ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Lee-Penagos, Alejandro

Working Paper

Modelling contributions in public good games with punishment

CeDEx Discussion Paper Series, No. 2016-15

Provided in Cooperation with:
The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx)

Suggested Citation: Lee-Penagos, Alejandro (2016) : Modelling contributions in public good games with punishment, CeDEx Discussion Paper Series, No. 2016-15, The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx), Nottingham

This Version is available at:
https://hdl.handle.net/10419/163016

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Discussion Paper No. 2016-15

Alejandro Lee-Penagos

October 2016

Modelling Contributions in Public Good Games with Punishment

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
Fax: +44 (0) 115 95 14159
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

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.

1 INTRODUCTION

Experiments with the Voluntary Contribution Mechanism (VCM)[1] 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"[2].

In order to close this gap, Arifovic and Ledyard (2012) (AL from now on) have developed IELORP[3], 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

[1] 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.

[2] pp. 541

[3] 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[4]. 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), we attempt to replicate AL's model and test its previously reported results (section 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 4), we formally present our model and defend its methodology (section 5). Here is worth mentioning our calibration strategy (section 6.1). One of the main reasons for using IELORP as our starting point for modelling behaviour, is the previously reported

[4] 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 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 heuristics, can account for the most relevant facts observed in the repeated public goods games experimental data.

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.

2.1 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 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[5]. The specific values of the parameter triplet (P, B, G) are discussed in section 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 inner traits of the agents. An agent with the same "inner" parameters (i.e. β^i

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

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)[6].

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.

2.3 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.

2.3.1 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

[6] The exact conditions for each type are presented in Appendix 8.2.

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

1. *Experimentation*[7]*:* 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 $a_{j,t}^i$ is replaced by a new contribution strategy selected at random from X^i. The distribution used for this replacement is normal $\sim N(a_{j,t}^i, \sigma)$. 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 $v^i(a_{jt}^i | I^i(x_t))$ 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 $v^i(a_{jt}^i | I^i(x_t))$ 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 $j = 1,\dots,J$, $a_{j,t+1}^i$ is chosen as follows. From a uniform distribution, pick randomly (with replacement) two members of A_t^i. Let such two members be $a_{k,t}^i$ and $a_{l,t}^i$. Then

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

   Replication in period $t+1$ favors alternatives with many replicates in A_t^i 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_t^i. A_t^i 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

[7] 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.

2.3.2 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:

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

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.

2.4 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. Both ideas are well stablished in the literature and have been treated in previous models.

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 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 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.

3 TESTING IELORP PREVIOUS EVIDENCE

AL's claim is that IELORP can track several stylized facts in VCM experimental data (introduced in section 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[8]. 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.

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 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 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 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.

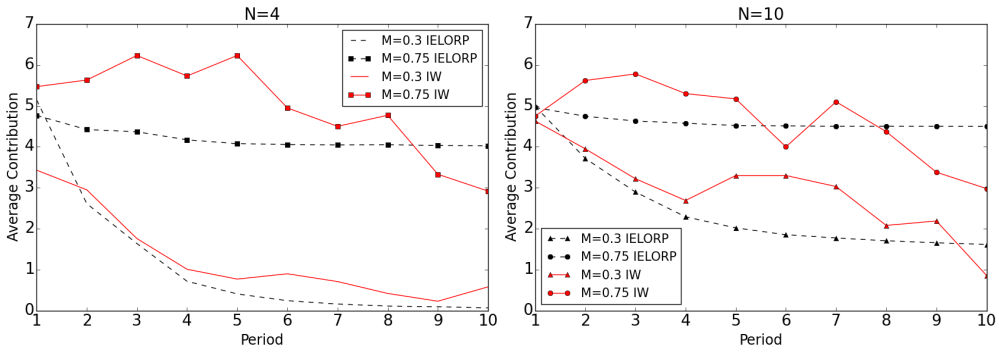


Figure 1: Comparison of our independent replication of IELORP (simulations) versus experimental data of Isaac and Walker (1988) *(IW). Includes four treatments: group size taking values N=4 (left panel) and N=10(right panel), and marginal productivity of the public good taking values M= (0.3,0.75).*

[8] 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).

