Navarra Center for International Development

Working Paper nº 02/2023

Structural Identification of Social Preferences: Heterogeneity Matters for Incentives

David Echeverry
University of Navarra

María Cristina Figueroa
University of Amsterdam

Sandra Polanía-Reyes
University of Navarra

Structural Identification of Social Preferences: Heterogeneity Matters for Incentives

David Echeverry* María Cristina Figueroa† Sandra Polanía-Reyes‡

October 13, 2023

Abstract

Using a common pool resource (CPR) game with villagers whose livelihood depends on an actual CPR, we estimate a structural model of preferences for altruism, reciprocity and inequity aversion. Latent class estimates show that preferences for equity are widespread. Sociodemographic and attitudinal drivers of preference types provide internal validity. The evidence suggests that incentives to reduce individual extraction (a fine, a subsidy or a non monetary instrument) exert heterogeneous effects across types: a subsidy crowds in inequity aversion and reciprocity, while a fine crowds out the latter (but not the former). We illustrate our type classification using data from a gift exchange game designed to elicit reciprocity. We show that reciprocity is present, but preferences for equity remain essential to explain the data.

Keywords: Reciprocity, altruism, inequity aversion, latent class models, policy intervention.

JEL classification: C51, C93, D63, H41, Q20.

*Universidad de Navarra. Email: `decheverryp@unav.es`.

†University of Amsterdam. Email: `c.figueroa@uva.nl`.

‡Corresponding author. Universidad de Navarra. Email: `spolania@unav.es`. We thank Orazio Attanasio, Sam Bowles, Syngjoo Choi, Antonio Cabrales, Juan Camilo Cárdenas, Simon Gaechter, seminar participants at Advances in Field Experiments, EDepo seminar, UC Berkeley, University of Ǵottingen and University of Notre Dame. We also thank the Kellogg Institute for International Studies and the Center for Ethics and Culture at the University of Notre Dame.

1 Introduction

Economic settings where other-regarding behavior matters are often characterized by non-homogeneous preferences (Kurzban and Houser, 2005; Fischbacher and Gächter, 2006). This heterogeneity can significantly add to the explanatory power of a theory (Erlei, 2008) and achieve more efficient policies. Yet methods to account for the mixture of types in a population are often subject to one of two shortcomings. Either they follow a unifunctional specification, considering only one preference type at a time and thus shedding no light on the type composition of the sample, or they rely on finite mixture models where data-driven labels do not fit the types that are developed by the theory literature. We overcome both shortcomings using a structural approach.

A common pool resource (CPR) is held by a collective of individuals. Each of them can extract some of it to derive an individual payoff, but profit-maximizing behavior can lead to depletion of the resource. Other-regarding behavior can reduce social inefficiencies in the absence of complete contracts and thus solve social dilemmas (Ostrom, 1990), making it an important policy topic. Policy interventions can increase cooperation levels (Delaney and Jacobson, 2016; Cárdenas, 2011). However, such interventions are not neutral to social preferences, and their interaction can have unintended consequences (Bowles and Polania-Reyes, 2012). Incentives can either boost or crowd out social preferences, but much remains to be understood about the heterogeneous effects of policy interventions on different prosociality types.

The key issue, as well as the main contribution of this paper, is identification. We derive a structural model of other-regarding types and apply it in a setting where social preferences are crucial to avoid the tragedy of the commons. We estimate a latent class model on the experimental data gathered by Cárdenas (2011) among CPR users across villages in Colombia. Partner-matched groups play a common pool resource game for 20 rounds. From the output of the first 10 rounds, we simultaneously back out the parameters of the different utility functions, using socioeconomic covariates as type determinants. Our type taxonomy follows a robust literature featuring altruism (Charness and Rabin, 2002), inequity aversion (Fehr and Schmidt, 1999; Bérgolo, Burdin, Burone, De Rosa, Giaccobasso, and Leites, 2022) and reciprocity (Malmendier, te Velde, and Weber, 2014; Malmendier and Schmidt, 2017), applying structural restrictions to characterize four latent classes.

We find that villagers exhibit strong preferences for equity over outcomes, with over 70% class share, compared to reciprocity, altruism and self-regard (which is our baseline specification). We also quantify the parameter for all these utility functions, which is important for equilibrium considerations.[1] Sociodemographic variables drive individual type probabilities. We find that individuals whose main income activity is fishing are more likely to be inequity averse or altruistic rather than selfish or reciprocating. Years of education are linked with a lower probability of being self-regarding.

A growing literature debates whether incentives can crowd out social preferences (Narloch, Pascual, and Drucker, 2012; Handberg and Angelsen, 2019; Kaczan, Swallow, and Adamowicz, 2019; Blanco, Moros, Pfaff, Steimanis, Velez, and Vollan, 2023). In the context of ecosystem services, which is close to the subsidy to preserve the CPR in our framework, we provide evidence that a policy intervention, monetary or otherwise, is likely to have heterogeneous effects across types. We interact the type probabilities, estimated from observations up to round 10, with the introduction of a fine, subsidy or nonmonetary intervention after

[1] For example, Fehr and Schmidt (1999) show that the dominant strategy in equilibrium depends critically on the value of the coefficient governing the preference for fairness.

round 10, and look at the effect of the interaction on individual extraction levels. Our evidence suggests that a subsidy is likely to crowd in social preferences, relative to self-regard where such interventions are relatively less effective. Instead, a fine appears to crowd out reciprocating types relative to self-regarding players.

Assigning players to their most probable type, we can assess the within-type effect of a given instrument. We find that the most effective instrument depends on which type it is applied to. While a low subsidy appears to be more effective for altruists, reciprocators and selfish types, inequity averse individuals respond more to a high fine. In contrast, all types respond in roughly the same way to the ability to vote or decide on the treatment, so there does not appear to be any heterogeneity in the effect of giving institutional participation to players.

A unifunctional approach to a setting characterized by heterogeneous preferences (Burlando and Guala, 2005; Budria, Ferrer-i Carbonell, and Ramos, 2012) is unable to assess the relative weight of different types, let alone characterize its drivers. We illustrate the insights that can be drawn from our structural approach by taking it to the data by Malmendier and Schmidt (2017), who run a gift-giving game on sample of university students. We both replicate and qualify their results. Even as we confirm that reciprocity is present, we show that preferences for equity are more widespread. Moreover, the inequity aversion parameter we back out from the structural results is high enough that inequity averse agents in their setting actually *dislike* receiving a business present. A reduced-form approach would be unable to reach these observations.

A first step towards heterogeneous preferences is finite mixture models (Cappelen, Hole, Sørensen, and Tungodden, 2007; Cappelen, Sørensen, and Tungodden, 2010; Cappelen, Hole, Sørensen, and Tungodden, 2011; Cappelen, Moene, Sørensen, and Tungodden, 2013), calculate type distributions in a sample. A second step, latent type models without structural restrictions (Breffle, Morey, and Thacher, 2011; Morey, Thacher, and Breffle, 2006; Varela, Jacobsen, and Soliño, 2014; Farizo, Joyce, and Soliño, 2014). In these settings, types are inferred ex post from the sign of taste parameters. Because they often do not match those studied by the theory, it makes both economic predictions and external validity enduring issues. Another type of models with structural identification includes random coefficient models (Vélez, Stranlund, and Murphy, 2009; Rodriguez-Sickert, Guzmán, and Cárdenas, 2008; Polania-Reyes, 2015) where structural restrictions are incorporated through multiple binary specifications. To the best of our knowledge, ours is the first structural identification of social preferences where more than two types are simultaneously identified (ex ante) from structural restrictions.

Our common pool resource game provides a rich strategy profile where reciprocity, altruism, inequity aversion and selfishness can be distinguished. Simpler games, such as dictator games, are not rich enough to do so. For instance, Fréchette, Kagel, and Morelli (2005) does not tell apart selfishness from conditional reciprocity. A gift exchange game like the one in Malmendier and Schmidt (2017) cannot distinguish between altruism and inequity aversion. To address this "lack of discrimination" of ORP, some studies use a combination of games (Cox, 2004; Blanco, Engelmann, and Normann, 2011). Combinations of games can also lead to empirical type classifications (Nelson, Schlüter, and Vance, 2018) which make direct inference from the output of a battery of games without resorting to an econometric classification.[2]

[2] Trust and reciprocity have been amply studied in trust games (Berg, Dickhaut, and McCabe, 1995; McCabe, Rigdon, and Smith, 2003), though results are not free of confounding factors (Cox, 2004).

The structural restrictions allow us to control what preference types we end up with, avoiding a data-mined output that may put in question the external validity of the results. But it does raise the question of whether our four-preference approach can satisfy a goodness-of-fit comparison to models with more free parameters. We compare our baseline specification to alternatives with either different structural restrictions or purely data-driven. Based on measures of Bayesian Information (BIC) and Akaike Information Criterion (AIC), we claim that our main specification performs relatively well from a pure data fit perspective vis-à-vis a specification without structural restrictions.

