Conte, Anna; Levati, M. Vittoria

Working Paper

Use of data on planned contributions and stated beliefs in the measurement of social preferences

Jena Economic Research Papers, No. 2011,039

Provided in Cooperation with:
Max Planck Institute of Economics

Jena Economic Research Papers

2011 – 039

Use of data on planned contributions and stated beliefs in the measurement of social preferences

by

Anna Conte
M. Vittoria Levati

www.jenecon.de

ISSN 1864-7057

The Jena Economic Research Papers is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Use of data on planned contributions and stated beliefs in the measurement of social preferences

Anna Conte[a,b], M. Vittoria Levati[a,c,*]

[a] *Max Planck Institute of Economics, Kahlaische Str. 10, 00745 Jena, Germany*

[b] *University of Westminster, EQM Department, 35 Marylebone Road, London NW1 5LS, UK*

[c] *University of Verona, Department of Economics, Via dell'Artigliere 19, 37129 Verona, Italy*

Abstract

In a series of one-shot linear public goods game, we ask subjects to report their contributions, their contribution plans for the next period, and their first-order beliefs about their present and future partner. We estimate subjects' preferences from plans data by a finite mixture approach and compare the results with those obtained from contribution data. Our results indicate that preferences are heterogeneous, and that most subjects exhibit conditionally cooperative inclinations. Controlling for beliefs, which incorporate the information about the other's decisions, we are able to show that plans convey accurate information about subjects' preferences and, consequently, are good predictors of their future behavior.

JEL classification: C35; C51; C72; H41

Keywords: Public goods games; Experiments; Social preferences; Mixture models

*Corresponding author. Max Planck Institute of Economics, Strategic Interaction Group, Kahlaische Str. 10, 07745 Jena, Germany. Tel.: +49 3641 686629; fax: +49 3641 686667.

E-mail address: levati@econ.mpg.de (M. V. Levati).

1 Introduction

A vast amount of experimental evidence has shown that individuals contribute voluntarily to public goods, even if self-interest implies that free riding should be their dominant strategy. Several researchers explain this finding in terms of social preferences, which are mostly discussed in the economics literature under the rubric of altruism (Levine, 1998), efficiency concerns (Charness and Rabin, 2002), inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), and conditional cooperation (Fischbacher et al., 2001). Our paper represents a first attempt to use data on planned contributions and stated beliefs about others' future contributions in order to assess the relative importance of competing types of social preferences.

A handful of experimental studies have investigated the empirical validity of the various models of social preferences. The majority of these studies have relied on distribution games (i.e., games where players make decisions about the allocation of payoffs between themselves and others). Andreoni and Miller (2002), for instance, examine a series of dictator games which are characterized by different costs of giving and confirm that, for most subjects, altruistic choices are consistent with the axioms of revealed preferences. Charness and Rabin (2002) conduct a wide range of dictator and response games and observe that people are especially concerned with increasing social welfare. Cox et al. (2007) introduce a parametric model of reciprocity and fairness, and estimate it for, among others, the dictator and ultimatum games. Iriberri and Rey-Biel (2009) collect choices in dictator games similar to those implemented by Andreoni and Miller (2002) and classify a high fraction of their subjects as inequity averse.[1]

In the context of public goods games, the identification of social preferences seems to require a careful exploration not only of the decision maker's behavior, but also of his beliefs about the others' behavior. In a piece of early work, Offerman et al. (1996) use a step-level public goods game with binary contributions to provide insights into the relationship between expectations and behavior. The authors find that expectations are rather reasonable

[1]Other studies aimed at identifying and quantifying different types of preferences are Fisman et al. (2007), and Blanco et al. (2011). This line of work finds substantial heterogeneity in individual behavior.

than rational (individuals are too optimistic about the others' behavior). They also argue on the existence of a relationship between the subjective probability of being decisive and the propensity to contribute. In linear public goods games, Croson (2007) detects a significant and positive relationship between an individual's own contribution and his beliefs about the contributions of the other members of his group. More recently, Fischbacher and Gächter (2010) and Ambrus and Pathak (2011) independently explain the decline of contributions in repeated public goods settings by combining the role played by beliefs in influencing contributions with the presence of different types of players. Finally, González et al. (2011) model the relationship between contributions and elicited beliefs by means of a random utility approach that allows for conditionally cooperative, selfish, and altruistic preferences. They find that most players are strongly conditional cooperators.[2]

We depart from the aforementioned literature in that we identify people's cooperative preferences by looking at the relationship between planned contributions and stated beliefs regarding the others' actions in the following period. We consider a sequence of fifteen one-shot two-person linear public goods games (i.e., our participants play fifteen periods in the so-called 'perfect stranger' design which ensures that nobody meets the same person more than once). In each game/period, subjects make two contribution decisions and specify two distributions of first-order beliefs: (a) they choose their contribution amount from a given set of 11 elements, and state their subjective probabilities that the participant they are currently matched with has opted for any element of the same set, and (b) they choose from an identical discrete set the amount that they plan to contribute in the following period, and state their subjective probabilities that the participant they will be matched with then will opt for any element of the set.

The intuition here is that, if other-regarding preferences are idiosyncratic and beliefs play a role in contribution determination, then a person should reveal his preferences not only in the way his contribution relates to his current beliefs, but also in the way his planned contribution relates to his beliefs about the other's contributions one period ahead. A major

[2]Recently, experimental economists have begun to elicit subjects' beliefs about the likely play of others to investigate, besides issues related to social preferences, the consistency between stated beliefs and choice behavior (see, e.g., Costa-Gomes and Weizsäcker, 2008; Rey-Biel, 2009).

contribution of this paper is to verify this intuition and thus to test whether plans convey accurate information about preferences and behavior. We believe this issue to be important because intentions to contribute (mostly gathered through 'consequential' survey questions)[3] are frequently used by businesses and governments to determine which product to offer and which policy to adopt.

It is somehow surprising that economists tend to ignore data on plans, entrusting for example the study of purchase intentions to marketing (see Morwitz, 1997, and references therein). The typical claim is that people cannot plan ahead,[4] with the empirical divergence between plans and eventual behavior being interpreted as evidence of how poor predictors individuals are of their future acts. However, as Manski (1990, p. 934) points out, such a conclusion is unwarranted: planned and eventual behavior may differ as a consequence of events occurring between the time plans are elicited and the time actions take place. As we consider public goods games in which each individual receives feedback about the other's contribution after each period, what may cause one's own final behavior to differ from one's own earlier plan is a revision of first-order beliefs (which in turn captures the effect of the new information acquired in the interim period between elicitation of plans and final choices). Hence, following Manski's contention, to assess whether the preference types estimated from plans are similar to those estimated from final contributions we condition on beliefs that incorporate such changes.

We focus on a) purely selfish preferences, b) preferences for altruism or unconditional giving, and c) preferences for conditional cooperation. Selfish types contribute zero regardless of their beliefs. Altruists always contribute the same positive amount. Finally, conditional cooperators contribute conditionally on the other's expected contributions. We estimate the proportion of individuals in the population who behave according to each of these three preference types, separately from data on planned contributions and from data on final contributions, using a finite mixture approach. This allows us to test whether subjects' cooperative preferences as estimated from plans data are consistent with their cooperative preferences as

[3]Survey questions are consequential if the "survey's results are seen as potentially influencing an agency's actions and the agent cares about the outcomes of those actions" (Carson and Groves, 2007, p. 183).

[4]On this topic, see, e.g., Bone et al. (2003) or Hey (2005), who study dynamic decision problems under risk involving Nature moves in a probabilistic, rather than strategic, way.

revealed in final contributions, both at a population level and at an individual level.

We consider two experimental treatments that only differ in the level of information supplied to the participants. In one treatment, labeled Plan-Info, subjects are informed at the end of each period about both the final decision and the planned-in-the-previous-period decision of the person that they are currently matched with. With plans made known, “rational” individuals − who are dynamically consistent − have no reason to change their preferences across planned and final decisions. Thus, under the hypothesis of “rationality”, we would expect to find similar proportions of types comparing the players’ erstwhile planning behavior to their present-period behavior.

In the other treatment, labeled Plan-NoInfo, participants receive feedback only on the final decision of their current co-player. Since their plans remain concealed until the end of the game, subjects may uninhibitedly express their “true” preferences in the plan dimension. Selfish forward-looking players, for example, have strategic incentives to contribute positive amounts in the early periods so as to manipulate the beliefs of the conditional cooperators. But, because this kind of behavior is justified only if it can be observed by the others, in the Plan-NoInfo treatment such selfish persons can formulate their plans without concealing their real preferences. Moreover, some altruists may dislike being played for a sucker (see Orbell and Dawes, 1981, on the so-called *sucker effect*) and consequently avoid to contribute in the belief that their fellow player will withhold contributions (see, e.g., Schnake, 1991). But sucker aversion would diminish (or even vanish) if the other is not informed of one’s own contributions. Thus, the Plan-NoInfo treatment provides a novel way to estimate the extent of both strategic forward-looking behavior and sucker aversion in our sample.

The rest of the paper is organized as follows. After introducing the basic games, Section 2 details our experimental treatments and procedures. Section 3 describes the data and reports preliminary statistical tests. In Section 4 we define the mixture model, and in Section 5 we present and discuss the estimates of the model. Section 6 summarizes our central findings and concludes.

2 The experiment

2.1 The public goods games

The basic decision situation is a standard linear public goods game. Let $N = \{1, \ldots, 30\}$ stand for a population of 30 individuals who interact in pairs for $t = 1, \ldots, 15$ periods according to a perfect-stranger matching design.[5] At the beginning of every period, each individual $i \in N$ is endowed with 100 ECU (Experimental Currency Units) which he can either keep for himself or contribute to a public good. We discretize the choice set of each individual i to eleven alternatives: $\mathcal{A} \in \{(0, 100), (10, 90), (20, 80), (30, 70), \ldots, (80, 20), (90, 10), (100, 0)\}$, where the first and second amounts denote the number of ECU that i contributes to the public good and keeps for himself, respectively. More synthetically, we can denote each alternative by a ($a = 0, \ldots, 10$), so that each element of $\mathcal{A}$ can be expressed as $(a \times 10,\ 100 - a \times 10)$. For example, opting for $a = 0$ means contributing nothing and keeping everything for oneself. Let $c_{i,t}$ be i's contribution in period t. Likewise, let $c_{j,t}$ define player i's partner's (player j's) chosen contribution in t.[6] In the standard voluntary contribution mechanism, participants make choices only for the present period and the monetary payoff of player i (for all $i \in N$) in each period $t = 1, \ldots, 15$ is given by:

$$\pi_{i,t} = 100 - c_{i,t} + 0.8\,(c_{i,t} + c_{j,t}), \qquad (1)$$

where the public good is equal to the sum of the contributions of i and j.

The game we consider deviates from this usual practice in that we require subjects not only to choose for the present period, but also to plan what they intend to do in the following period. We do not want to force the participants to commit themselves to the plan, while at the same time we want an incentive-compatible way of eliciting plans so that players are motivated to honestly report their intentions. To this aim, we let the public good in periods 2 to 15 be based on either the sum of i's and j's final contributions or the sum of the contribution plans they formulated beforehand, with both possibilities being equally likely.[7] Let $p^t_{i,t-1}$ and

[5] We chose this protocol to minimize strategic effects from repeated play and to allow for revisions in beliefs only at the population level.

