

# Logit estimation of conditional cooperation in a repeated public goods experiment\*

Luis G. González<sup>‡</sup>      Graciela González-Farías<sup>†</sup>

M. Vittoria Levati<sup>‡</sup>

February 2005

## Abstract

A conditional cooperator in a public goods game wants to match his partners' expected contribution. We investigate theoretically and empirically whether (and to what extent) conditional cooperation can explain how individual contributions evolve in a repeated two-person public goods experiment using a perfect strangers design. To identify a random utility model including non-pecuniary preferences we elicit participants' beliefs. Our econometric results show that the distribution of preferences in the population can be captured by a latent-class mixed logit specification with three subpopulations, and that 55% of participants can be regarded as conditional cooperators. Thus, the decline in average contribution levels may be attributed to the presence of conditional cooperators who have to revise their expectations about the others' behavior.

KEYWORDS: Conditional cooperation; Quantal response equilibrium; Latent class logit; Belief learning

## 1 Introduction

The robust experimental evidence of positive contributions in linear public goods game when zero-contribution is a dominant strategy<sup>1</sup> has motivated an enormous research effort to identify and explain patterns of cooperation. Theories of

---

\*We would like to thank participants at the 2004 meetings of the Mexican Statistical Association and the Mexican Mathematics Society, and seminar participants at the Universities of Mannheim and Tilburg for their comments and suggestions. Financial support through CICYT Refs. no. SEC 2001-08090 and SEJ 2004-04101-ECON is gratefully acknowledged. Torsten Weiland wrote the computer program used in the laboratory experiment.

<sup>‡</sup>Centro de Investigación en Matemáticas (CIMAT), Área de Probabilidad y Estadística.

<sup>†</sup>Max Planck Institute for Research into Economic Systems, Strategic Interaction Group.

<sup>1</sup>See Davis and Holt (1993) and Ledyard (1995) for extensive surveys on public goods experiments.

cooperative behavior in public goods games can be classified, broadly speaking, in two categories. The first one, which we call the *random utility* approach, focuses on the dynamics of learning and the stochastic nature of deviations from equilibrium predictions. Here, the influential idea of quantal response equilibria (McKelvey and Palfrey, 1995), combining the familiar tools of game theory with models of random choice, has been particularly successful in explaining the heterogeneity of observed individual decisions (see, e.g., Offerman and Sonnemans, 1998). The second category, which we refer to as the *neoclassical* approach, incorporates motives different from own monetary income like pure altruism, fairness, and efficiency concerns.

By combining the neoclassical and the random utility approaches we test whether intrinsic preferences for conditional and unconditional cooperation can account for observed behavior in a public goods experiment. Unconditional cooperation may include, for instance, altruism and efficiency concerns, whereas conditional cooperation is more closely related to reciprocity (what can be expressed by fairness and inequity aversion).<sup>2</sup> Thus we do not intend to disentangle among the various motivations underlying cooperation. However, since the basic game is repeated many times and beliefs typically adapt to experience, we do examine how the belief dynamics govern the choice dynamics. In a unified framework we study: first, the stochastic nature of individual behavior; second, how individuals differ in their preferences; and third, how people *learn* from previous experiences and adapt their behavior accordingly.

The rest of the paper is organized as follows. After introducing the basic game, Section 2 proposes a static random utility model that allows for nonmonetary concerns via a quadratic specification similar to Anderson et al. (1998). We derive different equilibrium behavioral patterns depending on the relative weights given to unconditional and conditional cooperation by individual decision makers. Section 3 extends the static game to the multi-period setting with belief adaptation, and discusses its solution based on both rational expectations (McKelvey and Palfrey, 1998) and a naive belief formation rule. Section 4 describes the experimental design and the collected data. Section 5 presents the estimation procedure used to account for heterogeneity of preferences. Section 6 concludes with a discussion of our findings.

---

<sup>2</sup>Fischbacher, Gächter and Fehr (2000) suggest that conditional cooperation can be considered as a motivation in its own or result from fairness preferences like inequity aversion or reciprocity. These “non-standard motivations” have received a lot of attention in the recent literature (e.g., Sugden, 1984; Andreoni, 1995; Palfrey and Prisbrey, 1997; Anderson, Goeree and Holt, 1998; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

## 2 The static model

### 2.1 Representation of preferences and beliefs

The basic game is the voluntary contribution mechanism as introduced by Isaac, Walker and Thomas (1984). Two individuals are endowed with an amount  $W$ , and must privately and simultaneously decide on how much to contribute to a public good. Let  $c_i$  denote player  $i$ 's contribution to the public good, with  $c_i \in [0, W]$ . The monetary payoff  $m_i$  of player  $i = 1, 2$  is given by

$$m_i \equiv m(c_i, c_{i'}) = W + (\mu - 1)c_i + \mu c_{i'}, \quad (1)$$

where  $\mu \in (\frac{1}{2}, 1)$  is the marginal per capita return from the public good, and  $c_{i'} \in [0, W]$  is the contribution of player  $i$ 's partner. Due to  $\frac{\partial m_{i,t}}{\partial c_{i,t}} = \mu - 1 < 0$ , if players are monetary payoff maximizers, the only dominant strategy is to contribute nothing.

We generalize Eq. (1), and suppose that the player  $i$ 's preferences about all possible outcomes<sup>3</sup> can be represented by the following utility function:

$$\begin{aligned} \bar{u}_i(c_i, c_{i'}) &= \alpha_i m(c_i, c_{i'}) - \theta_i (c_i - c_{i'})^2 + \gamma_i c_i \\ &= W + [\alpha_i(\mu - 1) + \gamma_i] c_i + \mu c_{i'} - \theta_i (c_i - c_{i'})^2. \end{aligned} \quad (2)$$

Here,  $\alpha_i \geq 0$  represents the weight that  $i$  gives to his own monetary income. Similarly,  $\theta_i \geq 0$  and  $\gamma_i \in \mathbb{R}$  capture  $i$ 's intrinsic preference for conditional and unconditional cooperation, respectively.<sup>4</sup> Utility function (2) allows for multiple, partially conflicting goals (including monetary payoffs, but also stylized notions of fairness and social attitudes). So, for example, player  $i$  gives a negative weight  $\theta_i$  to deviations from his partner's contribution. Similarly, the parameter  $\gamma_i$  captures the net weight that player  $i$  puts on unconditional cooperation. Since both players contribute simultaneously, in deciding about his contribution player  $i$  has to rely on his ex ante *beliefs* about his partner's contribution.

#### DEFINITION 1 (FIRST-ORDER BELIEFS ABOUT PARTNER'S CONTRIBUTION)

Player  $i$ 's first-order belief about his partner's contribution is a probability distribution function  $F_{i'}^{(i)}(c)$  with support  $[0, W]$  and expected value  $b_{i'}^{(i)}$ , specifying

<sup>3</sup>An outcome is a pair  $(c_i, c_{i'}) \in [0, W] \times [0, W]$ .

<sup>4</sup>In principle, the parameter  $\theta_i$  could assume negative values; in this case player  $i$  would have a preference for inequality of contributions.

the (subjective) probability that player  $i$  attributes to the event  $c_{i'} \leq c$ .