So Figure 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 can replicate such findings. For this, simulations were run with our independent implementation of the model[9]. 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[10]. 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 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?

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 $\bar{c}^{10}_{sim}(r)$ 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 $\bar{c}^{3}_{sim}(r)$ be the analogous but only for the average of the last three periods, and $\bar{c}^{10}_{IW}(r)$ and $\bar{c}^{3}_{IW}(r)$ 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

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

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

[9] N artificial agents are created for each simulation, endowing them with Other Regarding Preferences parameters as explained in section 2.2, and playing for 10 periods. A new draw of parameters is done for each different run.

[10] $J = 100, \rho = 0.033, \sigma = \frac{w}{10} = 1, P = 0.48, B = 22, G = 8$. 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.

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

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[11].

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[12]. Some other important experimental results will be referred to indirectly, but such discussion if left for section 5.3.

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

4.1 The punishment experimental setup

A traditional public goods game experiment with punishment works in the following way[13]. After players have decided on their contributions as they would if punishment is not allowed (i.e. in the VCM environment presented in section 2.1), 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

[11] 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 8.1.

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

[13] This notation follows that of Nikiforakis and Normann (2008), since their data set is the one used later for testing the model (section 6).

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 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.

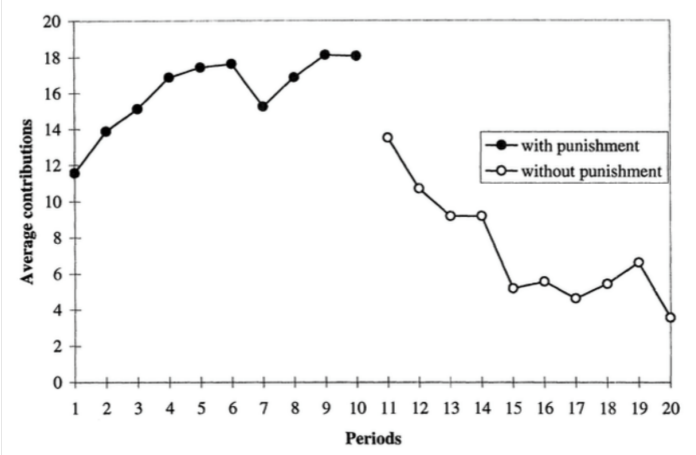
4.2 Stylized facts

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 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)

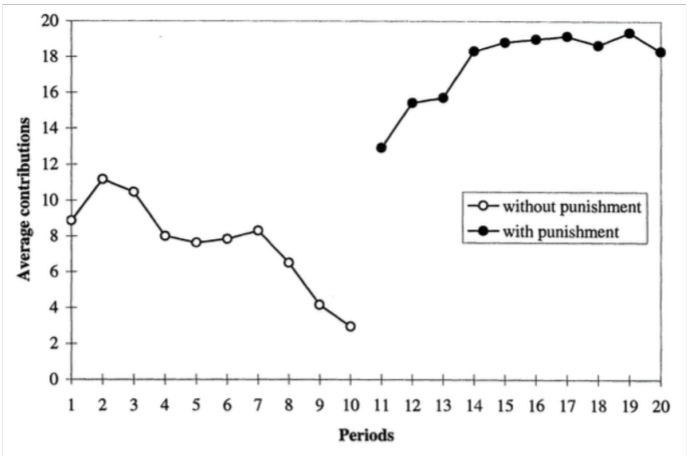


Figure 2: Average contributions over time (partners setup). Sequence 1 had subjects play 10 rounds with punishment allowed, followed by 10 rounds without punishment. Sequence 2 reversed that order. Source: Fehr and Gächter (2000)

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 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.