[6] To simplify notation, we always refer to player i's partner as j, although this is a different person in each period.

[7] A similar procedure for incentivizing subjects to state a carefully considered, truthful plan has been applied

$p^t_{j,t-1}$ denote the amounts that, respectively, player i and player j (player i's partner) plan at time $t-1$ to contribute in t, with $t = 2, \ldots, 15$.[8] The payoff function in the game with plans elicitation is given by (1) in the first period. Afterwards (i.e., in $t = 2, \ldots, 15$), it can be either (1) with 50% probability or

$$\pi_{i,t} = 100 - p^t_{i,t-1} + 0.8 \left(p^t_{i,t-1} + p^t_{j,t-1}\right), \quad (2)$$

with 50% probability. In what follows, for each player $i \in N$, we shall call $c_{i,t}$ the "final contribution" of i and $p^t_{i,t-1}$ the "planned contribution" of i.

2.2 Treatments, decisions, and scoring rule for beliefs

Using a between-subjects design, we study three treatments. In the control (C) treatment, subjects play the standard public goods game with payoff function (1). In every period $t = 1, \ldots, 15$, each participant i chooses one of the eleven alternatives in $\mathcal{A}$, thereby making a contribution choice $c_{i,t}$, and reports a first-order belief vector $\mathbf{b}^t_{i,t}$, i.e., a probability distribution over the eleven possible choices of his current partner j.

In the other two treatments, Plan-Info (P_I) and Plan-NoInfo (P_{NI}), subjects play the public goods game with plans elicitation described above where, in all periods but the first, the payoff function can be either (1) or (2), each with probability 1/2. In every period $t = 1, \ldots, 14$, besides choosing $c_{i,t}$ and stating $\mathbf{b}^t_{i,t}$, each participant i decides on the alternative that he plans to select in the next period, thereby providing a planned contribution $p^{t+1}_{i,t}$, and specifies his beliefs $\mathbf{b}^{t+1}_{i,t}$ about the alternative that his next-period partner will choose.[9]

In all three treatments, at the end of each period, participants receive feedback about the final contribution decision of their current-period partner, namely $c_{j,t}$.[10] Participants in P_I are also informed about the planned-in-the-previous period contribution decision of the

by Barkan and Busemeyer (1999). The aim of the authors was to test for dynamic consistency in an experiment in which participants had to make a planned choice and a final choice about a second gamble within a sequence of two gambles.

[8] When convenient, we will equivalently use the notation $p^{t+1}_{\cdot,t}$, $t = 1, \ldots, 14$, to indicate contribution plans made in t for $t+1$.

[9] The instructions make clear that subjects have to predict the decisions of two different persons: the current-period partner ($\mathbf{b}^t_{i,t}$) and the next-period partner ($\mathbf{b}^{t+1}_{i,t}$).

[10] To simplify presentation, players' contributions in treatment C will be sometimes referred to as "final" even though no distinction between final and planned contributions is made in C.

person they are currently matched with, i.e., they also learn about $p^t_{j,t-1}$ with $t = 2, \ldots, 15$. In all treatments, no information about the realized public good and the period monetary payoff is provided until the end of the experiment. It is therefore impossible for participants in P_{NI} to infer their partner's plans during the game.

The control treatment is used to ensure that eliciting plans does not influence subjects' behavior. The two treatments with plans elicitation serve the main purposes of this paper. They enable us (i) to estimate subjects' types based on their elicited plans, and (ii) to compare these estimates with those obtained using final contributions so as to test the hypothesis that there is no difference between them. This "consistency" hypothesis may, however, not be confirmed in treatment P_{NI} since non-disclosed plans could lead subjects to reveal their "genuine" preferences, free of strategic forward-looking reasoning and sucker aversion.

We ask for beliefs because we want to assess the relationship between them and contributions, which we presume to differ across types. Previous research in experimental economics has shown that the mere act of eliciting beliefs about the others' contributions can affect behavior in finitely repeated public goods games (see, e.g., Croson, 2000; Gächter and Renner, 2010), albeit the evidence regarding the undesirable effects of beliefs elicitation procedures is far from being conclusive, and it does not concern stranger matching protocols.[11] As participants state their first-order beliefs in all our treatments and, in P_I and P_{NI}, for both the present and next periods, the unintended effects of beliefs on behavior (if any) would occur in all our treatments and apply to all our variables of interest.

Beliefs are elicited by endowing participants with 100 tokens and asking them to allocate these tokens on the 11 alternatives available to their partner. Participants are asked to assign tokens to each alternative in a way that reflects the probability they attach to the event that their partner chooses that alternative. We can think of each token as representing one percentage point.

We give subjects proper incentives for accurate predictions by using a quadrating scoring rule. The rule is defined as follows. Assume that $\hat{c}_{j,t}$ is the alternative actually chosen by subject j (i's partner) in period t. Let i's beliefs in period $t - \tau$ be $\mathbf{b}^t_{i,t-\tau}$ with τ equals 0 in

[11] Wilcox and Feltovich (2000), for instance, find that contributions are not affected by whether beliefs are elicited or not.

treatment C and either 0 or 1 in treatments P_I and P_{NI}. Indicate the generic element of the belief vector as $b^t_{i,t-\tau}(a)$, which denotes the probability (in percentage points) that in period $t-\tau$ subject i assigns to the event that his partner in period t chooses alternative a. In other words, $\mathbf{b}^t_{i,t-\tau} \equiv \left(b^t_{i,t-\tau}(0), b^t_{i,t-\tau}(1), \ldots, b^t_{i,t-\tau}(10)\right)$ with $\sum_{a=0}^{10} b^t_{i,t-\tau}(a) = 100$. Subject i's payoff for accuracy of predictions is:

$$v_{i,t} = 100 - 0.005 \times \sum_{a=0}^{10} \left(b^t_{i,t-\tau}(a) - 100 \times \mathbb{1}\left(a = \hat{c}_{j,t}\right)\right)^2, \tag{3}$$

where $\mathbb{1}(\cdot)$ is an indicator function taking on the value 1 if the statement in brackets is true and 0 otherwise.[12] Note that since beliefs are elicited in percentage points, they have to be divided by 100 to get probabilities.

In the instructions, we use a verbal description of the rule and give numerical examples. Problems of the quadrating scoring rule are that incentives are flat at the maximum and that it may be difficult to understand. To avoid this latter problem, our instructions emphasize that the more accurate the beliefs, the higher the payment.

2.3 Procedures

The experiment was programmed in z-Tree (Fischbacher, 2007) and conducted in the experimental laboratory of the Max-Planck Institute of Economics in Jena (Germany). The subjects were undergraduate students from the University of Jena, who had never participated in public goods and prisoner dilemma experiments before. They were recruited using the ORSEE (Greiner, 2004) software. Upon entering the laboratory, the subjects were randomly assigned to visually isolated computer terminals. The instructions (which are reproduced, translated from German, in Appendix B) were distributed and then read aloud to establish public knowledge. All subjects' questions were answered individually at their seats. Before starting the experiment, subjects had to answer a control questionnaire which tested their comprehension of payoff functions (1) and (2) and thus of their monetary incentives. The experiment did not

[12] A similar rule has been used by, e.g., Offerman et al. (1996), Nyarko and Schotter (2002), Costa-Gomes and Weizsäcker (2008), and Rey-Biel (2009), although there exists no consensus among experimentalists about the optimal incentive mechanism for eliciting beliefs. Huck and Weizsäcker (2002) compare beliefs elicited via a quadratic scoring rule with beliefs elicited via a Becker-DeGroot-Marshak pricing rule, and find that the quadratic scoring rule yields more accurate beliefs.

start until all participants had answered the questionnaire correctly. We can therefore safely assume that the participants understood the game.

Overall, we ran thirteen sessions: three for the C treatment, and five for each of the two plan treatments (P_I and P_{NI}). In each session, we had 30 participants so that, in total, our analysis relies on 90 individuals observed in treatment C and 150 individuals observed in each of the other two treatments.

Participants in treatment C were paid according to their contributions in one randomly chosen period t_1, at a rate of €0.15 per ECU, and according to the accuracy of their belief statements in another randomly chosen period $t_2 \neq t_1$, using incentive rule (3). In treatments P_I and P_{NI} two further random draws determined (i) whether the public good in t_1 was based on either the sum of the final contributions or the sum of the contribution plans formulated beforehand, and (ii) which of the two reported belief vectors ($\mathbf{b}_{i,t_2}^{t_2}$ or $\mathbf{b}_{i,t_2-1}^{t_2}$) counted for payment in t_2.[13] Sessions lasted, on average, two hours with most of the time being used up for reading the instructions and answering the control questionnaire. Average earnings per subject were €26.82 (inclusive of a €2.50 show-up fee), ranging from €14.50 in treatments C and P_{NI} to €43.30 in treatment P_{NI}.

3 Data description and preliminary tests

We will approach the description of our data with the intention to assess whether the elicitation of plans induces distortions in individual attitude to contribute.

Figure 1 shows the time path of the average final contributions (solid lines) for each of the three treatments as well as the average planned contributions (dashed lines) for the two experimental treatments. The figure shows three things. First of all, consistent with previous experimental results, all five time series of average contributions begin high and decrease over time. Second, the average final contributions in period 1 are quite similar across treatments. However, while the average final contributions in treatments C and P_{NI} proceed side by side, those in treatment P_I eventually depart. Finally, planned contributions lie, on average, always slightly above the respective final contributions. The gap seems to reduce in P_I toward the

[13] See the instructions in Appendix B for a description of the random procedures that were used.

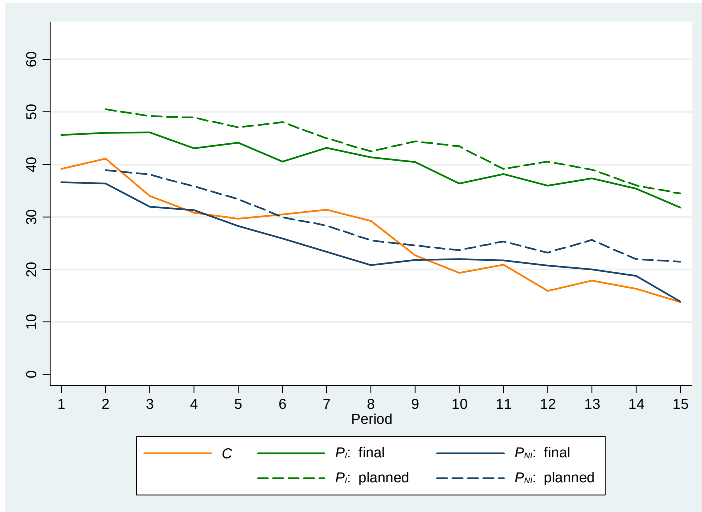

Figure 1: Comparison of average (final and planned) contributions over time across treatments.