Defining  $s_{i'}^{(i)} \equiv \text{Var}_i c_{i'} = \int (c - b_{i'}^{(i)})^2 dF_{i'}^{(i)}(c)$ , and noticing that  $\text{E}_i c_{i'}^2 = (b_{i'}^{(i)})^2 + s_{i'}^{(i)}$ , we can write player  $i$ 's (subjective) expected utility as

$$\begin{aligned} \bar{u}_i^e(c_i) &= \text{E}_i \bar{u}_i(c_i, c_{i'}) \\ &= W + [\alpha_i(\mu - 1) + \gamma_i] c_i + \mu b_{i'}^{(i)} - \theta_i (c_i - b_{i'}^{(i)})^2 - \theta_i s_{i'}^{(i)} \\ &= \bar{u}_i(c_i, b_{i'}^{(i)}) - \theta_i s_{i'}^{(i)}. \end{aligned} \quad (3)$$

Expected utility (3) can be further extended to allow for a random term accounting for other idiosyncratic, not further specified aspects of individual preferences. In particular, we write

$$u_i(c_i) = \bar{u}_i^e(c_i) + \varepsilon_i(c_i), \quad (4)$$

where  $\varepsilon_i(\cdot)$  is a stationary max-stable process such that, for all  $c_i \in [0, W]$ ,  $\varepsilon_i(c_i)$  is *iid* Type I extreme value with  $\text{Var} \varepsilon_i(c_i) = \lambda^2$ . As pointed out by Resnick and Roy (1991), each  $\varepsilon_i(c_i)$  can be thought of as being generated by a collection of information signals regarding the desirability of  $c_i$ , which player  $i$  processes “in a boundedly rational fashion, recalling only the most ‘significant’ signal for each alternative” (p.273). Multiplying (4) by  $1/\lambda$ , collecting terms, and subtracting all terms that do not depend on  $c_i$ , we can reduce the random utility model to the following quadratic expression:

$$u_i^*(c_i) = \gamma_i^* c_i - \theta_i^* (c_i - b_{i'}^{(i)})^2 + \varepsilon_i^*(c_i), \quad (5)$$

where  $\gamma_i^* = [\alpha_i(\mu - 1) + \gamma_i]/\lambda$ ,  $\theta_i^* = \theta_i/\lambda$ , and  $\text{Var} \varepsilon_i^*(c_i) = 1$  for all  $c_i \in [0, W]$ .

Given the realization of  $\varepsilon_i^*(c_i)$ , the reduced form (5) implies that, if two different first-order beliefs  $F_{i'}^{(i)}$  and  $\check{F}_{i'}^{(i)}$  have the same expected value  $b_{i'}^{(i)}$ , both yield the same pairwise preference ordering on the set of alternatives  $[0, W]$ . Therefore, two such beliefs are said to be equivalent, and we will refer to  $b_{i'}^{(i)}$  and  $F_{i'}^{(i)}$  as synonymous.

## 2.2 Logit choice model with exogenous beliefs

Given player  $i$ 's belief about his partner's contribution,  $b_{i'}^{(i)}$ , and given the distributional assumptions regarding  $\varepsilon_i^*(c_i)$ , there is a.s. a unique value of  $c_i$  maximizing player  $i$ 's utility. In particular, from the viewpoint of an outside observer, the

probability density of player  $i$  choosing any specific contribution level  $c \in [0, W]$  (conditional on  $b_{i'}^{(i)}$ ) can be written as a continuous generalization of the logit choice model (Ben-Akiva, Litinas and Tsunokawa, 1985):

$$\begin{aligned} f(c; b_{i'}^{(i)}, \gamma_i^*, \theta_i^*) &= \frac{\exp\{\gamma_i^* c - \theta_i^* (c - b_{i'}^{(i)})^2\}}{\int \exp\{\gamma_i^* c - \theta_i^* (c - b_{i'}^{(i)})^2\} dc} \\ &\propto \exp\{\gamma_i^* c - \theta_i^* (c - b_{i'}^{(i)})^2\}. \end{aligned} \quad (6)$$

This model predicts different patterns of behavior, depending on the values assumed by the preference parameters  $\alpha_i$ ,  $\gamma_i$ ,  $\theta_i$ , and  $\lambda$ .<sup>5</sup> More specifically, if the individual has no preferences for conditional cooperation (i.e., if  $\theta_i = 0$ ), (6) reduces to

$$f^I(c; \gamma_i^*) \propto \exp\{\gamma_i^* c\}.$$

Therefore, player  $i$ 's behavior does not depend on what he expects his partner will contribute, but only on  $\gamma_i^*$ , i.e., on the ratio between his net preference for unconditional cooperation and the variance of the random process  $\varepsilon_i(\cdot)$ . In this case, the likelihood of any given contribution  $c \in [0, W]$  is discontinuous at  $\gamma_i = \alpha_i(1 - \mu)$ :

$$f^I(c; \gamma_i^*) = \begin{cases} \frac{-\gamma_i^* \exp\{\gamma_i^* c\}}{1 - \exp\{\gamma_i^* W\}}, & \text{if } \gamma_i < \alpha_i(1 - \mu) \\ 1/W & \text{if } \gamma_i = \alpha_i(1 - \mu) \\ \frac{\gamma_i^* \exp\{\gamma_i^* (W - c)\}}{1 - \exp\{-\gamma_i^* W\}}, & \text{if } \gamma_i > \alpha_i(1 - \mu). \end{cases} \quad (7)$$

The first case in equation (7) is a truncated exponential density function, which is monotonically decreasing on its support  $[0, W]$ . The second case,  $\gamma_i = \alpha_i(\mu - 1)$  and  $\theta_i = 0$ , implies that the decision maker is indifferent between all possible outcomes, except for differences in perceived utility attributed to  $\varepsilon(\cdot)$ . Thus, the logit model predicts uniformly distributed contributions on  $[0, W]$ . Finally, if the preference for unconditional cooperation is strong enough

<sup>5</sup>Although  $f$  is frequently interpreted as a mixed or behavioral strategy in game theoretic contexts (McKelvey and Palfrey, 1995; McKelvey and Palfrey, 1998), it is important to emphasize that player  $i$  does not randomize, but chooses his unique dominant strategy after privately observing the realization of  $\varepsilon_i(c_i)$ , similar to Harsanyi (1973)'s purification approach for mixed strategy equilibria. Nevertheless, for an outside observer, behavior seems to be random.

(i.e., if  $\gamma_i > \alpha_i(1 - \mu)$ ), the resulting pattern of contributions is described by a truncated negative exponential density with more probability weight on higher contributions. In all three cases, however, (as pointed out above) the ex-ante likelihood of contributing a certain  $c \in [0, W]$  does not depend on the beliefs about the partner's behavior.

In contrast, following Anderson et al. (1998), if the decision maker has intrinsic preferences for conditional cooperation (i.e., if  $\theta > 0$ ), one can reparameterize the model by completing the square in equation (6) so as to obtain the truncated normal density function

$$f_i^{II}(c; b_{i'}^{(i)} + \eta_i, \sigma_i^2) \propto \exp\left\{-\frac{1}{2\sigma_i^2} \left(c - b_{i'}^{(i)} - \eta_i\right)^2\right\}, \quad (8)$$

whereby  $\sigma_i^2 \equiv \lambda/2\theta_i = 1/2\theta_i^*$ , and  $\eta_i \equiv [\alpha_i(\mu - 1) + \gamma_i]/2\theta_i = \gamma_i^*/2\theta_i^*$ . Here, a change in beliefs affects the distribution of  $i$ 's contributions: the more  $i$  believes that his partner will contribute, the higher the expected value of his own contribution.

Equations (7) and (8) make clear that the reduced random utility model (5) can account for a variety of individual behavioral patterns depending on the interplay of the two reduced form parameters,  $\theta_i^*$  and  $\gamma_i^*$ . If, for instance,  $\theta_i^* = 0$ , the contribution pattern of a decision maker can be classified as purely opportunistic ( $\gamma_i^* < 0$ ), purely altruistic ( $\gamma_i^* > 0$ ), or purely random ( $\gamma_i^* = 0$ ). On the other hand, if  $\theta_i^* > 0$ , the individual exhibits a propensity to conditionally cooperate, although with a relative bias towards opportunism (altruism) when  $\gamma_i^*$  is strictly negative (positive). These cases are illustrated in Figure 1.

### 2.3 Rationalizable expectations

So far we assumed exogenous individual beliefs  $F_{i'}^{(i)}$ . This reduces player  $i$ 's decision problem to a well defined (one-person) optimization task. However, common knowledge of rationality requires two interacting players to anticipate each other's likely behavior. For the analysis of a two-person, one-shot public goods game, the following definition will be useful:

**DEFINITION 2 (FIRST-ORDER PRIMITIVE BELIEF)**

Player  $i$ 's first-order primitive belief about his partner's preference parameters  $(\gamma_{i'}^*, \theta_{i'}^*)$  is a probability distribution function  $G_{i'}^{(i)}(\phi)$ , with support  $\mathbb{R}^2$ , giving the (subjective) probability that player  $i$  attributes to the event  $(\gamma_{i'}^* \leq \phi_1, \theta_{i'}^* \leq \phi_2)$ .

