Identification of Other-regarding Preferences: Evidence from a Common Pool Resource game in Colombia*

Sandra Polanía-Reyes
University College London†

David Echeverry
University of California Berkeley‡

March 27th, 2015

Abstract

Social preferences have been important to the explanation of deviations from Nash equilibrium in game outcomes. An enduring challenge in any model of other-regarding preferences is to identify heterogeneity within the population. Using data from a common pool resource (CPR) game in the field with $1,095$ individuals (21% students and 79% villagers, users of a CPR) we estimate a structural model including preferences for altruism, reciprocity and equity. We identify behavioral types using a latent class logit model. Exogenous determinants of type are examined such as socio-economic characteristics, perceptions on the CPR, perceived interest in cooperation among the community, whether the participant does volunteer work and whether the CPR is the household main economic activity of the household.

A competing explanation of deviations from Nash equilibrium is the existence of a cognitive factor: the construction of a best reply might make rational expectations about other players' mistakes (e.g. quantal response equilibrium). Whilst a cognitive aspect would help the model better fit the data, we do not find much evidence for cognitive heterogeneity, and instead a great deal of behavioral heterogeneity. Choice prediction based on types is robust out of sample.

JEL classification: Q2, C51, C23, C93, D64, H39, H41.

Keywords: Common-pool resources, social preferences, laboratory and field experiments, explicit incentives, inequity aversion, latent heterogeneity, finite mixture models.

*We thank Juan-Camilo Cárdenas, Shachar Kariv and César Mantilla for valuable comments and suggestions. We also thank workshop participants at BABEW (Stanford University) as well as conference participants at IAAE (Queen Mary University) and University of Göttingen for helpful comments.

†Email: sandra.polania-reyes.09@ucl.ac.uk. I am grateful to my advisors Orazio Attanasio, Sam Bowles and Syngjoo Choi. I thank Juan Camilo Cárdenas, Steffen Huck, Adriana Molina and Ana Ma. Roldán. I acknowledge financial support to the field work from the MacArthur Foundation, University of los Andes and Javeriana University in Colombia.

‡Email: david.echeverry@berkeley.edu.

1 Introduction

Common pool resources (CPR) are held by a collective of individuals for each of them to extract from in order to derive an individual payoff. At their heart lies the danger that profit-maximizing individual behavior leads to depletion of the common pool resource, and hence to a loss of utility for all the agents. The social dilemma that arises from the wedge between the Nash equilibrium (NE) and the social optimum is a key concern for the governance of the commons. The corresponding (positive) question of how agents respond to social dilemmas in real life is key to make inference about the motivants of individual behavior.

Consistent deviations from NE have been documented in the empirical literature (Rassenti et al., 2000). The existence of pro-social behavior has been widely documented, suggesting that social preferences are important influences on economic behavior Fehr et al. (1997); Bewley (1999); Fehr and Schmidt (1999); Fehr and Gächter (2000); Sobel (2005). An individual behaves pro-socially in order to help others -including himself- to achieve a common good. Social preferences are those concerns for the well-being of others and desires to uphold ethical norms. They reduce social inefficiency in the absence of complete contracts (Arrow, 1971; Becker, 1976; Akerlof, 1984) and thus are key to solve social dilemmas (Ostrom, 1990), in which the uncoordinated actions of individuals result in an outcome that is Pareto inefficient.

Charness and Rabin (2002) find empirical evidence about preferences for altruism (social welfare in their setting), for reciprocity and to a lesser extent against inequality. Similarly, Bolton and Ockenfels (2000) argue for other-regarding preferences in the sense of concern for total welfare. Fehr and Schmidt (1999) argue inequity aversion is important. Since then, structural models of pro-social behavior Falk and Heckman (2009); Manski (2011); Andreozzi et al. (2013) have been used to assess the prevalence of a given type of social preferences.

Social preferences have been studied in the context of public goods games (Isaac and Walker, 1988). Common Pool Resource games have been implemented by assuming homogeneous preferences either with only students (Fischbacher et al., 2001; Kurzban and Houser, 2001; Carpenter et al., 2009; Falk et al., 2002) based on Walker et al. (1990) or only real users (see evidence in forest management by Rustagi et al. (2010) or Margreiter et al. (2005); Vélez et al. (2009)). In our dataset, users deviate more from the NE than students (Cárdenas (2004), 2011) which is in line with other empirical findings (Carpenter and Seki, 2010; Molina, 2010; Cárdenas et al., 2013). Several of these papers explain their findings with the existence of social preferences but do not explore the role of heterogeneity.

Yet heterogeneity of social preferences is both a prevalent and a relevant phenomenon (Charness and Rabin, 2002; Leider et al., 2009; Manski and Neri, 2013; Kurzban and Houser, 2005; Goeree et al., 2002; Burlando and Guala, 2005). Using a random coefficient model, Polania-Reyes (2014) identifies types using the same classification as we do. Specifications similar to ours have been used before. Similarly, Vélez et al. (2009) use a random effects specification to assess the prevalence of different social preferences, finding evidence of preferences for conformity. Rodriguez-Sickert et al. (2008) use a model similar in spirit to derive preferences for selfishness, altruism and cooperation. They focus on the effect of incentives rather than on type identification, as we do here. Compared to these studies, our contribution lies on type identification, which remains largely unexplored among the empirical literature, and less so within a structural model.

This paper uses a structural approach to examine which types of social preferences individuals exhibit in a common pool resource environment, in which the CPR is collectively owned or shared (e.g. natural resources, land, software) and foregoing the overexploitation of the jointly used resource leads to a Pareto superior outcome. Understanding heterogeneity of individual preferences in this environment is the first step to the design of Pareto efficient incentives: we estimate simultaneously the distribution of types proposed by economic theory and the parameters of each type in our sample. We then examine determinants of social preferences as suggested by the empirical literature (Almås et al., 2010).

One recent development towards a structural model with heterogeneous preferences has been the simultaneous estimation of preference parameters and type composition in the sample by means of finite mixture models (Cappelen et al., 2007, 2010). Finite mixure models have been applied to estimate structural models such as Cappelen et al. (2011, 2013). But to our knowledge this is the first model to identify behavioral types with their preference parameters under heterogeneous social preferences. To do so we use a latent class model as described in Pacifico (2012), namely the expectation maximization (EM) algorithm.

A growing literature exploiting latent class models (Train, 2009) to identify social preferences (Breffle et al., 2011; Morey et al., 2006) has already proven fruitful in Ecological Economics. Varela et al. (2014) study the heterogeneity of social preferences for fire prevention in Europe; they argue for the existence of four types: typical, yea-saying, burnt-worried and against. Farizo et al. (2014) estimate a multilevel latent class model to capture five classes: typical, environmentalist, budgetary, futurist, against. Our estimation method closely follows those used in these papers within a CPR context. However, our use of the structural model is grounded in the theory of social preferences literature reported before gives external validity to our labeling of altruists, selfish, reciprocators and inequity averse.

Our sample is composed of both students and real users of the CPR and has different CPR environments (water, firewood and fish) under a rich set of (economic and non economic) incentive schemes. Second, we improve the type classification method currently used in the literature of social preferences (i.e. random coefficients model) and explain these types of social preferences within an economic model. Finally, we propose an alternative method based on a structural estimation of both the type preferences composition and their respective theoretical parameters.

An alternative to individuals deviating from the NE out of concern for others' outcomes or behavior is the incorporation of (foreseeable) errors into the best reply function. This is the principle of QRE - quantal response equilibrium (McKelvey and Palfrey, 1995). Close to the concept of QRE, another possible explanation of the consistent deviation from Nash behavior comes from the dynamic aspect of learning. Using the same data we use for the present study, Cárdenas et al. (2013) argue that students' behavior likely follows a payoff sampling equilibrium (PSE). It is a satisfactory explanation for several features of the distribution of outcomes.

Though QRE has been suggested as a competing explanation as opposed to social preferences, they can actually work as complements (Arifovic and Ledyard, 2012). In particular, it has been shown that QRE cannot account for social preferences (Ioannou et al., 2012; Hoppe and Schmitz, 2013). Largely due to the fact that it only uses one parameter, it cannot explain the large cross-sectional variation in the data. We compare the predictions from our model to a baseline QRE in order to highlight the tradeoff between parsimony and goodness of fit across the two specifications.