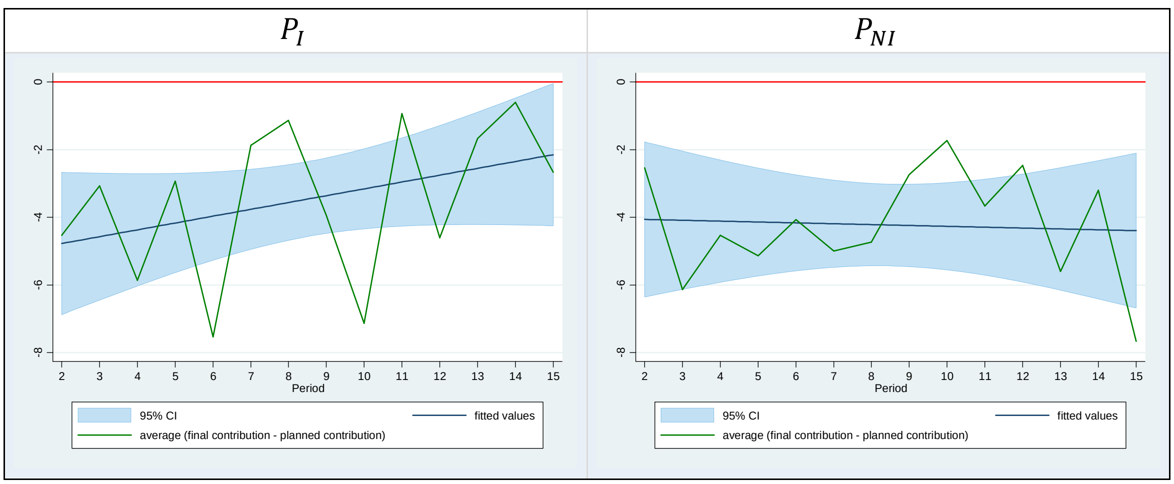

Figure 2: Difference between final and planned contributions over time: the green lines connect the average of this difference measured in each period; the blue lines are linear fits with individual random effects of this difference on period; the shaded areas represent 95% confidence intervals, computed using the standard error of prediction.

end of the game.

Confirmation of this last observation comes from Figure 2 that plots, separately for P_I and P_{NI}, the time evolution of the difference between final and planned contributions (namely, $c_{i,t} - p^t_{i,t-1}$, for $t = 2, \ldots, 15$). This graphical representation makes easy to see that in both treatments the average difference (green line) is always negative but very small. The blue line in each pane represents a linear fit of the difference (individual level) on period. While the gap between final and planned contributions tends to reduce in treatment P_I, it looks

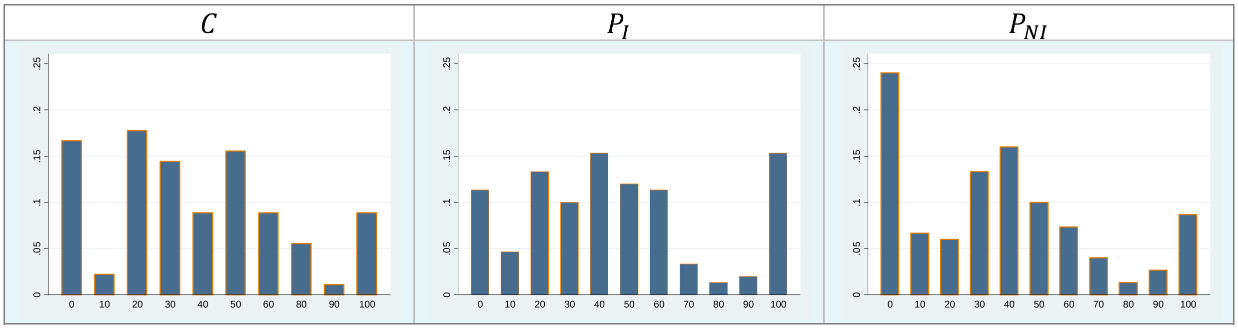

Figure 3: Bar-diagrams of period-1 final contributions. The bar height indicates the proportion of times the corresponding contribution has been chosen.

quite stable or even slightly increasing in treatment P_{NI}. Following Manski (1990), we argue that the gap between final and planned contributions may be explained by differences in the belief formation process. Disregarding for the time being the relationship between choices and beliefs (which will be investigated in detail in the next two sections by a structural approach), we concentrate on descriptive statistics of choices and beliefs separately.

Regardless of the treatment, at the beginning of the game all participants have no information about the others so that their period-1 final contribution, $c_{i,1}$, cannot be affected by the observation of others' behavior. If the mere elicitation of plans has an effect on subjects' responses regarding their final contributions, we would expect this to emerge from a comparison of the distribution of period-1 final contributions in the control and the two experimental treatments. Such distributions are reported in Figure 3. At a first sight, the three diagrams do not look alike, but they share some similarities: they all have some mass at the extreme contributions (0 and 100) and the remainder somewhat at the center. In effect, both the Wilcoxon rank-sum (WRS) and Kolmogorov-Smirnov (KS) tests reveal that there are no significant differences in $c_{i,1}$ between C and P_I (WRS: p-value = 0.15; KS: p-value = 0.43) as well as between C and P_{NI} (WRS: p-value = 0.51; KS: p-value = 0.36).

Turning to first-order beliefs, Figure 4 draws for each treatment box plot sets of $\mathbf{b}^t_{i,t}$. Each set of box plots refers to the period indicated on the vertical axes. Each box plot in a set represents the distribution of the probabilities in percentage points, i.e., the number of tokens (measured on the horizontal axes) assigned to the corresponding alternative in period t. Each alternative is characterized by a color. A legend, reported at the bottom of the figure,

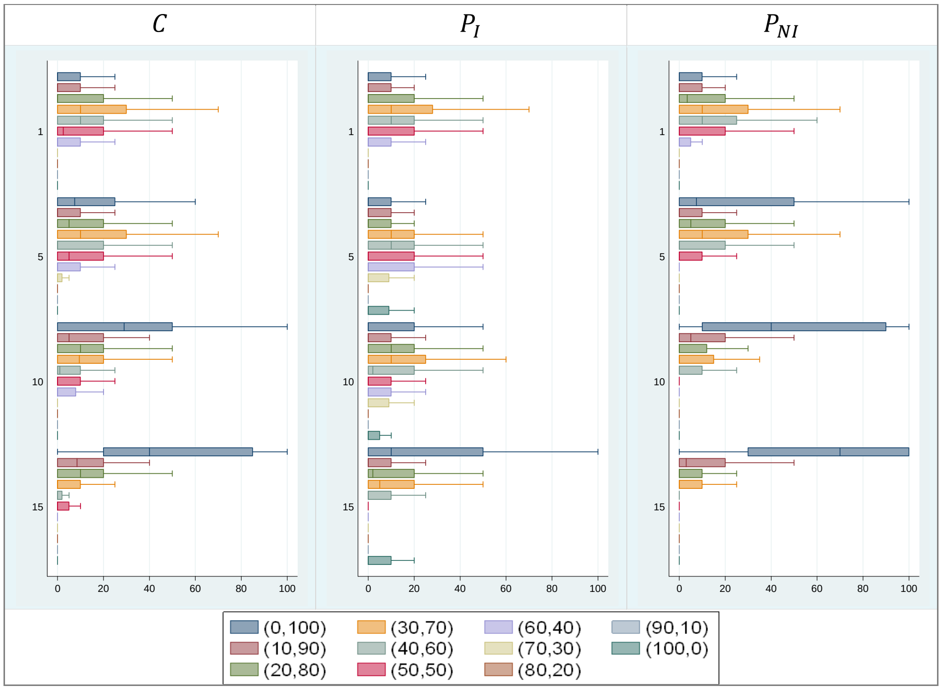

Figure 4: Sets of box plots of subjects' stated beliefs about the current-period partner's decision ($\mathbf{b}^t_{i,t}$), separately for each treatment and periods $t = 1, 5, 10, 15$ (reported on the vertical axes). Each alternative in the choice set is represented by a different color, listed in the legend at the bottom.

associates colors to alternatives. For each of the three treatments, box plot sets for periods 1, 5, 10, and 15 are juxtaposed to facilitate within- and between-treatment comparisons. Two features in the figure are worth noting: (a) the box plot sets of beliefs in period 1 are impressively similar in all three treatments; (b) beliefs tend to concentrate over time toward the alternative (0, 100), especially in treatments C and P_{NI}.

Figure 5 juxtaposes for each treatment with plans elicitation box plot sets of beliefs about the current-period partner as stated both in the previous period and in the current period ($\mathbf{b}^t_{i,t-1}$ and $\mathbf{b}^t_{i,t}$, respectively). This representation allows us to compare beliefs reported in the planning phase of the previous period with beliefs reported after knowing the contribution of another participant. Each pane consists of four sets of box plots, one for each of the following periods: 2, 6, 11, and 15. The figure shows that beliefs converge toward the lowest contributions as time proceeds; this tendency is slightly more marked for the current-period beliefs than for the one-period-ahead beliefs, and more pronounced in P_{NI} than in P_I.

Based on $\mathbf{b}^t_{i,t-1}$ and $\mathbf{b}^t_{i,t}$, we can compute the amount that in period $t-1$ subject i expects

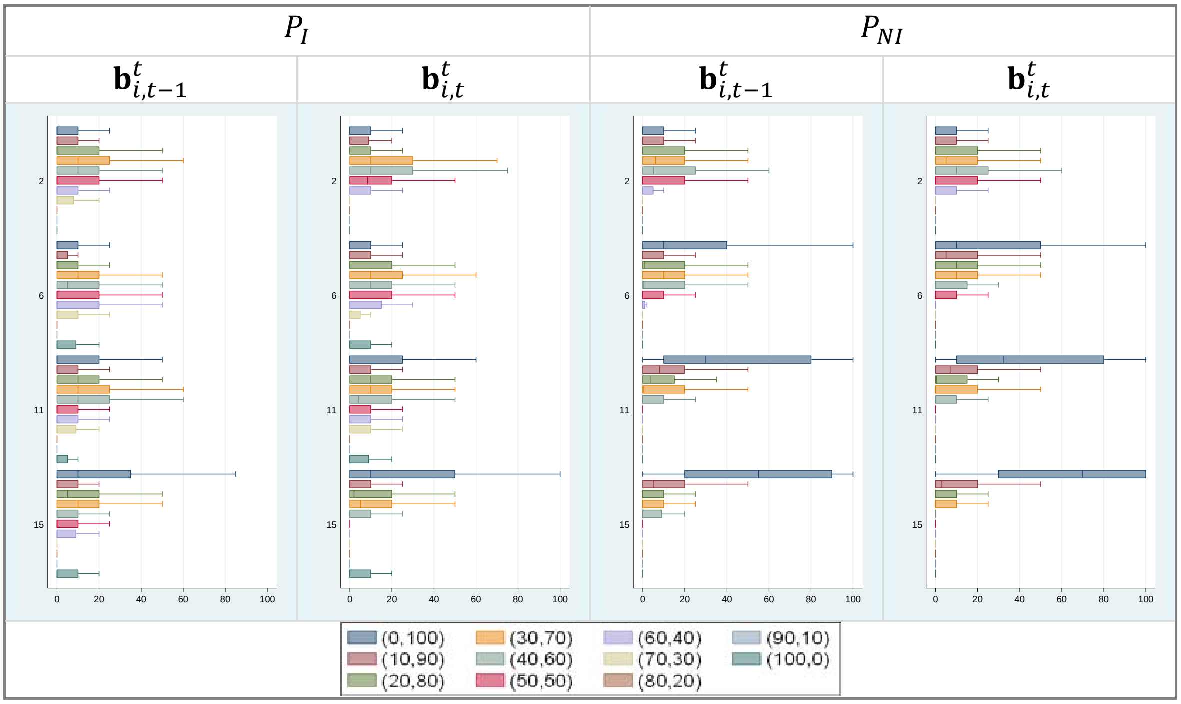

Figure 5: Sets of box plots of subjects' beliefs about their period-t partner's choice as stated in the previous period ($\mathbf{b}^t_{i,t-1}$) and in the current period ($\mathbf{b}^t_{i,t}$) separately for P_I and P_{NI} and periods $t = 2, 6, 11, 15$. See the caption of Figure 4 for details.

his partner j to contribute in t (one-period-ahead expected contribution, or $\mathrm{E}_{i,t-1}[c_{j,t}]$) and the amount that in period t i expects j to contribute in t (final expected contribution, or $\mathrm{E}_{i,t}[c_{j,t}]$). These amounts are calculated by averaging all the possible contributions, weighted for the corresponding beliefs.[14] The distributions of the difference between final and one-period-ahead expected contributions are plotted as a time series of box plots in Figure 6. The figure distinctly shows that, as time progresses, the box plots collapse to zero in P_{NI}, indicating a tendency for final beliefs to catch up with one-period-ahead beliefs. Such a tendency is less pronounced in P_I.

