Experience-Driven Learning in Public Goods Games

Anna Conte[a,b,*], M. Vittoria Levati[a,c,†], Natalia Montinari[a,d‡]

[a] *Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany*
[b] *WBS, University of Westminster, EQM Department, 35 Marylebone Road, NW1 5LS London, UK*
[c] *University of Verona, Department of Economics, Via dell'Artigliere 19, 37129 Verona, Italy*
[d] *Lund University, Department of Economics, Lund University, 22007 Lund, Sweden*

October 14, 2013

Preliminary version. Please, do not quote without the authors' permission.

Abstract

We analyze data from a sequence of one-shot linear public goods game, where subjects are asked to report their contributions and their first-order beliefs about their partner's contribution choice. Our aim is to study the effect of past participation in economics experiments on subjects' behavior in a public goods environment. Specifically, we aim at understanding whether (and how) the subjects' familiarity with the social dilemma decision environment affects their contribution behavior and its dynamics. We allocate the effect of subjects' previous participation into two components referred to as: i) experience designating previous participation in social dilemma-type experiments (e.g. prisoner's dilemma, public goods games, etc.); ii) history designating participation in experiments of a different class than the social dilemma (e.g. individual decision making, coordination games, etc.) We have three main results. First, at the aggregate level, the experienced subjects contribute a positive amount of their endowment which is significantly smaller than the amount contributed by the inexperienced ones. Similarly, the experienced hold lower expectations about the others' contribution. Second, when comparing the distribution of the cooperative preferences between two subjects pools (i.e. looking at the relative proportion of selfish, unconditional cooperator and conditional cooperator subjects), we find significant differences. Specifically, the proportion of unconditional cooperators in the experienced subject pool decreases, with respect to the inexperienced pool, in favor of the selfish type. Finally, our data reveals that history also influences subjects' behaviour in a public goods game, but such an effect is less trenchant than in the case of experience. In the light of our results, both these factors are entitled to be properly accounted for when conducting economic experiments and documented in the experimental procedures, which should always acknowledge subjects' experience of similar games and their participation history. We postulate that this may lead to an improvement in terms of both i) external validity and ii) replicability of experimental results.

JEL classification: C35; C51; C72; H41

Keywords: Public goods experiments; Social preferences; Mixture models

*Email: aconte@econ.mpg.de; a.conte@westminster.ac.uk
†Email: levati@econ.mpg.de; vittoria.levati@univr.it
‡Email: natalia.montinari@nek.lu.se; montinari@econ.mpg.de

1 Introduction

In this paper, we study the effect of past participation in economics experiments on subjects' behavior in a public goods environment. To the extent of our knowledge, no other research has specifically analyzed whether (and how) the subjects' familiarity/knowledge with the social dilemma decision environment affects their contribution behavior and its dynamics. The unique selling point of our paper is the attempt to capture this aspect in a laboratory experiment, exploiting the subjects' previous participation in other experiments both similar and dissimilar to the public goods environment. At the same time, we take the heterogeneity of subjects' preferences to heart. As well as subjects' beliefs about others' contributions, the heterogeneity of preferences seems to explain the pattern of cooperation decay observed in almost all the finitely repeated public goods experiments (see Fischbacher and Gächter, (2010)), which has been variously interpreted as "learning" triggered off by experience within an experimental session.

Our main focus is on the study of specific forms of *learning* which take place *across* different experimental sessions. We investigate, in particular, the effect of past experience of social dilemma games as opposed to past experience of other experimental environments. In doing so, we do not neglect the form of learning that occurs *within* the same experimental session. Ledyard (1995), in his review on public goods experiments, has already shown that much more attention has been devoted to learning *within* an experimental session rather than to learning *from* previous participation in similar games. All these forms of learning are potentially relevant to account for subjects' behavior in public goods games and, in general, social dilemma environments; though, only a few studies, distinguishable into three groups, have devoted their attention to such issues.

As already mentioned, a first group of studies, focusing on learning occurring *within* an experimental session, debate about the role of learning and confusion in explaining subject's (dominated) contribution choices rather than free riding. These take the lead from Andreoni, (1988) and Palfrey and Prisbrey, (1993). More recent contributions are due to Houser and Kurzban (2002), Ferraro and Vossler (2010) and Bayer et al. (2012), among others. The results from these studies are twofold. On the one hand, they seem to converge upon the idea that learning within the experimental session only partially explains the increasing choice of dominant free riding strategies over repetitions; on the other hand, they bring attention to the relevance of (the heterogeneity of) social preferences as a complementary explanation for the incomplete decay of contribution choices (see, for example, Goeree et al. (2002) and Fischbacher and Gächter (2010)).

Subjects' previous participation has also been considered by experiments which investigate the stability of preferences. Evidence in this research area comes from repeated observations on pools of subjects who are asked to participate repeatedly to an identical public goods experiment (spaced out or not with other games) within a certain time lapse. Results are mixed: Volk et al. (2011) measure cooperation preferences and discover that they are pretty invariant at the aggregate level, but such a result is violated at the individual level. Brosig et al. (2007) find that other regarding preferences fade away eventually, while a consistent behavior is only observed in selfish subjects.

A third group of studies ponder learning *from* previous participation in other public goods-like games. Here the debate is confined to the contributions of Isaac et al. (1984) and Zelmer (2003). Specifically, Isaac et al. (1984) report an increase in free riding due to subjects' previous participation in similar experiments, although their result is not strongly statistically supported.

Zelmer (2003), in a meta-analysis on public goods games, concludes that experience lowers the level of contributions. Despite the call by Isaac et al. (1984)[1] for additional research and the small number of inconclusive studies, no systematic research relative to the impact of previous participation on behaviour in public goods games has been undertaken.

Assessing whether subjects' previous participation in a social dilemma environment may somewhat determine their subsequent behavior in a public goods experiment sheds light on a form of learning deriving from the opportunity subjects have to reflect on their past choices (and outcomes), before revisiting a similar decisional situation. We believe that this form of learning captures an essential aspect of the real-world faced by individuals, who – differently than in the laboratory – are possibly familiar and/or experienced with the decision task. Outside the laboratory, when re-experiencing a specific environment, it is likely that individuals anchor to their past experience to make decisions in the new situation.[2]

Our experiment is motivated with this aim in mind. The main variable of investigation is represented by subjects' previous participation in economics experiments. Specifically, we allocate the effect of previous participation into two components referred to as: i) *experience*, which designates previous participation in social dilemma-type experiments (e.g. prisoner's dilemma, public goods games, etc.); ii) *history*, which designates participation in experiments of a different class than the social dilemma (e.g. individual decision making, coordination games, etc.) Following a common approach in the public goods literature (see Fischbacher et al. (2001)), we identify different types of player, defined on the basis of their cooperative preferences, by means of a finite mixture model (see Bardsley and Moffatt (2007) and Conte and Levati (forthcoming)). For each type we consider (in the specific, unconditional contributors, conditional contributors and selfish), the mixture approach together with our data set – which contains information about subjects' lab background – enable us to disentangle the effect of *experience* from the one of *history* on subjects' contribution behavior, controlling for first order beliefs. The final scope of using this approach is to assess whether the behavioral change due to experience and history (if any) is driven by a change in cooperative preferences, by a change in expectations about others' behavior or by a combination of the two.