Burlando and Guala (2005) discuss the learning process in repeated games and conclude that the 'decay of overcontribution' over time, which depends critically on the group composition. Group composition is indeed an important factor, which lends credency to the QRE approach. Empirical findings suggest a negative relationship between group heterogeneity and public goods provision (e.g. Alesina et al. 1999; Miguel and Gugerty 2005; Vigdor 2004, Lucas, Oliveira and Banuri 2012; Fischbacher and Gächter 2010; Gächter and Thöni 2005). In that sense, Arifovic and Ledyard (2012) combine other-regarding preferences and learning. Subjects have a utility function determined by their own payoff, the average group payoff (altruism, or welfare) and the level of disparity between their own payoff and the group average (envy). We use a similar specification[1].

One main difference between our model and theirs is that in their model, agents only have other-regarding preferences (ORP) over outcomes and not over intentions, which implies reciprocity arises as an equilibrium behavior and not as a type.[2] Because our empirical estimates allow to identify individual types, and given the novelty of their approach, one of our contributions is to assess the empirical soundness of introducing a 'cooperator' type.

Compared to Arifovic and Ledyard (2012) beliefs in our model are simplistic: they only take into account immediately preceding experience, whereas their model incorporates evolutionary learning. In such setting, the variables are a finite set of remembered strategies for each agent and a corresponding probability distribution; learning happens by experimentation, replication and learning. Now, just as preferences can be heterogeneous so can approaches to learning be. Because our focus is on heterogeneity and we cannot jointly identify heterogeneous social preferences and (indepedent) heterogeneous cognitive types, we shut down the learning channel allows to focus on the preferences one.

The paper is organized as follows. In section 2 we introduce the CPR framework for this study. In section 3 we introduce the different models of social preferences we will examine. Section 4 estimates the latent class model. Section 5 concludes the paper by summarizing the main results and suggesting points of future research.

[1]Our specification follows Fehr and Schmidt (1999) rather than Arifovic and Ledyard (2012), who nevertheless establish the equivalence between the two formulations.

[2]The model in Vélez et al. (2009) highlights the difference between playing reciprocally and being a reciprocator type. In their model, reciprocal behavior arises from preferences for conforming to what others are expected to do.

2 Common Pool Resource framework

i Description of the CPR game

The setting is well explained in Cárdenas (2004),(2011). Here we briefly describe the game.[3]

Each individual i is endowed with $e = 8$ units of effort (e.g. hours of extraction or investment in equipment) which he can use to extract $x_i \in \{1, \dots, 8\}$ units from the CPR. Given the group size $n = 5$ and the other players' extraction decisions $x_{-i} \in \{4, ..., 32\}$, the individual payoff is given by

$$\pi_i = \pi(x_i, x_{-i}) = ax_i - \frac{1}{2}bx_i^2 + \varphi(40 - (x_i + x_{-i})) \tag{1}$$

In our setting $(a, b, \varphi) = (60, 5, 20)$. The payoff features direct benefits from extraction $60x_i - \frac{5}{2}x_i^2$ and the indirect costs from depletion $\varphi(40 - (x_i + x_{-i}))$ following from i's as well as others' extraction level.

Participants play a finitely repeated ($T = 10$)[4] partner matching game. At the beginning of period t individuals decide simultaneously (x_{it}, x_{-it}). At the end of period t, the experimenter announces aggregate extraction $(x_{it} + x_{-it})$ and players are informed about other players' aggregate behavior. That is i does not know individual extraction by $-i$. She only knows the *average* extraction by $-i$: $\overline{x}_{-it} = \frac{\sum_{j \neq i}^{n-1} x_{jt}}{n-1}$. The lack of detail about individual extractions favors the simplification of learning aspects in order to focus on the identification of preferences.

The composition of the group remains the same during the following T rounds $t = 11, \dots, 20$. At the beginning of round 11 the experimenter announces (and implements) an incentive. The incentive could be monetary (fine or subsidy) or non-monetary (e.g. affecting reputation or other considerations rather than payoffs).

The efficient outcome or social optimum (SO) maximizes the aggregate payoff of the group

$$(x_1^{SO}, \dots, x_5^{SO}) = \underset{(x_1, \dots, x_5) \in \{1, \dots, 8\}^5}{\arg\max} \sum_{i=1}^{5} \pi_i.$$

Our socially optimal decision of extraction $x_i^{SO} = 1$ of extracting the minimum level possible. There is a conflict between the Nash equilibrium and the socially efficient strategies (see table 8). The Unique Nash Equilibrium (NE) for a self-regarding individual is given by

$$x_i^{NE} = \underset{x_i}{\arg\max}\, \pi_i \,\forall\, i$$

[3] Apart from the game outcomes, survey data was gathered covering sociodemographic (idiosyncratic) factors, though only for villagers. For this reason we can only estimate our social preferences model for villagers.

[4] Individuals did not know how many rounds they would play. There were 2 example rounds and 1 practice round and the game started once the experimenter assured the participants understood the procedure.

which gives $x_i^{NE} = e = 8$.[5]

The subgame perfect Nash equilibrium of the repeated game is the same for all the rounds and is equal to the Nash equilibrium, i.e. the self-regarding outcome. Individuals might behave pro-socially in the presence of reputation effects (Kreps et al., 1982; Bohnet and Huck, 2004; Mailath and Samuelson, 2006) (see figure A.1 in the appendix). However, our setting precludes such a reputation channel.

ii Game outcome in the field

The final sample is composed of 230 students and 705 villagers.[6] The first result to motivate this study is that individuals who consistently play the NE strategy are a small proportion of the sample. Figure 5.1 presents how consistently (between 1 and 10 times) individuals extracted 8 units.

Overall, 35% of the players never played the NE strategy, a quarter of the players chose the NE strategy only once among the ten rounds and only 2% of 1000 individuals played the NE consistently. Figure 5.2 shows the path of average extraction.

Students extract consistently more than villagers (though the difference does not appear to be significant). Also students seem more prone to the last-round effect: between rounds 9 and 10 the proportion of the sample extracting 8 units goes from 18% to 28% among students while in the villager sample it remains at 18%.[7] This raises the question of whether students and villagers have differing levels of rationality. Following Cárdenas et al. (2013) we estimate a QRE model and compare the outcome for both samples.

iii Static quantal response equilibrium

We estimate a logit QRE specification following Cárdenas et al. (2013) and extend it to the sample of villagers. Suppose players make errors in choosing from $\{1, \dots, e\}$ but the distribution of choices $P(x = k)$, $k \in \{1, \dots, e\}$ is common knowledge. If $\pi(x_i, x_{-i})$ is the payoff for x_i given others' pure strategy x_{-i}, let $\pi(x_i, P)$ be the expected payoff of x_i given others' are mixing strategies according to $P(.)$. Then the logistic QRE[8] associated to the parameter $\lambda \in [0, \infty)$[9] is a stable outcome of a belief and choice formation process given

[5] Both the social optimum and the Nash equilibrium are corner solutions, which constitutes a potential drawback for identification (especially under incentives). Cárdenas et al. (2013) point out that QRE outperforms payoff sampling equilibria under corner solutions.

[6] Though the full sample contains 865 villagers, some of them ended up participating more than once. The second observation for those who did have been removed. See Polania-Reyes (2014).

[7] If looking at the last-round effect in terms of the cooperative strategy (extract 1 unit) we see a slight reduction for students (8% to 7%) and an increase for villagers (12% to 14%), which speaks to the stylized finding that CPR users have the habit of "not finishing everything on the table".

[8] Logit is the most common specification for a QRE. Assuming a symmetric equilibrium, errors ϵ_{ik} of individual i adopting strategy k are independent and identically distributed according to a type I extreme value distribution.