In order to provide a formal test of the hypothesis that beliefs are not affected by plans elicitation, we proceed similarly to the analysis of contribution choices and compare the distributions of period-1 final expected contributions in the control and the two experimental treatments. Figure 7 lends first visual support to the fact that the distributions are similar

[14] More specifically, the one-period-ahead and final expected contributions are computed, respectively, as:

$$\mathrm{E}_{i,t-1}[c_{j,t}] = \frac{\sum_{a=0}^{10} (a \times 10) \times b^t_{i,t-1}(a)}{100} \qquad \text{and} \qquad \mathrm{E}_{i,t}[c_{j,t}] = \frac{\sum_{a=0}^{10} (a \times 10) \times b^t_{i,t}(a)}{100}.$$

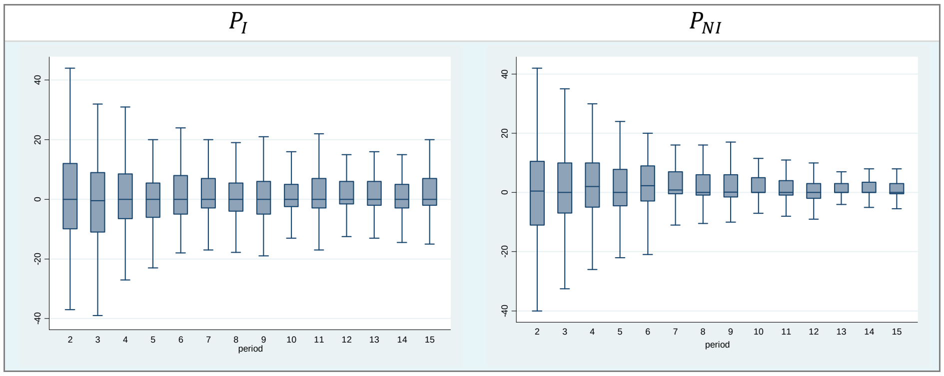

Figure 6: Time series of box plots of the difference between final expected contributions and one-period-ahead expected contributions.

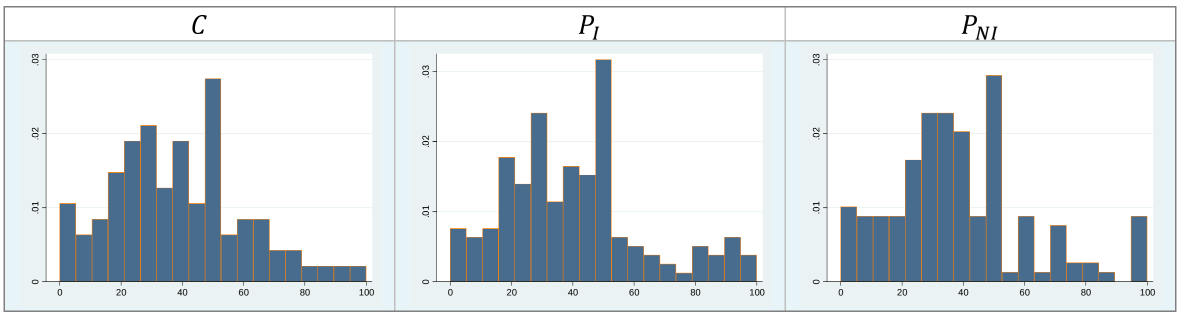

Figure 7: Histograms of period-1 final expected contributions.

across the three treatments. Further corroboration stems from non-parametric hypothesis testing: both the Wilcoxon rank-sum test and the Kolmogorov-Smirnov test confirm a lack of significant difference in the distribution of first-period expected contributions between C and P_I (WRS: p-value = 0.68; KS: p-value = 0.91) and between C and P_{NI} (WRS: p-value = 0.96; KS: 0.97, respectively).

The analysis presented here suggests that eliciting plans does not affect behavior. Yet, there appears to be a divergence between planned and final contributions. In the next section, we will examine whether a joint evaluation of choices and stated beliefs can account for such divergence.

4 Distinguishing contribution motives: the mixture model

The following empirical analysis is focused on the interaction of choices and beliefs in revealing individual cooperative decisions. It is based on the assumption that there are different types of individual in the population and that each type is defined by a rule that describes his decision process. Each behavioral rule has a peculiar content in terms of preferences and beliefs. We consider selfish agents and non-selfish agents, concentrating our attention, in particular, on altruists and conditional cooperators. Our mixture assumption states that each player is of one of these three types and that he cannot switch type between periods.

In order to distinguish between types, we have to define the rule that each type uses to compute final and planned contributions. Let us assume that i's final contribution, $c_{i,t}$, and i's planned contribution, $p_{i,t}^{t+1}$, at time t are empirical realizations of the random variable $Y_{i,t}$. Note that we assume that players cannot change type across periods, but that they can behave according to one type's rule when choosing their final contributions and according to another type's rule when making their contribution plans. In other words, the random variable that generates final contributions and planned contributions is not necessarily unique. While we expect the distributions of types estimated from final contribution data and from planned contribution data to be alike under P_I, such a relationship might not hold under P_{NI}.

Let us now proceed by defining the behavioral rule of each type. The first type we consider is the *selfish* type (SEL). This individual is only interested in the maximization of his own monetary payoff. Since the marginal per capita return in payoff functions (1) and (2) is less than unity, the dominant strategy for such a type is to contribute and plan to contribute nothing. Hence, the behavior of a selfish player is described by the following equation:

$$Y_{i,t} = 0 \qquad \forall t. \tag{4}$$

An *altruist* (ALT) contributes from unconditional concern for others. His final or planned contribution is a fixed positive amount that is dictated by his preferences over his own and his partner's payoff. An altruistic agent is expected to behave according to the following rule:

$$Y_{i,t} = m_i \qquad \forall t, \qquad m_i > 0. \tag{5}$$

Following Bardsley and Moffatt (2007), we take m_i to equal alternatively the median of i's 15 final contributions and the median of i's 14 planned contributions.

A *conditional cooperator* (CC) dislikes contributing different amounts than others (Fischbacher et al., 2001; Fischbacher and Gächter, 2010). His behavior is described by:

$$Y_{i,t} = Y_{j,t} \quad \forall t. \tag{6}$$

Since player i is unaware of $Y_{j,t}$ when deciding on his own contributions, i's conditional choices can only be based on his first-order beliefs about j's final (or planned) contributions. We assume that conditional cooperator i contributes an amount equal to his partner's most likely contribution, namely an amount equal to the mode of the distribution of first-order beliefs. If this distribution is multimodal, we assume that conditional cooperator i chooses an alternative corresponding to any of the modes. More formally, $Y_{i,t} \in \text{mode}\left(\mathbf{b}_{i,t}^{t+\tau}\right) \ \forall t$, with $\tau \in \{0, 1\}$.[15] This process of conditioning on beliefs together with repeated observations per subject allow us to distinguish conditional cooperators from altruistic and selfish types.[16]

We do not expect a player to faithfully comply with what is dictated by the behavioral rule corresponding to his type. As argued by, e.g., Andreoni (1995), Palfrey and Prisbrey (1996, 1997), Anderson et al. (1998), and Houser and Kurzban (2002), subjects may be confused and make mistakes. We allow for the possibility of sub-optimal behavior by introducing a tremble parameter, $w_{i,t} \in [0, 1]$.[17] This represents the probability that player i – whatever the reason – chooses completely at random between the alternatives. We also assume that each player is characterized by an individual-specific probability of trembling, and that the tremble probability $w_{i,t}$ is distributed $Beta\,(\beta_t, \gamma_t)$ over the population. The Beta distribution is the most natural candidate to represent the distribution of probabilities that are framed within the interval $[0, 1]$. It is a pretty flexible distribution whose shape is determined by two parameters that we allow to depend on time, $\beta_t > 0$ and $\gamma_t > 0$. We assume the following

[15] To characterize the behavior of conditional cooperators, we could have used either a utility function *à la* Fehr and Smith (1999) or a different rule for $Y_{j,t}$ like, e.g., the final (one-period-ahead) expected contribution. We opted for our simple rule for three reasons: (a) finding the functional form that fits the data best is not one of the objectives of this paper; (b) we wanted for the conditional cooperator type a behavioral rule as straightforward as those used for the other two types; (c) finally, but most importantly, our data analysis suggests that about 50% of the contributions, both final and planned, comply with such a rule.

[16] Nevertheless, identification fails to achieve in the following cases: when one of the modes of i's distribution of beliefs always corresponds to $a = 0$ and i always chooses to contribute 0 (in this case, a conditional cooperator is indistinguishable from a selfish subject); when one of the modes of i's distribution of beliefs always corresponds to the median of i's contributions and i always chooses to contribute exactly that amount (in that case, a conditional cooperator is indistinguishable from an altruist); when subjects change preferences over time.

[17] See Moffatt and Peters (2001) and Loomes (2005).

simple functional forms: $\beta_t = \exp(b_0 + b_1 (t-1))$ and $\gamma_t = \exp(g_0 + g_1 (t-1))$. With this hypothesis we want to capture the possibility that players *learn* during the game such that they are more firm in their decisions and tremble less toward the end. Hence, we expect to see the Beta distribution more and more concentrated toward zero as the game goes by.