Consistent deviations from the Nash Equilibrium (NE), as documented in the empirical literature (Rassenti, Reynolds, Smith, and Szidarovszky, 2000), can be due to to individuals deviating from payoff-maximizing behavior through some degree of quantal response (McKelvey and Palfrey, 1995; Goeree, Holt, and Palfrey, 2016), experimentation and sampling (Cárdenas, Mantilla, and Sethi, 2015) or learning over time (Burlando and Guala, 2005). Using a quantal response equilibrium (QRE), we compare the coefficient of rationality inferred from our main sample, composed of villagers whose livelihood is connected to a CPR, to that of university students.[3] Our QRE estimates of the rationality parameter are broadly similar across the two samples, so heterogeneity on a cognitive level does not appear to be a first order question. Thus our focus is on heterogeneity in social preferences, shedding light on how this heterogeneity interacts with policy interventions.

[3] Other-regarding preferences have been mostly studied either with only students (Fischbacher, Gächter, and Fehr, 2001; Kurzban and Houser, 2001; Carpenter, Bowles, Gintis, and Hwang, 2009; Falk, Fehr, and Fischbacher, 2002; Walker, Gardner, and Ostrom, 1990) or only real users (Rustagi, Engel, and Kosfeld, 2010; Margreiter, Sutter, and Dittrich, 2005; Vélez et al., 2009). Our sample includes both. In our dataset, users deviate more from the self-regarding outcome than students (Cárdenas 2004; 2011), which is in line with other empirical findings (Carpenter and Seki, 2010; Molina, 2010).

2 Structural model

A general specification of preferences takes into account own payoff, others' payoff and others' behavior. Each period, the individual chooses a level of extraction in order to solve[4]

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

$E^i_{t-1}[\overline{\pi}_{-it}]$ 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 $\overline{x}_{-it}$). We will consider four of the most popular types in the behavioral economics literature: i) self-regarding, ii) altruistic, iii) reciprocator and iv) inequity averse.

Self-regarding individuals care only about their own monetary cost and benefits. Their utility function is given by $U^S = \pi_i$. By construction, the self-regarding optimum coincides with the Nash equilibrium level of extraction $x_i^S = x_i^{NE} = 8$. In contrast, altruistic individuals (Leider, Möbius, Rosenblat, and Do, 2009; Goeree, Holt, and Laury, 2002) see a positive utility component from others' benefit. We adapt our CPR framework to the models proposed by Levine (1998) and Casari and Plott (2003). Individuals that exhibit altruistic preferences 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* has a utility given by

$$U^A = \pi_i + \rho \overline{\pi}_{-i}. \tag{2}$$

[4] For simplicity, we will be assuming linear individual utility functions, which translates expected payoffs into expected utilities. However, neutrality is an important matter measuring other-regarding preferences. The analysis becomes more complicated with other functional forms.

In equation (2), ρ is the parameter of altruism, the positive weight an altruist puts on other's payoff, as long as it is positive. If negative, we can no longer interpret ρ as a parameter for altruism. Without loss of generality, specification (2) can be normalized from a general regression model (3), putting weights on both variables, as $\rho = \tilde{\rho}/\tilde{\eta}$. Applying a similar ratio to models featuring reciprocity and inequity aversion, we reach our estimates for μ and β for Table 4, respectively.

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

A purely altruistic solution, giving a large weight to others' payoff, is equivalent to disliking own payoff. If altruism is seen as an extreme form of concern for social efficiency, the sign of $\tilde{\eta}$ is helpful in making a distinction between altruism and concern for efficiency (Charness and Rabin, 2002): negative $\tilde{\eta}$ is a sign of altruism, positive $\tilde{\eta}$ points to preferences for efficiency. This extends to a negative value for $\tilde{\rho}$, which in the context of a social dilemma can be seen as a strong form of protection for the common resource. While a combination of positive $\tilde{\eta}$ and negative $\tilde{\rho}$ could be seen as a form of spitefulness, when both coefficients are negative this is still altruism, in the form of preference for social efficiency.

Our model for inequity aversion is based on Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). We use the adaptation to a CPR model by Falk et al. (2002). An *inequity averse* individual i has a utility given by

$$U_i^I = \pi_i + \alpha \max(\overline{\pi}_{-i} - \pi_i, 0) + \beta \max(\pi_i - \overline{\pi}_{-i}, 0) \ \forall\ i \tag{4}$$

The second term in equation 4 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, $\beta \leq 0$.[5] Disadvantageous inequality can only be identified under interior solutions (Falk et al., 2002; 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 ($\alpha = 0$). The sign on β identifies preferences for inequity, if positive, and for equity otherwise. The magnitude is also important, as it affects best response functions and hence the equilibrium (Fehr and Schmidt, 1999).

Reciprocators cooperate if only if others do so (Rabin, 1993; Bowles, 2004; Levine, 1998). They abide by a social norm, i.e. a pattern of behavior such that individuals prefer to conform to it on the condition that they believe that most people in their reference network i) conform to it (i.e. empirical expectations) and ii) think they ought to conform to the norm (i.e. normative expectations) (Bicchieri, 2005, 2014). We assume that both are at work in our setting, but our utility specification implicitly assumes that all the effect of a social norm is channeled through its empirical aspect.

A predetermined social norm x_i^* is internalized by player i. A reciprocator conditions her behavior on the perceived cooperation by others according to the utility function

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

[5] In addition, i is loss averse in social comparisons: i suffers more from inequality that is to his disadvantage (Loewenstein, Thompson, and Bazerman, 1989): $\alpha_i \geqslant \beta_i$.

Extraction level x^{*i} is a norm, relative to which i judges extractions from others, deriving more utility if others' extraction is below the norm, and less otherwise. A positive value of μ indicates a desire to uphold the social norm. In fact we use the average number of extracted units in the last practice round, which is arguably the best measure of ex ante expectations.[6]

Empirical expectations are typically anchored on what individuals in the reference group have done in the past (Bicchieri, 2014). In repeated encounters, people have an opportunity to learn from each other's behavior, and to secure a pattern of reciprocity that minimizes the likelihood of misperception (Bicchieri and Muldoon, 2014). This form of experimentation over time is bound to affect positive expectations (Cárdenas et al., 2015) as well as normative ones. We make the simplifying assumption that social norms in our specific setting precede the game and are captured by the average extractions by other players during the trial round, i.e. before any in-game interaction.

In equations (2), (4) and (5), actions and payoffs of other players are only known (in aggregate) with a lag, and thus agents form guesses when computing the value of their given utility function. In Arifovic and Ledyard (2012), expectations are formed from a finite set of remembered strategies and a corresponding probability distribution; learning happens by experimentation, replication and learning. We take into account only the immediately preceding round to say that expectations are fully based on what the experimenter revealed to them at the end of the previous round.

$$E^i_{t-1}[\overline{\pi}_{-it}] = \overline{\pi}_{-i,t-1} \text{ and } E^i_{t-1}[\overline{x}_{-it}] = \overline{x}_{-i,t-1} \tag{6}$$

[6] Polania-Reyes (2015) estimates the structural parameters ρ and μ using a random coefficients model. Selfish behavior is identified as the opposite of selfless behavior as given by the value of ρ.

Because individuals might behave pro-socially in the presence of reputation (Kreps, Milgrom, Roberts, and Wilson, 1982; Bohnet and Huck, 2004; Mailath and Samuelson, 2006), there is a difference between intrinsic and extrinsic reciprocity (Malmendier et al., 2014; Arifovic and Ledyard, 2012; Vélez et al., 2009).[7] Our type identification relies on observing individual behavior over multiple rounds. Agents make their decisions anonymously, which precludes reputation motives.

3 The common pool resource game

We follow Cárdenas (2004) and Cárdenas (2011) and use their data (see also Cárdenas (2009) for the full experimental protocol). Individual i can extract $x_i \in \{1, \ldots, 8\}$ units from the common resource. The individual payoff function is common knowledge and is given by

$$\pi_i = \pi(x_i, x_{-i}) = ax_i - \frac{1}{2}bx_i^2 + \varphi(8n - (x_i + x_{-i})). \tag{7}$$

The payoff features direct benefits from extraction $60x_i - \frac{5}{2}x_i^2$, reflecting a convex cost of effort. In our setting, $a = 60$ and $b = 5$, which places the cost-reward tipping point above the $x_i \leq 8$ limit. Thus, the economic tradeoff lies not in the cost of effort but in the indirect cost of depletion $\varphi(40 - (x_i + x_{-i}))$ following from aggregate extraction. The value $\varphi = 20$ makes the depletion externality salient. The Pareto efficient outcome -or social optimum

[7] Agents in Arifovic and Ledyard (2012) only have other-regarding preferences over outcomes and not over intentions, which implies reciprocity arises as an equilibrium behavior and not as a type. Similarly, Vélez et al. (2009) distinguishes between playing reciprocally and being a reciprocator. In their model, reciprocal behavior arises from preferences for conforming to what others are expected to do.

(SO)- maximizes the aggregate payoff of the group

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

The socially optimal extraction $x_i^{SO} = 1$ corresponds to the minimum level possible. Instead, the Unique Nash Equilibrium (NE) is given by the corner solution $x_i^{NE} = 8$. The wedge between the Pareto optimum and the Nash equilibrium gives rise to the social dilemma.

Participants play a finitely repeated ($T = 10$) game with partner matching. The subgame perfect Nash equilibrium of the repeated game coincides with the one-shot Nash equilibrium.[8] In 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 other players but only the *average* extraction

$$\overline{x}_{-it} = \frac{\sum_{j\neq i}^{n-1} x_{jt}}{n-1}.$$

The experiment was conducted in 8 Colombian villages between 2001 and 2002. The sample is composed of 230 students and 705 real CPR users.[9] Villages and resources, outlined in Table 12, exhibit a good deal of geographic dispersion and diversity of economic activities. Though the full sample contains 865 real CPR users, we exclude those who participated more than once. Participants are paid in COP per the payoff matrix in Table 13, the

[8] 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.