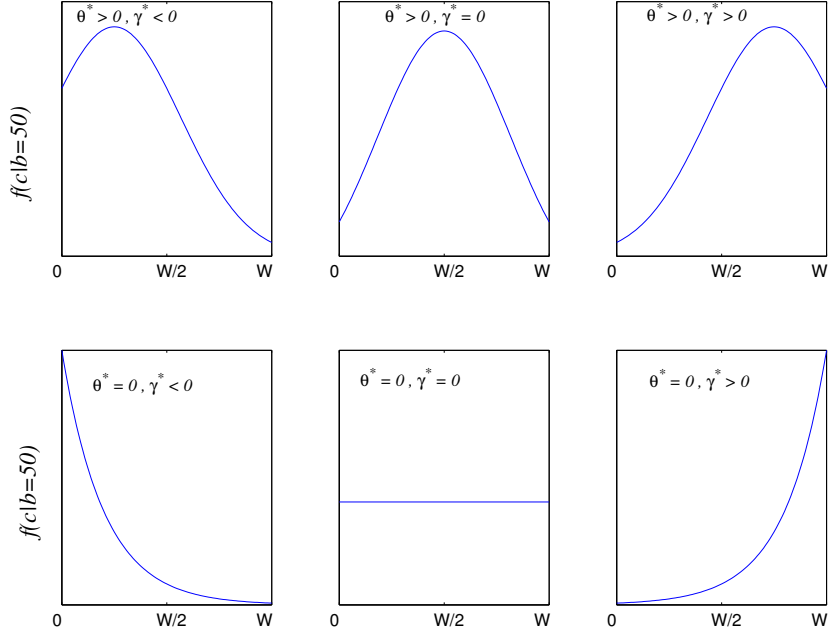


Figure 1: Behavioral strategies generated by a logit choice function under different utility parameter values  $\theta_i^*$  and  $\gamma_i^*$  (assuming fixed beliefs  $b_{i'}^{(i)} = W/2$ )

We consider primitive beliefs to be exogenously determined, e.g., via some non-specified, pre-game processes. However, common (knowledge of) rationality requires beliefs to be *rationalizable* (Bernheim, 1984; Pearce, 1984) in the following sense:<sup>6</sup>

**DEFINITION 3 (RATIONALIZABLE BELIEFS ABOUT PARTNER'S CONTRIBUTION)**

Player  $i$ 's belief about his partner's contribution in a one-shot, two-person public goods game,  $b_{i'}^{(i)}$ , is rationalizable if there is an infinite sequence of higher order

<sup>6</sup>Notice that the implicit assumption that the distribution of  $\varepsilon_i^*(c_i)$  is common knowledge already implies that  $\frac{d}{dc} F_{i'}^{(i)}(c) = \int f(c; b_i^{(i')}(i), \phi) dG_{i'}^{(i)}(\phi)$ , i.e., player  $i$ 's (subjective) belief must assume the functional form of a mixed-logit probability distribution of  $c$ , with mixing distribution  $G_{i'}^{(i)}(\phi)$  and appropriate choice of second-order belief  $b_i^{(i')}(i)$ .

beliefs about outcomes,  $\{b_i^{(i')(i)}, b_{i'}^{(i)(i')}, \dots\}$ , such that:

$$\begin{aligned}
b_{i'}^{(i)} &\equiv \mathbf{E}_i c_{i'} &= \int \int c f(c; b_i^{(i')(i)}, \phi) dG_{i'}^{(i)}(\phi) dc, \\
b_i^{(i')(i)} &\equiv \mathbf{E}_i \mathbf{E}_{i'} c_i &= \int \int c f(c; b_{i'}^{(i)(i')}, \phi) dG_i^{(i')(i)}(\phi) dc, \\
b_{i'}^{(i)(i')(i)} &\equiv \mathbf{E}_i \mathbf{E}_{i'} \mathbf{E}_i c_{i'} &= \int \int c f(c; b_i^{(i')(i)(i')}, \phi) dG_{i'}^{(i)(i')(i)}(\phi) dc, \\
&&\vdots
\end{aligned}$$

Evidently, if it is common knowledge that (after choosing appropriate values of  $\gamma^*$  and  $\theta^*$ ) preferences can be represented by a utility function like (5), then any first-order belief  $b_{i'}^{(i)} \in (0, W)$  can be rationalized, regardless of the fixed sequence of primitive beliefs  $\{G_{i'}^{(i)}, G_i^{(i')(i)}, G_{i'}^{(i)(i')}, \dots\}$  that one assumes.

The fact that beliefs can be “almost anything” without contradicting players’ rationality creates a serious identification problem. Thus behavioral data does not suffice when testing “stochastically rational” behavior (McFadden, 1999). This is true even if one assumes (for the sake of tractability) that the *distribution* of preference parameters within the population of potential partners is common knowledge, yielding homogeneous primitive beliefs for all players of the form  $G = G_i^{(i)} = G_{i'}^{(i')(i)} = \dots$ .

However, it is much more important for practical purposes<sup>7</sup> that one cannot statistically infer the relative importance of conditional cooperation, at least not without assuming further structure on the data generating process (i.e., on the beliefs that players rely on). We confront here a particular case of the general identification problem discussed by Manski (2004), forcing the analyst to restrict the feasible set of beliefs (or to elicit them directly).

## 2.4 Rational expectations

Restricting beliefs usually relies on so-called rational expectations equilibria. A rational expectations equilibrium for a two-person, one-shot public goods game with a random utility specification yields a logit quantal-response equilibrium (McKelvey and Palfrey, 1995). This benchmark solution allows the econometrician to pin down the parameters of the underlying utility function by *postulating* that primitive beliefs and beliefs about contributions are correct, and that this

<sup>7</sup>Arguably, it makes little sense to build a random utility model in order to test rationality in the first place.



is common knowledge. Specifically, one first requires primitive beliefs to exactly match the average preferences in the population of possible partners (i.e., each player  $i$  knows the preferences of the “representative” player). Letting  $G$  be the commonly known bi-variate distribution of possible types  $(\gamma^*, \theta^*)$  in the population (with support  $\mathbb{R}^2$ ), one can then write

$$\frac{d}{dc} F_{i'}^{(i)}(c) = \int f(c; b_i^{(i')}, \phi) dG(\phi) \quad \text{for all } i.$$

Common knowledge of  $G$  means that all primitive beliefs of higher order are equal to  $G$ .

The second requirement of a logit quantal-response equilibrium is that player  $i$ 's subjective expectation about his partner's contribution,  $b_{i'}^{(i)}$ , exactly corresponds to the actual average behavior. This implies that first-order beliefs about outcomes are homogeneous, i.e.,  $b_{i'}^{(i)} = b_i^{(i')}$ , and therefore equilibrium beliefs are a fixed point  $b^*$  in the sense of

$$b^* = \mathbb{E}[c|b^*] = \int \int c f(c; b^*, \phi) dG(\phi) dc. \quad (9)$$

The rational expectations or logit quantal-response equilibrium closes the model because, by assuming a functional form  $G(\phi|\beta)$ , it is *in principle* possible to estimate, using only choice data, the set of parameters  $\beta$  that describe the distribution of preferences across individuals in the population. This estimation problem is non-trivial as the functional form of the likelihood function is highly non-linear due to the recursive inclusion of  $b^*$  on the right hand side of

$$f(c|\beta) = \int f(c; b^*, \phi) dG(\phi|\beta).$$

Once the estimates of  $\beta$  are available, one can recover via maximum likelihood estimates the preference parameters  $\gamma^*$  and  $\theta^*$  for each individual, which is a usual goal of econometric analysis.

### 3 The dynamic model

We now examine expected patterns of behavior in a dynamic setting, where interaction with different partners allows a player to update his beliefs about the distribution of preferences and the likelihood of outcomes.