[9] λ indicates the degree of rationality: when $\lambda \to \infty$ (the error rate tends to zero) subjects are rational and when $\lambda = 0$ subjects are acting randomly according to a uniform probability function.

by

$$P(x_i = k) = \frac{\exp(\lambda\pi(k, P))}{\sum_{j=1}^{8}\exp(\lambda\pi(j, P))}\,\forall k \in \{1, \ldots, 8\}.$$

λ is chosen to match the QRE distribution, which derived from the payoff function alone, to the empirical distribution. Like Cárdenas et al. (2013) we choose λ in order to minimize mean squared error (MSE). Figure 5.4 shows the outcome:

The value of λ minimizing MSE is very close across the samples: 0.03 for students and slightly lower for villagers at 0.02. Though this suggests a somewhat higher level of rationality among the student sample, the order of magnitude is the same. We take this as indicative evidence that using a constant λ across the population and across types is an adequate assumption.

Figure 5.5 compares the predicted and realized distributions. As Cárdenas et al. (2013) point out, a slightly better fit is achieved within the student sample ($MSE_s = 0.053\%$) than that of villagers ($MSE_v = 0.065\%$). In particular, the higher (respectively lower) incidence of payoff-maximizing (resp. socially efficient) behavior among students seem better matched by the QRE. However, and in spite of the overall constant trend over time (see figure 5.2) a lot of cross-sectional variation remains that cannot be explained using symmetric strategies. We now turn to a model of other-regarding preferences.

3 A structural model of social preferences

We introduce a structural model to estimate preferences for altruism, selfishness, reciprocity and inequity aversion.

Individual i with preferences type q has a utility function U_i^q where $\mathbf{\Theta} = \{\rho, \mu, \beta\}$. We will consider the most popular types of individuals in the behavioral economics literature: i) self-regarding, ii) altruist, iii) reciprocator and iv) inequity averse. All these other-regarding preferences are defined over payoffs: they incorporate concerns over outcomes (as captured by the payoffs). Reciprocators instead exhibit preferences over behaviors (as captured by others' extraction levels).

As discussed in section 1, a general specification of preferences takes into account own payoff, others' payoff and others' behavior. Consequently, at each point in time each individual chooses a level of extraction in order to solve[10]

$$\max_{x_{it}} U^i(\pi_{it}, E_{t-1}[\overline{\pi}_{-it}], E_{t-1}[\overline{x}_{-it}]; \Theta) \tag{2}$$

where $E_{t-1}^i[\overline{\pi}_{-it}]$ denotes individual expectations about others' strategy, $\overline{\pi}_{-i} = \frac{\sum_{j\neq i}\pi_j}{n-1}$, given their information at hand (and similarly for $\overline{x}_{-it}$). Our previously discussed simpliying

[10]For simplicity, we will be assuming linear individual utility functions, which makes our formulation in terms of expected payoffs equivalent to one in terms of expected utilities. However, neutrality is an important matter measuring social preferences. The analysis becomes more complicated with other functional forms of the utility function.

assumption about beliefs reads as

$$E^i_{t-1}[\overline{\pi}_{-it}] = \overline{\pi}_{-i,t-1} \text{ and } E^i_{t-1}[\overline{x}_{-it}] = \overline{x}_{-i,t-1}.$$

i Baseline: self-regarding preferences

Individuals that exhibit self-regarding preferences care only about their own monetary cost and benefits and are usually called in the literature as free-riders, selfish or defectors. A *self-regarding* individual i has a utility function given by $U_i^S = \pi_i$. Note that the (self-regarding) best reply is the maximum extraction level $x_i^S = 8$.[11]

ii Altruistic preferences

We adapt our CPR framework to the models proposed by Levine (1998) and Casari and Plott (2003). Individuals that exhibit these preferences are those who care about others' utility - i.e. altruists in Andreoni and Miller (2002); Carpenter et al. (2009), unconditional cooperators in Fischbacher et al. (2001) or pure cooperators in Rabin (1993).

An *altruist* i has a utility given by

$$U_i^A = \pi_i + \rho_i \overline{\pi}_{-i} \tag{3}$$

The specification above is rescaled from a general regression model which would put weights on both variables.

$$U_i^A = \eta_A \pi_i + \rho \overline{\pi}_{-i} \tag{4}$$

A challenge to interpretation arises if $\eta < 0$ (and similarly in subsequent models). In the case of altruism this deserves special discussion: in the context of a social dilemma, a purely altruistic solution (i.e. to give a large weight to others' payoff) is equivalent to foregoing own payoff. If altruism is seen as a particular (extreme?) form of concern for efficiency, we argue that the sign of η is helpful in making a distinction between altruism and concern for efficiency (Charness and Rabin, 2002), the presence of a negative coefficient calling for the former label rather than the latter.

[11] By construction the Nash equilibrium of the game is the stable strategic outcome from a game between self-regarding players.

iii Reciprocity

Our reciprocators are individuals that cooperate only if others cooperate and present similar behavior to conformism (Rabin, 1993; Bowles, 2004; Levine, 1998). A *reciprocator* i has a utility given by

$$U_i^R = \pi_i + \mu(x^{*i} - \overline{x}_{-i})\overline{\pi}_{-i} \; \forall \; i \tag{5}$$

where x^{*i} is a norm based on which i rates extractions from others, deriving more utility if others' extraction is below the norm and less otherwise. A positive value of μ would indicate a desire to uphold the social norm.

Polania-Reyes (2014) estimate the structural parameters ρ and μ by means of a random coefficients model, which assumes idiosyncratic coefficients for each individual. Selfish behavior is identified as the opposite of selfless behavior as given by the value of ρ (ρ_i in her specification).

iv Fairness and inequity aversion

This model is based on Fehr and Schmidt (1999); Bolton and Ockenfels (2000). An *inequity averse* individual i has a utility given by

$$U_i^I = \pi_i + \alpha \max(\overline{\pi}_{-i} - \pi_i, 0) + \beta \max(\pi_i - \overline{\pi}_{-i}, 0) \; \forall \; i \tag{6}$$

The second term in equation 6 measures the utility loss from disadvantageous inequality, and the third term measures the loss from advantageous inequality. It is assumed that the utility gain from i's payoff is higher than her utility loss for advantageous inequality and her utility loss from disadvantageous inequality is larger than the utility loss if player i is better off than other players, $0 \leq -\beta < 1$. In addition, i is loss averse in social comparisons: i suffers more from inequality that is to his disadvantage (Loewenstein et al., 1989): $\alpha_i \geqslant \beta_i$.

Disadvantageous inequality can only be identified under interior solutions (Fehr and Schmidt, 1999; Vélez et al., 2009). Because our CPR setting yields boundary solutions for both the Nash equilibrium and social optimum, our regression specification only incorporates advantageous inequality:

$$U_i^I = \eta_I \pi_i + \beta \max(\pi_i - \overline{\pi}_{-i}, 0) \; \forall \; i \tag{7}$$

The sign on β will identify preferences for inequity, if positive, and for equity otherwise.

v Beliefs

The formulation of beliefs is as important as that of preferences. In fact one of the basic insights behind QRE is that if agents make errors, they expect others to make the same mistakes. The formulation of beliefs raises an identification challenge. Expectations are closely linked to learning. Arifovic and Ledyard (2012) provide a model that incorporates both social preferences (altruism, selfishness and inequity aversion) and learning (through an Individual Evolutionary Model, IEM). An IEM is characterized by experimentation, replication and learning (each of these adding one free parameter to the model).

We assume agents only take into account other players' immediately preceding action. This simplification, which allows us to focus on the classification of behavioral types, is warranted by the fact that agents only learn previous round average extraction. A more detailed model of belief formation such as Arifovic and Ledyard (2012) might add precision to the model, in return for more free parameters to be estimated, but it wouldn't help the identification procedure itself because of its reliance on symmetric gognitive profiles across players.

vi Summary: a mixture model without type identification

We suppose the population comprises 4 homogeneous (unobservable) types. On each round $t \in \{1, \dots, T\}$, individual i makes her extraction decision x_{it} in order to maximize their utility, given the other 4 player's previous behavior in the group, $\overline{x}_{-it-1}$. We then define the structure of the error term as we introduce errors in decisions for each type and use a random utility specification in this choice environment. The expected utility takes the linear form for an individual type q, being self-regarding, inequity averse, reciprocator or altruist, $q \in \{S, I, R, A\}$. At time t, agent i chooses an action $j \in \{1, \dots, J\}$ to derive utility

$$\tilde{U}^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) = U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) + \varepsilon^q_{ijt} \quad \forall j \in \{1, \dots, J\} \tag{8}$$

The choice probability, conditional on type q, is then determined by the logit function

$$\tilde{f}_q(x_{ijt}; \theta_q, \lambda_q, \overline{x}_{-it-1}) = \frac{\exp[\lambda_q U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1})]}{\sum\limits_{m=1}^{J} \exp(\lambda_q U^q(x_{imt}; \theta_q, \overline{x}_{-it-1}))} \tag{9}$$

This logit function is reminiscent of the QRE specification of section iii. As we argued back then, we will drop λ_q, $q \in \{S, I, R, A\}$ from the problem assuming a constant parameter applies throughout.