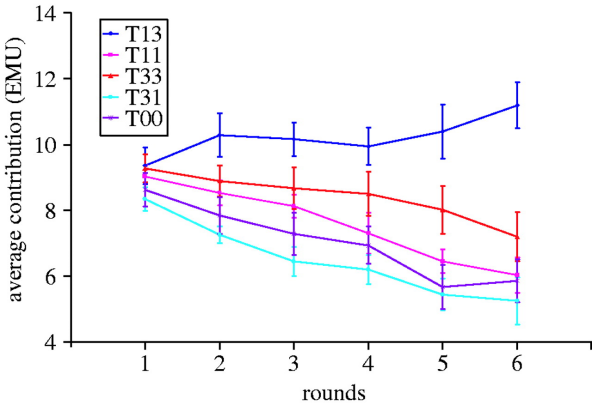


Figure 3: Average contributions in a public goods game. Each treatment T changes how much it costs one player to buy one token of punishment as well as how much that token deducts the payoffs of the punished player. For example, T31 means that one has to pay three tokens (cost) in order to deduct one 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 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.

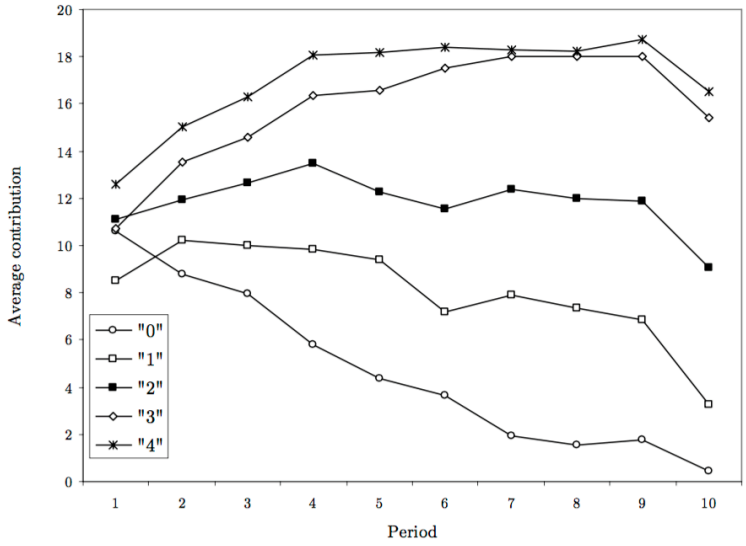


Figure 4: Average contributions in a public goods game with punishment. Each line presents a treatment with different effectiveness levels. For example, treatment "3" means that for each token of punishment assigned, the punished player payoffs are reduced by 3. Source: Nikiforakis and Normann (2008).

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.

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 3 and Figure 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.

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 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 established, punishment was unambiguously beneficial. The explanation is that once high levels of contributions are established, 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.

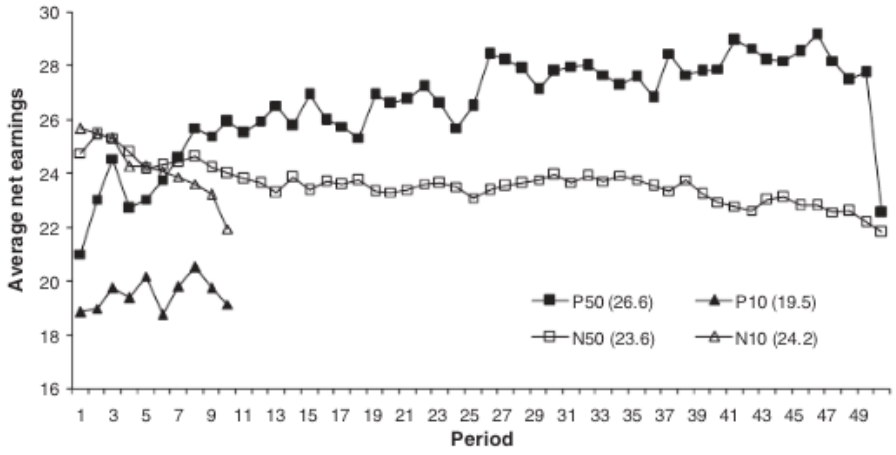


Figure 5: Average net earnings in a public goods game. Treatments included both no punishment and punishment setups ('N' and 'P' respectively), as well as time spans of both 10 and 50 rounds. For example, treatment P50 means that punishment was allowed and that the game lasted 50 rounds. Numbers in parenthesis 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.