The results of our analysis about the impact of *experience* on behaviour in public goods games are adamant. First, at the aggregate level, the experienced contribute a positive amount of their endowment which is significantly smaller than the amount contributed by the inexperienced. Similarly, the experienced hold lower expectations about the others' contribution. Second, when comparing the distribution of the cooperative preferences between two subjects pools (i.e. looking at the relative proportion of selfish, unconditional cooperator and conditional cooperator subjects), we find significant differences. Specifically, the proportion of unconditional cooperators in the experienced subject pool decreases, with respect to the inexperienced pool, in favour of the other two types. Finally, our data reveals that *history* also influences subjects' behaviour in a public goods game, but such an effect is less trenchant than in the case of *experience*. The individual analysis demonstrates that the decisions of all the types of player are highly and equally noisy at

[1] "Although the qualitative results more often suggest that 'free riding' behavior increases with experience, there are no statistically supported conclusions. A design focusing on the factor of experience could reveal whether this effect is sustained under more extensive replication" (Isaac et al. (1984), p. 140-141).

[2] "The critical assumption underlying the interpretation of data from lab experiments is that the insights gained can be extrapolated to the world beyond" (Levitt and List (2007a), p. 153) and there are "many reasons to suspect that these laboratory findings might fail to generalize to real markets" (Levitt and List (2008), p. 909).

the beginning of the experimental session, no matter if they have already faced a similar situation. Such a noise reduces eventually but more pronouncedly for the experienced. As far as the accuracy of beliefs is concerned, we do not note any improvement among experienced subjects with respect to their inexperienced peers, at the beginning of the experiment. The inexperienced do not seem to make the most of the information about the other's move they get in each round, since their predictions only mildly refine in the end. Differently, after only a few rounds, the experienced's accuracy ameliorates dramatically. There appear to be no notable difference in beliefs' accuracy among the three types, regardless of the treatment.

We believe that the present study may have important implications in terms of the experimental methodology in the context of public goods games. In fact, both *experience* and *history* can be easily controlled in the recruitment of participants for experimental sessions. In the light of our results, both these factors are entitled to be properly accounted for when conducting economic experiments and documented in the experimental procedures, which should *always* acknowledge subjects' experience of similar games and their participation history. We postulate that this may lead to an improvement in terms of both i) external validity and ii) replicability of experimental results. Guala (1999) defines the set of properties (e.g. the characteristics of participants, the rules used for their recruitment, etc.) a given experimental system necessitates to consent generalisation of the results it produces. Our proposal to enrich such a set with *experience* and *history* is meant to bring elements typical of the real-word – where experience matters – in the experimental background conditions. In the light of our results, controlling the composition of subject pools in terms of history and experience becomes essential when one attempts to replicate others' results, in that differences in experience and history of participants might be able to explain inconsistencies emerging from different samples (either located in different laboratories or not).

In conclusion, we want to emphasize that ours is not an isolated pursuit. Other experimental fields have already recognised, and more thoroughly assessed, the relevance of experience. In industrial organization, it is worth mentioning the studies by Harrison et al. (1985) and Benson and Faminov (1988), and the review by Holt (1993). Specifically, Harrison et al. (1985) find that experienced subjects are much more effective monopolists than inexperienced ones. Benson and Faminow (1988) notice that, when experiments are conducted with inexperienced subjects, collusion is rarely detected. The opposite holds when experienced subjects are recruited. Also, experienced subjects seem to achieve tacit cooperation (i.e. collusion) more often than inexperienced subjects. In a *threshold* public goods game, Isaac et al. (1989) and Marwell and Ames (1980) do not observe significant differences from the comparison of subjects who acknowledge having previously taken part in similar experiments and subjects who do not. In an alternating-offer bargaining setting, Bolton (1991) finds that previous participation in similar games does not lead to more frequent (equilibrium) play based on payoff maximization. Finally, in the context of allocation games (dictator and ultimatum games, for examples), Matthey and Regner (2012)'s analysis reveals that having previously participated in experiments tends to increase the amount subjects reserve for themselves, especially if they have already a knowledge of that particular sort of experiments.

The rest of the paper is organized as follow: Section 2 presents the experimental design. In Section 3 we present our summary statistics. Section 4 ... Section 5 concludes

2 The experiment

2.1 The public goods game

The basic decision situation is a 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 so that nobody meets the same person more than once.[3] At the beginning of any 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: $\mathcal{A} \in \{(0, 100), (10, 90), \ldots, (50, 50), \ldots, (90, 10), (100, 0)\}$, 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) contribution in t.[4] In each period $t = 1, \ldots, 15$, the monetary payoff of player i (for all $i \in N$) is given by:

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

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

In every period $t = 1, \ldots, 15$, each participant i other than choosing one of the eleven alternatives in $\mathcal{A}$, $c_{i,t}$, reports a first-order belief vector $\mathbf{b}_{i,t}$, i.e., a probability distribution over the eleven possible choices of his current partner j. We ask for beliefs because crucial to the identification of a subject's type is the relationship between contributions and beliefs.[5]

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.[6] The rule is defined as follows. Let i's beliefs in period t be $\mathbf{b}_{i,t}$. Let us indicate the generic element of the belief vector by $b_{i,t}(a)$, which denotes the probability (in percentage points) that in period t subject i assigns to the event that his partner in period t chooses alternative a, that is $c_{j,t} = a \times 10$. In other words, $\mathbf{b}_{i,t} \equiv \big(b_{i,t}(0), b_{i,t}(1), \ldots, b_{i,t}(10)\big)$ with $\sum_{a=0}^{10} b_{i,t}(a) = 100$. Assume that $\hat{c}_{j,t}$ is the alternative actually chosen by subject j (i's partner) in period t. Then, subject i's payoff for

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

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

[5] Previous research in experimental economics has shown that the mere act of eliciting beliefs can affect behavior in finitely repeated public goods games (see, e.g., Croson, 2000; Gächter and Renner, 2010), although the evidence regarding the undesirable effects of beliefs elicitation procedures is far from being conclusive (e.g., Wilcox and Feltovich, 2000), and it does not concern stranger matching protocols. As all participants, irrespectively from their history and experience, state their first-order beliefs, the unintended effects of beliefs on behavior (if any) would occur apply to all our variables of interest.

[6] See Selten (1998) for an axiomatic characterization of the rule, and Offerman et al. (2009) for an experiment investigating its behavioral properties.

accuracy of predictions is:

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

where $\mathbb{1}(\cdot)$ is an indicator function taking on the value 1 if the statement in brackets is true and 0 otherwise.[7] Note that since beliefs are elicited in percentage points, they have to be divided by 100 to get probabilities.[8]

At the end of each period, participants receive feedback about the contribution decision of their current partner, $\hat{c}_{j,t}$.

2.2 Treatments and Hypotheses

We compare two (between subjects) treatments which are defined on the basis of subjects' *experience.* Depending on whether or not they had previously participated in *at least* one social dilemma experiment (i.e. another public goods or prisoner's dilemma game), according to the information stored in our database, we invite to the lab two different groups of participants: the experienced (E) and the inexperienced (I). The experiment is administered to the two groups of participants in separate sessions.[9]

Except for the experience (and history) of the participants, the two treatments are identical: subjects are faced with the same basic decision situation described in section 2.1 and at the end of the experiment they are asked to disclose their biographical data and information about their previous participations, if any, in experimental sessions.