The individual contribution to the total likelihood function is the sum of the component densities $f_q(x_i; \theta_q, \overline{x}_{-i})$ weighted by the probabilities p_q that individual i belongs to type q such that $q \in Q = \{S, I, R, A\}$:

$$f(x_i; \Theta) = \sum_{q \in Q} p_q \prod_{t=1}^{T} \prod_{j=1}^{J} \left(f_q(x_i; \theta_q, \overline{x}_{-i})\right)^{d_{ijt}} \tag{10}$$

where d_{ijt} is a dummy for whether action j was indeed chosen at time t. This leads to the likelihood function

$$\ln L(\Psi; x) = \sum_{i=1}^{N} \ln f(x_i; \Psi) = \sum_{i=1}^{N} \ln \sum_{q \in Q} p_q f_q(x_i; \theta_q, \overline{x}_{-i}) \tag{11}$$

Assuming $U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) = U(x_{ijt}; \theta_q, \overline{x}_{-it-1})$ where $\theta_q = \theta \sim F(.)$ allows us to estimate $\mathbf{p} = \{p_S, p_I, p_A\}$, $\Theta = \{\theta_q\} = \{\rho, \beta, \mu\}$ by direct maximization of

$$\ln L(\Psi; x) = \sum_{i=1}^{N} \ln f(x_i; \Psi) = \sum_{i=1}^{N} \ln \sum_{q \in Q} p_q \int_{-\infty}^{\infty} \left(f(x_i; \theta_q, \overline{x}_{-i})\right) \mathrm{d}F(\theta) \tag{12}$$

Among the structural preference models that take into account agent heterogeneity, this continuous mixture model is the most commonly used ((Cappelen et al., 2007, 2010, 2011, 2013), (Cappelen et al., 2013), all of which assume a lognormal distribution for the parameters). In addition to the need for a predefined functional form for the continuous mixture, the finite mixture model does not allow the estimation of separate parameters for the different preference functions, i.e. $U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) \neq U'^q(x_{ijt}; \theta_q, \overline{x}_{-it-1})$. Because this is precisely what we intend to do, we refine the formulation of p_q following a latent class model.

4 Type identification using a latent class model

In order to identify individual types, we use a latent class estimated using the Expectation Maximization (EM) algorithm (Dempster et al., 1977; Train, 2008). More specifically we follow an implementation of (Train, 2008) by Pacifico (2012)[12] using the specification in section vi.[13]

The simultaneous estimation of types and parameters relies on an iteration of two steps: one where likelihood conditional on types is maximized (the M-step) and one where idiosyncratic type distribution is updated.

i The E-step

During the E-step, we take the conditional expectation of the complete-data log likelihood, $\ln L^c(\Psi)$ given the observed extraction profiles x, using the current fit for Ψ. Let $\Psi^{(0)}$ be the value specified initially for Ψ. Then on the first iteration of the EM algorithm, the E-step requires the computation of the conditional expectation of $\ln L^c(\Psi)$ given x, using $\Psi^{(0)}$ for Ψ:

$$G(\Psi, \Psi^{(0)}) = \mathbb{E}_{\Psi^{(0)}}[\ln L^c(\Psi) | X = x] \tag{13}$$

[12] We use the Stata module developed by Pacifico (2012) called lclogit

[13] The model specification is time-invariant, which implies that $\upsilon_{qt} = \upsilon_q$. Kasahara and Shimotsu (2009) study type identification in finite mixture models with panel data.

On the *(k + 1)*th iteration the E-step requires the calculation of $G(\Psi, \Psi^{(k)})$ where $\Psi^{(k)}$ is the value of Ψ after the *k*th EM iteration. Since $\ln L^c(\Psi)$ is linear in the unobservable υ_{iq}, it requires that $\mathbb{E}_{\Psi^{(k)}}(V_{iq}|X = x) = \tau_{iq}^{(k+1)}(x; \Psi^{(k)})$ [14], where V_{iq} is the random variable corresponding to υ_{iq} and[15]

$$\tau_{iq}^{(k+1)}(x; \Psi^{(k)}) = \frac{p_q^{(k)} f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})}{\sum_{q \in Q} p_q^{(k)} f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})} \tag{14}$$

are the *a posteriori* probabilities that the ith member of the sample with observed value x_i belongs to the *q*th component of the mixture, computed according to Bayes law given the actual fit to the data, $\Psi^{(k)}$. Then

$$G(\Psi, \Psi^{(k)}) = \sum_{i=1}^{N} \sum_{q \in Q} \tau_{iq}^{(k+1)}(x_i; \Psi^{(k)}, \overline{x}_{-i})[\ln p_q^{(k)} + \ln f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})] \tag{15}$$

ii The M-step

The M-step on the *(k + 1)*th iteration, the complete-data log likelihood function 15 is maximized with respect to $\Psi^{(k)}$ to provide the updated estimate $\Psi^{(k+1)}$.[16]

As the E-step involves replacing each υ_{iq} with its current expectation $\tau_{iq}^{(k+1)}(x; \Psi^{(k)})$ in the complete-data log likelihood, the updated estimate of p_q is giving by replacing each υ_{iq} in (23):

$$\widehat{p_q}^{(k+1)} = \sum_{i=1}^{N} \frac{\tau_{iq}^{(k+1)}(x_i; \Psi^{(k)}, \overline{x}_{-i})}{N} \tag{16}$$

Dempster et al. (1977) show that the sequence of likelihood values $\{L(\Psi^{(k+1)})\}$ is bounded and non-decreasing from one iteration to the next, so the EM algorithm converges monotonically to its maximum. The E- and M-steps are thus alternated repeatedly until the difference $L(\Psi^{(k+1)}) - L(\Psi^{(k)})$ changes by a -previously fixed- arbitrarily small amount.

Note that these posterior probabilities of individual group membership are not only used in the M-step, but they also provide a tool for assigning each individual in the sample to one of the Q types. Thus, finite mixture models may serve as statistically well grounded tools for endogenous individual classification (Bruhin et al., 2010).

iii Testing for the number of types

An important aspect of the contribution by (Arifovic and Ledyard, 2012) is that reciprocity arises not as a type but as an equilibrium behavior. This raises the empirical question of

[14] $\mathbb{E}_{\Psi^{(k)}}(V_{iq}|X = x) = Pr_{\Psi^{(k)}}[V_{iq} = 1|X = x]$ is the current conditional expectation V_{iq} of given the observed data $X = x$

[15] $f(x_i; \Psi^{(k)}, \overline{x}_{-i}) = \sum_{q \in Q} p_q^{(k)} f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})$

[16] For the FMM the updated estimates $p_q^{(k+1)}$ are calculated independently of the update estimate $\boldsymbol{\xi}^{(k+1)}$ of the parameter vector containing the unknown parameters in the component densities. See (Cappelen et al., 2007, 2010, 2011, 2013,?)

whether reciprocity can be thought of as an attribute. We provide some empirical information to this question by testing for the optimal number of types using a latent class model to fit the data.

Table 1 summarizes the performance of a different number of factors for the sample of villagers along the dimensions of information (as measured by the Consistent Akaike Information Criterion and Bayesian Information Criterion) and of likelihood (as measured by the likelihood ratio).

Table 1 provides evidence that the optimal model to describe the data is either one with 4 classes, the information criteria such as the CAIC or the BIC being less prone to overparametrization than the likelihood criterion. Table 2 presents similar results for the student sample.