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 5.3 will discuss this and other points, after PH is presented.

5.1 Motivation for punishment as a simple rule of thumb

How do subjects across experiments decide on which other players to punish? Figure 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 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.

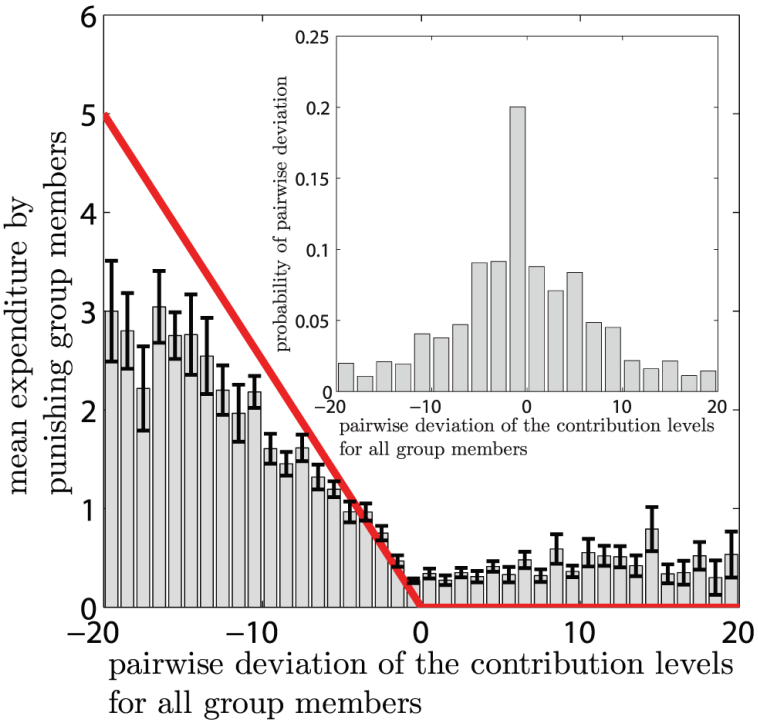


Figure 6: Mean expenditure of a given punishing member as a function of the deviation between that member's contribution and that of the punished member (for all pairs of subjects within a group). The straight line crossing zero shows the average decision rule for punishment: the more negative the deviations, 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[14].

The above rule of thumb is closely related to how different effectiveness levels affect punishment and the sustainability of cooperation (stylized fact in section 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 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*

[14] Figure 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"[15]. 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 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*"[16]. 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.

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[17]. L represents how the tolerance T of players changes in response to different values of e.

---

[15] pp.875, italics added.

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

[17] 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[18]. 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 $z_t^i = \left(R_i^t - a_t^i\right)K$, 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

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

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 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:

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

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

[18] 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[19]. 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 [20]. 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.

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

[19] 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.

[20] 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.[21]

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

[21] 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 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.

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 4.2 that are not addressed with the main data.

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[22]. So here the approach followed is to test PH using exactly the same parameter numbers estimated by AL. The main goal of doing

[22] As mentioned in Appendix 8.1.

this is that the amount of degrees of freedom is reduced to only two (PH's free parameters).[23]

Table 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 1: Parameter values for PH. IEL and ORP values taken from Arifovic and Ledyard (2012). *PH values estimated*

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 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)[24]. 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.

[23] 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.

[24] As discussed in section 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.

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. A grid search was conducted for the duplet (K, L) selecting the values with better fit to NN data[25]. 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 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[26]

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

where R is the total number of treatments. For the present case of NN, $R = 4$ (treatments 1,2,3 and 4)[27]. 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 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.