Let us recall that the indicator function $\mathbb{1}(\cdot)$ takes the value 1 if the statement into brackets holds and 0 otherwise. Let T_s, $s \in \{final, planned\}$, indicate the last period, that is $T_{final} = 15$ and $T_{planned} = 14$, and let $f(w; \beta, \gamma)$ represent the Beta density function. Given our assumptions, we can now define the individual likelihood contribution for each subject type. For a *selfish* player, the likelihood contribution is:

$$\begin{aligned} l_i^{SEL} &= Prob\,(Y_{i,1} = 0,\ \ldots,\ Y_{i,T_s} = 0 | i = SEL) \\ &= \int_0^1 \prod_{t=1}^{T_s} \left\{ (1 - w_{i,t}) \times \mathbb{1}(Y_{i,t} = 0) + \frac{w_{i,t}}{11} \right\} f(w; \beta, \gamma)\, dw. \end{aligned} \tag{7}$$

For an *altruistic* player, the individual contribution to the likelihood is:

$$\begin{aligned} l_i^{ALT} &= Prob\,(Y_{i,1} = m_i,\ \ldots,\ Y_{i,T_s} = m_i | i = ALT) \\ &= \int_0^1 \prod_{t=1}^{T_s} \left\{ (1 - w_{i,t}) \times \mathbb{1}(Y_{i,t} = m_i) + \frac{w_{i,t}}{11} \right\} f(w; \beta, \gamma)\, dw. \end{aligned} \tag{8}$$

Finally, the likelihood contribution for a *conditional cooperator* is:

$$\begin{aligned} l_i^{CC} &= Prob\left(Y_{i,1} = mode\left(\mathbf{b}_{i,1}^{t+\tau}\right),\ \ldots,\ Y_{i,T_s} = mode\left(\mathbf{b}_{i,T_s}^{t+\tau}\right) | i = CC\right) \\ &= \int_0^1 \prod_{t=1}^{T_s} \left\{ (1 - w_{i,t}) \times \mathbb{1}\left(Y_{i,t} = mode\left(\mathbf{b}_{i,t}^{t+\tau}\right)\right) + \frac{w_{i,t}}{11} \right\} f(w; \beta, \gamma)\, dw. \end{aligned} \tag{9}$$

The use of a mixture approach is suggested by the observation that different individuals may behave differently in a public goods game. We then proceed by assuming that a proportion π_{SEL} of the population from which the experimental sample is drawn behaves selfishly; a proportion π_{ALT} shows altruistic attitudes; and finally a proportion $\pi_{CC} = 1 - \pi_{SEL} - \pi_{ALT}$ behaves conditionally cooperative. Accordingly, the likelihood contribution of player i is:

$$L_i = \pi_{SEL} \times l_i^{SEL} + \pi_{ALT} \times l_i^{ALT} + \pi_{CC} \times l_i^{CC}. \tag{10}$$

The full sample log-likelihood for the set I of individuals is given by:

$$LogL\,(\beta, \gamma, \pi_{SEL}, \pi_{ALT}, \pi_{CC}) = \sum_{i \in I} logL_i. \tag{11}$$

5 Results

Here we present and discuss the estimates of the mixture model defined in the previous section. The model is estimated using data (choices and beliefs) from each treatment separately, discriminating between final contribution data and planned contribution data. Our samples I consist of 150 subjects for each of the two treatments P_{NI} and P_I, where each subject's final (planned) contribution is observed 15 (14) times. To estimate the model, we use the method of Maximum Simulated Likelihood. Integration over w (equations (7), (8) and (9)) is performed by simulation using two sequences of 100 Halton draws per subject.[18]

Table 1 displays the parameter estimates from the maximization of (11). The estimates of the mixing proportions (π_{SEL}, π_{ALT}, and π_{CC}) show that, under any treatment and sample used, the conditional cooperator type is the most common, representing about half of the population. The conditional cooperator type is followed by the selfish (altruistic) type under treatment P_{NI} (P_I); the estimated mixing proportion of selfish people ranges between 16% and 32%, and that of altruists between 16% and 38%.

Table 1 also shows estimates of the parameters in the distribution of the tremble probability, which – as explained in Section 4 – we assume to be distributed Beta. The table displays the estimated values of the two parameters characterizing the Beta distribution in the first and last periods of the game. Under each treatment, for both final and planned contributions, the effect of time is strongly significant.[19] Figure 8 shows the distributions of the tremble probabilities in the first and last periods of the game based on the estimates in Table 1. It is worth noting that, when the two parameters which characterize the Beta distribution equal

[18] Details can be found in Train (2003).

[19] In unreported analysis, we estimate the four models in Table 1 without time effects (i.e., constraining b_1 and g_1 to equal zero). Likelihood-ratio tests strongly reject the null hypothesis of no time effects (in all cases the p-values of the tests are < 0.000). The regression results of these models are available from the authors upon request. We do not report the results here for two reasons: none of the conclusions concerning the main hypothesis under investigation change when time effects are added to the mixture model; the models *with* time effects showed to be far superior on statistical grounds.

	P_I		P_{NI}	
	final	planned	final	planned
π_{SEL}	0.216	0.164	0.316	0.294
	(0.040)	(0.036)	(0.048)	(0.051)
π_{ALT}	0.326	0.382	0.157	0.186
	(0.042)	(0.045)	(0.034)	(0.036)
π_{CC}	0.458	0.454	0.527	0.519
	(0.049)	(0.052)	(0.052)	(0.058)
$\beta_1 = exp(b_0)$	0.562	0.723	1.234	0.816
	(0.120)	(0.165)	(0.306)	(0.202)
$\gamma_1 = exp(g_0)$	0.459	0.535	1.416	0.807
	(0.102)	(0.133)	(0.394)	(0.208)
$\beta_{T_s} = exp\big(b_0 + b_1(T_s - 1)\big)$	0.636	0.331	0.155	0.142
	(0.139)	(0.068)	(0.035)	(0.032)
$\gamma_{T_s} = exp\big(g_0 + g_1(T_s - 1)\big)$	1.982	0.856	0.614	0.503
	(0.562)	(0.200)	(0.155)	(0.126)
I (number of subjects)	150	150	150	150
T_s (observations per subject)	15	14	15	14
LogL	-3207.34	-3193.81	-2945.63	-2842.28

Table 1: Maximum likelihood estimates of the mixture model's parameters. The log-likelihoods are maximized using two sequences of 100 Halton draws. The mixture models are estimated separately for treatments P_I and P_{NI} and, for each treatment, separately on final and planned contributions.

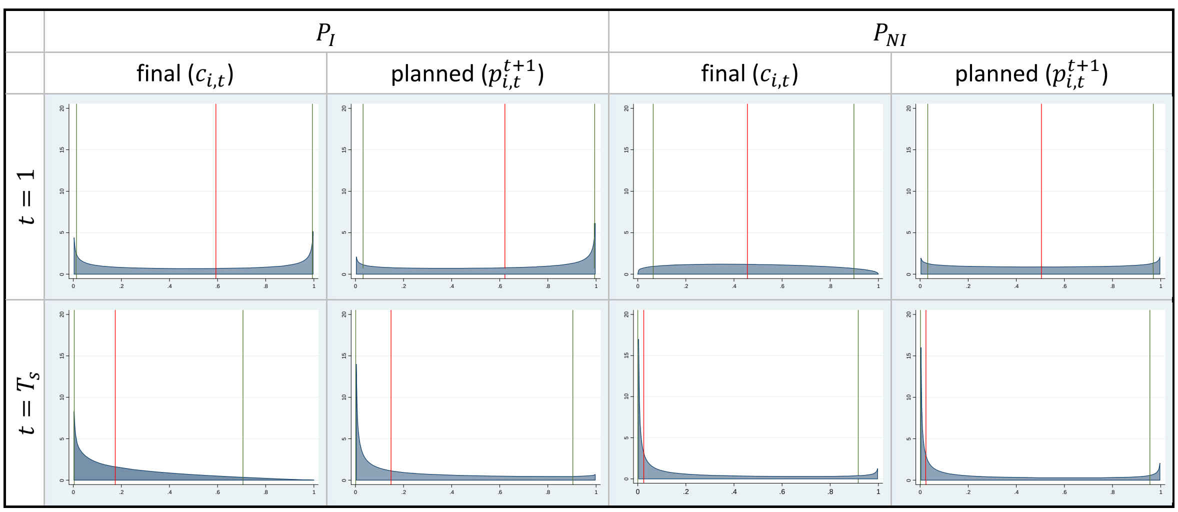

Figure 8: Distributions of the tremble probability from the estimates reported in Table (1), separately for each sample and for the first and the last period. The red vertical line indicates the median of the distribution. The two green vertical lines on the left and on the right of the red line represent the 5^{th} and the 95^{th} percentile, respectively.

one, the Beta distribution collapses to a Uniform distribution. We cannot reject the null hypothesis that the tremble probabilities are uniformly distributed in the first period of P_{NI} for both final and planned contributions. In the first period of P_I, instead, the distribution of the tremble, for both final and planned contributions, shows that more than half of the decisions are extremely noisy, being concentrated towards 1; the remainder, though, shows a rather small tremble probability.

Moving to the last period, the fact that the four bottom graphs in Figure 8 share a single pattern becomes immediately evident: all the distributions are concentrated towards zero. Such result suggests – in line with previous research – that players are rather confused about their type and the actions to take at the beginning, but they get used to the game period by period, thereby learning how to play.

A further conclusion we can get from model (11)'s estimates is that there is substantial heterogeneity across the population, confirmed by the statistically significant estimates of the mixing proportions of types and the parameters of the tremble that warrant a non-degenerate distribution. These results vindicate the reasons for choosing our approach.

The estimates reported in Table 1 allow us to verify whether there is any difference in subjects' behavior between planned and final contributions, having controlled for beliefs that bear the information about the other's action acquired at the end of each period. Our null hypothesis is that, conditional on beliefs, there is no difference between the mixing proportions estimated from planned contributions and those estimated from final contributions, against the alternative that those two distributions actually differ. As already pointed out, when plans are not disclosed (treatment P_{NI}), people might reveal different preferences in the plan dimension: on the one side, they have no strategic incentives to contribute; on the other side, they do not need to worry about being played for suckers. The Wald tests for the null hypothesis result in a $\chi^2(2) = 1.39$ in the P_I case and in a $\chi^2(2) = 0.79$ in the P_{NI} case (the p-values of these tests equal, respectively, 0.500 and 0.675). This implies that we cannot reject the consistency hypothesis under both treatments, even if we might have reasonably expected a rejection under P_{NI}.

This is quite a strong result at a population level. Yet, it does not make certain that individual players do not behave according to one type's rule when they plan and according to another type's rule in their final deed, in a sort of reshuffling-of-types process that keeps invariant the proportion of the population who are of each type. Therefore, to test the consistency hypothesis we have to go a little further and verify whether it holds at an individual level as well.

Bayes' rule comes in our aid for this purpose. This rule allows us to calculate the posterior probability of each individual of being of a certain type. Once determined each subject's type when he plans and when he contributes, we can establish whether he is consistent or not. The posterior probability for subject i of being of type $k \in \{SEL, ALT, CC\}$ is computed as follows:

$$\begin{aligned} Pr\left[i = \text{type } k \mid \text{obs}_i\right] &= \frac{Pr\left[i = \text{type } k\right] \times Pr\left[\text{obs}_i \mid i = \text{type } k\right]}{Pr\left[\text{obs}_i\right]} \\ &= \frac{\lambda_k \times Pr\left[\text{obs}_i \mid i = \text{type } k\right]}{Pr\left[\text{obs}_i\right]} = \frac{\pi_k \times l_i^k}{L_i}, \end{aligned} \tag{12}$$

where obs_i represents the observations collected from i (including both choices and stated beliefs), and l_i^k is the component of the likelihood function resulting from type k's behavior, alternatively defined by (7), (8), and (9).