Based on the studies focusing on the role of experience mentioned in the Introduction (Isaac et al. (1984), Ledyard (1995) and Zelmer (2003)), we state the following three hypotheses at the aggregate level.

Hypothesis 1: **Experience and Contribution Choice**.
The experienced contribute, on average, smaller amounts than the inexperienced.

Hypothesis 2: **Experience and Beliefs about Others' Contribution Choices**.
Compared to inexperienced subjects, experienced subjects expects the other participants to contribute smaller amounts.

Hypothesis 3: **Experience and Accuracy of Beliefs**.
The experienced hold more accurate beliefs about others' contributions than the inexperienced.

If, when making their decisions, subjects recall free riding by other participants and -for those who participated to a session with repeated interactions- the dynamics of contributions (and, in

[7] A similar rule has been used by, e.g., Offerman et al. (1996), 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.

[8] In the instructions, we use a verbal description of the rule and give numerical examples. Recognised 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.

[9] Details on the experimental procedures can be found in the Appendix.

particular, the decay in cooperation), then it should be the case that, compared to the inexperienced, the experienced contribute and expect the others to contribute smaller amounts. The structure of our data set enables us to test whether the effect of experience (if any) is proportional to the number of public goods experiments undertaken by the subjects, or whether, alternatively, it is determined by the mere fact of having been exposed once again to the social dilemma environment. Anyhow, if experience from previous attendances at similar experiments (whatever the number) improves subjects' understanding of the environment of interaction and its dynamics, and of other subjects' behaviour, compared to the inexperienced, we should observe that, the experienced: i) expect the other participants to contribute smaller amounts; ii) are more able than the inexperienced to predict the others' contribution behavior. If all this is true, then it should result in the substantiation of the hypotheses above from our data.

The first three hypotheses have been formulated at the aggregate level and based on the experimental findings of the few other studies which scrutinise the role of experience on public goods contribution. Several subsequent experiments, however, have documented the existence of heterogeneity in cooperative preferences and classified types of contributors (see Fischbacher et al. (2001), Bardsley and Moffatt (2007) and Conte and Levati (forthcoming)). These studies have essentially identified, in the context of social dilemmas, different types of player that can be brought back to the following three categories: selfish, unconditional cooperator and conditional cooperators. Along similar lines, we recognise the importance of dealing with individual heterogenity and, therefore, formulate specific hypotheses for any type of cooperative preferences.

Hypothesis 4a: **Selfishs**.

The proportion of selfishly-behaving subjects is larger among the experienced than among the inexperienced.

Selfish individuals simply maximise their own payoff and, as a consequence, choose the free riding (dominant) action. If the contribution observed in many public goods experiment is the result of subjects' confusion or mistakes (Andreoni (1995), Palfrey and Prisbrey (1996, 1997)) and if learning from previous experience plays a role, then we should observe an increase in the number of free riding (dominant) actions chosen by experienced subjects. If this holds, then this should also emerge from their beliefs: with experience, selfishs' expectation about others' contribution should appear more accurate.

Hypothesis 4b: **Unconditional Cooperators**.

The proportion of unconditional cooperators is smaller among experienced than inexperienced. Experienced unconditional cooperators contribute smaller amounts than inexperienced unconditional cooperators.

Unconditional contributors choose to contribute to the public good irrespectively from the others' contribution. This attitude has been frequently associated with lack of understanding of the decision situation or with altruism. In both cases, unconditional contributors are exploited by free riders, and, disregarding others' behavior, are likely to obtain low payoff when they interact with other players. For this reason, we expect that, with experience, the proportion of unconditional

cooperators decreases. If the behavior of unconditional contributors is attributable to mistakes, then learning from experience should induce such subjects to revise their choices. Alternatively, if unconditional contributors are motivated by altruism, then we may observe a reduction in the level of contributions by the unconditional cooperators who are still keen to contribute to the public good. Concerning this type's beliefs, we do not have any provisional hypothesis different from those formulated at the aggregate level (see Hypotheses 2 and 3).

Hypothesis 4c: **Conditional Cooperators**.

Experienced conditional cooperators expect lower contribution by others' and, consequently, contribute less than inexperienced conditional cooperators.

Conditional cooperators *condition* their behavior on what others do or are believed to do (Fischbacher and Gächter, 2010, p.542). With regard to the relative popularity of this type among inexperienced and experienced subjects, we are unable to formulate any prior hypothesis, neither do previous studies help in this respect. Nevertheless, the behavioral rule which characterises conditional cooperators makes their presence crucial to the decay of cooperation observed in public goods experiments. For this reason, if experienced subjects have a better understanding of the heterogeneity of cooperation preferences in the population of participants and if they have already experienced the decay of cooperation in other experiments, then they should expect lower contributions by other participants. This should, in turn, induce them to contribute smaller amounts.

3 Description of Data and Aggregate Results

3.1 Biographical information and previous participation in experiments

In this section, we compare our two groups of participants on the basis of the information provided in the post-experimental questionnaire. Out of the 420 participants, only 3 subjects in treatment E refused to provide their additional information.

In our samples, the inexperienced are aged 23.452 years (s.d. 3.887, min 18, max 65, N=210), on average; the experienced, instead, 22.807 years (s.d. 2.981, min 18, max 36, N=207). According to a chi-squared test, treatments are strongly balanced with respect to gender: females represent 52.38% of participants in sample I and 53.14% of participants in sample E ($\chi^2(1)$=0.024, p-value=0.877). Similarly, there are no significant between-treatment differences in the participants' field of study ($\chi^2(3)$=3.367, p-value=0.338).

Table 1 contains summary statistics about history and experience of our participants discriminated by treatment. The experienced have participated on average in 2.2 public goods experiments or other experiments of sort. The experienced show to have, in general, a larger number of attendances at other experiments than the inexperienced (7.4 vs. 1.9).

Table 2 shows the percentage of subjects with at least one participation in four groups of experiments different than public goods and prisoner's dilemma games. Group A includes experiments which do not involve strategic interactions, as, for example, risk elicitation experiments. Group B includes experiments inspired to the principles of the trust game and gift-exchange experiments, implemented with or without a labor market framing. Group C includes experiments classifiable within the class of dictator and ultimatum games. Finally, Group D comprises auctions,

number of lab experiments		obs	mean	std. dev.	min	max
I		210	1.871	2.689	0	14
E	*total*	207	7.415	5.342	1	36
	$\neq$ *from public goods games*	207	5.241	4.423	0	29
	public goods games only	207	2.174	1.371	1	8

Table 1: Participants' history and experience by treatment

Group of experiments:	I	E
Group A: Individual decision making	38.10%	21.74%
Group B: Trust game and labour market	31.90%	78.26%
Group C: Dictator and Ultimatum	20.48%	67.15%
Group D: Auction, Bargaining, Coordination, other experiments	31.90%	72.46%

Table 2: Percentage of subjects with at least one participation in the group of experiments as classified.

bargaining, coordination game and some other experiments which do not enter in the previous three categories. It can be noted that, except for the experiments in Group A, the percentage of experienced subjects who participated in at least one experiment of other classes is much larger than that of inexperienced subjects.

Given these between-treatment differences in subjects' history, if experience shows to have an effect, then it must be purged from the possible influence other type of experiments may have on individual behaviour. Therefore, a conclusive evidence on the determinants of the observed differences in both contributions and beliefs across treatments needs to pass through a discriminating analysis of all their plausible causes that include the participation to non-public-goods-like experiments. This will be object of investigation in Section 5.