[25] This procedure follows AL calibration for IELORP.

[26] 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.

[27] 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.

6.2 *Simulations and statistical tests*

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 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.

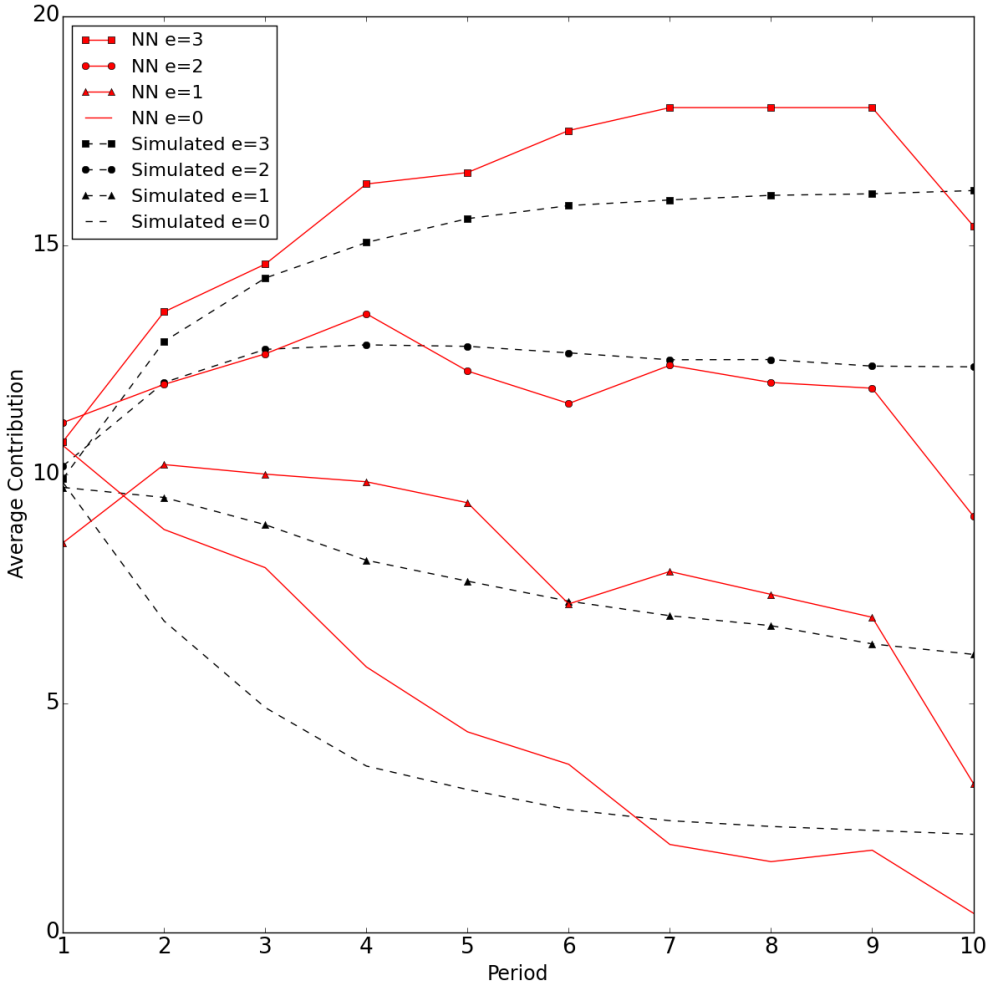


Figure 7: Experimental vs. simulated data for treatments e=0, e=1, e=2 and e=3. Simulations generated with PH (blacked dashed lines). Experimental data source is* Nikiforakis and Normann (2008) *(NN, red lines).*

Figure 7 shows how close each treatment is replicated by the model[28]. 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 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 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 6.3.

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 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 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.*

[28] 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 less than 5% (or 10%) significance level. What is happening in the data? Observing Figure 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[29], 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.