The picture arising from the student sample (table 2) is not exactly the same as from that of villagers, as it seems to suggest the use of a fifth class. In the absence of theoretical support for the additional class, our observation is that the results found across the two populations are in broad agreement.

The results so far support the use of four types, which according to our theoretic model are the self-regarding, altruistic, inequity averse and reciprocators.

iv Latent class model results

iv.1 Utility parameters: coefficients and labels

Table 3 provides the results for the class share determinants model estimated with the villager sample.

Inequity aversion occupies a large share within the villager sample: most villagers are affected negatively by advantageous inequality in their payoffs. Pure selfish and pure altruists make up a smaller share of the sample, very close to the random coefficients model outcome in Polania-Reyes (2014) using a (10% altruists, 7% selfish).

As discussed before, the negative sign on the weight to own payoff is at first sight unsettling, but highlights the nature of the social dilemma. In making a distinction between altruism and concern for efficiency, the presence of a negative coefficient argues for altruism in the present case.

Only a small percentage in our sample are reciprocators (to the point of Arifovic and Ledyard (2012)). Here we are far from the results in Polania-Reyes (2014) where a high incidence of reciprocating behavior is found. We note that her random coefficient model cannot accommodate inequity aversion and has a high share of unidentified types (32%). This limits the interpretability and comparability of results across the two studies. The negative sign on the concern for the norm is counterintuitive and suggests a specification issue in our function, possibly in how the social norm itself is defined.

In order to compare estimates across populations, we constrain the coefficients on the student sample so that the weight on own payoff matches the one from the villager sample.[17] The results are recorded in table 4.

Again we observe a large number of inequity averse individuals (with a similar magnitude for the utility parameter), similar to the results from the villager sample. In stark contrast, when trying to match altruistic behavior we end up with a negative coefficient. Interpreted directly this coefficient points to spiteful behavior, whereby agents are affected negatively by both their outcomes and those of others. Polania-Reyes (2014) does not provide a point of comparison on the student sample.

Our latent class estimate allows to make choice prediction. In order to understand the relative performance of each model, we document the choice prediction outcome below. So far we haven't taken advantage of the data from rounds 11 to 20. We do so now by comparing the model performance in-sample (rounds 1 to 10) and out-of-sample (rounds 11 to 20).

A naive model (e.g. our static QRE) can only attain a 1/8 choice probability, which is improved within all classes except that of reciprocators, in line with the concern expressed previously about this category. The out-of-sample performance is comparable (sometimes slightly higher) than in-sample, something we take as an important sign of internal validity. In terms of relative performance, the altruistic-spiteful category performs better than the rest, and better still than the self-regarding category. This is a surprising finding, given that the Nash strategy is expected to be more stable (hence a priori more predictable) than others.

In order to understand the type classification above, we now examine the drivers for the probability of belonging to each type.

iv.2 Class share determinants

Table 3 reports an estimation of the class share model for villagers.

Those users whose income depends 100% on the CPR are more likely altruistic or inequity averse than those whose income doesn't. The belief that the community has no need of an external authority to rule them increases the likelihood of altruistic or inequity averse classification, and decreases that of the self-regarding one. The perception that the resource will remain still greatly decreases the likelihood of being altruistic as opposed to self-regarding. Voluntary participation, instead, shows a counterintuitive role, leading to a lower probability of being inequity averse and instead a higher probability of being self-regarding.

5 Conclusion

This is a study on type classification for social preferences from a CPR game. We bring a novel method to identify types in a unique sample including villagers and students. We

[17] See alternative specifications in appendix B

examine the most popular types of social preferences in the theory literature, testing for the optimal number of types. Our structural estimation relies on four types, which the data supports. The most salient feature is the prevalence of aversion to inequity across both samples. There is evidence both of pure altruistic behavior in the villagers' sample and of spiteful behavior among students. The lack of empirical evidence for reciprocal types sheds doubt on our specification, but also gives an indirect signal that reciprocity arises not as a type but as an equilibrium behavior across types.

A key feature of heterogeneity is the role of individual background. For example, the use of CPR in real life by the participants. Figure 5.3 shows the fraction of players that extract 8 units according to their dependence to the CPR. Those users whose income depends 100% on the CPR extract significantly less whereas those users whose income depends 0% on the CPR extract significantly more. Those who in real life depend more on the common pool resource have a lower probability of being allocated to the selfish type.

Using an RCM classification, (Polania-Reyes, 2014) finds that non-monetary incentives are more effective in groups where other-regarding preferences are prevalent and only the subsidy is effective in promoting behavior among self-regarding individuals. While we leave aside the treatment of incentives, we note that types are likely to be state dependent. Our finding that in-sample and out-of-sample model outcomes are comparable provide internal validity to our findings. This is particularly important in the latent class literature where, as previously discussed, labels are commonly found to be driven by data rather than theory.

We acknowledge the importance of beliefs in the decision making progress. Facing the possibility of heterogeneous preferences as well as that of heterogeneous learning, we shut down the latter to focus on the former. We assume an overly simplistic system of beliefs, namely that agents only take into account what others did in the previous round. While an IEM type model would take into account the likely higher complexity of the thought process (at the cost of parsimony), an identification challenge remains in terms of the two types of heterogeneity (cognitive and behavioral). The development of heterogeneous QRE under cognitive hierarchies proposed by Rogers et al. (2009) might be helpful in that sense. Our conjecture is that cognition and social preferences are correlated, suggesting the importance of identifying such correlations.

Testing and identification remain a challenge, both for a model of social preferences or for a pure model of bounded rationality such as QRE (McCubbins et al., 2013). On one hand, classical competive behavior might obtain in an economy subject to social preferences (Dufwenberg et al., 2011). On the other hand, there is evidence that social preferences are subject to framing effects (Dariel, 2013; Ackermann et al., 2014) or the institutional setting (Cassar et al. (2013)).

Group composition is indeed a key feature. While we restrict ourselves to variables at the individual level, Polania-Reyes (2014) performs a regression analysis with a probit model where the probability of being type q depends on socioeconomic characteristics at the individual and village level. She finds community level drivers are important, and in particular that types are somewhat correlated. If types are robust over time (as the evidence discussed here suggests) yet at the same time context- or group-dependent, an evolutionary approach might be fitting not only for the learning but also the behavioral aspect of choice in CPR settings as well as similar collective action problems.