The results of this exercise are shown in the graphs displayed in Figure 9. The posterior probabilities obtained using (12) combined with the estimates shown in Table 1 are represented on 2-simplexes. Each vertex of a simplex corresponds to one type. Each subject is indicated by a point in the simplex: the closer this point is to a vertex, the higher the subject's posterior probability of being of the type represented on that vertex. The size of the circles indicates the number of individuals concentrated in the same area of the simplex: the larger the circle, the higher the concentration of subjects in that area. We have rounded all posterior probabilities to the nearest 0.05, to create the graphs. The triangle inscribed in each simplex gives a measure of the strength of the model in assigning players to types. A player whose posterior probabilities result in a point situated within the inner triangle cannot be classified to any type with reasonable confidence; players located outside the inner triangle are assigned to a type with success. All the models appear highly successful at segregating subjects: most subjects are close to the vertices and very few fall in the center of the triangle. This result provides a further confirmation of the validity of our modeling approach.

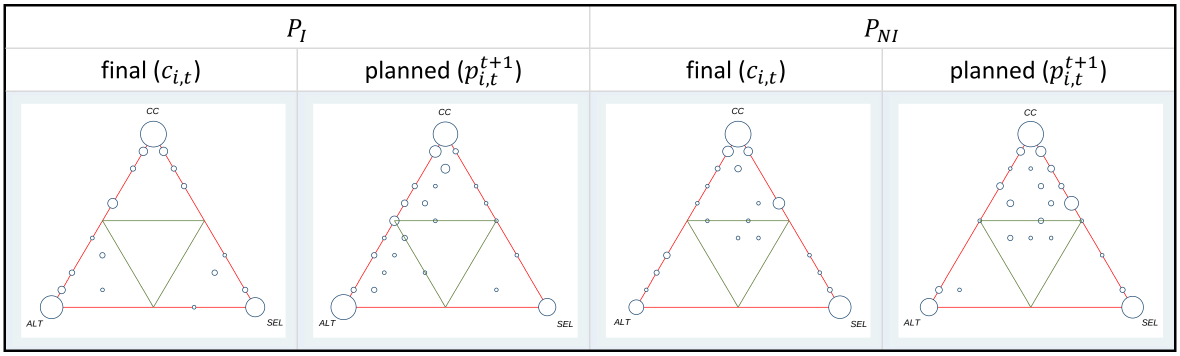

Figure 9: Posterior probabilities distribution of the three types from the estimates reported in Table 1.

In order to establish whether subjects change type between the planning and the contributing phase we resort to a cross tabulation of their types. The results are shown in a 3×3 matrix format in Table 2. Each cell reports the frequency counts of subjects who are of the type indicated by the corresponding row label when planning and of the type indicated by the corresponding column label when finally contributing. Obviously, subjects are assigned to types according to the posterior probabilities of types based on Table 1's estimates. To create the matrices of frequency counts subjects are only assigned to a type if their posterior probability of being of that type is larger than 0.5. If none of the posterior probabilities is larger than 0.5, he is not classifiable to any type, i.e., he is represented by a point in the inner triangle of the 2-simplexes. The main diagonal of the two matrices shows the frequency counts of subjects assigned to the same type with both samples. Ideally, for the consistency hypothesis to be confirmed, all the frequency counts should be arrayed within the three cells on the main diagonal.

We can see that the vast majority of subjects are, in fact, on the main diagonals (79% in P_I and 77% in P_{NI}). Respectively, 32 and 33 subjects change type between plans and final contributions in treatment P_I and treatment P_{NI}.[20] None of the subjects who are most likely to be selfish in their plans seems to be altruistic in their final contributions under both treatments. Instead, under P_{NI}, all subjects with selfish preferences when planning become conditional cooperators when finally contributing. This indicates that there exists a minority of selfish forward-looking individuals (4.17%). There are as well 10 participants (6.94%)

[20]The 3×3 matrix in the right pane (treatment P_{NI}) excludes 6 subjects who cannot be assigned to any type with reasonable confidence.

	P_I					P_{NI}				
		final ($c_{i,t}$)					final ($c_{i,t}$)			
		SEL	***ALT***	***CC***	Tot.		***SEL***	***ALT***	***CC***	Tot.
planned ($p_{i,t}^{t+1}$)	***SEL***	21	0	1	22	***SEL***	28	0	6	34
	ALT	7	37	10	54	***ALT***	3	14	7	24
	CC	4	10	60	74	***CC***	11	6	69	86
	Tot.	32	47	71	150	Tot.	42	20	82	144

Table 2: Two-way matrices of frequency counts classified according to the two categorical variables: type based on posterior probabilities from planned contribution data (rows) and type based on posterior probabilities from final contribution data (columns).

who, in P_{NI}, shift from being altruistic in their plans to being either selfish or conditionally cooperative in their actual contributions, thereby exhibiting sucker aversion.

The few changes in non-selfish preferences under P_I are more difficult to justify. They may be attributable to the incapacity of these subjects to plan ahead or to the presence of preferences that are more sophisticated than those allowed here. We speculate, however, that such inconsistencies may be caused by the larger amount of information that participants in P_I must process. Receiving feedback about both the partner's final contribution and his planned-in-the-previous period contribution may have rendered some individuals more uncertain about the others' behaviour.

To corroborate this conjecture, we look at the accuracy of beliefs in both experimental treatments. For this purpose, Figure 10 shows – separately for the two treatments – bar graphs of the proportion of times each alternative in $\mathcal{A}$ has been chosen in periods 2 and 15, and superimposes on these bar graphs connected dots representing the expected probabilities that the others choose the corresponding alternative (as computed from belief data). We refer to these expected probabilities as "expected beliefs" and denote them by $\mathrm{E}[b_a]$, $a = 0, \ldots, 10$.[21] In period 2, participants in both P_I and P_{NI} underestimate the probability that the others will choose extreme contributions (0 and 100). Inspection of the graphs for the last period

[21]Details on the derivation of the expected beliefs are reported in Appendix A.

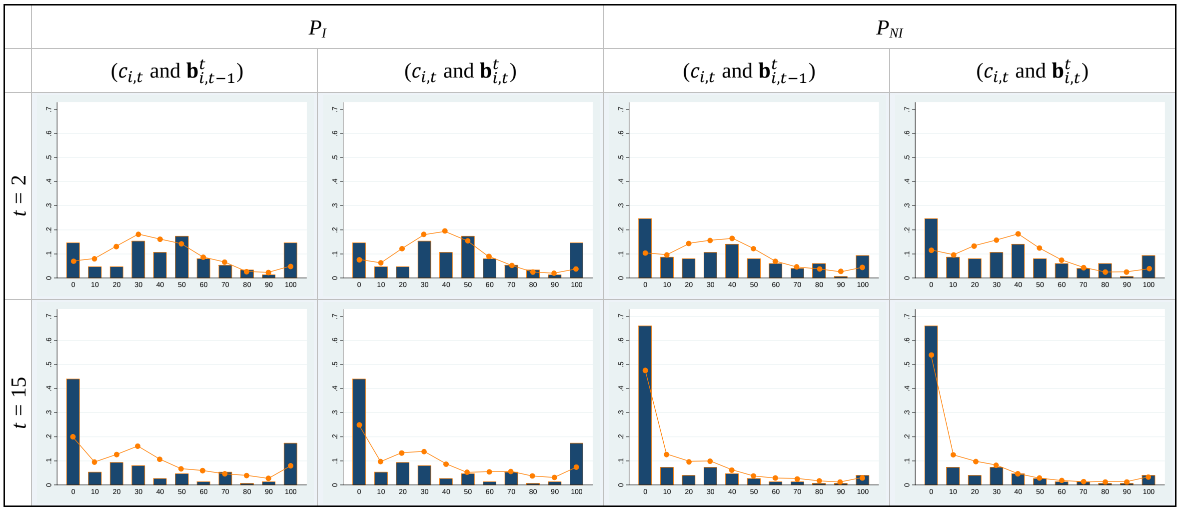

Figure 10: Comparison between contribution choices and expected beliefs. The bar height indicates the proportion of times the corresponding alternative has been chosen, calculated from final contribution data. The connected red dots represent the expected beliefs about that alternative.

demonstrates how the beliefs updating process is much slower in P_I than in P_{NI}. In fact, while beliefs in P_{NI} clearly incorporate the tendency of final contributions to move towards zero, beliefs in P_I are still unable to capture about 50% of the mass concentrated at the two extreme contributions.

The slower movements of beliefs in P_I compared to P_{NI} can also be appreciated in Figure 11, displaying the time evolution of expected beliefs about current and future contributions separately for the two treatments. Each color represents an alternative and the associations color-alternative are listed in the legend located at the bottom. The two pairs of graphs look quite different. In P_{NI}, there is a clear tendency for $\mathrm{E}[b_0]$ to grow almost steadily over time and for the other expected beliefs to reduce, exception being $\mathrm{E}[b_{10}]$ that sees a slight increase. Conversely, in P_I, the real take-off of $\mathrm{E}[b_0]$ occurs around period 10 and the other expected beliefs patterns do not show any clear modification, except for a slight decline in $\mathrm{E}[b_{40}]$ and $\mathrm{E}[b_{50}]$, and a slight increase in $\mathrm{E}[b_{20}]$. We argue that this phenomenon can be due to a problem of signal extraction: receiving feedbacks about both others' final and planned contributions might have rendered the signal concerning the others' behaviour more difficult to read.

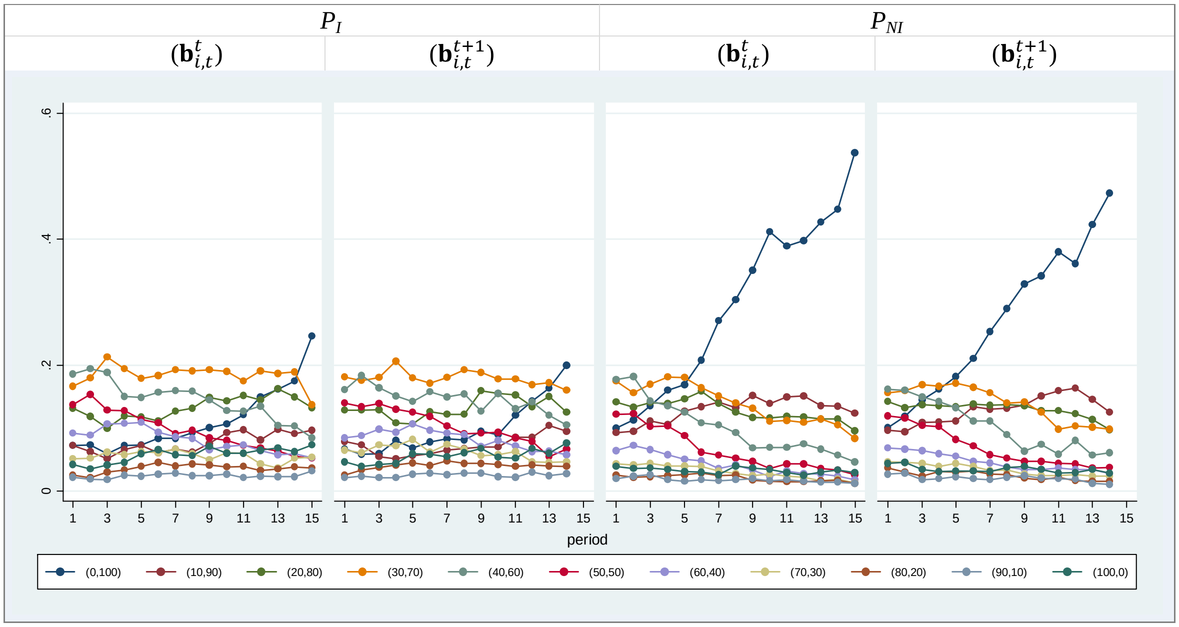

Figure 11: Time evolution of expected beliefs for each alternative. Different alternatives are represented by different colors. The legend with associations alternative-color is displayed at the bottom.