[9] Only villagers were given the survey, students were not.

average payment per player being commensurate with the minimum wage in Colombia at the time of the experiment.

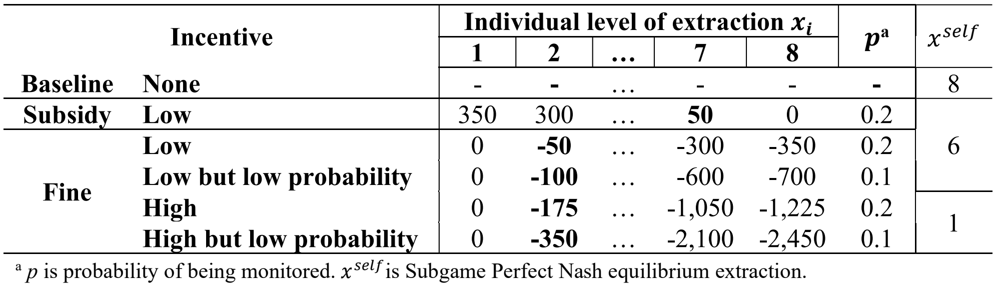
Table 1: Summary Statistics: Sociodemographic Variables

	Observations	Mean	SD
Age	652	34.2	13.8
Female	650	0.5	0.5
Main income activity is CPR	665	0.5	0.4
Main activity is agriculture	664	0.3	0.5
Main income activity is cattle	664	0.1	0.3
Volunteer work	625	0.6	0.5
Education level (years)	598	6.2	3.8
No. people living in the house	658	4.8	2.7
Village with sea view	705	0.4	0.5
Land owner	639	0.8	0.4
Gas kitchen	609	0.7	0.5
Electric kitchen	609	0.2	0.4
years in the zone	628	26.4	16.6
Perceive interest in cooperating	559	0.7	0.3
Perceive community is best guardian	640	0.4	0.5
Thinks CPR will remain still	571	0.3	0.4
Thinks resource is now depleted or will be	571	0.8	0.4

We report summary statistics for our main sociodemographic variables in Table 1. It shows that our sample of villagers consists of a gender-balanced middle aged group of individuals with 6 years of education on average. The majority of these villagers depend on the CPR for their sustenance and are aware that the group extractions deteriorate it. Thus, many engage in volunteer activities or in active cooperation with the attendance to community meetings. Taking the last three variables as a proxy for wealth, we see that, though many individuals are landowners, there is dispersion in wealth levels.[10]

[10] In our sample, only the villagers had their survey data recorded (not the students).

The composition of the group remains the same in rounds $t = 11, \dots, 20$. After round 10, the experimenter announces (and implements) an incentive, which could be monetary (fine or subsidy) or non-monetary (e.g. communication, affecting reputation or other considerations rather than payoffs). Under monetary incentives, an economic incentive s is introduced, proportional to the level of extraction. Each round, one player is chosen randomly with a probability of inspection and face one monetary incentive, either a fine or subsidy, as described in Figure 1.



Incentive		Individual level of extraction x_i					p^a	x^{self}
		1	**2**	**…**	**7**	**8**		
Baseline	**None**	-	-	…	-	-	-	8
Subsidy	**Low**	350	300	…	**50**	0	0.2	6
Fine	**Low**	0	**-50**	…	-300	-350	0.2	
	Low but low probability	0	**-100**	…	-600	-700	0.1	
	High	0	**-175**	…	-1,050	-1,225	0.2	1
	High but low probability	0	**-350**	…	-2,100	-2,450	0.1	

[a] p is probability of being monitored. x^{self} is Subgame Perfect Nash equilibrium extraction.

Figure 1: Description of monetary treatments.

Individuals facing non-monetary incentives are subject to three possibilities. The first one is one-shot communication, a single 5 min face-to-face communication only once prior to making all ten decisions (corresponding to rounds 11 to 20). The second option, repeated communication, allows a single 5 min conversation before each round. Finally, public announcement is as follows: The rule of extracting the social optimum level is announced. For each round, one player is chosen randomly with $p = 0.2$; if violating the rule ($x^{SO} = 1$), he or she pays no fine but must show the monitor his or her extraction level, which is then announced publicly to the group.

The full list of treatments, described in Table 2, shows that both fines and subsidies work effectively induce a reduction in individual extraction, and that the effect of the instrument is increasing in its cost. The summary statistics also show the substantial level of variation across players' extraction levels tends to slightly drop in rounds 10-20, but remains sizable. Cárdenas, Ahn, and Ostrom (2004); Cárdenas (2004) analize these incentives and Cárdenas (2011) compares them without accounting for type heterogeneity.

Table 2: Summary statistics: Extraction levels by treatment, as a percentage of the maximum possible extraction. For each group of players undergoing the same treatment, the table provides average and standard deviation for the first 10 rounds, then for rounds 11-20, and an unpaired t-statistic for the difference in means between the two groups.

	Rounds 1-10		Rounds 11-20		Difference		
	Mean	SD	Mean	SD	Diff.	T-stat	Obs
High fine	59.1	27.6	29.2	24.6	-29.9***	(-19.8)	600
Communication each round	56.8	29.9	32.3	25.3	-24.6***	(-16.0)	650
Low subsidy	59.5	28.5	34.9	26.1	-24.5***	(-17.4)	750
Low fine	58.0	30.1	33.6	26.6	-24.4***	(-19.2)	1000
Low fine but low probability	57.3	31.4	36.5	28.8	-20.8***	(-7.7)	250
High fine but low probability	54.3	27.7	35.8	26.9	-18.6***	(-5.9)	150
High fine, voted each round	61.5	28.9	43.3	29.4	-18.2***	(-10.8)	600
Communication one shot	55.5	28.9	39.3	26.1	-16.3***	(-10.7)	650
Low subsidy assigned by players	48.2	29.3	32.5	24.6	-15.7***	(-7.1)	300
Low fine, voted each round	60.6	27.7	46.5	28.8	-14.2***	(-8.7)	600
Low fine, voted once	52.5	29.3	40.2	28.4	-12.3***	(-7.4)	600
Low fine assigned by players	45.5	28.1	34.9	26.9	-10.6***	(-6.1)	500
Controls	58.2	27.0	59.2	27.5	1.0	(0.5)	400

The socially optimal extraction level is 12.5%, and the Nash equilibrium extraction 100%. T-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

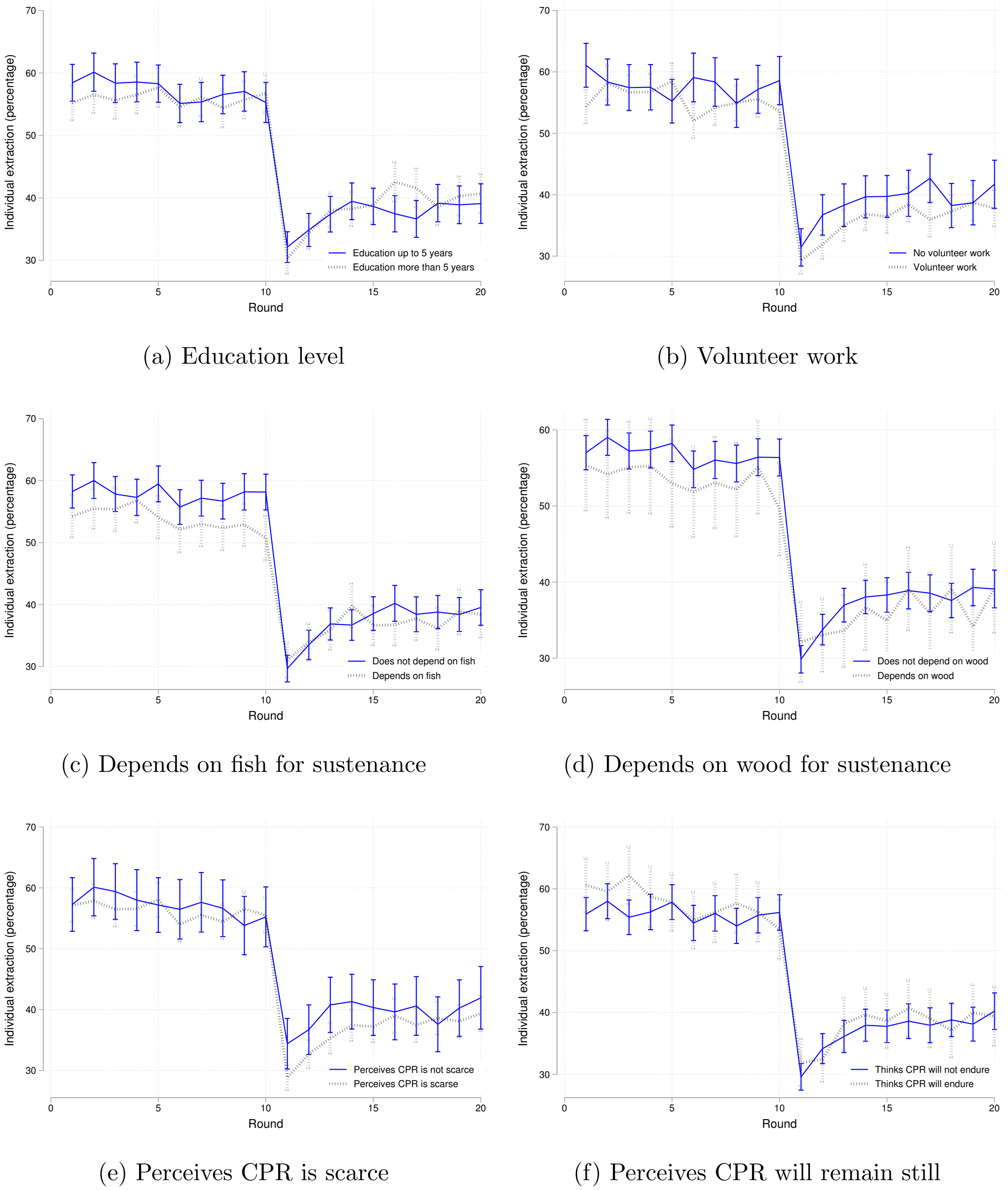
Individuals who consistently play the Nash equilibrium strategy -maximal extraction- are a small proportion of the sample, as are individuals who consistently play the social optimum -minimal extraction-. Figure 2 shows the evolution of extraction over rounds. We notice that cooperation does not seem to collapse in the last round (round 10 or round 20). The graphic evidence suggests that volunteering, level of education, CPR dependence for sustenance and perception of scarcity lower the level of extraction, while the perception that the CPR will remain still tend to increase it. The differences are often not statistically significant at 95% confidence, which shows there is considerable variation not accounted for by individual observables. This motivates the latent class model we develop in the next section.

