.5 Additional Joint Automata (JM)

3.7.5.1 No-Signal JMs

Compared to the JMs with the signal, the ones without are much simpler due to their deterministic nature. Fig. E presents three typical JMs that evolved, one under each regime. All JMs without the signal have a very similar "lollipop" shape: they visit several states in order, and at their end (since the automata are finite) they transition back to one that was previously visited. This last transition marks the beginning of a cycling behavior, meaning that the machine will forever repeat its actions. In our analysis, usually the JM takes one or more states that can include some miscoordination, but then enters the cycle and coordinates in a way reflected by the corresponding regime.

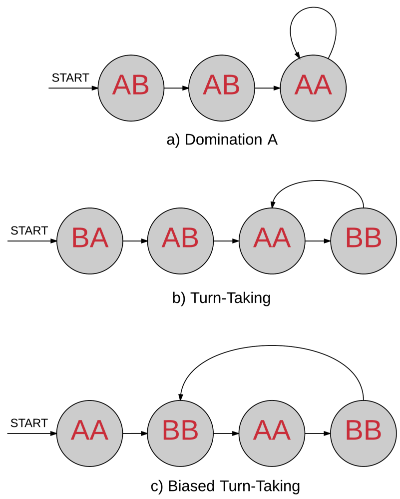


Fig. E: Joint Automata under No-Signal treatment. Transitions are deterministic. The machines will eventually come back to an already visited state, cycling forever into a subset of states. Each Joint Automata corresponds to the indicated regime.

REFERENCES

ANBARCI, N., FELTOVICH, N., GÜRDAL, M.Y., (2015). "Payoff Inequity Reduces the Effectiveness of Correlated-Equilibrium Recommendations." *Work. Pap. Monash Univ.,.*

ANDREONI, J., MILLER, J.H., (1995). "Auctions with Artificial Adaptive Agents." *Games Econ. Behav.*, Vol. 10, pp. 39–64.

ARIFOVIC, J., (1994). "Genetic algorithm learning and the cobweb model." *J. Econ. Dyn. Control*, Vol. 18, pp. 3–28.

ARIFOVIC, J., BOITNOTT, J.F., DUFFY, J., (2015). "Learning Correlated Equilibria: An Evolutionary Approach." *Work. Pap. Simon Fraser Univ.,.*

AUMANN, R.J., (1974). "Subjectivity and correlation in randomized strategies." *J. Math. Econ.*, Vol. 1, pp. 67–96.

AUMANN, R.J., (1987). "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica*, Vol. 55, pp. 1–18.

AXELROD, R., (1980). "Effective Choice in the Prisoner's Dilemma." *J. Conflict Resolut.*, Vol. 24, pp. 3–25.

AXELROD, R., (1986). "An Evolutionary Approach to Norms." *Am. Polit. Sci. Rev.*, Vol. 80, pp. 1095–1111.

BONE, J., DROUVELIS, M., RAY, I., (2013). "Co-ordination in 2 x 2 Games by Following Recommendations from Correlated Equilibria." *Work. Pap. 12-04R, Univ. Birmingham,.*

BROWNING, L., COLMAN, A.M., (2004). "Evolution of coordinated alternating reciprocity in repeated dyadic games." *J. Theor. Biol.*, Vol. 229, pp. 549–57.

CAMERER, C.F., (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press.

CASON, T.N., SHARMA, T., (2007). "Recommended play and correlated equilibria: an experimental study." *Econ. Theory*, Vol. 33, pp. 11–27.

DUFFY, J., (2006). "Chapter 19 Agent-Based Models and Human Subject Experiments," in: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics. Elsevier, pp. 949–1011.

DUFFY, J., FELTOVICH, N., (2010). "Correlated Equilibria, Good and Bad: An Experimental Study." *Int. Econ. Rev. (Philadelphia).*, Vol. 51, pp. 701–721.

DUFFY, J., LAI, E.K., LIM, W., (2014). "Language and Coordination: An Experimental Study." *Work. Pap. Univ. Pittsbg.,.*

FOSTER, D.P., VOHRA, R. V., (1997). "Calibrated Learning and Correlated Equilibrium." *Games Econ. Behav.*, Vol. 21, pp. 40–55.

FUDENBERG, D., LEVINE, D.K., (1999). "Conditional Universal Consistency." *Games Econ. Behav.*, Vol. 29, pp. 104–130.

GIGERENZER, G., HERTWIG, R., PACHUR, T., (2011). Heuristics: The foundations of adaptive behavior. Oxford University Press, New York.

GIGERENZER, G., TODD, P., GROUP, A.B.C.R., (2002). Simple Heuristics That

Make Us Smart. Oxford University Press.

GODE, D.K., SUNDER, S., (1993). “Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality.” *J. Polit. Econ.*, Vol. 101, pp. 119–137.