6 Conclusions

In this paper, we include plans and beliefs about the opponent's future actions into the modeling of behavior in public goods games. Whereas there exists a large body of literature that has attempted to classify subjects according to their cooperative dispositions based on their choices in various games (see, e.g., Andreoni and Miller, 2002; Charness and Rabin, 2002; Engelmann and Strobel, 2004; Cox et al., 2007; Fisman et al., 2007; Blanco et al., 2011), to our knowledge no study has measured people's preferences toward cooperation on the basis of their plans and beliefs about the other's future contributions.

We introduce an incentive-compatible way to elicit plans, and specifically design our experimental treatments to address two main questions: (1) Do planned contributions convey accurate information about people's preferences and final behavior? (2) To what extent are final contribution decisions explicable in terms of strategic forward-looking reasoning and/or sucker aversion? In addition, we test whether eliciting plans induces distortions in subjects' contribution decisions, and find that this is not the case.

We concentrate on three preference types (selfishness, altruism, and conditional cooperation), and define for each one of them a simple behavioral rule. Notwithstanding its simplicity, our modeling approach is able to accommodate the behavior of almost all experimental subjects (only 6 out of 300 participants are not classifiable to type with high posterior probability). Our data replicates previous public goods experiments' findings in two important aspects. First, our model's estimates indicate that players are heterogeneous in their preferences, and that about half of them exhibit some inclination toward conditional cooperation (see, e.g., the survey by Gächter, 2007). Second, the estimates of the tremble probabilities suggest that players – both when they plan and when they finally contribute – are initially confused about their type and the actions to take, but they get used to the game period by period so that their choices are, in the end, less noisy.[22]

Let us now briefly recap the findings on our main questions of interest. Concerning the "consistency" hypothesis, our conclusion is clear-cut: plans are predictors of future behavior if, following Manski (1990), we control for beliefs, which capture the effect of the new information acquired in the interim period between elicitation of plans and final decisions. Such a result holds both at the population level and at the individual level. We detect in fact consistency of preferences for nearly 80% of our subjects. This result is quite striking, for it implies that the three straightforward behavioral rules that we consider are able to account for an impressive amount of consistent behavior.

The finding that plans are good predictors of behavior stands against the existing empirical evidence which has frequently observed divergence between stated intentions and subsequent behavior (e.g., Barkan and Busemeyer, 1999; Hey, 2002; Bone et al., 2009). The reason for this divergence may lie in the fact that these former studies simply compare plans and final choices, discounting events that may occur between the time plans are elicited and the time behavior is determined. Barkan and Busemeyer (1999, p. 547), for instance, acknowledge that the inconsistencies they observe could be due to "the effect of actual experience on the reference point used for the evaluation of the decision problem", experience that the authors disregard.

Turning to the extent to which final contributions can be explained in terms of selfish

[22] For studies on the significant role played by errors and confusion in contribution decisions see, e.g., Andreoni (1995), Palfrey and Prisbey (1996; 1997); Anderson et al., (1998).

forward-looking reasoning and sucker-aversion, our results show that (a) there is a small number of participants (4.2%) who tries to manipulate the others' beliefs by contributing positive amounts when their choices can be observed, and (b) there are as well participants (6.9%) who withdraw observable contributions for fear of being played for a sucker.

A further interesting result is that participants in the Plan-Info treatment, who are informed at the end of each period about both the final and the planned-in-the-previous period contribution decisions of the person they are currently matched with, are more confused about the distribution of types in the population compared to participants in the Plan-NoInfo treatment, who receive periodical feedback only on final contribution decisions. We find, indeed, that stated beliefs are less accurate in the Plan-Info treatment. This indicates, in line with previous work on cognitive limitations (Costa-Gomes and Crawford, 2006; Gabaix et al., 2006) and information overload (Jacoby et al., 1974; Edmunds and Morris, 2000; Kaiser, 2011), that too much information can actually backfire.

Overall, our data corroborate past findings, but we go further because we show that people's stated plans are good predictors of subsequent behavior if one controls for events not yet realized at the time in which plans were elicited. The ultimate lesson is that researchers can and should expect much from intentions data if they treat them appropriately.

Appendix A. Estimating expected beliefs.

Here, we describe the rationale behind the econometric approach we use to estimate (from belief data) the expected probability that each alternative in $\mathcal{A}$ is being chosen – referred to as "expected beliefs". In this sense, it can be helpful to picture a subject with an urn containing colored balls. Each color corresponds to an alternative. The composition of the urn reflects the subject's beliefs concerning his partner's actions. When asked to report his beliefs, we can imagine that this subject draws 100 balls from his urn with replacement and reports the number of times each color/alternative has been drawn. At the end of each period, the composition of the urn is updated as a consequence of the new acquired information about the other's action. We allow each subject to be characterized by his own urn. More technically, the composition of any of these urns can be interpreted as a point on a 10-simplex. Each urn is located in a different point of the simplex that reflects its composition, i.e., player's beliefs.

We analyze data on beliefs with this picture in mind, and, consequently, we estimate the distribution of the different points (one for each player) on the 10-simplex. The natural choice to model a distribution over a simplex or the composition of an urn from which we observe several draws per subject (100 in the specific) is the Dirichlet-multinomial distribution.[23] This distribution is characterized by the 11-dimensional vector of parameters $\boldsymbol{\lambda} \equiv (\lambda_0, \ldots, \lambda_a, \ldots, \lambda_{10})$ and has the nice and useful property that, for each alternative, the expected probability of being selected (expected belief) is:

$$\mathrm{E}[b_a] = \frac{\lambda_a}{\sum_{\tilde{a}=0}^{10} \lambda_{\tilde{a}}}, \qquad \text{with} \quad a = 0, \ldots, 10.$$

We can think of $\mathrm{E}[b_a]$ as the composition of the average urn. We use this property to estimate, period by period, the vector of parameters $\boldsymbol{\lambda}$ and to calculate the expected beliefs.

[23]See Hausman et al. (1984) and Guimarães and Lindrooth (2007) for full details about the distribution and the model estimated here.

Appendix B. Experimental instructions (Not for publication)

In this appendix we report the instructions (originally in German) that we used for the P_I treatment. The instructions for the control and the P_{NI} treatments were adapted accordingly. They are available upon request.

Welcome! You are about to participate in an experiment funded by the Max Planck Institute of Economics. Please switch off your mobile and remain quiet.

You will receive €2.50 for showing up on time. By reading these instructions carefully you can make profitable decisions and earn more. The €2.50 show-up fee and any additional amounts of money you may earn will be paid to you in cash at the end of the experiment. Payments are carried out privately, i.e., with the others unaware of the extent of your earnings.

During the experiment we shall speak of ECUs (Experimental Currency Unit) rather than euros. The conversion rate between them is 1 ECU = 0.15 euro (i.e., for each ECU you earn, you will receive at the end of the experiment €0.15).

It is strictly forbidden to talk to the other participants. In the event of communication, the session will be terminated automatically and no payments will be made. Please raise your hand whenever you have a question and one of the experimenters will come to your aid.

Detailed information on the experiment

Group formation

The experiment lasts 15 periods. In every period you are placed in a group of two people (a pair). Pair composition is subject to variation: you will be matched with a different person in each period. There is no chance of interacting with the same participant more than once, and you will never learn the identity of the participants you are going to be matched with.

Decisions

At the beginning of each period, you (as well as the other member of your pair) receive an endowment of 100 ECUs. Then, you have to make two distinct decisions concerning two projects which are set in different time periods. More specifically, you must decide:

1. how much of the endowment that you have just received you want to contribute to a project that is taking place in the current period and how much of it you want to keep for yourself;

TABLE 1

In this period		
I contribute	I keep	
0	100	○
10	90	○
20	80	○
30	70	○
40	60	○
50	50	○
60	40	○
70	30	○
80	20	○
90	10	○
100	0	○

TABLE 2

In the next period		
I plan to contribute	I plan to keep	
0	100	○
10	90	○
20	80	○
30	70	○
40	60	○
50	50	○
60	40	○
70	30	○
80	20	○
90	10	○
100	0	○

2. how much of the endowment that you will receive at the beginning of the next period you plan to contribute to a project that will take place in the next period and how much of it you plan to keep for yourself.

Thus, your decisions in periods 1 to 14 concern the present as well as the following period. Of course, in period 15 (the last period) you are not required to provide a future plan.

The distribution of your endowment between what you contribute to the project and what you keep for yourself is limited to a set of 11 possible choices. More specifically, you indicate your decision in relation to the current period by choosing one of the 11 options shown in Table 1. Similarly, you indicate your plan for the next period by choosing one of the 11 options shown in Table 2.

Your period earnings

Your period earnings consists of two parts:

a) "Income from the project" = 0.8 × *size of the project*
(you will shortly learn more about the "size of the project");

b) "ECU you keep" = 100 − *your contribution to the size of the project.*

Thus, formally,

Your period earnings = **Income from the project** + **ECU you keep**

The "*size of the project*" is defined as follows:

- In period 1, it is the sum of contributions that you and your fellow pair member make to the (first-period) project.
- In periods 2 to 15, it can be either

(a) the sum of contributions that you and your fellow pair member make to the current-period project, or

(b) the sum of contribution plans that you and your fellow pair member made in the previous period (i.e., the plans that both of you made in the previous period about your contributions in the current period).

Possibilities (a) and (b) are equally likely: the probability that the size of the project is based on current-period contribution decisions is equal to 50%, and the probability that the size of the project is based on previous-period contribution plans is again 50%.

Example: Suppose that we are in period 2 and (following a random mechanism that will be explained shortly) the size of the project depends on the contribution plans that you and the participant you are currently matched with made in period 1. If your planned contribution in the first period was 40 ECUs and your fellow pair member's planned contribution in the same period was 30 ECUs, then the size of the project is $(40+30=)$ 70 ECUs and the income from the project amounts to $(0.8 \times 70 =)$ 56 ECUs. Furthermore, in period 1 you planned to keep for yourself $(100 - 40 =)$ 60 ECUs. It follows that your earnings in period 2 are $(56 + 60 =)$ 116 ECUs.

Behavioral predictions

In periods 1 to 14, besides setting your current-period contribution and planning your next-period contribution, you will be asked to predict the behavior of the participant with whom you are matched in the current period and the behavior of the participant with whom you will be matched in the next period. More specifically, in each period you will have to assess:

1. how likely it is that the participant you are currently matched with will choose each one of the 11 options available to him/her;
2. how likely it is that the participant you will be matched with in the next period will choose each one of the 11 options available to him/her.

In period 15 (the last period), you will have to predict only the behavior of the participant with whom you are matched in that period.

When making such predictions, you have 100 points and you must distribute them across the options available to the other member of your pair. You should assign more points to the options that you think more likely to be chosen (*original instructions included two screen-figures here*).