4 Identification: latent class model

We estimate a latent class logit model (Dempster, Laird, and Rubin, 1977; Train, 2008), using only the first 10 rounds of observations to identify individual types. A set of N agents ($N_{villagers} = 705$ and $N_{students} = 230$) face $J = 8$ alternatives in each of the $T = 10$ (initial) rounds of the game. Given a number of $C = 4$ classes, a latent class logit algorithm simultaneously estimates taste parameters, referring to the different possible levels that an individual chooses, and a set of class membership parameters associated with sociodemographic characteristics. In Appendix A.2 we provide more detail about the iterative process involved in the simultaneous computation of these two sets of parameters, which we estimate using the Stata command developed by Rabe-Hesketh and Skrondal (2008) to estimate generalized linear latent and mixed (GLLAM) models.[11]

Structural restrictions define a latent class (i.e. a preference type) by making all parameters equal to zero except for the two that characterize the class. Thus, an inequity averse individual will exhibit a taste for her own payoff and for advantageous inequality, for instance, as summarized in Table 3. Table 4 summarizes our latent class logit results, where

[11] 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. To initialize values for the iterative process, we use a wrapper of gllamm called lclogit citeppacifico12.



(a) Education level

(b) Volunteer work

(c) Depends on fish for sustenance

(d) Depends on wood for sustenance

(e) Perceives CPR is scarce

(f) Perceives CPR will remain still

Figure 2: Average extraction level over the 20 rounds of play across subsamples of players. Vertical whiskers plot 95% confidence intervals around the average.

Table 3: Summary Statistics: Experimental Variables (Rounds 1-10). We calculate individual means for each of the taste variables used in (5), (4) and (2) and their standard deviation. Payments are provided in COP.

	Villagers			Students		
Rounds 1-10	count	mean	sd	count	mean	sd
Own Payoff	7,050	554.0	105.8	2,300	515.4	92.9
Others' Payoff	7,050	554.5	78.2	2,300	515.1	68.4
Advantageous Inequality	7,050	31.3	45.2	2,300	27.4	39.1
Deviation from the norm	7,050	-1595.8	570.8	2,300	-1832.9	438.9
	Villagers			Students		
Rounds 11-20	count	mean	sd	count	mean	sd
Own Payoff	7,050	619.8	119.5	2,300	567.7	137.9
Others' Payoff	7,050	633.1	100.3	2,300	572.5	124.5
Advantageous Inequality	7,050	23.5	46.2	2,300	34.1	61.6
Deviation from the norm	7,050	-935.9	767.5	2,300	-1187.5	796.0

self-regarding individuals are the baseline type and are thus omitted. The first four rows report taste parameters for each type, and the remaining rows report membership likelihood variables for each class. The constant directly translates into the baseline share of the given type, a positive value implying a higher likelihood of that type with respect to self-regard.

Class membership parameters are given by using the individual sociodemographic characteristics from Table 1 as membership drivers. The link between sociodemographics and membership varies by type. In all three specifications, there is strong evidence about the prevalence of inequity aversion. Taste parameters appear to be robust across specifications. Specification I is the raw model without membership variables. Specification II adds perception that the CPR will hold, social capital proxies and income source. Specification III, which we use as our main model for type identification, includes education and age. The main driver of type probabilities, years of education, is overall linked with a lower probability of being self-regarding. Fishermen appear to be more likely to be altruistic or inequity averse.

We compute membership probabilities in Table 5, where we also compute utility preference parameters as the quotient of the relevant coefficients. Though the theory suggests four types, we contrast the goodness of fit of such a framework with a data mining approach with no structural constraints. A latent class logit model takes as an input the number of classes, C. Both, our structural model developed in the previous section and BIC/AIC analysis of model fit provide us with 4 distinct classes (or types) within our sample. Table 5 provides the model fit comparison for the sample of villagers in the first ten rounds.

To elaborate Table 5, we use the same membership variables as in specification III from Table 4. The first column specifies the number of classes considered and is followed by the information criteria values as well as the number of estimated parameters. The columns on *utility weight* provide the specific taste parameters for reciprocity (μ), inequity aversion (β) and altruism (ρ), which we back out by normalizing the taste coefficient with the coefficient for the utility of own payoff.

The taste parameters of our other-regarding preferences are robust for reciprocity, and especially so for inequity aversion. Both coefficients are consistently negative, clearly showing preferences for adhering to the social norm and for equity, respectively. The altruistic type we uncover is more of a concern for social efficiency, as discussed in Section 3. The last four columns, which present average type composition probabilities, show that the share of inequity aversion across specifications is robust and large, with reciprocators and altruists making up a lower share (lower also than self-regarding individuals).

Specification	I			II			III		
	Reciprocity	Inequity Aversion	Altruism	Reciprocity	Inequity Aversion	Altruism	Reciprocity	Inequity Aversion	Altruism
Own Payoff	-0,030 ***	0,018 ***	-0,070 ***	-0,029 ***	0,019 ***	-0,067 ***	-0,030 ***	0,020 ***	-0,048 ***
	(0,002)	(0,001)	(0,009)	(0,002)	(0,001)	(0,011)	(0,002)	(0,001)	(0,009)
Advantageous inequality	0	-0,023 ***	0	0	-0,022 ***	0	0,000	-0,022 ***	0
	(.)	(0,001)	(.)	(.)	(0,001)	(.)	(.)	(0,001)	(.)
Deviation from the norm	-0,003 *	0	0	-0,003 **	0	0	-0,004 **	0	0
	(0,001)	(.)	(.)	(0,001)	(.)	(.)	(0,002)	(.)	(.)
Others' payoff	0	0	-0,076 ***	0	0	-0,079 ***	0,000	0	-0,056 ***
	(.)	(.)	(0,009)	(.)	(.)	(0,010)	(.)	(.)	(0,008)
Main income activity is fish				-0,525	0,578	-1,276	0,746	1,448 **	1,733 **
				(0,629)	(0,422)	(0,896)	(0,805)	(0,581)	(0,832)
Main income activity is wood				1,130	0,430	1,285	0,741	0,093	-0,585
				(0,826)	(0,719)	(0,952)	(0,958)	(0,765)	(1,236)
Thinks CPR will remain still				0,641	-0,448	-1,049	0,627	-0,170	-0,211
				(0,732)	(0,456)	(0,675)	(0,782)	(0,549)	(0,827)
Number of community meetings attended				0,004	-0,005	-0,015	0,000	-0,003	-0,003
				(0,004)	(0,006)	(0,024)	(0,005)	(0,006)	(0,007)
Education level (years)							0,303 **	0,240 **	0,340 **
							(0,126)	(0,117)	(0,138)
Age							-0,003	-0,014	-0,046
							(0,020)	(0,016)	(0,031)
Gas stove							1,383 *	0,178	1,285
							(0,707)	(0,461)	(0,941)
Constant	-0,726 ***	1,535 ***	-1,024 ***	-1,124	1,836 ***	-0,129	-3,195 **	1,135	-1,812
	(0,224)	(0,160)	(0,335)	(0,737)	(0,465)	(0,638)	(1,346)	(0,961)	(1,792)
Observations	56,400			40,880			33,280		

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: GLLAMM specifications for villagers in the first ten rounds with just one visit.
Results from the latent class logit model with four types: inequity averse, reciprocator, altruist and self-regarding, the latter being the reference type. The model is estimated by maximum likelihood on the villager sample on a dummy variable denoting which of the 8 possible extraction levels was chosen for each player and round. Class membership predictors include sociodemographic and attitudinal variables from the survey. The estimation uses the first ten rounds of the game. A coefficient of 0 in regressions corresponds to a structural restriction applied to the latent class.