### 3.1 The extensive game in agent form

The logit choice model of the supergame results from repeating the public goods stage-game over several periods, using a perfect-stranger matching protocol. We adopt a sequential-rationality approach to represent the extensive game. In particular, we model the different moves of the same player as decisions on different “agents”, whose ex-post payoff functions are identical, and restrict attention to agents who are actually called on to move.

Let  $c_i = (c_{i,1}, \dots, c_{i,T})$  be now the vector of  $i$ 's contribution decisions in periods  $t = 1, \dots, T$ . Similarly, denote by  $c_{i'} = (c_{i',1}, \dots, c_{i',T})$  the vector of contribution decisions by the  $T$  different players  $i$  is matched with during the game,  $i'_1, \dots, i'_T$ .<sup>8</sup> All agents of player  $i$ , indexed by  $(i, t)$ , have the same *ex-post* payoff function, namely,

$$\bar{u}_{i,t}(c_i, c_{i'}) = \sum_{\tau=1}^T [\alpha_i m(c_{i,\tau}, c_{i',\tau}) - \theta_i (c_{i,\tau} - c_{i',\tau})^2 + \gamma_i c_{i,\tau}]. \quad (10)$$

Players are assumed to have perfect recall. This means that in each period  $t = 1, \dots, T$  players are fully aware of the previous interactions in which they participated. In other words, all decisions made before  $t$  by player  $i$ , and by each of  $i$ 's previous partners *when interacting with him* are known to agent  $(i, t)$ .<sup>9</sup> For each  $t = 1, \dots, T$ , such information is summarized in the history vector  $h_{i,t} = ((c_{i,1}, c_{i'_1,1}), \dots, (c_{i,t-1}, c_{i'_{t-1},t-1}))$ , where  $h_{i,1}$  is the null history.

In the multi-period public goods game, each player makes decisions before observing the contributions of his current or future partners. Therefore, agent  $(i, t)$  optimizes his choice of  $c_{i,t}$  by taking into account only his *current beliefs* about the contributions of his current and future partners, denoted by  $b_{i'}^{(i,t)} = (b_{i',t}^{(i,t)}, \dots, b_{i',T}^{(i,t)})$ , and the optimal contributions of his own future agents, denoted by  $b_i^{(i,t)} = (b_{i,t+1}^{(i,t)}, \dots, b_{i,T}^{(i,t)})$ .

The sequential structure of the extensive game implies that each agent of the same player has different information on which he can base his contribution decision (different histories imply different information sets). Thus, expectations about what other players will do as well as about one's own optimal future choices may change in response to *both* the observed history of play *and* the course of action chosen for the current period. Let the set of  $i$ 's possible histories

<sup>8</sup>We abuse notation by writing  $c_{i',t}$  instead of  $c_{i'_t,t}$ .

<sup>9</sup>Of course, player  $i$  does not know the decisions made by other players in interactions in which he was not involved.

in  $t$  be  $H_{i,t} = \times_{\tau=1}^{t-1} [0, W]^2$ . Similarly, define the sets of possible beliefs regarding future choices as  $B_i^{(i,t)} = \times_{\tau=t+1}^T [0, W]$  and  $B_{i'}^{(i,t)} = \times_{\tau=t}^T [0, W]$ . In general, we postulate the existence of a subjective belief-learning function  $b^{(i)}$  that maps each possible history-action pair  $(h_{i,t}, c_{i,t})$  into a belief profile  $(b_i^{(i,t)}, b_{i'}^{(i,t)})$ ,

$$b^{(i)} : H_t \times [0, W] \rightarrow B_i^{(i,t)} \times B_{i'}^{(i,t)}.$$

Hence, we can express each agent's (subjective) expected utility from choosing  $c \in [0, W]$  as the sum of payoffs already obtained by previous agents, the expected payoff of the own decision, and the expected payoff of future agents' decisions:

$$\begin{aligned} \bar{u}_{i,t}^e(c | h_{i,t}, b^{(i)}) &= \sum_{\tau=1}^{t-1} \left[ \alpha_i \cdot m(c_{i,\tau}, c_{i',\tau}) - \theta_i \cdot (c_{i,\tau} - c_{i',\tau})^2 + \gamma_i \cdot c_{i,\tau} \right] \\ &+ \left[ \alpha_i \cdot m(c, b_{i',t}^{(i,t)}) - \theta_i \cdot (c - b_{i',t}^{(i,t)})^2 + \gamma_i \cdot c \right] \\ &+ \sum_{\tau=t+1}^T \left[ \alpha_i \cdot m(b_{i,\tau}^{(i,t)}(c), b_{i',\tau}^{(i,t)}(c)) - \theta_i \cdot (b_{i,\tau}^{(i,t)}(c) - b_{i',\tau}^{(i,t)}(c))^2 + \gamma_i \cdot b_{i,\tau}^{(i,t)}(c) \right], \end{aligned} \quad (11)$$

where, for notational simplicity, we write  $b_{j,\tau}^{(i,t)}(c)$  instead of  $b_{j,\tau}^{(i,t)}(h_{i,t}, c)$ ,  $j = i, i'$ , and  $b_{i',t}^{(i,t)}$  instead of  $b_{i',t}^{(i,t)}(h_{i,t}, \cdot)$ .

Observing that, at time  $t$ , the terms included in the first sum on the right hand side of (11) are fixed, and since the representation of preferences in a utility function is invariant to the addition of a constant, it is possible to reexpress (11) as

$$\begin{aligned} \bar{u}_{i,t}^e(c | h_{i,t}, b^{(i)}) &= [\gamma_i - \alpha_i(1 - \mu)]c - \theta_i (c - b_{i',t}^{(i,t)})^2 \\ &+ \sum_{\tau=t+1}^T \left[ (\gamma_i - \alpha_i(1 - \mu))b_{i,\tau}^{(i,t)}(c) + \alpha_i \mu b_{i',\tau}^{(i,t)}(c) - \theta_i (b_{i,\tau}^{(i,t)}(c) - b_{i',\tau}^{(i,t)}(c))^2 \right]. \end{aligned} \quad (12)$$

Finally, if we allow for a random error term in the objective function of each agent, we obtain the random utility model in agent form,

$$u_{i,t}(c | h_{i,t}, b^{(i)}) = \bar{u}_{i,t}^e(c | h_{i,t}, b^{(i)}) + \varepsilon_{i,t}(c), \quad (13)$$

where for all  $c \in [0, W]$ ,  $\varepsilon_{i,t}(c)$  is a Type I Extreme Value random variable with  $\text{Var} \varepsilon_{i,t}(c) = \lambda^2$ .

### 3.2 Logit choice model with exogenous beliefs

Assume, as in the static case, that the subjective belief-learning rule  $b^{(i)}$  is fixed. Then, setting  $\alpha_i^* = \alpha_i \mu / \lambda$ , the logit choice function for agent  $(i, t)$  is given by

$$f_{i,t}(c) \propto \exp \left\{ \gamma_i^* c - \theta_i^* \left( c - b_{i',t}^{(i,t)} \right)^2 \right\} \cdot R_{i,t}(c), \quad (14)$$

where the factor

$$R_{i,t}(c) \equiv \prod_{\tau=t+1}^T \exp \left\{ \gamma_i^* b_{i,\tau}^{(i,t)}(c) + \alpha_i^* b_{i',\tau}^{(i,t)}(c) - \theta_i^* \left( b_{i,\tau}^{(i,t)}(c) - b_{i',\tau}^{(i,t)}(c) \right)^2 \right\} \quad (15)$$

captures the “shadow of the future” that results from player  $i$  having a specific belief-learning function  $b^{(i)}$ . Notice that, since we consider a perfect-strangers design,  $R_{i,t}(c)$  cannot stand for *individual* strategic reputation building (in the sense of trying to directly influence future partners’ beliefs and behavior by adjusting one’s own current behavior). Instead, it captures how general beliefs regarding future behavior of a randomly chosen member of the population can be influenced by contributing higher amounts, e.g., by promoting a “cooperative atmosphere” within the population. Hence, we refer to  $R_{i,t}(c)$  as *social reputation* concerns.