Example: Suppose that you were sure that the other member of your pair will contribute 30 ECUs and keep 70 ECUs for him/herself. Then, you should enter 100 in the space below that specific option and 0 in all other spaces (you enter numbers below the horizontal axis of the barplot and the corresponding bars are adjusted accordingly, see Screenshot 1). Alternatively, suppose that the above option is the most probable, but there is a 20% chance that he/she will contribute 40 (keeping 60) and a 10% chance that he/she will contribute 50 (keeping 50). In this case, you would enter 70, 20 and 10 in the spaces below options (30, 70), (40, 60), and (50, 50), respectively, and 0 in the remaining spaces (see Screenshot 2).

When you specify your predictions, you can use only integer numbers (like 0, 1, 2, 3, ...). In any case, the sum of the eleven numbers that you enter must be 100.

Notice that in periods 1 to 14 you are required to predict the contribution decisions of two different participants: the participant with whom you are matched in the current period and the participant with whom you will be matched in the next period. You *do not* have to predict the plan of the participant that is currently in the same pair with you.

Payment for your behavioral predictions

Only one of your two predictions can count for your payment, where each prediction is equally likely to be selected by a random mechanism (which will be explained shortly). Regardless of which one of the two predictions is randomly selected, your payment will be determined according to the accuracy of your predictions.

Specifically, the amount of money you will be paid depends on the difference between your predictions of the other's behavior and the actual option chosen by the other. Your payment will be higher the closer is your prediction to the "true" option chosen by the other. Likewise, your payment will depend on how well you predicted which options were *not* chosen by him/her.

The exact payment calculation will proceed as follows: for *each one* of the 11 options available to the other, we will calculate a number which reflects how well you predicted whether or not he/she would choose this option. Using these eleven numbers, we will calculate your payment.

First, we will look how well you predicted the option that was actually chosen by the other. For example, let us say that he/she actually chose to contribute 20 and to keep 80, i.e., the option (20, 80). We will compare your prediction of this option (a number between 0 and 100) with the number 100, and calculate the difference between the two. This difference will then be squared (multiplied by itself) and the resulting number will be multiplied by 0.005. Hence, if you assigned a 90% chance that the other would contribute 20 and keep 80, the resulting number will be smaller as compared to the case in which you assigned a 50% chance that the other would contribute 20 and keep 80; in fact, in the first case the square of the difference between your prediction and 100 is $(100 - 90)^2 = 100$, and in the second case it is $(100 - 50)^2 = 2500$.

On the other hand, we will also take into account how well you predicted that the remaining ten options (which were *not* chosen by the other) would not be chosen. For example, still assuming that the other chose the option (20, 80) − i.e. contributed 20 and kept 80 − this means that none of the remaining ten options were chosen. For each of these ten options, we will apply a similar procedure as we did above, for the option (20, 80). For example, for the option (10, 90) we will take your prediction of this option (a number between 0 and 100) and multiply it by itself. Again, the resulting number will be multiplied by 0.005. The same procedure will be used for all the other options that were not chosen by the other.

We will then take the eleven numbers so computed and subtract them from the number 100. This will determine the number of ECUs that you will receive for predicting the other's decision.

<u>Example 1</u>: Suppose that the other chose to contribute 30 and to keep 70 − i.e., the option (30, 70). Suppose that you assigned a 90% chance that he/she would choose this option, a 10% chance the he/she would choose the option (10, 90), and a 0% chance to the remaining nine options. Then, you would earn $100 - 0.005(0-10)^2 - 0.005(100-90)^2 = 100 - 0.5 - 0.5 = 99$ ECUs.

<u>Example 2</u>: Suppose that the other still chose the option (30, 70), but now you assigned a 50% chance to this option, a 10% chance to the option (0, 100), and a 40% chance to the option (10, 90). Then, you would receive a number of ECUs equal to $100 - 0.005(0-10)^2 - 0.005(0-40)^2 - 0.005(100-50)^2 = 100 - 0.5 - 8 - 12.5 = 79$.

These examples should illustrate that you will earn more ECUs the more accurate your predictions are.

Caution: The numbers used in all the examples were selected arbitrarily. They are NOT intended to suggest how anyone might decide.

The information you receive at the end of each period

At the end of each period, you will be informed about the decision concerning the current-period project made by your current pair member. At the end of periods 2 to 15, you will also be informed about his/her contribution plan in the previous period.

Your final payoff

Your final payoff will be calculated as follows.

At the end of the experiment one of the fifteen experimental periods will be randomly selected. This will be done by randomly drawing a ball from an urn containing 15 balls numbered 1–15. Your payoff from this randomly selected period will be determined by your decisions concerning the projects.

- If ball with number 1 is randomly drawn, you will be paid according to your current-period decision.
- If one of the remaining balls (i.e., balls with number 2–15) is randomly drawn, the "size of the project" will be determined by randomly drawing a ball from another urn containing 2 balls, one red and one blue. If the red ball is drawn, then you will be paid according to your current-period decision. If the blue ball is drawn, then you will be paid according to your previous-period plan.

Out of the remaining fourteen periods, another period will be randomly selected by drawing a second ball from the first urn. Your payoff from this second randomly selected period will be determined by your behavioral predictions.

- If ball with number 1 is randomly drawn, you will be paid according to your predictions of the participant with whom you were matched in period 1.
- If one of the remaining balls is randomly drawn, another ball will be drawn from the urn containing one red ball and one blue ball. If the red ball is drawn, then you will be paid according to your predictions of the participant with whom you were matched in the selected period. If the blue ball is drawn, then you will be paid according to your previous-period prediction of the participant with whom you were matched in the selected period.

Your payoffs in these two periods will be added up. The resulting sum will be converted to euros and paid out to you in cash.

One participant in the experiment will be randomly selected to make the four draws from the urns. The outcome of each draw will apply to all the pairs. To select the participant, one experimenter will draw a ball from an urn containing as many balls as there are participants in the experiment.

Before the experiment starts, you will have to answer some control questions to ensure your understanding of the rules of the experiment.

Please remain quietly seated during the whole experiment. If you have any questions, please raise your hand now. When you have finished reading the instructions for this part of the experiment and if there are no questions, please click "ok".

References

Ambrus, A., Pathak, P.A., 2011. Cooperation over finite horizons: A theory and experiments. Journal of Public Economics 95, 500–512.

Anderson, S.P., Goeree, J.K., 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.

Andreoni, J., Miller, J., 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. Econometrica 70, 737–753.

Bardsley, N., Moffatt, P.G., 2007. The experimetrics of public goods: Inferring motivations from contributions. Theory and Decision 62, 161–193.

Barkan, R., Busemeyer, J.R., 1999. Changing plans: Dynamic inconsistency and the effect of experience on the reference point. Psychonomic Bulletin & Review 6, 547–554.

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

Bolton, G., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity and competition. American Economic Review 90, 166–193.

Bone, J.D., Hey, J.D., Suckling, J.R., 2003. Do people plan ahead? Applied Economics Letters 10, 277–280.

Carson, R.T., Groves, T., (2007). Incentive and informational properties of preference questions. Environmental and Resource Economics 37, 181–210.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics 117, 817–869.

Costa-Gomes, M.A., Crawford, V.P., 2006. Cognition and behavior in two-person guessing games: An experimental study. American Economic Review 96, 1737–1768.

Costa-Gomes, M.A., Weizsäcker, G., 2008. Stated beliefs and play in normal-form games. Review of Economic Studies 75, 729–762.

Cox, J.C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games and Economic Behavior 59, 17–45.

Croson, R., 2000. Thinking like a game theorist: Factors affecting the frequency of equilibrium play. Journal of Economic Behavior & Organization 41, 299–314.

Croson, R., 2007. Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. Economic Inquiry 45, 199–216.

Edmunds, A., Morris, A., 2000. The problem of information overload in business organisations: A review of the literature. International Journal of Information Management 20, 17–28.

Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency and maximin preferences in simple distribution experiments. American Economic Review 94, 857–869.

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

Fischbacher, U., 2007. Zurich toolbox for readymade economic experiments. Experimental Economics 10, 171–178.

Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71, 397–404.

Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. American Economic Review 100, 541–556.

Fisman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. American Economic Review 97, 153–158.

Gabaix, X., Laibson, D., Moloche, G., Weinberg, S., 2006. Costly information acquisition: Experimental analysis if a boundedly rational model. American Economic Review 96, 1043–1068.

Gächter, S., 2007. Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. In: Frey, B.S., Stutzer, A. (Ed.), Economics and Psychology. A Promising New Cross-Disciplinary Field. MIT Press, pp. 19–50.

Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public good experiments. Experimental Economics 13, 364–377.

González, L.G., González-Farías, G., Levati, M.V., 2011. Identifying preferences for conditional cooperation using individual beliefs. Communications in Statistics - Theory and Methods 40, 3099–3118.

Greiner, B., 2004. An online recruitment system for economic experiments. In: Kremer, K., Macho, V. (Eds.), Forschung und wissenschaftliches Rechnen 2003. Ges. für Wiss. Datenverarbeitung, Göttingen, pp. 79–93.

Guimarães, P., Lindrooth, R.C., 2007. Controlling for overdispersion in grouped conditional logit models: A computationally simple application of Dirichlet-multinomial regression. Econometrics Journal 10, 439–452.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Economic models for count data with an application to the patents-R&D relationship. Econometrica 52, 909–938.

Hey, J.D., 2005. Do people (want to) plan? Scottish Journal of Political Economy 52, 122–138.

Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. American Economic Review 92, 1062–1069.

Huck, S., Weizsäcker, G., 2002. Do players correctly estimate what others do? Evidence of conservatism in beliefs. Journal of Economic Behavior & Organization 47, 71–85.

Iriberri, N., Rey-Biel, P., 2009. Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do? Department of Economics and Business, Universitat Pompeu Fabra, Economics Working Papers No. 1137.

Jacoby, J., Speller, D., Berning, C., 1974. Brand choice behavior as a function of information load: Replication and extension. The Journal of Consumer Research 1, 33–42.

Kaiser, K., 2011. Variety aversion and information overload: An experimental approach. Banco de México Working Papers No. 2011-01.

Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1, 593–622.

Loomes, G., 2005. Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. Experimental Economics 8, 301–323.

Manski, C.F., 1990. The use of intentions data to predict behavior: A best-case analysis. Journal of the American Statistical Association 85, 934–940.

Moffatt, P.G., Peters, S.A., 2001. Testing for the presence of a tremble in economic experiments. Experimental Economics 4, 221–228.

Morwitz, V.G., 1997. Why consumers don't always accurately predict their own future behavior. Marketing Letters 8, 57–70.

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70, 971–1005.

Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations, and voluntary contributions in public goods. Economic Journal 106, 817–845.

Orbell, J., Dawes, R., 1981. Social dilemmas. In: Stephenson, G.M., Davis, J.H. (Eds.), Progress in Applied Social Psychology, vol. 1. Wiley, Chichester, England, pp. 37–65.

Palfrey, T.R., Prisbrey, J.E., 1996. Altruism, reputation and noise in linear public goods experiments. Journal of Public Economics 61, 409–427.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: How much and why? American Economic Review 87, 829–846.

Rey-Biel, P., 2009. Equilibrium play and best response to (stated) beliefs in normal form games. Games and Economic Behavior 65, 572–585.

Schnake, M.E., 1991. Equity in effort: the "sucker effect" in co-acting groups. Journal of Management 17, 41–55.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.

Wilcox, N.T., Feltovich, N., 2000. Thinking like a game theorist: Comment. University of Houston Department of Economics Working Paper.