3.2 Contributions and expected contributions

In this Section, we introduce the main characteristics of the two treatments at the aggregate level. Section 4 contains a structural analysis of contributions and beliefs at the individual level.

The following descriptive analysis can be subsumed into three results which correspond to the first three hypotheses formulated in Section 2.2.

Result 1 *On average, the experienced systematically contribute smaller amounts than the inexperienced.*

Result 2 *On average, the experienced systematically expect the other participants to contribute smaller amounts than the inexperienced.*

Result 3 *On average, the experienced's beliefs tend to be more accurate than the inexperienced's.*

For each of the two treatments, Figure 1 displays the evolution of average contributions (solid lines) as well as average expected contributions (dashed lines).[10] A first glance at the figure reveals several striking features of the data at hand: both time series of average contribution start off from

[10]With "expected contribution" we mean the amount that subject i expects his partner j to contribute in each period t. These amounts are calculated by averaging all the possible contributions, weighted for the corresponding

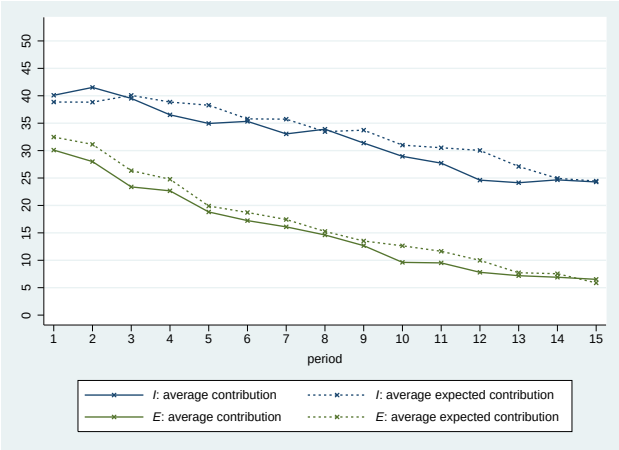

Figure 1: Average contribution ($\sum_{i=1}^{210} c_{i,t}/210$) and average expected contribution ($\sum_{i=1}^{210} \mathrm{E}_{i,t}(c_{j,t})/210$) against period, $t = 1, \ldots, 15$.

quite a high level in the first period and steadily decrease; in each treatment, average expected contributions lie mostly above average contributions, even if they remain pretty close and almost coincide in the last couple of rounds; both average contributions and expected contributions in treatment E begin at a lower level and decrease more rapidly than in treatment I.

Figure 2 magnifies the situation in period 1. It shows bar-graphs of contributions (top panes) and histograms of expected contributions (bottom panes). In both treatments, the distribution of contributions appears tri-modal, with two of the modes at 0 and 100, and the third at 40 in I and 30 in E, even if the mass at the 0-level contribution in treatment E doubles that in treatment I. According to the Kolmogorov-Smirnov test, we cannot reject the null hypothesis that contributions in period 1, $c_{i,1}$, from treatment E are smaller than those from treatment I (p-value=.001). Similarly, the distribution of expected contributions in period 1, $\mathrm{E}_{i,1}(c_{j,1})$, from sample E appears more concentrated towards the low-level contributions in treatment E than in treatment I (p-value=.001). We get exactly the same conclusions, and statistical significance, when we perform those tests using session averages of contributions and expected contributions (aggregated over all 30 players and 15 periods) as independent observation units.[11] These additional tests ensure that the differences in contributions and expected contributions between the two treatments are not confined to the first period.

The average accuracy of prediction of others' contribution, based on reported beliefs, by period and by treatment, can be looked up in Figure 3.[12] In the first four periods, both treatments share

beliefs. More in detail, expected contributions in period $t = 1, \ldots, 15$ are computed as

$$\mathrm{E}_{i,t}(c_{j,t}) = \frac{\sum_{a=0}^{10} (a \times 10) \times b_{i,t}(a)}{100}.$$

[11] Given our re-matching protocol, the numbers of statistically independent observations are 7 in both treatments.

[12] In order to assess the accuracy of beliefs, along the lines of Eq. 2, for each individual i in the two samples, we derive the following index:

$$\delta_{i,t} = \sqrt{\sum_{a=0}^{10} \left[\frac{b_{i,t}(a)}{100} - \sum_{j=1}^{210} \frac{\mathbb{1}(\hat{c}_{j,t} = a \times 10)}{210} \right]^2 \Big/ 11}.$$

It represents the square root of a quadratic deviation of subject i's beliefs from the empirical distribution of

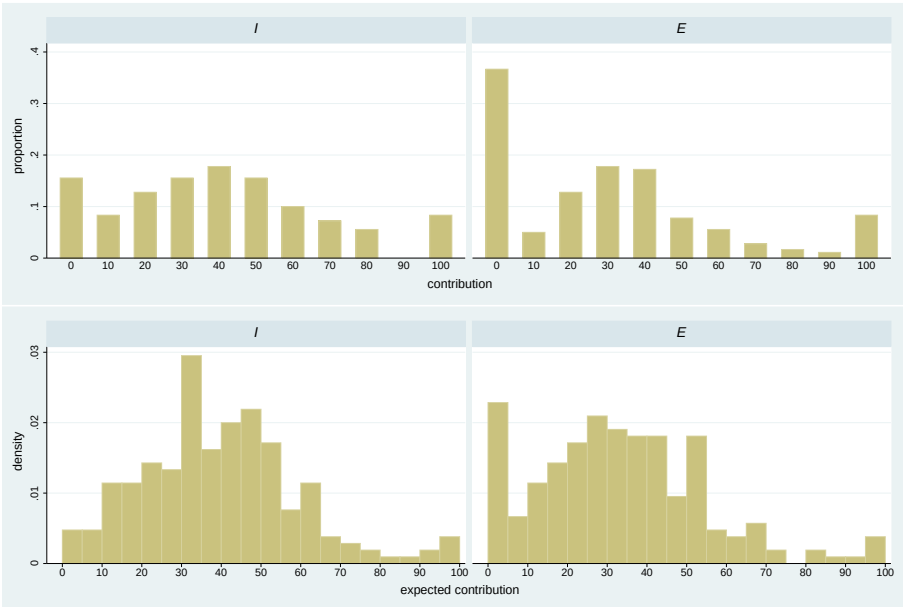

Figure 2: Top panes display bar-diagrams of period-1 contributions, $c_{i,1}$, (bar heights indicate the proportion of times the corresponding contribution is chosen); bottom panes display histograms of period-1 expected contributions, $\mathrm{E}_{i,1}(c_{j,1})$.

a similar pattern and a decreasing trend. Anyhow, starting from period 5, but more markedly from period 8 on, the experienced's beliefs become more and more accurate, so that the average accuracy index, in the last period, halves with respect to the beginning of the game. The same does not occur to the inexperienced, whose average measure of beliefs' accuracy reduces in the end but only marginally so.