HANAKI, N., (2006). “Individual and Social Learning.” *Comput. Econ.*, Vol. 26, pp. 31–50.

HARRISON, M.A., (1965). “Introduction to switching and automata theory.” McGraw-Hill, New York.

HART, S., MAS-COLELL, A., (2000). “A Simple Adaptive Procedure Leading to Correlated Equilibrium.” *Econometrica*, Vol. 68, pp. 1127–1150.

HOLLAND, J.H., (1975). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. U Michigan Press.

HOLLAND, J.H., (1992). “Genetic Algorithms.” *Sci. Am.*, Vol. 267, pp. 66–72.

HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E., THAGARD, P.R., (1986). Induction: processes of inference, learning, and discovery. MIT press, Cambridge, MA.

HOLLAND, J.H., MILLER, J.H., (1991). “Artificial Adaptive Agents in Economic Theory.” *Am. Econ. Rev.*, Vol. 81, pp. 365–370.

IOANNOU, C.A., (2013). “Coevolution of finite automata with errors.” *J. Evol. Econ.*, Vol. 24, pp. 541–571.

IOANNOU, C.A., ROMERO, J., (2014a). “A generalized approach to belief learning in repeated games.” *Games Econ. Behav.*, Vol. 87, pp. 178–203.

IOANNOU, C.A., ROMERO, J., (2014b). “Learning with repeated-game strategies.” *Front. Neurosci.*, Vol. 8, pp. 212.

KANDORI, M., MAILATH, G.J., ROB, R., (1993). “Learning, Mutation, and Long Run Equilibria in Games.” *Econometrica*, Vol. 61, pp. 29–56.

KIRMAN, A., VRIEND, N., (2000). “Learning to Be Loyal. A Study of the Marseille Fish Market,” in: Gatti, D., Gallegati, M., Kirman, A. (Eds.), Interaction and Market Structure. Springer Berlin Heidelberg, pp. 33–56.

KOLLMAN, K., MILLER, J.H., PAGE, S.E., (1992). “Adaptive Parties in Spatial Elections.” *Am. Polit. Sci. Rev.*, Vol. 86, pp. 929–937.

KOLLMAN, K., MILLER, J.H., PAGE, S.E., (1997). “Political Institutions and Sorting in a Tiebout Model.” *Am. Econ. Rev.*, Vol. 87, pp. 977–992.

LAU, S.-H.P., MUI, V.-L., (2008). “Using Turn Taking to Mitigate Coordination and Conflict Problems in the Repeated Battle of the Sexes Game.” *Theory Decis.*, Vol. 65, pp. 153–183.

LEYTON-BROWN, K., SHOHAM, Y., (2008). Essentials of game theory: Synthesis Lectures on Artificial Intelligence and Machine Learning, 1st editio. ed, Synthesis Lectures on Artificial Intelligence and Machine Learning (book 3). Morgan and Claypool Publishers.

MILLER, J.H., (1988). “The Evolution of Automata in the Repeated Prisoner’s Dilemma.” *Two Essays Econ. Imperfect Information, Ph.D.University Michigan*,.

MILLER, J.H., (1996). “The coevolution of automata in the repeated Prisoner’s

Dilemma." *J. Econ. Behav. Organ.*, Vol. 29, pp. 87–112.

MILLER, J.H., BUTTS, C.T., RODE, D., (2002). "Communication and cooperation." *J. Econ. Behav. Organ.*, Vol. 47, pp. 179–195.

MILLER, J.H., MOSER, S., (2004). "Communication and coordination." *Complexity*, Vol. 9, pp. 31–40.

MITCHELL, M., (1998). An introduction to genetic algorithms. MIT press.

MOORE, E.F., (1956). "Gedanken-experiments on sequential machines." *Autom. Stud.*, Vol. 34, pp. 129–153.

MYERSON, R.B., (1997). Game Theory: Analysis of Conflict (1st paperback edition). Harvard University Press.

POTEETE, A.R., JANSSEN, M.A., OSTROM, E., (2010). Working Together: Collective Action, the Commons, and Multiple Methods in Practice. Princeton University Press.

RUBINSTEIN, A., (1986). "Finite automata play the repeated prisoner's dilemma." *J. Econ. Theory*, Vol. 39, pp. 83–96.

SOLAN, E., VOHRA, R. V., (2002). "Correlated equilibrium payoffs and public signalling in absorbing games." *Int. J. Game Theory*, Vol. 31, pp. 91–121.

SOLÉ, R., (2016). "Synthetic transitions: towards a new synthesis." *Philos. Trans. R. Soc. London B Biol. Sci.*, Vol. 371.

SORNETTE, D., (2004). Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press.

ZHANG, W., (2015). "Can Errors Make People More Cooperative? Cooperation in an Uncertain World." *Work. Pap. Purdue Univ.*,.