Figure 5.1: Number of times individuals behaved as self-regarding

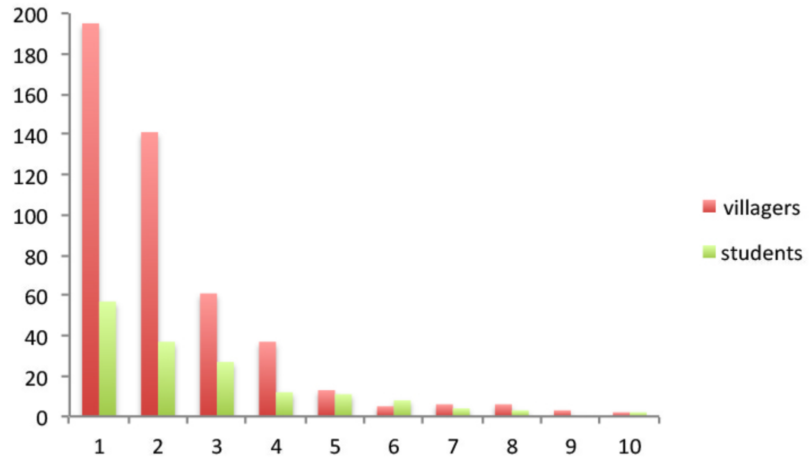


Figure 5.2: Average individual extraction over time

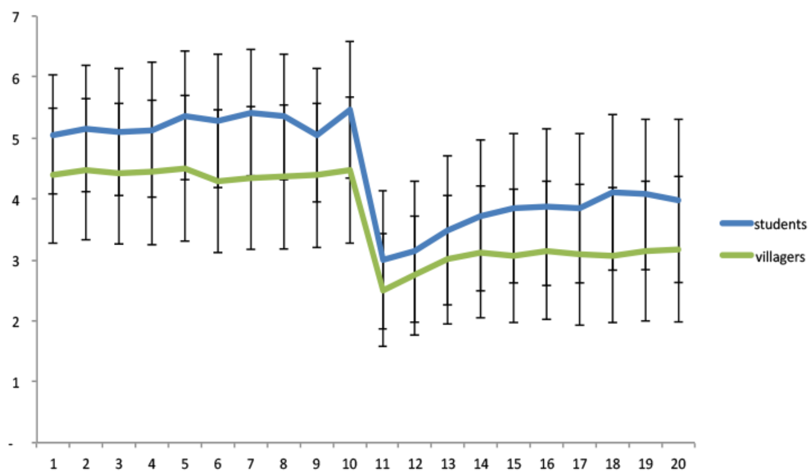


Figure 5.3: Heterogeneity of real level extraction of the CPR in the game all CPR users vs. students ($N = 1095$). The solid line shows the % time that the Self-regarding NE was chosen in the game by the Students sample. The round-dot line shows the case with individuals who use 0% of the real CPR. The square-dot line shows the average level of extraction in the game by individuals who use 50% of the real CPR. The long-dashed line the average level of extraction in the game by individuals who use 100% of the real CPR. The difference in means in the last round is significant at 10%.

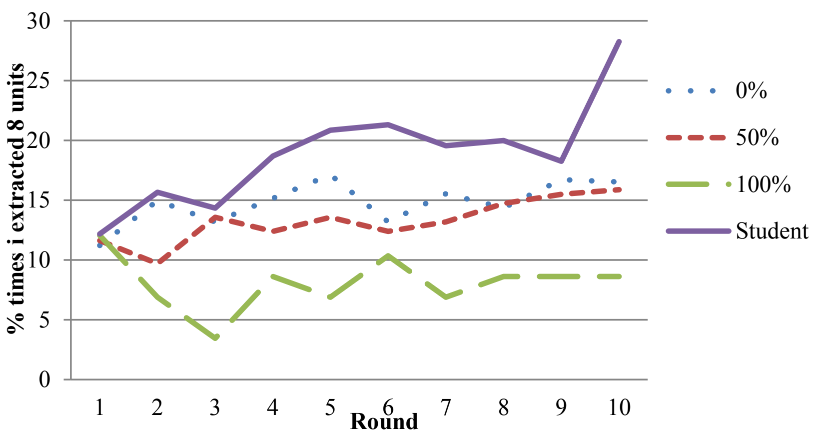


Figure 5.4: log(MSE) as a function of λ.

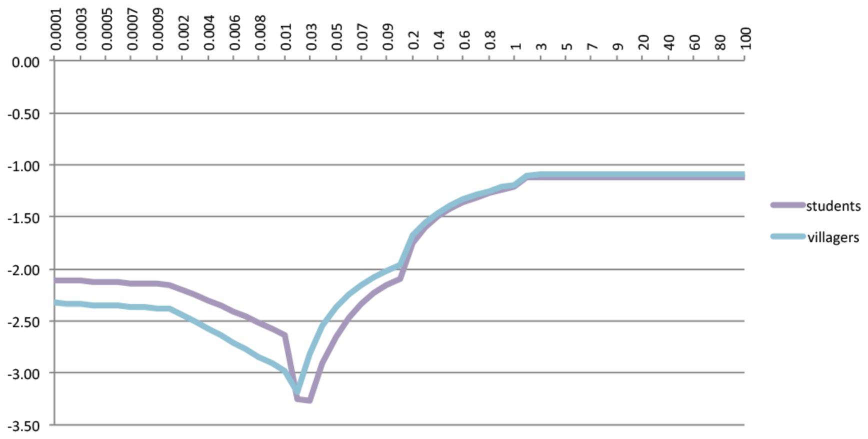
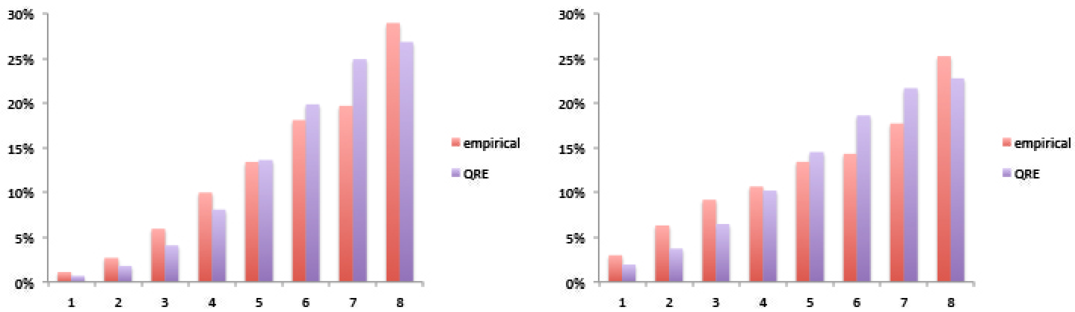


Figure 5.5: Empirical distribution of choice outcomes and QRE distribution



(a) Students: $\lambda = 0.03$

(b) Villagers: $\lambda = 0.02$

Table 1: Comparison of model performance by number of classes - villager sample

No. classes	LL	No. parameters	BIC	CAIC
2	-6992.937	21	14144.6	14123.6
3	-6887.851	38	14062.91	14024.91
4	-6821.428	55	**14058.56**	**14003.56**
5	-6783.39	72	14110.97	14038.97
6	**-6761.148**	89	14194.98	14105.98

Table 2: Comparison of model performance by number of classes - student sample

No. classes	LL	No. parameters	BIC	CAIC
2	-4045.653	9	8149.25	8140.25
3	-3999.159	14	8088.451	8074.451
4	-3981.109	19	8084.542	8065.542
5	-3956.442	24	**8067.399**	**8043.399**
6	**-3948.002**	29	8082.708	8053.708

Table 3: Class share determinants - villager sample

Variable	Self-regarding	Altruistic	Inequity averse	Reciprocator	Std. error[b]
π_i	0.019	-0.026	0.019	0.235	0.001
$\overline{\pi}_{-i}$	0[a]	0.003	0[a]	0[a]	0.000
$\max(\pi_i - \overline{\pi}_{-i}, 0)$	0[a]	0[a]	-0.023	0[a]	0.000
$\overline{\pi}_{-i}(x^{i*} - \overline{x}_{-i})$	0[a]	0[a]	0[a]	-0.015	0.000
Class Share	0.185	0.107	0.687	0.021	-

[a] Constrained to 0 in estimation
, [b] Standard errors are computed from the covariance matrix of regression coefficients over the full sample. Implied variances and covariances of choice model coefficients are averaged across individuals in the prediction sample.

Table 4: Class share determinants - student sample

Variable	Self-regarding	Spiteful	Inequity averse	Reciprocators	Std. error[d]
π_i	0.034	-0.026[b]	0.019[c]	0.045	0.002
$\overline{\pi}_{-i}$	0[a]	-0.047	0[a]	0[a]	0.001
$\max(\pi_i - \overline{\pi}_{-i}, 0)$	0[a]	0[a]	-0.025	0[a]	0.001
$\overline{\pi}_{-i}(x^{i*} - \overline{x}_{-i})$	0[a]	0[a]	0[a]	-0.038	0.000
Class Share	0.122	0.129	0.731	0.018	-

[a] Constrained to 0 in estimation
,[b] Constrained to -0.026 in estimation
,[c] Constrained to 0.019 in estimation
, [d] Standard errors are computed from the covariance matrix of regression coefficients over the full sample. Implied variances and covariances of choice model coefficients are averaged across individuals in the prediction sample.