Players who are already acquainted with a social dilemma situation should have apprehended that contributions tend to converge towards the dominant strategy, in the long run. In principle, then, in later occasions, these players should contribute and expect the others to contribute 0. For this reason, we might have expected to observe an even larger mass at the 0-level contribution and an improved beliefs' accuracy in the early periods of treatment E. This is not the case in our data, and there are at least three conjectures (which are *not* mutually exclusive) we can make about this finding. A first reading key lies in the so-called re-start effect, documented by Andreoni (1998) during a public goods experiment as a result of an unannounced start-afresh of the experimental session. Another possible explanation may be associated to the fact that participants in treatment E are not facing the exact environment experienced in previous experiments, so that they need some time to trace the new situation back to the already experienced one(s) and uncover their similarities. Finally, it is plausible that other-regarding preferences matter, so that, regardless of their awareness of the 0-level contribution as being the dominant strategy, subjects are still willing to contribute positively to the public good and also expect other people to do so. According to Fischbacher and Gächter (2010)'s voluntary contributions hypothesis, free riding would comes out mostly as a consequence of people's imperfect cooperative propensity. In this authors' view this is sufficient to enact the decreasing reciprocal contributions/beliefs dynamics. This hypothesis is

contributions: the lower $\delta_{i,t}$ is the closer the subject's beliefs are to such a distribution.

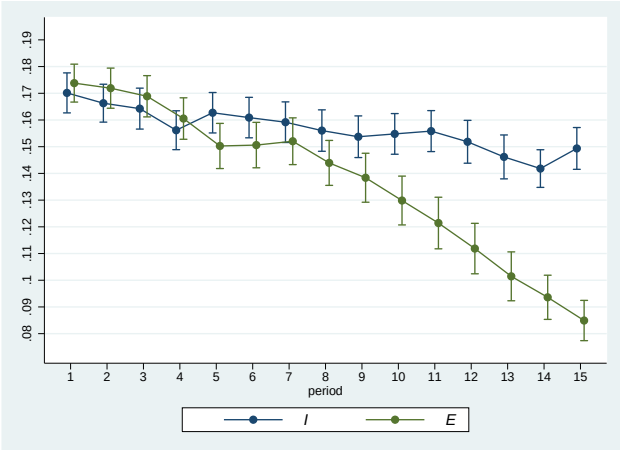

Figure 3: Average accuracy of individual beliefs distribution from the all-sample distribution of contributions by period and by treatment ($\sum_{i=1}^{210} \delta_{i,t}/210$). The range plots represent 95% confidence intervals.

supported by Fig. 1 in that it shows the experienced as willing to cooperate, but with amounts mildly smaller than those expected from the other participants, exactly as the inexperiences seem to do.

Anyhow, all these results and conjectures leave several issues open to interpretation. They do not provide sufficient clues for the reasons of the observed differences in both contributions and expected contributions across treatments. Are these explained by revisions to first-order beliefs alone induced by participants' previous experience in similar games, by a change in subjects' preferences or both is still to be determined. All this will be object of investigation in the following sections.

4 The mixture assumption

In this Section, we introduce a mixture model. The finite mixture approach adopted here pools data over subjects and allows them to be of different types, so dealing with subjects' heterogeneity. This point is taken care of by assuming that each subject is of one type, and that he does not change type throughout the experiment, and by estimating the proportions of the population who are of each type, termed the "mixing proportions". The point of view provided by an analysis of our data per type of player, conditional on first-order beliefs, should be able to clarify: i) whether the causes of the observed differences in contributions between treatments can be solely imputed to a change in beliefs (in that case, we should obtain a similar distribution of preferences from the two samples); ii) whether a leading role is played by a change in subjects' preferences (in that case, we should obtain a dissimilar distribution of preferences from the two samples); iii) whether there is room for both these explanations (and, in that case, we will try to extricate the different effects experience has on preferences and beliefs).

A common practice in the analysis of public goods game data is to consider selfish agents and non-selfish agents, distinguished in unconditional cooperators and conditional cooperators.

For each of these types, we have to define a behavioral rule with its peculiar content in terms

of preferences and beliefs.

The *selfish* type (SE)'s target is maximising his own monetary payoff. Given the payoff function (1) with a marginal per capita return smaller than one, the dominant strategy for such a type is to contribute 0. Hence, the behavior of a selfish player can be described by the following equation:

$$c_{i,t} = 0, \quad \forall t. \tag{3}$$

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 mostly attributable to lapses of concentration, distraction and confusion, or, more simply, take some time to understand which is the dominant strategy. Along similar lines to Moffatt and Peters (2001) and Loomes et al. (2002), we capture cases of sub-optimal behavior by a tremble that represents the probability that a selfish player – whatever the reason – chooses completely at random between the alternatives: $w_t^{SE} = \theta^{SE} \exp\left(\tau^{SE} \times (t-1)\right)$, $t = 1, \ldots, 15$. Here θ^{SE} represents the tremble probability of selfish players at the beginning of the experiment, while τ^{SE} represents the rate at which such a probability changes throughout the experiment. A negative τ^{SE} may be interpreted as the rate at which selfish players learn how to play their optimal strategy (contributing 0, in the specific). Given these assumptions, the individual likelihood contribution for a *selfish* player is:

$$l_i^{SEL}(\theta^{SE}, \tau^{SE}) = \prod_{t=1}^{15} \left\{ \left(1 - w_t^{SE}\right) \times \mathbb{1}\left(c_{i,t} = 0\right) + \frac{w_t^{SE}}{11} \right\}, \tag{4}$$

where the indicator function $\mathbb{1}(\cdot)$ takes the value 1, if the statement into brackets holds, and 0 otherwise.

An *unconditional cooperator* (UC) is a player who is willing to contribute positive amounts, overlooking the dominant strategy and his/her own beliefs on the others' move (see, among the others, Andreoni, (1993), Goeree et al. (2002) and Eckel et al. (2005)). Since, by definition, unconditional contributors do not anchor contributions to beliefs, there is no reason to expect that they will change their level of contribution throughout the game. Therefore, such a type of player is modelled so as to contribute according to the following rule:

$$c_{i,t} = m_i, \quad m_i > 0, \quad \forall t. \tag{5}$$

Following Bardsley and Moffatt (2007), we take m_i to equal the median of i's 15 contributions observed during the experiment. Similarly to the SE type, we allow for the possibility of sub-optimal behavior by introducing a tremble probability: $w_t^{UC} = \theta^{UC} \exp\left(\tau^{UC} \times (t-1)\right)$, $t = 1, \ldots, 15$. θ^{UC} and τ^{UC} have the same interpretation here provided for the selfish-type player. Given these assumptions, the individual likelihood contribution for an *unconditional cooperator* is:

$$l_i^{UC}(\theta^{UC}, \tau^{UC}) = \prod_{t=1}^{15} \left\{ \left(1 - w_t^{UC}\right) \times \mathbb{1}\left(c_{i,t} = m_i\right) + \frac{w_t^{UC}}{11} \right\}. \tag{6}$$

It is worth noting that neither the behavioural equation of selfish agents nor that of unconditional cooperators depend in any way on their beliefs about their partner's actions.

Differently, conditional cooperators *condition* their choices on the other' actions in the way it is explained in the next paragraphs. We assume that a *conditional cooperator* (CC) dislikes inequitable outcomes. To characterise the behavior of a CC, we build on a Fehr and Smith (1999) utility function, which depends both on subject i's payoff, $\pi_{i,t}$, and his partner j's payoff, $\pi_{j,t}$,

$$\begin{aligned} \mathrm{U}_i(c_{i,t}, c_{j,t}) &= \pi_{i,t} - \alpha_i \max\{\pi_{j,t} - \pi_{i,t}, 0\} - \beta_i \max\{\pi_{i,t} - \pi_{j,t}, 0\} \\ &= (100 - c_{i,t} + 0.8(c_{i,t} + c_{j,t})) - \alpha_i \max\{c_{i,t} - c_{j,t}, 0\} - \beta_i \max\{c_{j,t} - c_{i,t}, 0\}, \\ &\qquad \forall t. \end{aligned} \tag{7}$$

Here, α_i represents the intensity of the inequity experienced by i when he is worse off (or contributes more) than j; β_i represents the intensity of the inequity i experiences when he is better off (or contributes less) than j.