Applying a similar argument to higher order beliefs as in the static case, one can see from equations (14) and (15) that almost any pattern of behavior is rationalizable. On the other hand, obtaining a rational expectations equilibrium would require finding a belief-learning rule that is consistent with actual behavior, which seems to be at least an impractical task. In what follows, therefore, we propose a simplified social reputation model that, together with direct elicitation of first-order beliefs  $b_{i',t}^{(i,t)}$ , will allow us to identify the preference parameters  $\gamma_i^*$  and  $\theta_i^*$ . In particular, rather than deriving a rational expectations equilibrium for the social reputation function,  $R_{i,t}(c)$ , we develop a simplified parametric form which can be used to estimate the distribution of  $\gamma^*$  and  $\theta^*$  in the population via maximum-likelihood.

### 3.3 Social reputation concerns

Just as in the one-shot game, if  $\theta_i = 0$  the optimal decision of agent  $(i, T)$  is independent of his beliefs. However, in contrast to the static case, social reputation concerns can induce higher initial cooperation by an opportunistic agent

as compared to what would be predicted using choice function (7). This means that, even in the absence of intrinsic motivations to conditionally cooperate, the hope of indirectly encouraging high cooperation levels by other members of the population (who may be conditional cooperators themselves) may affect the optimal choices of player  $i$ 's "younger agents". Only if it is commonly known that  $\theta_i = 0$  for all players, backwards induction yields a stationary behavioral pattern of the form (7) for all players.

Although in principle there are many belief-learning functions  $b^{(i)}$  from which player  $i$  may anticipate the effect of his current decision on the future evolution of play, some minimal consistency requirements seem to be obvious. First of all, recall that for any period  $t = 1, \dots, T$ , agents  $(i, t)$  and  $(i', t)$  contribute simultaneously. This implies that, whatever agent  $(i, t)$  believes about his partner's choice, this belief is independent of his own choice. For this reason we can write  $b_{i',t}^{(i,t)}(c) = b_{i',t}^{(i,t)}$  in equation (11). In the last period, therefore, player  $i$ 's behavior can be described by the static logit equilibrium (6), namely,

$$f_{i,T}(c) \propto \exp \left\{ \gamma_i^* c - \theta_i^* \left( c - b_{i',T}^{(i,T)} \right)^2 \right\}. \quad (16)$$

Consider now player  $i$ 's beliefs about how his contribution in  $t$  will affect his partner's behavior in  $t + 1$ . Under perfect-strangers matching, the choice made by  $i$  at time  $t$  remains unknown to the person he is matched with in the next period. Thus,  $i$ 's choice in  $t$  cannot affect the level of cooperation of agent  $(i', t + 1)$ , and  $b_{i',t+1}^{(i,t)}(c) = b_{i',t+1}^{(i,t)}$  for all  $t = 1, \dots, T - 1$ . Moreover, since  $i$ 's expectation about the behavior of the person he is matched with in the next period is not dependent on  $i$ 's own current action, the same lack of dependency holds for  $i$ 's optimal action in the next period. Thus, we can write  $b_{i',t+1}^{(i,t)}(c) = b_{i',t+1}^{(i,t)}$ . As a result, it is straightforward to see that social reputation concerns in the last two periods are not consistent with backward-induction reasoning, and we can write the logit choice function for period  $T - 1$  as

$$f_{i,T-1}(c) \propto \exp \left\{ \gamma_i^* c - \theta_i^* \left( c - b_{i',T-1}^{(i,T-1)} \right)^2 \right\}. \quad (17)$$

The above logic, however, does not extend to earlier periods. The reason is that a player might well expect his current behavior to trigger some contagion in the population after 2 or more periods. Put differently, initial contribution levels might be reasonably expected to establish a "cooperative atmosphere" which,

in turn, can affect future partners' behavior. So, for instance, player  $i$  may think that his contribution in period  $t$  will induce a higher level of cooperation from agent  $(i'_t, t + 1)$ , which in turn may trigger higher cooperation from his respective partner in periods  $t + 2, t + 3$ , etc. Hence, it may be reasonable for player  $i$  to assume a causal relationship between his choice at time  $t$  and the expected contribution of his partner in periods  $t + 2, t + 3$ , etc.

More specifically, if player  $i$  thinks that there are at least some conditional cooperators in the population, he will reasonably expect his choice of  $c$  at time  $t$  to have some positive impact on the behavior of his future partners after *a couple of periods*. We thus assume that each player  $i$  has beliefs such that

$$\frac{\partial b_{i',t+\tau}^{(i,t)}(c)}{\partial c} > 0 \quad \text{for } \tau = 2, \dots, T - t,$$

irrespective of whether he has opportunistic or conditionally cooperative preferences himself.

Now, if player  $i$  is a conditional cooperator, then we would also expect

$$\frac{\partial b_{i,t+\tau}^{(i,t)}(c)}{\partial c} > 0 \quad \text{for } \tau = 2, \dots, T - t,$$

since a conditional cooperator can anticipate his own response to his partner's cooperation, which was (indirectly) induced by himself via early contributions. In contrast, if player  $i$  has only opportunistic preferences, this derivative should be near zero, especially for larger values of  $\tau$ .

Following the above considerations, we assume that, for  $\tau = 2, \dots, T - t$ ,

$$\left[ \frac{\partial b_{i,t+\tau}^{(i,t)}(c)}{\partial c} - \frac{\partial b_{i',t+\tau}^{(i,t)}(c)}{\partial c} \right] \approx 0, \text{ if } \theta_i > 0, \quad (18)$$

and  $< 0$  in case  $\theta_i = 0$ .

From (18), it is possible to argue that neglecting the quadratic term in (15) is not too problematic, since this term tends to be small even if player  $i$  is a conditional cooperator himself. The intuition behind this assertion is that a conditional cooperator will anyway try to match the expectations he has about his future partners' behavior, and therefore does not need to take this into account when choosing early levels of cooperation.

Thus, to model  $R_{i,t}(c)$ , we only take into account the expected cost of creating a cooperative environment, represented by the terms  $\{\gamma_i^* b_{i,t+\tau}^{(i,t)}(c)\}$ , and the

benefit derived from it in the form of higher contributions by future partners, represented by the terms  $\{\alpha_i^* b_{i',t+\tau}^{(i,t)}(c)\}$ . As discussed before, the cost effect should be smaller for opportunistic players (since they do not reciprocate in later periods, while conditional cooperators do), whereas the benefit effect is similar for both types of preferences. Therefore, early social reputation building should be much stronger in the case of opportunistic players. Put differently, conditional cooperators should not engage in cooperation *in order to establish a good social reputation*, but mainly because of their intrinsic motivation to “reciprocate”. In this sense, early cooperation from opportunistic players should tend to vanish because the shadow of the future shrinks. On the other hand, early cooperation from conditional cooperators should remain relatively constant *if* their beliefs about what others will do remain constant.

In order to keep the estimation problem tractable, we can assume that  $R_{i,t}$  takes on a very simple parametric form, only dependent on the number of periods left to be played, such as

$$R_{i,t}(c) = \exp \left\{ \frac{\rho_i}{t} c \right\}, \quad t = 1, \dots, T, \quad (19)$$

where  $\rho_i \geq 0$ . While such a social reputation function has the desirable property of vanishing as the game proceeds, it does not allow  $\rho_i$  to be smaller the stronger the preference for conditional cooperation. However, nothing changes as long as  $R_{i,t}(c)$  is assumed to be a function separable in  $\theta_i$ , such as

$$R_{i,t}(c) = \exp \left\{ \frac{\rho_i}{t} c - g(\theta_i^*) \right\}, \quad t = 1, \dots, T,$$

which is equivalent to multiplying (19) by a constant.

Finally, we reached a manageable specification of the dynamic logic choice function from which we can estimate the preference parameters of interest,  $\gamma_i^*$  and  $\theta_i^*$ :

$$f_{i,t}(c) \propto \exp \left\{ \left( \gamma_i^* + \frac{\rho_i}{t} \right) c - \theta_i^* \left( c - b_{i',t}^{(i,t)} \right)^2 \right\}.$$

## 4 The experiment

In order to obtain empirical evidence about the relative importance of conditional and unconditional cooperation, we implemented the public goods game in the laboratory of the Max Planck Institute in Jena (Germany). The way in which this experiment was conducted, as well as a description of the experimen-

tal data observed, is explained below.