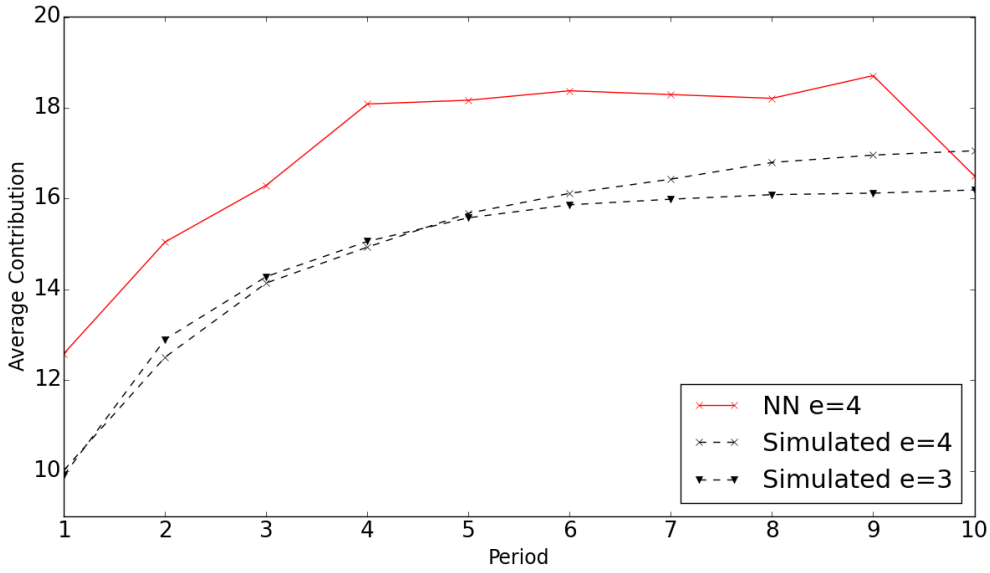


Figure 9: Experimental vs. simulated data for Treatment e=3 (simulations) and e=4 (simulations and experimental data). Simulations are generated with PH (dashed, black lines). Experimental data source is* Nikiforakis and Normann (2008) *(NN, red line).*

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

[29] 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 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.

6.3 Further analysis of stylized facts

As seen above (Figure 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[30].

[30] 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 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.

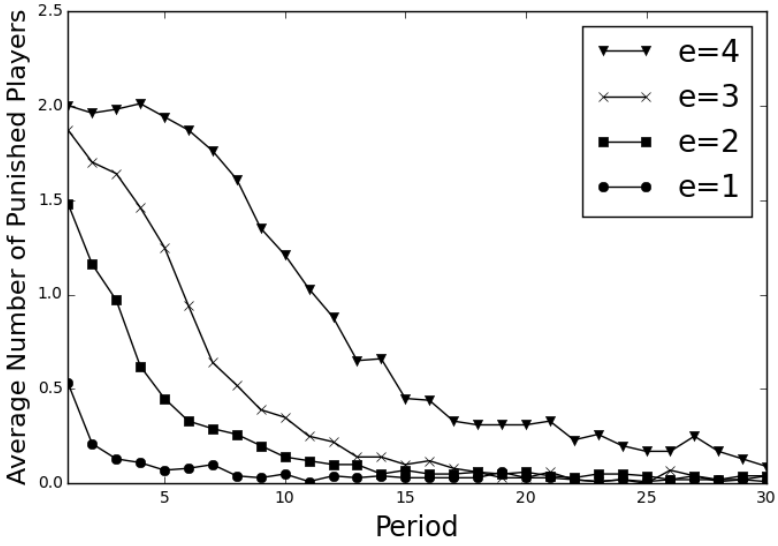


Figure 10: Average number of agents punished for different effectiveness levels (across 100 simulations per treatment). Punishment is modeled: an agent punishes any other in the group that contributes below the group average minus the tolerance level T.

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[31]. 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 established, 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

[31] 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.

8 APPENDIX

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[32].

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

[32] 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.

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 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.

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.*,.