Table 5: Class-conditional probability of choice

Class	Villagers		Students	
	In-sample	Out-of-sample	In-sample	Out-of-sample
Self-regarding	0.173	0.166	0.205	0.189
Altruistic / Spiteful	0.3136	0.437	0.336	0.341
Inequity averse	0.164	0.210	0.173	0.217
Reciprocator	0.136	0.064	-	-

Table 6: Drivers of class share - villager sample

Variable	Self-regarding	Altruistic	Inequity averse	Reciprocator
HH size	-0.479	-0.417	-0.478	0
Age	-0.021	-0.011	-0.019	0
Sex	-1.942	-1.95	-1.609	0
Education Level	-0.26	-0.05	-0.166	0
Land owner	0.869	0.814	1.134	0
Interest_com_perceived	-0.686	-0.414	-0.73	0
Belief_local_govern	1.762	1.925	1.943	0
Voluntary_part	0.22	0.104	-0.346	0
CPR_Still_perceived	-1.773	-2.485	-1.844	0
HH_Income_CPR	33.482	34.475	33.874	0

References

Ackermann, K. A., J. Fleiß, and R. O. Murphy (2014). Reciprocity as an individual difference. *Journal of Conflict Resolution*.

Akerlof, G. A. (1984). *An Economic Theorist's Book of Tales*. Cambridge, UK: Cambridge University Press.

Almås, I., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2010). Fairness and the development of inequality acceptance. *Science 328*(5982), 1176–1178.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Andreozzi, L., M. Ploner, and I. Soraperra (2013). Justice among strangers. On altruism, inequality aversion and fairness. CEEL Working Papers 1304, Cognitive and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia.

Arifovic, J. and J. Ledyard (2012, June). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. Discussion Papers wp12-01, Department of Economics, Simon Fraser University.

Arrow, K. J. (1971). *Political and Economic Evaluation of Social Effects and Externalities*, pp. 3–23. Amsterdam: North Holland.

Becker, G. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature 14*(3), 817–826.

Bewley, T. (1999). *Why wages don't fall during a recession*. Harvard Univ Press.

Bohnet, I. and S. Huck (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *The American Economic Review 94*(2), 362–366.

Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *The American Economic Review 90*(1), 166–193.

Bowles, S. (2004). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press. microbook.

Breffle, W. S., E. R. Morey, and J. A. Thacher (2011). A joint latent-class model: Combining likert-scale preference statements with choice data to harvest preference heterogeneity. *Environmental and Resource Economics 50*(1), 83–110.

Bruhin, A., H. Fehr-Duda, and T. Epper (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica 78*(4), 1375–1412.

Burlando, R. and F. Guala (2005). Heterogeneous agents in public goods experiments. *Experimental Economics 8*(1), 35–54.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review 97*(3), 818–827.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2011). The importance of moral reflection and self-reported data in a dictator game with production. *Social Choice and Welfare 36*(1), 105–120.

Cappelen, A. W., J. Konow, E. Ø. Sørensen, and B. Tungodden (2013). Just luck: An experimental study of risk-taking and fairness. *The American Economic Review 103*(4), 1398–1413.

Cappelen, A. W., K. O. Moene, E. Ø. Sørensen, and B. Tungodden (2013). Needs Versus Entitlements - An International Fairness Experiment. *Journal of the European Economic Association 11*(3), 574–598.

Cappelen, A. W., E. Ø. Sørensen, and B. Tungodden (2010). Responsibility for what? fairness and individual responsibility. *European Economic Review 54*(3), 429–441.

Cárdenas, J. C. (2004). Norms from outside and inside: an experimental analysis on the governance of local ecosystems. *Forest Policy and Economics 6*, 229–241.

Cárdenas, J. C. (2011). Social norms and behavior in the local commons as seen through the lens of field experiments. *Environmental and Resource Economics 48*(3), 451–485.

Cárdenas, J. C., C. Mantilla, and R. Sethi (2013, September). Commons without Tragedy: Sampling Dynamics and Cooperative Resource Extraction. DOCUMENTOS CEDE 011892, UNIVERSIDAD DE LOS ANDES-CEDE.

Carpenter, J., S. Bowles, H. Gintis, and S.-H. Hwang (2009). Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behavior & Organization 71*(2), 221–232. doi: DOI: 10.1016/j.jebo.2009.03.011.

Carpenter, J. P. and E. Seki (2010). Do social preferences increase productivity? field experimental evidence from fishermen in toyama bay. *Economic Inquiry in press*.

Casari, M. and C. Plott (2003). Decentralized management of common property resources: experiments with a centuries-old institution. *Journal of Economic Behavior & Organization 51*(2), 217–247.

Cassar, A., G. d'Adda, and P. Grosjean (2013, October). Institutional Quality, Culture, and Norms of Cooperation: Evidence from a Behavioral Field Experiment. Discussion Papers 2013-10, School of Economics, The University of New South Wales.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics 117*(3), 817–869. doi: 10.1162/003355302760193904.

Dariel, A. (2013, February). Cooperation preferences and framing effects. Diskussionsschriften dp1302, Universitaet Bern, Departement Volkswirtschaft.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2011). Other-regarding preferences in general equilibrium. *The Review of Economic Studies 78*(2), 613.

Falk, A., E. Fehr, and U. Fischbacher (2002). *Appropriating the Commons: A Theoretical Explanation*. National Academy Press.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

Farizo, B. A., J. Joyce, and M. Solio (2014). Dealing with Heterogeneous Preferences Using Multilevel Mixed Models. *Land Economics 90*(1), 181–198.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods games. *American Economic Review 90*(4), 980–994.

Fehr, E., S. Gächter, and G. Kirchsteiger (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica 65*(4), 833–860.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404. doi: DOI: 10.1016/S0165-1765(01)00394-9.

Goeree, J. K., C. A. Holt, and S. K. Laury (2002). Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *Journal of Public Economics 83*(2), 255–276.

Hoppe, E. I. and P. W. Schmitz (2013). Contracting under incomplete information and social preferences: An experimental study. *The Review of Economic Studies*.

Ioannou, C. A., S. Qi, and A. Rustichini (2012). A test of social preferences theory. Technical report, University of Southampton, Economics Division, School of Social Sciences.

Isaac, R. M. and J. M. Walker (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 179–199.

Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica 77*(1), 135–175.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory 27*(2), 245–252.

Kurzban, R. and D. Houser (2001). Individual differences in cooperation in a circular public goods game. *European Journal of Personality 15*(S1), S37–S52.

Kurzban, R. and D. Houser (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America 102*(5), 1803–1807.

Leider, S., M. M. Mbius, T. Rosenblat, and Q.-A. Do (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics 124*(4), 1815–1851.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics 1*(3), 593–622. Department of Economics, UCLA.

Loewenstein, G. F., L. Thompson, and M. H. Bazerman (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology 57*(3), 426–441.

Mailath, G. and L. Samuelson (2006). *Repeated games and reputations: long-run relationships*. Oxford University Press, USA.

Manski, C. (2011). Policy analysis with incredible certitude*. *The Economic Journal 121*(554), F261–F289.

Manski, C. F. and C. Neri (2013). First- and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior 81*(0), 232 – 254.

Margreiter, M., M. Sutter, and D. Dittrich (2005). Individual and collective choice and voting in common pool resource problem with heterogeneous actors. *Environmental and Resource Economics 32*(2), 241–271.

McCubbins, M. D., M. Turner, and N. Weller (2013). Testing the foundations of quantal response equilibrium. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 144–153. Springer.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and economic behavior 10*(1), 6–38.

Molina, A. (2010). *Teachings from the field to the lab: the role of real common pool resources dependance on experimental behavior*. Ph. D. thesis.

Morey, E., J. Thacher, and W. Breffle (2006). Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model. *Environmental & Resource Economics 34*(1), 91–115.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

Pacifico, D. (2012). Estimating nonparametric mixed logit models via em algorithm. *The Stata Journal Forthcoming*(5), 1281–1302.

Polania-Reyes, S. (2014). *Pro-social behavior, Heterogeneity and Incentives: Experimental evidence from the local commons in Colombia*. Ph. D. thesis.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rassenti, S., S. S. Reynolds, V. L. Smith, and F. Szidarovszky (2000). Adaptation and convergence of behavior in repeated experimental cournot games. *Journal of Economic Behavior & Organization 41*(2), 117 – 146.

Rodriguez-Sickert, C., R. A. Guzmán, and J. C. Cárdenas (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization 67*(1), 215–227.

Rogers, B. W., T. R. Palfrey, and C. F. Camerer (2009). Heterogeneous quantal response equilibrium and cognitive hierarchies. *Journal of Economic Theory 144*(4), 1440–1467.