#### 4.1 Experimental design and procedures

Participants were undergraduate students from different disciplines at the University of Jena, and the experiment was computerized using the z-Tree software (Fischbacher, 1999). The experimental procedures were as follows. After being seated at a computer terminal, participants received written instructions. A control questionnaire was used in order to assure understanding of the rules before the experiment started.<sup>10</sup> Overall, we ran 3 sessions with a total of 72 subjects (24 subjects per session). Each session took about 60 minutes.

In all experimental sessions, subjects played 21 rounds of the standard linear two-person public good game described in the previous section. The pecuniary payoff function explained to the subjects was (1) with  $\mu = 1.4$  and  $W = 100$  tokens (where 100 tokens = 0.40 Euro).

We used a perfect-stranger matching scheme, meaning that subjects never met the same partner more than once. Moreover, subjects were never informed (neither during nor after the experiment) about the identity of their partners; at the end of each period they were only told their own period-earnings and the contribution made by their current partner to the public good.

In all sessions we elicited subjects' expectations about the current partner's contribution. More specifically, besides choosing their contribution level, participants were asked to predict how many tokens their current partner would have contributed, and to state on a scale from 0% to 100% how sure they were about this prediction. To encourage subjects to report their beliefs truthfully and seriously, we paid an additional bonus of 15 Euro to the participant with the most accurate predictions in each session. The general form of the rule used to calculate subject  $i$ 's prediction accuracy was

$$\text{Score}_i = - \sum_{t=0}^{20} (\hat{b}_{i',t}^{(i)} - c_{i',t})^2,$$

where  $\hat{b}_{i',t}^{(i)}$  denotes the beliefs stated by  $i$  in period  $t$ . Although participants were not informed about the exact content of this rule, it was explained to them that the closer their predictions were to the actual contributions of partners, the higher their chances of receiving the bonus.

---

<sup>10</sup>English translations of instructions and control questionnaire are included in the Appendix.



We ran one additional experimental session in which expectations were not elicited and participants had to supply only a contribution decision. Such a “control” session allowed us to check whether the mere act of eliciting beliefs changes individual behavior and changes the way in which the models of belief learning under consideration fit the data.

## 4.2 Exploratory data analysis

Figure 2 shows the evolution of average contributions throughout the experiment. Our results are in line with most experimental evidence from public good games with perfect stranger design (see, e.g., Andreoni, 1995; Keser and van Winden, 2000): although average contributions in initial periods are high, they show a clear downward trend.

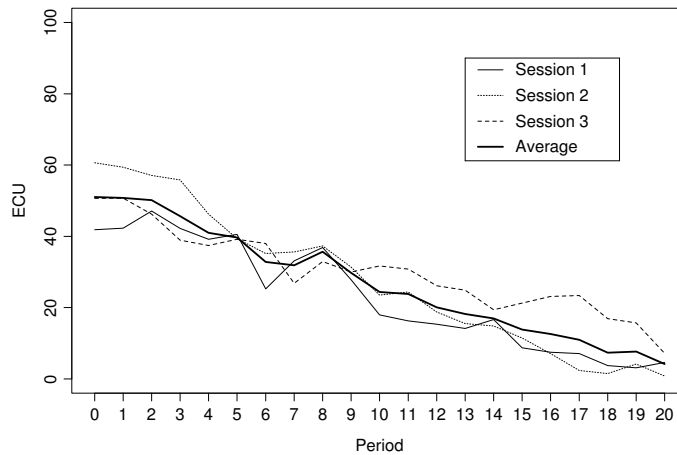


Figure 2: Average contributions by experimental session

The average deviation (by session and by period) between actual contribution levels and stated beliefs appears in Figure 3. This deviation fluctuates around zero, which is in line with the hypothesis of conditional cooperation. Table 1 also hints toward the existence of conditional cooperators. It shows the result of fitting a logit regression where the dependent variable is the proportion of endowment contributed in each period,  $c_{i,t}/W$ . Both the estimated deviance and the Akaike Information Criterion reveal a significant contribution of stated

	Model 1	Model 2
(Intercept)	0.3635*	-1.7195**
	(-0.1369)	(0.2063)
Time Period	-0.1369*	-0.0556**
	(0.0111)	(0.0133)
Stated Belief		0.0381**
		(0.0030)
Deviance	752.12	563.00
AIC	1394.16	1155.86

Dependent variable is the proportion of endowment contributed in each period. Standard deviations appear in parenthesis.

Table 1: Exploratory Logit Regression.

beliefs,  $b_{i,t}^{(i)}$ , in explaining variation in contribution levels.

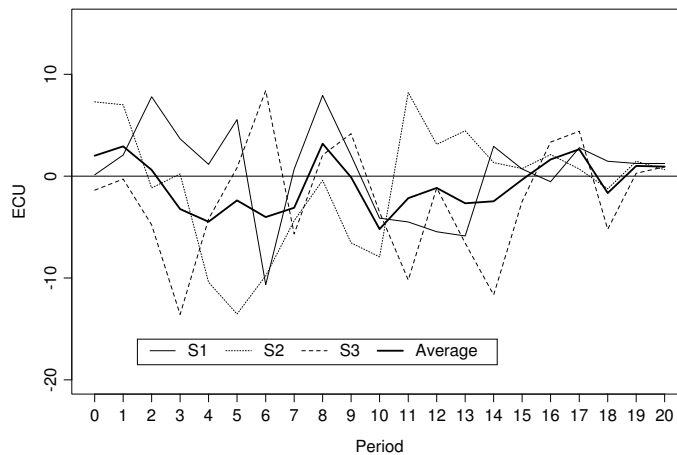


Figure 3: Average deviations of contributions with respect to stated beliefs, by session

Nevertheless, one must refrain from drawing conclusions by looking only at aggregate behavior, since heterogeneity in individual preferences may hide more subtle structural features in the data. Indeed, as Table 2 shows, allowing for random effects in the logit regression exhibits large variation between subjects.

The purpose of the next section is to investigate in more detail what factors are responsible for the apparent downward trend in contributions, and whether participants can be classified as either conditional cooperators or opportunists.

	Model 1		Model 2	
	<i>Fixed effects</i>	<i>Random effects</i>	<i>Fixed effects</i>	<i>Random effects</i>
(Intercept)	0.5125* (0.2574)	(2.0763)	-1.9006** (0.2589)	(1.8504)
Time Period	-0.1899** (0.0185)	(0.1412)	-0.0871** (0.0123)	(0.0756)
Stated Beliefs			0.0437** (0.0437)	(0.0234)
AIC	6087.698		5972.282	
log Likelihood	-3037.849		-2976.141	

Dependent variable is the proportion of endowment contributed in each period. Standard deviations appear in parenthesis.

Table 2: Linear mixed-effects logit regression

In order to do this, we implement maximum-likelihood estimation of models  $f^I$  (truncated exponential) and  $f^{II}$  (truncated normal) outlined in the previous section, allowing for heterogeneous social-reputation concerns.

## 5 Estimation of conditional cooperation

In view of the exploratory results from the previous section, it seems necessary to account for heterogeneity in preferences among the population. As outlined in Section 3, it is precisely the interaction between conditional cooperators and opportunistic players that induces a decrease in initially high cooperation levels as the game approaches its final period.

To handle heterogeneity we propose a latent-class maximum-likelihood estimation procedure, based on the idea that there exist three types of players, namely: “conditional cooperators”, “opportunistic players”, and “others.”<sup>11</sup> In particular, we assume that in addition to each individual’s observed data (which consists of a vector of contributions,  $c_i$ , and a vector of beliefs,  $b_i$ ), there is a non observable, latent-class vector  $Z_i = (Z_i^1, Z_i^2, Z_i^3)$ , whose elements are dummy variables indicating the membership of individual  $i$  to the type  $k = 1, 2$ , and  $3$ , respectively.<sup>12</sup>

If one could observe the value of  $z_i$ , it would be possible to write the log-likelihood function of each type of player’s preference (and social-reputation)

<sup>11</sup>A different approach would be to rely on a mixed-effects specification.