C	LL	Par	BIC	CAIC	Utility weight (normalized)			Class shares			
					Ineq. Averse	Reciprocal	Altruist	P(R)	P(IA)	P(A)	P(SR)
					$\hat{\beta}$	$\hat{\mu}$	$\hat{\rho}$				
2	-8328.78	12	16740.71	16729.71	0.104	0	0	0.233	0	0	0.767
2	-7975.82	12	16034.78	16023.78	0	-1.866	0	0	0.754	0	0.246
2	-7975.81	14	16042.31	16030.31	0.006	-1.865	0	0.246	0.754	0	0
2	-8331.40	12	16745.94	16734.94	0	0	-0.309	0	0	0.778	0.222
2	-7978.94	14	16048.57	16036.57	0	-0.961	0.811	0	0.844	0.156	0
3	-8210.61	25	16579.95	16558.95	-0.566	0	1.700	0.728	0	0.160	0.112
3	-7752.97	25	15664.67	15643.67	0.122	-1.137	0	0.099	0.730	0	0.171
3	-7740.85	25	15640.42	15619.42	0	-1.126	1.830	0	0.733	0.102	0.165
3	-7740.69	28	15647.66	15625.66	0.022	-1.124	1.828	0.165	0.733	0.102	0
4	-7702.47	40	15639.24	15608.24	0.116	-1.119	1.165	0.099	0.735	0.067	0.099
4	-7276.34	40	14855.01	14815.01	N/A			Unconstrained			
5	-7228.36	52	14849.74	14797.74	N/A			Unconstrained			

Table 5: Model comparison for different number of classes. For different combinations of types, we estimate the GLLAMM model using taste variables as in specification III from Table 4.

5 Heterogeneous effects of incentives, crowding in and crowding out

The type identification derived from the first 10 rounds of data yields idiosyncratic type probabilities. As illustrated in Table 2, a rich set of instruments is applied after round 10. We now run a linear regression of the individual extraction level (as a percentage of the maximum possible) on the interaction of type probability, expressed in percentage points, and a dummy for each incentive (fine, subsidy or communication). We use a saturated specification, controlling for type probabilities and the isolated effect of incentives. In our main specification we use the data on all rounds, which gives 20 observations per player.

The evidence, shown in Table 6, suggests that incentives exert heterogeneous effects across types. For instance, if the probability of being inequity averse goes up by 10 percentage points, the effect of introducing the subsidy is to reduce extraction levels by 20.4 percent.

This is almost twice the effect it would have if the 10% increase was applied to the probability of being a reciprocator. Overall, non-monetary incentives appear to be most effective for reciprocators. Both the subsidy and the non-monetary instrument appear to crowd in social preferences, relative to self-regard (where such interventions appear to be less effective), the crowding in being strongest for inequity aversion.

Table 6 exhibits some evidence on the extensive margin that a fine is most effective on self-regarding individuals (the baseline category) relative to the other three types, for which the interaction coefficient is either positive or insignificant. Another way of interpreting the result is a crowding out effect, concerning reciprocators in particular, relative to self-regarding types. For inequity averse individuals, the overall effect is unclear as intensive margin effect, provided under (2), stands somewhat in contrast with the extensive margin effect from (1).

For a robustness check, in Table 15, we average all observations from rounds 1-10 and rounds 11-20, which gives two observations per player (before the instrument and after its application). Though all coefficients are directionally the same, and close in magnitude to those from Table 6, their statistical significance drops with the drastic reduction in number of observations. The coefficients that remain statistically significant are the interaction between subsidy (fine) and the probability of inequity aversion (reciprocity), indicative of crowding in (out) of the said preference.

The previous analysis compared what happens to different types conditional on each incentive. To isolate types and compare different incentives, we now allocate individuals into the most probable type from their idiosyncratic distribution. Table 7 examines the relative effect of the different instruments from a t-statistic test of the difference in means before and

Table 6: Effect of incentives on the extraction level. This table reports the outcome of a linear regression of extraction level (as a percentage of the maximum possible) on the individual type probability, where type probability is derived with the classification given by specification III in Table 4.

	Fine		Subsidy		Non-monetary
	(1)	(2)	(3)	(4)	(5)
Instrument	-37.321***		-28.978***	141.647***	-34.136**
	(8.75)		(5.48)	(38.16)	(15.94)
× Prob(inequity averse)	0.182*			-2.041***	0.299
	(0.10)			(0.47)	(0.19)
× Prob(reciprocator)	0.526***		0.258	-1.188***	-0.792**
	(0.13)		(0.29)	(0.44)	(0.31)
× Prob(altruist)	0.171		0.069	-0.471	-0.658**
	(0.14)		(0.41)	(0.38)	(0.28)
Fine amount		0.081			
		(0.06)			
× Prob(inequity averse)		-0.002***			
		(0.00)			
× Prob(altruist)		0.003***			
		(0.00)			
× Prob(reciprocator)		0.002			
		(0.00)			
Prob(inequity averse)	Yes	Yes	Yes	Yes	Yes
Prob(reciprocator)	Yes	Yes	Yes	Yes	Yes
Prob(altruist)	Yes	Yes	Yes	Yes	Yes
Obs	6,160	8,320	1,240	1,240	1,760
R squ	0.12	0.12	0.13	0.15	0.18

The variable Instrument takes a value of zero for controls in all rounds, as well as treated individuals during the first 10 rounds; it takes a value of one for all treated individuals in rounds 11 to 20. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

after the treatment in question. Among inequity averse individuals a high fine appears to be the most effective instrument, while a low subsidy achieves the highest impact for all other types. The results reinforce the previous findings that type matters for a given treatment by showing that the relative effectiveness of a given instrument is likely to be type-dependent.

Table 7: Effect of incentives on individual extraction level, conditional on most probable player type. This table reports the outcome of a T-test of difference in average extraction level (as a percentage of the maximum possible) by treatment and type. Types are assigned according to the maximum probability within each agent's distribution, as derived from the GLLAMM algorithm on the first 10 rounds of data. We compare the mean before (rounds 1-10) and after (rounds 11-20) the given treatment.

	Inequity averse		Altruists		Reciprocators		Self-regarding	
	Diff	T-stat	Diff	T-stat	Diff	T-stat	Diff	T-stat
High fine	-23.4	(-21.7)	-22.4	(-8.9)	-22.4	(-8.9)	-22.4	(-8.9)
Low subsidy	-22.4	(-18.8)	-25.2	(-16.9)	-24.5	(-16.2)	-25.2	(-16.9)
Communication	-20.4	(-18.8)	-17.7	(-11.3)	-17.7	(-11.3)	-17.7	(-11.3)
Low fine	-17.2	(-22.8)	-19.1	(-16.5)	-19.1	(-16.6)	-19.1	(-16.5)
Controls	1.0	(0.5)	1.8	(0.7)	1.8	(0.7)	1.8	(0.7)
Observations	14,040		5,780		5,820		5,800	

T statistics in parentheses.

Another characteristic of treatments that could matter lies in their governance. While experimental treatments are typically designed and enforces by the experimenter, several of the treatments described in Table 2 involve the active participation of players, through voting or enforcement. For each of the instruments, we now distinguish those that involve participation from those that don't. The outcome, shown in Table 8, suggests that types are not differentially affected by the ability to decide on the application of the instrument. The gap between the effect of an instrument (fine or subsidy) that was voted, and one that wasn't, is roughly constant across types.

Table 8: Effect of voting on the extraction level, conditional on most probable player type. This table reports the outcome of an unpaired T-test of difference in average extraction level (as a percentage of the maximum possible) by whether a given treatment was agreed to through a voting system, including instruments assigned by players. Types are identified from the maximum probability of each agent's distribution, as derived from the GLLAMM algorithm on the first 10 rounds of data. We compare the mean before (rounds 1-10) and after (rounds 11-20) the treatment.

	Inequity averse		Altruists		Reciprocators		Self-regarding	
	Diff	T-stat	Diff	T-stat	Diff	T-stat	Diff	T-stat
Not voted	-21.8	(-30.6)	-21.8	(-21.0)	-21.5	(-20.7)	-21.8	(-21.0)
Voted	-14.1	(-17.6)	-14.6	(-10.4)	-14.7	(-10.5)	-14.7	(-10.4)
Observations	11,440		4,520		4,560		4,540	

T statistics in parentheses.

6 Social preferences or cognition?

Decision-makers are often subject to heuristics including Bayesian updating (El-Gamal and Grether, 1995; Houser, Keane, and McCabe, 2004) and sampling behavior (Cárdenas et al., 2015). We estimate a logit QRE specification (Goeree et al., 2016), which Cárdenas et al. (2015) apply to the student sample from Cárdenas (2004), and we to that of villagers.[12]

Players choose their effort decision following a distribution $P(x = k)$, $k \in \{1, \ldots, e\}$ that 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 playing x_i given others are mixing strategies according to $P(.)$. The QRE[13] associated to the error parameter $\lambda \in [0, \infty)$[14] is given by

[12] Cárdenas et al. (2015) point out that QRE outperforms payoff sampling equilibrium under corner solutions. Given that both the social optimum and Nash equilibrium are corner solutions, this favors the use of QRE in our setting.

[13] 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.

[14] 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.