Rustagi, D., S. Engel, and M. Kosfeld (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Science 330*(6006), 961–965.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature 43*(2), 392–436.

Train, K. E. (2008). Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling 1*(1), 40–69.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Varela, E., J. B. Jacobsen, and M. Soliño (2014). Understanding the heterogeneity of social preferences for fire prevention management. *Ecological Economics 106*(C), 91–104.

Vélez, M. A., J. K. Stranlund, and J. J. Murphy (2009). What motivates common pool resource users? experimental evidence from the field. *Journal of Economic Behavior and Organization 70*(3), 485–497.

Walker, J. M., R. Gardner, and E. Ostrom (1990). Rent dissipation in a limited-access common-pool resource: Experimental evidence. *Journal of Environmental Economics and Management 19*(3), 203–211. doi: DOI: 10.1016/0095-0696(90)90069-B.

Appendix

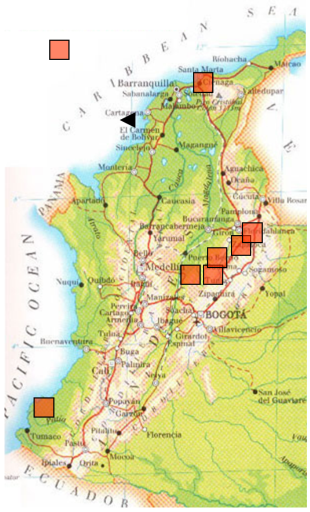
A Labs in the field Data

The experiments were conducted in 8 Colombian villages (see Figure 1) during 2001 and 2002 and a university in Bogotá. A total of 1095 participants attended the sessions, 230 undergraduate students and 865 real users of a CPR. Every village may depend on a different CPR (see Table **??**).

All data were collected using standard procedures in experimental economics in the laboratory: no deception, no field referents, fully salient choices. We collected information on individual characteristics of the villagers only.

Table 7: Labs in the field

Villages	*CPR*
Providencia	Coral reefs
	Coastal fisheries
	Crab gatherers
Gaira	Coastal fisheries
Sanquianga	Clamps
	Fisheries
	Shrimp
	Mangroves
Barichara	Andean Forests
Chaina	Firewood
Tabio	Andean Forests
	Water
La Vega	Water
Neusa	Damn reservoir
	Trout fishing



Note: the red squares are the villages.

Table 8: Table points of the CPR game.

Total Level of the extraction by others

	My Level of Extraction from the Resource								
Total Level of the extraction by others	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Average Level of extraction by others**
4	758	790	818	840	858	870	878	880	1
5	738	770	798	820	838	850	858	860	1
6	718	750	778	800	818	830	838	840	2
7	698	730	758	780	798	810	818	820	2
8	678	710	738	760	778	790	798	800	2
9	658	690	718	740	758	770	778	780	2
10	638	670	698	720	738	750	758	760	3
11	618	650	678	700	718	730	738	740	3
12	598	630	658	680	698	710	718	720	3
13	578	610	638	660	678	690	698	700	3
14	558	590	618	640	658	670	678	680	4
15	538	570	598	620	638	650	658	660	4
16	518	550	578	600	618	630	638	640	4
17	498	530	558	580	598	610	618	620	4
18	478	510	538	560	578	590	598	600	5
19	458	490	518	540	558	570	578	580	5
20	438	470	498	520	538	550	558	560	5
21	418	450	478	500	518	530	538	540	5
22	398	430	458	480	498	510	518	520	6
23	378	410	438	460	478	490	498	500	6
24	358	390	418	440	458	470	478	480	6
25	338	370	398	420	438	450	458	460	6
26	318	350	378	400	418	430	438	440	7
27	298	330	358	380	398	410	418	420	7
28	278	310	338	360	378	390	398	400	7
29	258	290	318	340	358	370	378	380	7
30	238	270	298	320	338	350	358	360	8
31	218	250	278	300	318	330	338	340	8
32	198	230	258	280	298	310	318	320	8

Note: The Self-regarding Nash Equilibrium produces an individual payoff of 320MU whereas the social optimum leads to an individual payoff of 758 MU.

Table 9: Real Users' Socio-economic Characteristics

Variable		**Mean**	**Median**	**Min.**	**Max**	**SD**	**%N**
HH Size		5.59	5	1	51	3.1	87
Age average		34.0	32	7	85	13.9	88
Woman(==1)		46.9	0			49.9	88
Years of education (average)		6.0	5	0	18	3.7	81
Landowners %		75.0	1			43.3	87
Membership %		46.3	0			0.5	95
Meetings Attendance %		11.3	1	0	2080	89.9	77
Perception cooperation %		46.5	50	0	75	28.0	82
Perception interest in CPR	%	62.5	30			37.6	76
Community should control	%	59.7	50	-1	1	42.6	85
Fraction of players with	100%	22.0					
Extraction of the CPR as	50%	65.2					88
main economic activity	0%	12.8					

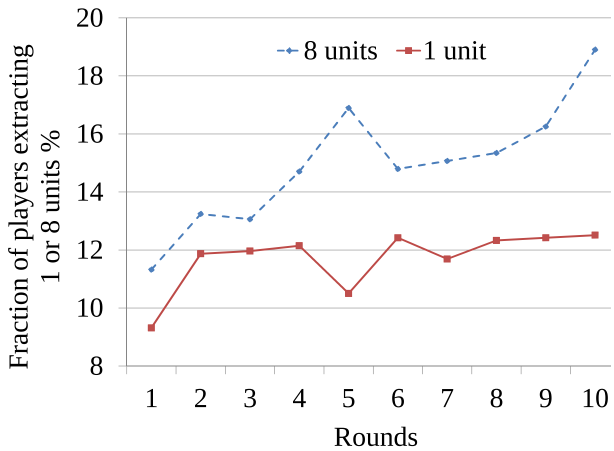


Figure A.1: Baseline: behavior over rounds for Pure Self-regarding and Pure cooperator

B Results of the latent class model under alternative specifications

Below we provide the estimates of class share for students without the restrictions coming from the villagers' model.

Now we present an alternative formulation of the model for students without any restrictions.

Now we present an alternative formulation of the model for villagers without any restrictions.

Table 10: Class share determinants (student sample)

Variable	Self-regarding	Spiteful	Inequity averse	Reciprocators	Std. error
π_i	-0.017	-0.049	0.026	0.043	0.002
$\overline{\pi}_{-i}$	0^a	-0.064	0^a	0^a	0.000
$\max(\pi_i - \overline{\pi}_{-i}, 0)$	0^a	0^a	-0.029	0^a	0.001
$\overline{\pi}_{-i}(x^{i*} - \overline{x}_{-i})$	0^a	0^a	0^a	0.001	0.000
Class Share	0.240	0.107	0.740	0.130	-

[a] Constrained to 0 in estimation

Table 11: Class share determinants - student sample

Variable	Class 1	Class 2	Class 3	Class 4	Std. error
π_i	-0.017	0.058	-0.042	0.010	0.002
$\overline{\pi}_{-i}$	-0.039	0.014	-0.123	-0.053	0.002
$\max(\pi_i - \overline{\pi}_{-i}, 0)$	-0.044	-0.054	-0.056	-0.048	0.000
$\overline{\pi}_{-i}(x^{i*} - \overline{x}_{-i})$	-0.006	-0.018	-0.014	-0.015	0.000
Class Share	0.216	0.137	0.193	0.453	-

[a] Constrained to 0 in estimation

Table 12: Class share determinants - villager sample

Variable	Class 1	Class 2	Class 3	Class 4	Std. error
π_i	-0.005	-0.018	-0.203	0.030	0.001
$\overline{\pi}_{-i}$	-0.045	-0.028	-0.196	-0.104	0.001
$\max(\pi_i - \overline{\pi}_{-i}, 0)$	-0.037	-0.043	-0.071	-0.073	0.000
$\overline{\pi}_{-i}(x^{i*} - \overline{x}_{-i})$	-0.011	-0.012	-0.026	-0.019	0.000
Class Share	0.631	0.243	0.027	0.098	-

[a] Constrained to 0 in estimation