Since player i is unaware of $c_{j,t}$ when deciding on his own contributions, i's conditional choices can only be based on his first-order beliefs about j's contributions. We assume that i computes the expected utility function, based on $\mathbf{b}_{i,t}$ and (7):

$$\mathrm{EU}_i(c_{i,t}; \mathbf{b}_{i,t}) = \sum_{a=0}^{10} \mathrm{U}_i(c_{i,t}, a \times 10) \times b_{i,t}(a)/100, \qquad \forall t. \tag{8}$$

In each period, $t = 1, \ldots, 15$, subject i is asked to choose his contribution to the public good between the 11 alternatives $c_{i,t} \in \{0, 10, \ldots, 100\}$. An error term, independent between alternatives and between tasks, is added to the utility of each alternative. The i.i.d. error term, $\epsilon_{c_{i,t}}$, is taken to follow a Type I extreme value distribution, so that across the alternatives the difference between any two $\epsilon_{c_{i,t}}$ is distributed logistic. Each subject i, draws a value for α_i and a value for β_i in (7) from a bivariate lognormal distribution, and these two values apply to all tasks faced by i in the fifteen periods of the game. These assumptions in combination with the expected utility function defined in (8) give rise to the model:

$$\begin{aligned} \mathrm{V}_i(c_{i,t}; \mathbf{b}_{i,t}) &= \mathrm{EU}_i(c_{i,t}; \mathbf{b}_{i,t}) + \epsilon_{c_{i,t}} = \left[\sum_{a=0}^{10} \mathrm{U}_i(c_{i,t}, a \times 10) \times b_{i,t}(a)/100 \right] + \epsilon_{c_{i,t}}, \qquad \forall t \\ & \begin{pmatrix} \ln(\alpha_i) \\ \ln(\beta_i) \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right]. \end{aligned} \tag{9}$$

Subject i in period t chooses the alternative that maximises (9).

The likelihood contribution of subject i, choosing alternative $c_{i,t}$, $t = 1, \ldots, 15$, given that subject i is of type CC, is:[13]

[13] To estimate the model, we divided contributions by 10.

$$
\begin{aligned}
l_i^{CC}(\theta^{CC}, \tau^{CC}, \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho) &= \int_0^\infty \int_0^\infty \left\{ \prod_{t=1}^{15} \left[(1 - w_t^{CC}) \right.\right. \\
&\times \left.\left. \frac{\exp\left[\mathrm{V}_i(c_{i,t}; \mathbf{b}_{i,t})\right]}{\sum_{c \in \{0,10,\ldots,100\}} \exp\left[\mathrm{V}_i(c; \mathbf{b}_{i,t})\right]} + \frac{w_t^{CC}}{11} \right] \right\} \\
&\times f(\alpha, \beta; \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho) \mathrm{d}\alpha \mathrm{d}\beta. \qquad (10)
\end{aligned}
$$

Here, $f(\alpha, \beta; \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho)$ is the density function of the bivariate lognormal distribution evaluated at α and β, with μ_α, μ_β, σ_α, σ_β and ρ being the parameters of the underlying bivariate normal distribution. Similarly to the two previously defined types, the tremble $w_t^{CC} = \theta^{CC} \exp\left(\tau^{CC} \times (t-1)\right)$, $t = 1, \ldots, 15$, deals with sub-optimal choices.

As already noted, we allow each subject to be of one of the three types just defined. Therefore, the likelihood contribution of subject i is:

$$
\begin{aligned}
&L_i(\pi_{SE}, \pi_{UC}, \pi_{CC}, \theta^{SE}, \tau^{SE}, \theta^{UC}, \tau^{UC}, \theta^{CC}, \tau^{CC}, \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho) \\
&= \pi_{SE} \times l_i^{SE} + \pi_{UC} \times l_i^{UC} + \pi_{CC} \times l_i^{CC}, \qquad (11)
\end{aligned}
$$

where π_{SE}, π_{UC} and π_{CC} are the mixing proportions of type SE, UC and CC, respectively, which are estimated along with the parameters of models (4), (6), (10).

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

$$
(12) \quad LogL\left(\pi_{SE}, \pi_{UC}, \pi_{CC}, \theta^{SE}, \tau^{SE}, \theta^{UC}, \tau^{UC}, \theta^{CC}, \tau^{CC}, \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho\right) = \sum_{i \in \mathcal{S}} logL_i.
$$

The model is estimated using data (choices and beliefs) from each treatment separately. Our samples $\mathcal{S} \in \{I, E\}$ consist of 210 subjects (N) for each of the two treatments I and E; each subject's contribution and vector of beliefs are observed 15 times. To estimate the model, we use the method of Maximum Simulated Likelihood. The integrations in (10) are performed by two sets of Halton sequences.[14]

4.1 Mixture estimation results

A mixture model approach together with the process of conditioning on beliefs and repeated observations per subject allow us to distinguish conditional cooperators from unconditional cooperators and selfish players, and to estimate the proportion of the population who are of each type.[15]

Our findings are summarized in Result 3:

Result 3 *The distribution of types varies between treatment. Experienced players, with respect to inexperienced, are less likely to be unconditional cooperators and more likely to be both selfish*

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

[15] Identification fails in the following cases: when, given i's distribution of beliefs and eq. (9), i's optimal contribution is always $c_{i,t} = 0$ (in that case, a conditional cooperator is indistinguishable from a selfish agent); when, given i's distribution of beliefs and eq. (9), i's optimal contribution always corresponds to the median of i's observed contributions, $c_{i,t} = m_i$, (in that case, a conditional cooperator is indistinguishable from an unconditional cooperator); when subjects' preferences are not stable throughout the game.

	I	E
π_{SE}	0.139 (0.026)***	0.218 (0.039)***
π_{UC}	0.209 (0.032)***	0.063 (0.018)***
π_{CC}	0.652 (0.039)***	0.719 (0.040)***
θ^{SE}	0.427 (0.175)**	0.327 (0.155)**
τ^{SE}	-0.472 (0.209)**	-0.342 (0.169)**
θ^{UC}	0.637 (0.108)***	0.661 (0.136)***
τ^{UC}	-0.110 (0.035)***	-0.200 (0.057)***
θ^{CC}	0.120 (0.053)**	0.120 (0.070)*
τ^{CC}	-0.069 (0.050)	-0.305 (0.073)***
μ_α	0.112 (0.206)	0.824 (0.164)***
σ_α	1.246 (0.186)***	1.461 (0.176)***
μ_β	0.101 (0.162)	0.499 (0.155)***
σ_β	1.077 (0.174)***	1.052 (0.171)***
ρ	0.539 (0.121)***	0.353 (0.149)**
N (number of subjects)	210	210
T (observations per subject)	15	15
log-likelihood	-4768.33	-3098.87

Table 3: Maximum likelihood estimates of the mixture model's parameters (the log-likelihoods are maximized using two sequences of 50 Halton draws). Standard errors, in parentheses, are bootstrapped (200 replications). ***, ** and * denote significance at 1%, 5% and 10%, respectively.

and conditional cooperators.