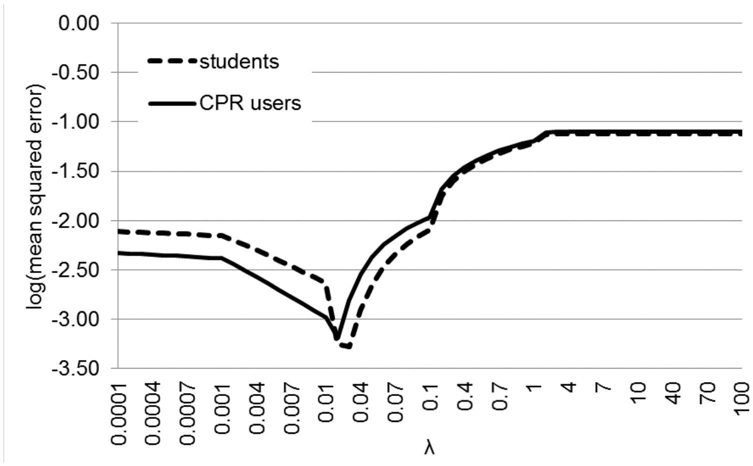


Figure 3: log(MSE) as a function of λ.

$$P(x_i = k) = \frac{\exp(\lambda\pi(k,P))}{\sum_{j=1}^{8}\exp(\lambda\pi(j,P))},\ k \in \{1,\ldots,8\} \tag{9}$$

λ is chosen to match the QRE distribution, which derived from the payoff function alone, to the observed distribution. Like Cárdenas et al. (2015) we choose λ in order to minimize mean squared error (MSE), which Figure 3 depicts for each value of λ.

The value of λ minimizing MSE is very close across the samples: 0.03 for students and slightly lower for real CPR users at 0.02. Though this suggests a somewhat higher level of rationality among the student sample, the order of magnitude is the same. Figure 4 compares the predicted and realized distributions. A slightly better fit is achieved within the student sample ($MSE = 0.053\%$) than that of real CPR users ($MSE = 0.065\%$).

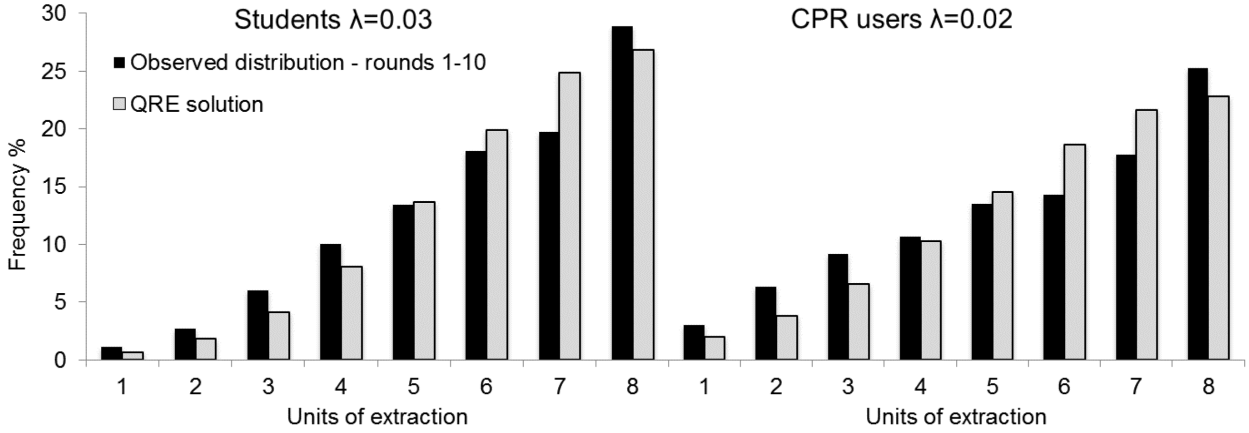


Figure 4: Observed distribution of choice outcomes and QRE distribution for the given value of λ.

7 External validity of our type classification

Malmendier and Schmidt (2017), henceforth MS, take a unifunctional approach to a problem of multiple social preferences. Though their game is not rich enough to distinguish between certain pairs of preference types, which we discuss in this section, it is sufficient to illustrate the insights from our multifunctional approach. MS use a laboratory experiment with university students to study gift giving and reciprocity. We show that though reciprocity is indeed present, preferences for equity remain preponderant.

A decision maker (DM) receives a fixed wage of 20 tokens to buy a product from one of two producers on behalf of a client. The products are lotteries with different expected values (qualities). The client receives the outcome of the lottery for the chosen product. The producer that was chosen receives 16 tokens and the other one receives 0. Under the gift treatment (GT), the first one is endowed with an extra token which can be sent as an unconditional gift to the DM, in which case the DM receives 2 extra tokens. Before deciding which product to buy, the DM learns whether the gift was sent or not by the potential gift giver as well as the quality of each product.

MS provide evidence for reciprocity, which we confirm and also qualify by weighing its share against that of inequity aversion. Using a reduced-form specification, they observe that the DM favors the gift giver when the gift is sent and punishes him (by favoring the competing producer) when the gift is not sent. Gift-induced reciprocity can generate negative externalities when the DM favors a worse quality producer who gave a gift, or punishes a better quality producer in favor of a worse alternative if the gift was not sent. MS note that the classical social preference theories (altruism, inequity aversion, maximin preferences) fail to explain the observed behavior in their gift-giving setup.

The setting by MS only allows the identification of three distinct types: inequity averse, maximin and reciprocator.[15] Self-regarding DMs cannot affect their payoff and thus they are indifferent about buying from either producer, so we cannot identify this type. Regarding equity concerns, only advantageous inequality can be observed because the DM is always better off than the rest. Altruists are thus indistinguishable from inequity averse types. Column (2) in Table 9 presents the utility functions associated to each type in the GT.

We estimate a model with the same utility functions used by MS: inequity aversion, altruism, self-regard, maximin and reciprocity. By extending their unifunctional framework, we can both confirm and assess the relative weight of reciprocity among their sample. As Table 10 suggests, reciprocity does not have a higher share of the sample compared to inequity aver-

[15] Maximin preferences imply favoring the producer which cannot send gifts if the gift is not sent. If the gift is sent, the choice of whom to buy from does not change the minimum payoff of the group, someone always receives 0 independently of the DM choice, thus the DM favors the client in this case. A reciprocating DM will favor the gift-giver conditionally on sending the gift and punish him by favoring the other producer otherwise.

Type	Utility function
Self-regarding (SR)	NA
Altruistic (A)	$U_i = \frac{\rho}{2}(\pi_{gg} + \pi_{ngg})$
Inequity Averse (IA)	$U_i = \beta \max(\pi_i - \frac{1}{2}(\pi_{gg} + \pi_{ngg}), 0)$
Maximin (MM)	$U_i = min(\pi_i, \pi_{gg}, \pi_{ngg}, \pi_C)$
Reciprocator (R)	$U_i = giftgiver * gift$

Table 9: Utility functions for each type in each treatment. The payoffs of the decision maker, the gift giving producer, the non gift giving producer and the client are denoted by $\pi_i, \pi_{gg}, \pi_{ngg}, \pi_C$, respectively. *gift* $= \{1, -1\} \equiv \{$gift given, gift not given$\}$;*giftgiver*$=$ $\{1, -1\} \equiv \{$potential gift giver chosen, potential gift giver not chosen$\}$; thus, for reciprocity, $U_i = \{1, -1\} \equiv \{$reciprocation, no reciprocation$\}$.

sion. The constant is not statistically significant for either type, and neither are the rest of sociodemographic variables MS measured, including age, gender and college major.

From a goodness of fit perspective, Table 11 shows that the models that best fit the data must include reciprocity. The best model explaining behavior in the GT treatment, looking at the Akaike and Bayesian information criteria, is a model including reciprocity, inequity aversion/altruistim and maximin preferences. The estimated shares, which are relatively stable across the different model specifications, suggest that equity/altruistic concerns explain the largest fraction of the behavior, followed by reciprocity.[16]

8 Conclusion

This study applies a latent class logit to identify other-regarding preferences in a common pool resource game. We bring a novel method to identify types in a unique sample, including real CPR users and students. Our structural estimation relies on four types, which the data

[16] Further analysis in alternative treatments also confirm that reciprocity explains behavior in a setting without third-parties, as MS suggest (see Appendix A.5).

	Inequity aversion	Reciprocity	MaxMin
Advantageous inequality	-3.095***	0	0
	(0.294)	(.)	(.)
Reciprocity	0	1.368***	0
	(.)	(0.115)	(.)
Maximin	0	0	-0.0666
	(.)	(.)	(0.363)
Age	0.252	0.237	
	(0.209)	(0.210)	
Female	-19.08	-19.08	
	(346.6)	(346.6)	
Economist	19.65	19.89	
	(639.4)	(639.4)	
Constant	13.39	13.27	
	(346.6)	(346.6)	
Observations	1920		

Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Results from the latent class logit model with three types: inequity averse, reciprocator, maximin, the latter being the reference type. The model is estimated by maximum likelihood on the gift treatment sample.

Types	C	LL	Nparam	CAIC	BIC	Utility weight			Class shares		
						R	IA(/A)	MM	P(R)	P(IA)	P(MM)
R+IA(/A)	2	-407.943	5	830.500	827.500	1.241	-2.141	0	0.384	0.616	0
R+MM	2	-522.226	5	1059.065	1056.065	1.428	0	0.800	0.345	0	0.655
IA(/A)+MM	2	-508.664	5	1031.942	1028.942	0	-2.336	2.954	0	0.581	0.419
R+IA(/A)+MM	3	-387.983	11	800.321	795.321	1.367	-3.390	-0.044	0.363	0.512	0.125

Table 11: We present all possible models in the gift treatment. Type $i = \{R, IA, MM, A\}$ stand for Reciprocator, Inequity Averse, Maximin and Altruist, respectively. Columns 7-10 provide baseline type probabilities (NA means that the share is not estimated).

supports based on information-based tests of the optimal number of types. The most salient feature is the prevalence of inequity aversion relative to altruistic behavior or reciprocity.