<sup>12</sup>Of course,  $Z_i^1 + Z_i^2 + Z_i^3 = 1$ .

parameters  $\phi_i^k = (\gamma_i^*, \theta^*, \rho^*)^k$  as

$$\mathcal{L} = \sum_{i=1}^N \sum_{k=1}^3 z_i^k \left[ \ln \pi^k + \ln \left( \prod_{t=1}^T f(\phi_i^k | c_{i,t}, b_{i,t}) \right) \right], \quad (20)$$

where  $\pi^k$  is the proportion of type- $k$  players in the population. Notice that since we are not being able to observe  $Z_i$ , one cannot estimate the parameters of (21) directly. However, as proposed by Dempster, Laird and Rubin (n.d.), it is possible to rely on the EM algorithm which iteratively estimates the *complete-data* likelihood function,

$$\mathcal{Q} = \sum_{i=1}^N \sum_{k=1}^3 \tau_i^k \left[ \ln \pi^k + \ln \left( \prod_{t=1}^T f(\phi_i^k | c_{i,t}, b_{i,t}) \right) \right], \quad (21)$$

where  $\tau_i^k = \mathbb{E}Z_i^k$  is the expected probability of player  $i$  belonging to class  $k$  (for a good introduction on the EM algorithm see, e.g., McLachlan and Peel (2000) and McLachlan and Krishnan (2005)).

The maximum likelihood estimation (20) of using the EM algorithm is very easy to implement because the logit specification allows us to express each component of the resulting mixture model as either a truncated normal or a truncated exponential density, whose integrals are implemented in almost any standard statistical software.

Table 3 shows the parameters estimates obtained from the data from the three experimental sessions. As can be seen from Figures 4 and 5, the classification resulting from the mixture model including one truncated normal component and two truncated exponentials (which yields the highest log-likelihood) does a very good job in classifying participants into “purely opportunistic” (sub-population 1, with a share of 11%), “conditional cooperators” (sub-population 2, with approximately 55% of the players), and “others” (sub-population 3).

## 6 Conclusions

A conditional cooperator in a public goods game wants to match (*ceteris paribus*) his partners’ expected contribution. This paper investigates theoretically and empirically whether (and to what extent) conditional cooperation can explain the evolution of individual contributions in a repeated two-person public goods experiment with a perfect strangers design. In the experiment, we elicit beliefs

Model		Subpopulation 1	Subpopulation 2	Subpopulation 3
NNN	$\pi$	15.4%	21.4%	63.6%
	$\gamma^*$	-0.398	0.018	-0.064
	$\rho$	0.246	0.041	0.112
	$\theta^*$	0.001	0.001	0.001
<i>log-likelihood = -6845.771</i>				
NEN	$\pi$	20.9%	32.3%	46.8%
	$\gamma^*$	-0.407	$8.4 \times 10^{-6}$	-0.067
	$\rho$	0.247	$-1.8 \times 10^{-4}$	0.123
	$\theta^*$	0.001		0.001
<i>log-likelihood = -7009.553</i>				
EEN	$\pi$	11.2%	33.0%	55.8%
	$\gamma^*$	-4.338	-0.008	-0.078
	$\rho$	4.254	0.008	0.134
	$\theta^*$			0.001
<i>log-likelihood = -6718.556</i>				

Table 3: Latent-class estimates of the continuous logit choice models

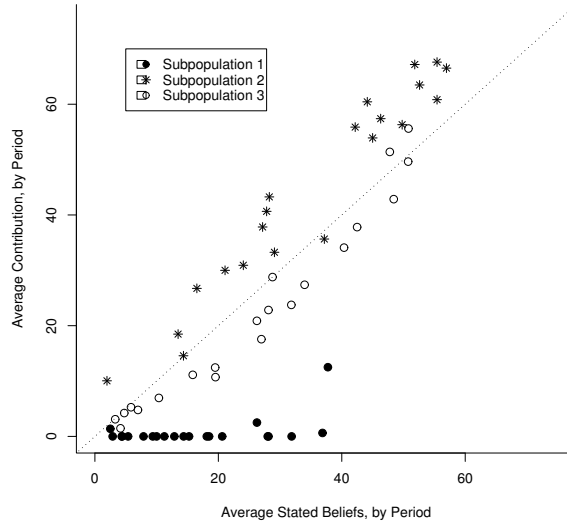
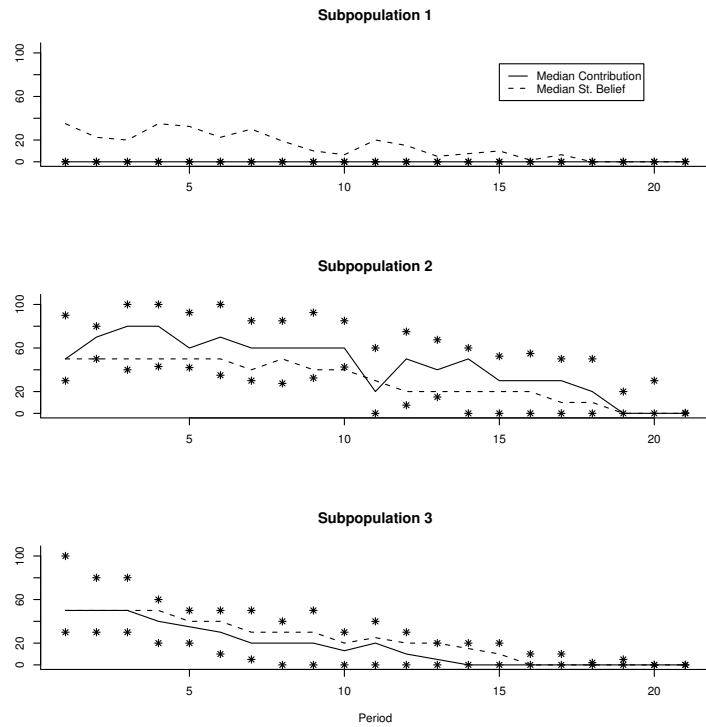


Figure 4: Average stated beliefs vs. average contributions, by period. (Classification of individuals according to the EEN latent-class model)



The symbol '\*' represents 1st. and 3rd. quartiles of contributions within sub-populations

Figure 5: Evolution of median stated beliefs and contributions  
(Classification of individuals according to the EEN latent-class model)

to obtain direct information about participants' expectations.

Within the framework of a random utility model, we derive a quantal-response equilibrium with rational expectations for the static, one-shot game. The equilibrium is characterized by either a truncated normal distribution or a truncated exponential distribution of contributions, depending on whether an agent has an intrinsic preference for conditional cooperation or not. By ruling out the assumption of rational expectations, we derive a 'naive' logit choice function, which is characterized either by a truncated exponential distribution (in the absence of conditional cooperation) or by a family of truncated normal distributions that vary only in their location parameter (if conditional cooperation is an argument of the utility function).

Allowing for (population) heterogeneity, the problem of statistically distinguishing whether conditional cooperation drives some individuals' behavior is not trivial in the extensive, multi-period game. Since beliefs are much more complex, having to anticipate not only the behavior of future partners, but also the one's own optimal response to such behavior, derivation of rational expectations equilibrium becomes untractable. However, we show that strategic reputation building in the sense of trying to induce cooperation from future partners by influencing their beliefs is more likely if a player has opportunistic preferences. Behavior of conditional cooperators, on the other hand, is more reactive, in the sense that current decisions mostly depend on beliefs about current behavior. This observation allows us to specify a set of parametric choice functions (in the context of a latent-class model), which can then be estimated via maximum likelihood using the EM algorithm.

Our results suggest that preferences differ among subgroups of individuals, with approximately 55% of the population being conditional cooperators, 11% being purely opportunistic, and the remaining 33% being relatively opportunistic albeit with a tendency to contribute higher average amounts. We conclude that the decline in contribution levels can be at least partially attributed to heterogeneity of preferences and initial beliefs in the population. Furthermore, the typical decline in contributions is not necessarily due to a reduction in the variance of the random utility specification (learning effect), but can also be explained in terms of forward-looking opportunistic players trying to trigger higher contributions from conditional cooperators using their own initial contributions strategically (social reputation effect).