Support to Result 3 comes from Table 3 which reports the estimation results of the mixture model described in the previous section. The mixing proportions deserve a particular attention, in view of the validation of our hypotheses. In both treatments, CC seems to be the most common type, representing 65% and 72% of the population estimated from I and E, respectively. The estimated mixing proportion of SE is 14% from sample I and 22% from sample E. Finally, the proportion of UC is estimated to be 21% and 6% from I and E, respectively. It appears that having already participated in public goods games drastically reduces the proportion of unconditional cooperators of about 15 percentage points and increases the proportion of selfish agents and conditional cooperators of 8 and 7 percentage points, respectively. Note that the composition of types estimated from treatment I reflects the one reported in Fischbacher et al. (2001), which is obtained with a different classification method based on the slope of the individual contribution schedule revealed via the strategy method. Such a study classifies 50% (22 out of 44) and 30% (13 out of 44) of subjects as conditional cooperators and as being selfish, respectively, while the remainder (6 subjects out of 44) as triangular contributors or 'others'.

Taken each mixing proportion singularly, we observe that the proportion of SE in E is significantly larger than the proportion of SE in I (z=2.047, p-value=0.020). The proportion of the population who are UC, estimated from E, is strongly significantly smaller from the estimate we get from sample I (z=-8.276, p-value=0.000). Finally, the estimated mixing proportion of CC in E shows to be statistically larger than in treatment I (z=1.673, p-value=0.047). A bootstrapped Wald test with asymptotic refinement (see Cameron and Trivedi (2005), among others) for the joint null hypothesis that the mixing proportions from E are not significantly different from the mixing proportions estimated from I strongly rejects the null ($\chi^2(2)$=75.490, p-value=0.000). Therefore, we can exclude that sample E is drawn from a population with characteristics similar to those

held by the population from which sample I is drawn.

In both treatments, selfish agents' choices are quite noisy in early rounds but sensibly less in treatment E (w_1^{SE} is about 43% in I and 33% in E) with quite a high rate of decay (signalled by the negative sign of τ^{SE}) in the following rounds, so that the probability of choosing at random already approaches zero at mid-game. Unconditional cooperators, at the beginning of the experiment, appear to be much more noisy than selfish agents (w_1^{UC} is about 65%, in both treatments). The decay rate of the tremble probability is higher in E than in I, even if not as rapid as in the selfish case. This makes UC in I still quite noisy at the end of the game (where w_{15}^{UC} is just above 13%), whilst the UC type is just moderately noisy (w_{15}^{UC} is around 4%) in E. Conditional cooperators appear to be the least noisy type –it is worth noting, though, that noise in the CC case is also captured by the additive error term in Eq. (9)– with a probability of trembling of 12% in the first round in both samples (even if, we have to stress, in treatment E, θ^{CC} is barely significant). What differs is the decay rate, which is absent in I and quite high and signficant in E. These findings imply that, in I, CC subjects keep the initial noisiness throughout the entire game, while, in E, such a noisiness completely disappears after only a few rounds.

The last five rows of Table 3 report estimation results from the CC-type model. To characterise the behaviour of a conditional cooperator, we have assumed that subjects are inequity averse. Thus, we have built on a Fehr and Smith utility function, that contains two parameters, α_i and β_i, which represent the relative importance subject i gives to distances (positive or negative) between his own and his partner's payoff (contribution). It appears that for the proportion 0.502(s.e. 0.053) of the population of I subjects and for the proportion 0.583(s.e. 0.056) of the population of E subjects, as predicted by Fehr and Smith (1999), the inequality $\alpha_i > \beta_i$ holds.[16] This is to say that being worse off (contributing more) than their partner reduces subjects' utility more than being better off (contributing less) for a little more than half of the population of both inexperienced and experienced.[17] We can also add that, as postulated by Fehr and Smith (1999), from both samples we get a positive and statistically significant correlation between α_i and β_i. Finally, the magnitude and statistical significance of σ_α and σ_β attest that there is substantial heterogeneity among conditional cooperators in both I and E.

The reduction in the proportion of unconditional cooperators among the experienced may be interpreted as a form of learning deriving from the participation to previous experimental sessions on social dilemma games. Inspired by Palfrey and Prisbey (1993), we may infer from this finding that subjects' confusion accounts for a large portion of the positive contributions observed in sample I. However, since we acknowledge an increase not only in the proportion of selfish subjects but also in the proportion of conditional cooperators, it is not obvious that, with experience, subjects are simply learning how to play their optimal strategy. As alternative to this interpretation, one may think that attendances at social dilemma-kind of experiments gain participants a more thorough sense of the dynamics of interaction and of the strategies chosen by the other subjects.

An interesting feature of these results is that noisiness characterises early choices of both inexperienced and experienced subjects, even if previous experience in similar games appears to reduce it sensibly, especially from mid-game on and regardless of the type of player. Inexperienced

[16] For details on identification and the calculation of these proportions the reader is referred to the Appendix.
[17] Blanco et al. (2011) also find that, at an individual level, the inequality $\alpha_i > \beta_i$ is often violated.

unconditional cooperators happen to be the most noisy type, as they keep on trembling until the very end of the game. We remark that even experienced subjects take time to understand how to play the optimal strategy dictated by their type's behavioural rule. As discussed in Section 3.2, we are inclined to attribute this to the so-called 're-start' effect and/or to having already faced a *similar* decision framing which does not mirror the current situation perfectly.

5 The motives of types

Combining Bayes' rule and the estimation results in Table 3, we can calculate the posterior probability of each individual in the two samples of being of each of the three considered types.For subject i, the posterior probability of being of type $k \in \{SEL, UC, CC\}$ is computed as:

$$
\begin{aligned}
pp_i^k = \Pr\left[i = \text{type } k \mid \text{obs}_i\right] &= \frac{\Pr\left[i = \text{type } k\right] \times \Pr\left[\text{obs}_i \mid i = \text{type } k\right]}{\Pr\left[\text{obs}_i\right]} \\
&= \frac{\pi_k \times \Pr\left[\text{obs}_i \mid i = \text{type } k\right]}{\Pr\left[\text{obs}_i\right]} = \frac{\pi_k \times l_i^k}{L_i}, \quad \forall k,
\end{aligned} \tag{13}
$$

where obs_i represents the observations collected from i (both contribution and stated beliefs data), and l_i^k is the component of the likelihood function resulting from type k's behavior, alternatively defined by (4), (6), and (10). In practice, π_k, l_i^k and L_i are replaced by their estimates obtained by maximizing Eq. (12). A graphical representation of the posterior probabilities from both sample is reported in the Technical Appendix.

As we said, we want to disentangle the effect of experience in public goods games from the effect of background experience in other kind of experiments on the probability of being of a certain type. Thereby, we proceed by verifying whether we can exclude that the observed differences in the mixing proportions estimated from I and E can be ascribed to factors different from subjects' previous experience of public goods games. As argued in Section 3.1, subjects who acknowledge at least one participation in a public goods game are more likely to have taken part in other experiments as well, which may have caused the observed change in preferences.