We assess the external validity of our type classification method by using it on a second dataset, based on a gift-exchange game designed to elicit reciprocity. We find that though reciprocity is indeed present, preferences for equity remain not only sizable but even prevalent across the sample. The results illustrate the comprehensive nature of our approach in weighing different preference types within a given population, not based on pairwise structural restrictions or ex post labels.

A quantal response equilibrium suggests that the rationality parameter of our villagers is of the same order of magnitude as that of our sample of students, so cognitive heterogeneity does not seem to be a channel of first order importance. However, equilibrium considerations apply. In particular, while we restrict ourselves to variables at the individual level, group composition is likely to be key feature. Because types are likely to be group-dependent (Polania-Reyes, 2015), an evolutionary approach could help better understand collective action problems in the long run.

Finally, types are likely to be endogenous to institutions. For that reason we rely only on the pre-treatment rounds for identification, and then consider the effect of various incentives on each type without looking at whether those incentives can affect the type. This remains a question for further research.

References

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.

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.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*, 122–142.

Bérgolo, M., G. Burdin, S. Burone, M. De Rosa, M. Giaccobasso, and M. Leites (2022). Dissecting inequality-averse preferences. *Journal of Economic Behavior & Organization 200*, 782–802.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bicchieri, C. (2014). Norms, conventions, and the power of expectations. *Philosophy of Social Science: A New Introduction*, 208.

Bicchieri, C. and R. Muldoon (2014). Social norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.).

Blanco, E., L. Moros, A. Pfaff, I. Steimanis, M. A. Velez, and B. Vollan (2023). No crowding out among those terminated from an ongoing pes program in colombia. *Journal of Environmental Economics and Management 120*, 102826.

Blanco, M., D. Engelmann, and H. T. Normann (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior 72*(2), 321–338.

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.

Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*.

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.

Budria, S., A. Ferrer-i Carbonell, and X. Ramos (2012). Personality and dislike for inequality: Typifing inequality aversion with respect to locus of control. *Mimeo*.

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., T.-K. Ahn, and E. Ostrom (2004). Communication and co-operation in a common-pool resource dilemma: A field experiment. *Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality*, 258–286.

Cárdenas, J. C., C. Mantilla, and R. Sethi (2015). Stable sampling equilibrium in common pool resource games. *Games 6*(3), 299.

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.

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

Cox, J. (2004). How to identify trust and reciprocity. *Games And Economic Behavior 46*(2), 260–281.

Cárdenas, J. C. (2009). *Dilemas de lo colectivo: Instituciones, pobreza y cooperación en el manejo local de los recursos de uso común* (1 ed.). Universidad de los Andes, Colombia.

Delaney, J. and S. Jacobson (2016). Payments or persuasion: common pool resource management with price and non-price measures. *Environmental and Resource Economics 65*, 747–772.

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.

El-Gamal, M. A. and D. M. Grether (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American Statistical Association 90*(432), 1137–1145.

Erlei, M. (2008). Heterogeneous social preferences. *Journal of Economic Behavior & Organization 65*(3-4), 436–457.

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

Farizo, B. A., J. Joyce, and M. Soliño (2014). Dealing with heterogeneous preferences using multilevel mixed models. *Land Economics 90*(1), 181–198.

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

Fischbacher, U. and S. Gächter (2006). Heterogeneous social preferences and the dynamics of free riding in public goods.

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.

Fréchette, G. R., J. H. Kagel, and M. Morelli (2005). Behavioral identification in coalitional bargaining: An experimental analysis of demand bargaining and alternating offers. *Econometrica 73*, 1893–1938.

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.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2016). *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press.

Handberg, Ø. N. and A. Angelsen (2019). Pay little, get little; pay more, get a little more: A framed forest experiment in tanzania. *Ecological Economics 156*, 454–467.

Houser, D., M. Keane, and K. McCabe (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica 72*(3), 781–822.

Kaczan, D. J., B. M. Swallow, and W. V. Adamowicz (2019). Forest conservation policy and motivational crowding: Experimental evidence from tanzania. *Ecological Economics 156*, 444–453.

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. Möbius, 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.

Malmendier, U. and K. M. Schmidt (2017). You owe me. *American Economic Review 107*(2), 493–526.

Malmendier, U., V. L. te Velde, and R. A. Weber (2014). Rethinking reciprocity. *Annu. Rev. Econ. 6*(1), 849–874.

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.

McCabe, K., M. Rigdon, and V. Smith (2003). Positive reciprocity and intentions in trust games. *Journal Of Economic Behavior & Organization 52*(2), 267–275.

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.

Narloch, U., U. Pascual, and A. G. Drucker (2012). Collective action dynamics under external rewards: Experimental insights from andean farming communities. *World Development 40*(10), 2096–2107.

Nelson, K. M., A. Schlüter, and C. Vance (2018). Distributional preferences and donation behavior among marine resource users in wakatobi, indonesia. *Ocean & Coastal Management 162*, 34–45. Coastal Systems in Transition.

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

Polania-Reyes, S. (2015). *Pro-social behavior, Heterogeneity and Incentives: Experimental evidence from the local commons in Colombia*. Ph. D. thesis.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

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.

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

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

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.

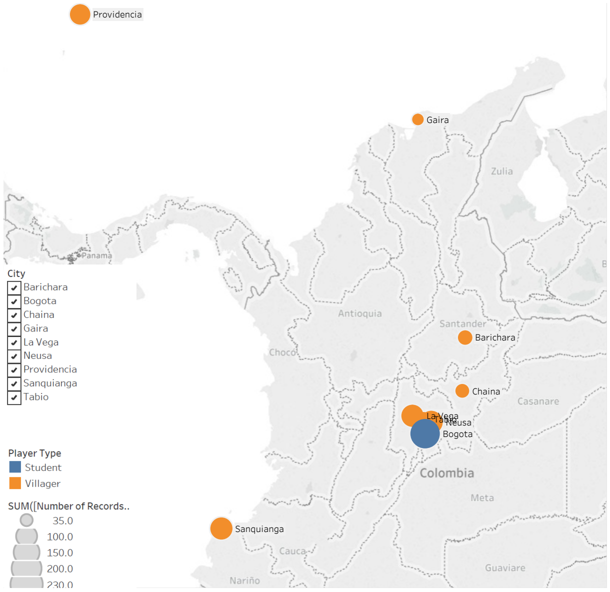
A Appendix

A.1 A finite mixture model without type identification

We suppose the population comprises 4 homogeneous (unobservable) types. On each round $t \in \{1, \ldots, T\}$, 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,

Table 12: Labs in the field

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



being self-regarding, inequity averse, reciprocator or altruist, $q \in \{S, I, R, A\}$. At time t, agent i chooses an action $j \in \{1, \ldots, J\}$ to derive utility

$$\tilde{U}^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) = U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1}) + \varepsilon^q_{ijt} \quad \forall j \in \{1, \ldots, J\} \tag{10}$$

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

$$\tilde{f}_q(x_{ijt}; \theta_q, \lambda_q, \overline{x}_{-it-1}) = \frac{\exp[\lambda_q U^q(x_{ijt}; \theta_q, \overline{x}_{-it-1})]}{\sum\limits_{m=1}^{J} \exp(\lambda_q U^q(x_{imt}; \theta_q, \overline{x}_{-it-1}))} \tag{11}$$

This logit function is reminiscent of the QRE specification of section 6. 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.

Table 13: Table points of the CPR game.

		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
Total Level of the 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

Specification	I			
	Reciprocal	Inequity Averse	Altruist	Selfish
Own payoff	0.017***	0.025***	-0.053***	0.048***
	(0.006)	(0.001)	(0.011)	(0.008)
Advantageous inequality	0	-0.028***	0	0
	(.)	(0.001)	(.)	(.)
Deviation from norm	0.008*	0	0	0
	(0.004)	(.)	(.)	(.)
Others' payoff	0	0	-0.067***	0
	(.)	(.)	(0.009)	(.)
Constant	-1.137	1.778***	-0.255	
	(0.753)	(0.269)	(0.415)	
Observations	18400			

Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: GLLAMM specification I for students in the first ten rounds with just one visit. Results from the latent class logit model with four types: inequity averse, reciprocator, altruist, and self-regarding, the latter being the reference type. The model is estimated by maximum likelihood on the student sample on a dummy variable denoting which of the 8 possible extraction levels was chosen for each player and round. Class membership predictors are not available for this sample.

Table 15: Effect of incentives on the extraction level. This table reports the outcome of a linear regression of average extraction level (as a percentage of the maximum possible) over rounds 1-10, and over rounds 11-20, on the individual type probability, where type probability is derived with the classification given by specification III in Table 4.