## References

- Anderson, S. P., Goeree, J. K. and Holt, C. A. (1998), ‘A theoretical analysis of altruism and decision error in public goods games’, *Journal of Public Economics* **70**, 297–323.
- Andreoni, J. (1995), ‘Cooperation in public goods experiments: Kindness or confusion?’, *American Economic Review* **85**, 891–904.
- Ben-Akiva, M., Litinas, N. and Tsunokawa, K. (1985), ‘Continuous spatial choice: the continuous logit model and distributions of trips and urban densities’, *Transportation Research* **19A**, 119–154.
- Bernheim, D. (1984), ‘Rationalizable strategic behavior’, *Econometrica* **52**, 1007–1028.
- Bolton, G. and Ockenfels, A. (2000), ‘ERC: A theory of equity, reciprocity and competition’, *American Economic Review* **90**, 166–193.
- Davis, D. and Holt, C. (1993), *Experimental Economics*, Princeton University Press, Princeton.
- Dempster, A., Laird, N. and Rubin, D. (n.d.), ‘Maximum likelihood from incomplete data via the EM algorithm’, *Journal of the Royal Statistical Society, Series B* **39**(1).
- Fehr, E. and Schmidt, K. (1999), ‘A theory of fairness, competition and cooperation’, *Quarterly Journal of Economics* **114**, 817–868.
- Fischbacher, U. (1999), Zurich toolbox for readymade economic experiments, Working paper no. 21, University of Zurich, Switzerland.
- Fischbacher, U., Gächter, S. and Fehr, E. (2000), ‘Are people conditionally cooperative? evidence for a public goods experiment’, *Economics Letters* **71**, 397–404.
- Harsanyi, J. (1973), ‘Games with randomly disturbed payoffs’, *International Journal of Game Theory* **2**, 1–23.
- Isaac, R. M., Walker, J. and Thomas, S. H. (1984), ‘Divergent evidence on free riding: An experimental examination of possible explanations’, *Public Choice* **43**, 113–149.



- Keser, C. and van Winden, F. (2000), ‘Conditional cooperation and voluntary contributions to public goods’, *Scandinavian Journal of Economics* **102**, 23–39.
- Ledyard, J. (1995), Public goods: A survey of experimental research, in J. Kagel and A.E.Roth, eds, ‘The Handbook of Experimental Economics’, Princeton University Press, Princeton, NJ.
- Manski, C. (2004), ‘Measuring expectations’, *Econometrica* **72**, 1329–1376.
- McFadden, D. (1999), ‘Rationality for economists?’, *Journal of Risk and Uncertainty* **19**(1-3), 73–105.
- McKelvey, R. and Palfrey, T. (1995), ‘Quantal response equilibria for normal form games’, *Games and Economic Behavior* **10**, 6–38.
- McKelvey, R. and Palfrey, T. (1998), ‘Quantal response equilibria for extensive form games’, *Experimental Economics* **1**, 9–42.
- McLachlan, G. and Krishnan, T. (2005), *The EM Algorithm and Extensions*, Wiley.
- McLachlan, G. and Peel, D. (2000), *Finite Mixture Models*, Wiley.
- Offerman, T., A. S. and Sonnemans, J. (1998), ‘Quantal response models in step-level public good games’, *European Journal of Political Economy* **14**, 89–100.
- Palfrey, T. and Prisbrey, J. (1997), ‘Anomalous behavior in public goods experiments: How much and why?’, *American Economic Review* **87**, 829–846.
- Pearce, D. (1984), ‘Rationalizable strategic behavior and the problem of perfection’, *Econometrica* **52**, 1029–1050.
- Resnick, S. I. and Roy, R. (1991), ‘Random USC functions, max-stable processes and continuous choice’, *The Annals of Applied Probability* **1**(2), 267–292.
- Sugden, R. (1984), ‘Reciprocity: The supply of public goods through voluntary contributions’, *Economic Journal* **94**, 772–787.

# Appendix. Experimental instructions

## General Instructions

Thank you for participating in the experiment. You receive 2.5 euro for having shown up on time. If you read these instructions carefully and follow all the rules, you can earn more money. The 2.5 euro and all additional amount of money will be paid to you in cash immediately after the experiment. During the experiment, we shall not speak of euros but rather of tokens. Tokens are converted to euros at the following exchange rate: 100 Tokens = 1 euro.

It is strictly forbidden to speak to other participants during the experiment. If you have any questions, please ask us. We will gladly answer your questions individually.

The experiment is divided in 21 periods. In every period, participants are matched in groups of two. You will therefore be interacting with another participant every period. *The composition of the groups will change after each period*, so that the person you are matched with will always be different from one period to the next.

*You have NO chance of interacting with the same participant more than once.*

The identity of the participant you are matched with will not be revealed to you at any time.

## Detailed Instructions

At the beginning of each period, each participant receives 100 tokens. In the following, we shall refer to this amount as *your endowment*. In each period, each of the two members of a group has to fulfill two tasks:

Task 1: Your first task is **to decide how to use your endowment**.

You have to choose an amount of tokens that you want to contribute to a “project”. Your payment for each period will then be determined in the following way:

- Your contribution and the contribution of the other participant will be added up and multiplied by 0.7:

*Your “income from the project”* =  $[0.7 \times (\text{total group contribution})]$  tokens

- You will keep for yourself all the tokens that you do not contribute to the project:

*“Tokens you keep”* =  $[100 - \text{your contribution}]$  tokens

$$\begin{aligned} \text{Your income} &= \text{Tokens you keep} + \text{Income from the project} \\ &= (100 - \text{your contribution}) + (0.7 \times \text{total group contribution}) \end{aligned}$$

The income from the project is determined in the same way for the two group members; this means that both receive the same income from the project, regardless

of the size of their individual contributions.

EXAMPLE 1: If the sum of the contributions of the two group members is 60 tokens, each group member receives an income from the project of  $(0.7 \times 60) = 42$  tokens.

EXAMPLE 2: If the total contribution to the project is 9 tokens, then each group member receives an income from the project of  $(0.7 \times 9) = 6.3$  tokens.

Each point that you keep for yourself raises your income by one point.

Each point that you contribute to the project raises your income *as well as* the income of the other member of your group by 0.7 tokens. Therefore, the group's total income from the project raises by  $0.7 \times 2 = 1.4$  tokens.

Similarly, you profit from the contributions made by the other member of your group: For each point that (s)he contributes, you earn 0.7 tokens.

At the end of each period, you will be informed about the number of tokens contributed by your current group member and about your period income.

Task 2: In every period, besides deciding about how to use your endowment, you have **to predict how many tokens the other member of your current group will contribute to the project.**

Additionally, you are required to state on a scale ranging from 0% to 100% how certain you are that your prediction is accurate.

- If you state 100%, this means that out of one hundred possible predictions you think that you are right all the times.
- Conversely, if you state 0%, this means that out of one hundred predictions you think that you are right 0 times.

At the end of the experiment, the participant with the most accurate predictions will receive an additional bonus of 7 euro. The closer your predictions are to the true contribution of the other participants you interact with, the higher are your chances of receiving the bonus.

*Please remain seated quietly until the experiment starts. If you have any questions please raise your hand.*

**Control Question:** Please answer the following questions. A wrong answer has no consequences. If you have any questions, please, raise your hand.

1. Each group member has an endowment of 100 tokens. Neither you nor the other member of your group contributes any token to the project. What is:  
Your income? ...  
The income of the other member of your group? ...

2. Each group member has an endowment of 100 tokens. You as well as the other member of your group contribute 100 tokens to the project. What is:  
Your income? ...  
The income of the other member of your group? ...
3. Each group member has an endowment of 100 tokens. The other member of your group contributes 50 tokens to the project. What is your income:  
If you contribute 0 tokens to the project? ...  
If you contribute 70 tokens to the project? ...
4. Each group member has an endowment of 100 tokens. You contribute 20 tokens to the project. What is your income:  
If the other member of your group contributes 0 tokens to the project? ...  
If the other member of your group contributes 60 tokens to the project? ...