	Fine		Subsidy		Non-monetary
	(1)	(2)	(3)	(4)	(5)
Instrument	-38.652**		-27.043***	145.952*	-33.052
	(18.91)		(9.64)	(75.34)	(31.13)
× Prob(inequity averse)	0.186			-2.073**	0.292
	(0.22)			(0.90)	(0.37)
× Prob(reciprocator)	0.525*		0.443	-1.168	-0.787
	(0.31)		(0.49)	(0.84)	(0.63)
× Prob(altruist)	0.169		0.154	-0.434	-0.656
	(0.32)		(0.73)	(0.76)	(0.56)
Fine amount		0.006			
		(0.15)			
× Prob(inequity averse)		-0.002			
		(0.00)			
× Prob(altruist)		0.001			
		(0.00)			
× Prob(reciprocator)		0.003			
		(0.00)			
Prob(inequity averse)	Yes	Yes	No	Yes	Yes
Prob(reciprocator)	Yes	Yes	Yes	Yes	Yes
Prob(altruist)	Yes	Yes	Yes	Yes	Yes
Obs	616	616	124	124	176
R squ	0.22	0.16	0.24	0.28	0.35

The variable Instrument takes a value of zero for controls in all rounds, as well as treated individuals during the first half of the 20 rounds; it takes a value of one for all treated individuals in the second half. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 $q \in Q = \{S, I, R, A\}$:

$$f(x_i; \Theta) = \sum_{q \in Q} p_q \prod_{t=1}^{T} \prod_{j=1}^{J} (f_q(x_i; \theta_q, \overline{x}_{-i}))^{d_{ijt}} \tag{12}$$

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

$$\ln L(\Psi; x) = \sum_{i=1}^{N} \ln f(x_i; \Psi) = \sum_{i=1}^{N} \ln \sum_{q \in Q} p_q f_q(x_i; \theta_q, \overline{x}_{-i}) \tag{13}$$

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

$$\ln L(\Psi; x) = \sum_{i=1}^{N} \ln f(x_i; \Psi) = \sum_{i=1}^{N} \ln \sum_{q \in Q} p_q \int_{-\infty}^{\infty} (f(x_i; \theta_q, \overline{x}_{-i})) \, \mathrm{d}F(\theta) \tag{14}$$

A.2 Latent class logit model

Our experiment provides us with data on each individual's extraction level per round. In our analysis we define our dependent variable as d_{njt} which equals 1 if agent n chose extraction level $j \in [1, .., 8]$ at round t and 0 otherwise. In this way we generate counterfactual extractions for every individual in every round.

A.3 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:

$$G(\Psi, \Psi^{(0)}) = \mathbb{E}_{\Psi^{(0)}}[\ln L^c(\Psi)|X = x] \tag{15}$$

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 kth EM iteration. Since $\ln L^c(\Psi)$ is linear in the unobservable υ_{iq}, it requires that $\mathbb{E}_{\Psi^{(k)}}(V_{iq}|X = x) = \tau_{iq}^{(k+1)}(x; \Psi^{(k)})$ [17], where V_{iq} is the random variable corresponding to υ_{iq} and[18]

$$\tau_{iq}^{(k+1)}(x; \Psi^{(k)}) = \frac{p_q^{(k)} f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})}{\sum_{q \in Q} p_q^{(k)} f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})} \tag{16}$$

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

$$G(\Psi, \Psi^{(k)}) = \sum_{i=1}^{N} \sum_{q \in Q} \tau_{iq}^{(k+1)}(x_i; \Psi^{(k)}, \overline{x}_{-i})[\ln p_q^{(k)} + \ln f_q(x_i; \theta_q^{(k)}, \overline{x}_{-i})] \tag{17}$$

A.4 The M-step

The M-step on the *(k + 1)*th iteration, the complete-data log likelihood function 17 is maximized with respect to $\Psi^{(k)}$ to provide the updated estimate $\Psi^{(k+1)}$.[19]

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

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

[19] 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; Cappelen, Konow, Sørensen, and Tungodden, 2013)

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):

$$\widehat{p_q}^{(k+1)} = \sum_{i=1}^{N} \frac{\tau_{iq}^{(k+1)}(x_i;\Psi^{(k)},\overline{x}_{-i})}{N} \tag{18}$$

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, Fehr-Duda, and Epper, 2010).

A.5 External validity: Further evidence of reciprocation in MS

MS also study whether the effect of the gift is moderated by the DM choosing on behalf of someone else, i.e., by him not bearing the consequences of its actions. To test for this, they compare the effect of gift giving in their Gift Treatment (GT) to the effect of the gift in the absence of third parties (No Externality Treatment/NET) where the DM is made full residual claimant of his decisions. MS find that even in this setting there is evidence for reciprocity concerns. Just as we did for the case of the GT, our algorithm also provides evidence of reciprocity in this setting, in agreement with MS.

In the NET, given the available strategies for each type, we are able to identify four types. Namely, a self-regarding type, an inequity averse type, a maximin type and a type motivated by reciprocity.

In this treatment, self-regarding DMs maximize their payoff by buying the product with the highest expected value. Inequity averse DMs minimize the distance between their own payoff and that of the producers. Due to the design of the game, disadvantageous inequality is only possible 2.5% of the time [20]. Given this small window for occurrence, we focus on identifying advantageous inequity aversion in this treatment. Altruist DMs care about the average payoff of the producers yet this is constant. Thus altruistic concerns do not generate a distinguishable strategy in this setting and we cannot identify them. We identify individuals playing maximin via those who favor the non gift giver if the gift is given and themselves otherwise, since in the latter case they cannot alter the payoff distribution among the producers. Finally, DMs motivated by reciprocity act as those described in the GT treatment and are thus identifiable in this treatment too. Table 16 column 2 presents the utility functions specified for each type in the NET.

Type	NET
SR	$U_i = \pi_i$
A	$U = \pi_i + \frac{\rho}{2}(\pi_{gg} + \pi_{ngg})$
IA	$U_i = \pi_i + \beta \max(\pi_i - \frac{1}{2}(\pi_{gg} + \pi_{ngg}), 0)$
MM	$U_i = \pi_i + min(\pi_i, \pi_{gg}, \pi_{ngg})$
R	$U_i = giftgiver * gift$

Table 16: Utility function specifications for each type in each treatment. The payoffs of the DM, the gift giving producer, the non gift giving producer and the client are denoted by $\pi_i, \pi_{gg}, \pi_{ngg}, \pi_C$, respectively.*gift* $= \{1, -1\} \equiv$ {gift given, gift not given};*giftgiver*$=$ $\{1, -1\} \equiv$ {potential gift giver chosen, potential gift giver not chosen}; thus, for reciprocity, $U_i = \{1, -1\} \equiv$ {reciprocation, no reciprocation}.

All in all, the maximum number of types that we can consider when estimating a model of heterogeneous preferences in the NET is 4. We can estimate SR, R, IA and MM types simultaneously. Thus, we estimate all the 11 possible models and we provide model fit and estimated shares in

[20] In only 16 out of 640 observations is the DM actually worse off than the producers.

Table 17 below.

Type	Desc	Classes	LL	Nparam	CAIC	BIC	P(R)	P(IA)	P(A)	P(SR)	P(MM)
IA+ R + MM	10	3	-115.775	14	261.730	253.730	0.618	0.314	0	0	0.068
SR + R	3	2	-121.479	5	258.049	254.049	0.521	0.479	0	0	0
SR+ IA + R	8	3	-117.441	11	261.290	254.290	0.556	0.066	0	0.378	0
R + MM	7	2	-120.549	7	259.962	254.962	0.723	0	0	0	0.277
SR + R + MM	11	3	-118.225	11	262.858	255.858	0.556	0	0	0.375	0.069
IA + R	5	2	-121.274	7	261.410	256.410	0.516	0.484	0	0	0
SR + IA + R + MM	1	4	-117.630	19	272.986	262.986	0.668	0.274	0	0.000	0.058
SR + MM	4	2	-135.814	5	286.718	282.718	0	0	0	0.092	0.908
IA + MM	6	2	-134.812	7	288.487	283.487	0	0.104	0	0	0.896
SR + IA	2	2	-136.643	5	288.376	284.376	0	0.365	0	0.635	0
SR+ IA + MM	9	3	-133.761	11	293.930	286.930	0	0.093	0	0.383	0.524

Table 17: We present the 11 possible models following the discussion on type identification in the NET. Type $i = \{SR, R, IA, MM\}$ stands for Self-Regarding, Reciprocator, Inequity Averse and Maximin, respectively. Columns 7-11 portray shares of type i. NA means that the share is not estimated.

Table 17 provides evidence regarding reciprocity in the NET condition. As in the GT, reciprocity explains a share of the observed behavior. Indeed, the best models in terms of fit include reciprocity as a type and estimate those concerns as representing a fairly stable share which varies in the interval 50-70%. Notice that the model that best describes the behavior in the NET coincides with the one that best describes behavior in the GT, a model of three types: inequity aversion, reciprocity and maximin preferences with a share of 31.4,61.8,6.8, respectively. MS find weaker effects of the gift in the NET versus the GT. We find that reciprocity, *relative* to other concerns, is higher in the NET versus the GT. These findings need not be contradictory as the shares are not directly comparable across treatments. In our model of heterogeneous preferences, we would need to keep constant the distribution of other possible concerns in order to evaluate whether reciprocity is higher in the GT than in the NET. It is clear that the setting does not allow us to do this as, to begin with, we cannot even identify the same types in both (recall, for example, that we cannot estimate SR in the GT or the confusion between A/IA in the GT or that we cannot estimate A in the NET). Thus, shares are not comparable across treatment but in both cases they provide evidence for reciprocity concerns driving a sizable share of behavior.