With this in mind, we deal with the posterior probabilities obtained from maximising Eq. (12) as the dependent variables in a three simultaneous equations (one for each type) model, as explained in the Appendix. Table 4 reports the marginal effects of a change in one of the regressors, say x_h, on the expected posterior probabilities, based on the estimation results of model (18). Details on the computation of marginal effects and their standard errors can be found in the Appendix. Here, we just need to know that x_h^b is the *base* value and x_h^f is the *final* value of the variable of interest, with respect to which we calculate the marginal effects on the expected posterior type-probabilties. The table reports different specifications of Eq. 18. The p-values of a likelihood-ratio test which compares alternative specifications of the model are displayed in the table. Those results reveal that experience of public goods games changes the expected posterior probability of being of type SE, UC and CC of +7.4 (p-value = 0.022), −15.3 (p-value = 0.000) and +7.9 (p-value = 0.032) percentage points, respectively.[18] These findings reproduce very closely the differences and tests' significance level between the mixing proportions resulting from Table 3. Having undergone other experiments seems to have a significant effect only on the

[18] All the p-values reported in brackets in this Section correspond to a one-sided Wald tests.

		(1)	(2)	(3)	(4)
experience of SDG	SE	0.074**	0.062*	0.056	0.062*
$x^b = 0$		(0.037)	(0.037)	(0.038)	(0.038)
$x^f = 1$	UC	-0.153***	-0.152***	-0.148***	-0.143***
		(0.029)	(0.029)	(0.029)	(0.029)
	CC	0.079*	0.090**	0.092**	0.081*
		(0.043)	(0.042)	(0.043)	(0.044)
experience of other	SE		0.039**		
experiments			(0.018)		
different from SDG	UC		-0.007		
$x^b = 0$			(0.024)		
$x^f = 1$	CC		-0.032		
			(0.036)		
experience of other	SE			0.044**	0.041**
experiments in				(0.017)	(0.018)
Group D (Tab. 2)	UC			-0.016	-0.017
$x^b = 0$				(0.019)	(0.020)
$x^f = 1$	CC			-0.028	-0.024
				(0.030)	(0.031)
gender	SE				0.049***
$x^b = 0$					(0.017)
$x^f = 1$	UC				0.004
					(0.021)
	CC				-0.053*
					(0.032)
test		L-R	L-R	L-R	L-R
spec. () vs. ()		(1) vs. (0)	(2) vs. (1)	(3) vs. (1)	(4) vs. (3)
p-value		0.0000	0.144	0.0298	0.0171

Table 4: Marginal effects of a change in one of the regressors on the expected posterior probabilities, based on model (18)' estimation results. 'SDG' stands for social dilemma games. Standard errors, in parentheses, are bootstrapped (200 replications). ***, ** and * denote significance at 1%, 5% and 10%, respectively. The bottom lines report p-values of the hypothesis test for alternative specifications of model (18). L-R indicates that a Likelihood-Ratio test is used; Specification (0) is intended as the constants-only model (not reported).

posterior probability of being of the SE type, which increases of 4 percentage points, but no clear effect on the other two types. Actually, specification (2), which includes the participation in experiments different from the public goods game, does not seem to be an improvement with respect to specification (1). Therefore, we tried to add to specification (1), one at a time, indicators of the participation in particular experiments and of groups of experiments as classified in Table 2.[19] This exercise shows that only having participated in experiments included in Group D (auctioning, bargaining, coordination and other experiments) seems to constitute an improvement with respect to specification (1). The fact that we observe an effect on the posterior type-probabilities only from such a large group of experiments makes us think that, perhaps, each experiment in that group has an effect that is too little to be detected and that emerges only when those experiments are pooled together (specifications (3) and (4)). A further improvement is represented by specification (4), which incorporates a binary indicator of gender. As a result of including such a variable, specification (4), which appears to be the most representative of the findings of

[19]Table 4 only reports the relevant results. We used several different controls, as, among the others, number of attendances at a particular type of experiment, time from the first and/or last participation in an experiment, both for public goods and other experiments, age, course of study, and so on. None of them seem to improve upon the specifications displayed in the table, but are available from the authors on request. In particular, we want to remark that a specification (not reported) alternative to specification (1) which includes the number of social dilemma experiments attended instead of a dummy accounting for having participated or not in a social dilemma experiment results in a BIC statistic larger than that produced by specification (1). Therefore, we infer from this that the number of attendances at a social dilemma experiment does not add any further detail.

this exercise, enables us to conclude that previous experience of public goods games drastically reduces the proportion of unconditional cooperators (-14.3 percentage points, p-value = 0.000), who split between selfish (+6.2 percentage points, p-value = 0.050) and conditional cooperators (+8.1 percentage points, p-value = 0.032). Having undergone one or more of the experiments included in Group D increases the percentage of selfish subjects (+4.1 percentage points, p-value = 0.011), but has no sharp effect on the other two types (UC: -1.7 percentage points, p-value = 0.191; CC: -2.4 percentage points, p-value = 0.221). Males are more likely to be selfish (+4.9 percentage points, p-value = 0.002) and less likely to be conditional cooperators (-5.3 percentage points, p-value = 0.050). There seems to be no relevant effect of gender on the probability of being unconditional cooperators (+0.4 percentage points, p-value = 0.422).

6 Aggregate Analysis by type

We can use posterior probabilities calculated according to Eq. (13) and the maximisation results of Eq. (12) as weights to calculate average contributions and average expected contributions by type as follows, respectively:

$$\overline{c_t}^k = \frac{1}{\sum_{i\in\mathcal{S}} pp_i^k} \sum_{i\in\mathcal{S}} pp_i^k \times c_{i,t}, \quad t = 1, \ldots, 15, \quad k \in \mathcal{K}; \tag{14}$$

$$\overline{\mathrm{E}_t(c_t)}^k = \frac{1}{\sum_{i\in\mathcal{S}} pp_i^k} \sum_{i\in\mathcal{S}} pp_i^k \times \mathrm{E}_{i,t}(c_{j,t}), \quad t = 1, \ldots, 15, \quad k \in \mathcal{K}. \tag{15}$$

Figure 4 displays average contributions and average expected contributions by type and by treatment so-calculated. The panes on the left(right) column represent average contributions (solid green lines) and average expected contributions (dotted green lines) for treatment $I(E)$ by type (top: selfish (SE); centre: unconditional cooperator (UC); bottom: conditional cooperators (CC)). For convenience, in each pane, a line representing the full-sample average contribution (see Fig. 1) is superimposed.

Consider first the subjects classified as selfish (SE). The experienced's and the inexperienced's average contributions follow a similar decreasing trend. As noted in Section 4.1, selfishly-behaving subjects' choices are quite noisy at the beginning of the experiment in both treatments. As a consequence, we observe that average contributions start from being positive, converge toward the 0-level contribution and stay at 0 steadily from mid-game on. With respect to average expected contributions, experienced selfish subjects match almost perfectly (with a negligible positive bias) the full-sample average contribution, while inexperienced selfish subjects tend to underestimate the full-sample average contribution level, even if both time series clearly resemble the declining trend of the full-sample average contributions. These findings give us no reason to believe that selfish subjects opt for the 0-level contribution because motivated by pessimistic expectations about the others' cooperative behavior, but they rather seem to be motivated by their own payoff maximisation. In the opposite case, we should have observed the average expected contributions from selfish subjects to match more closely their average contributions' pattern.

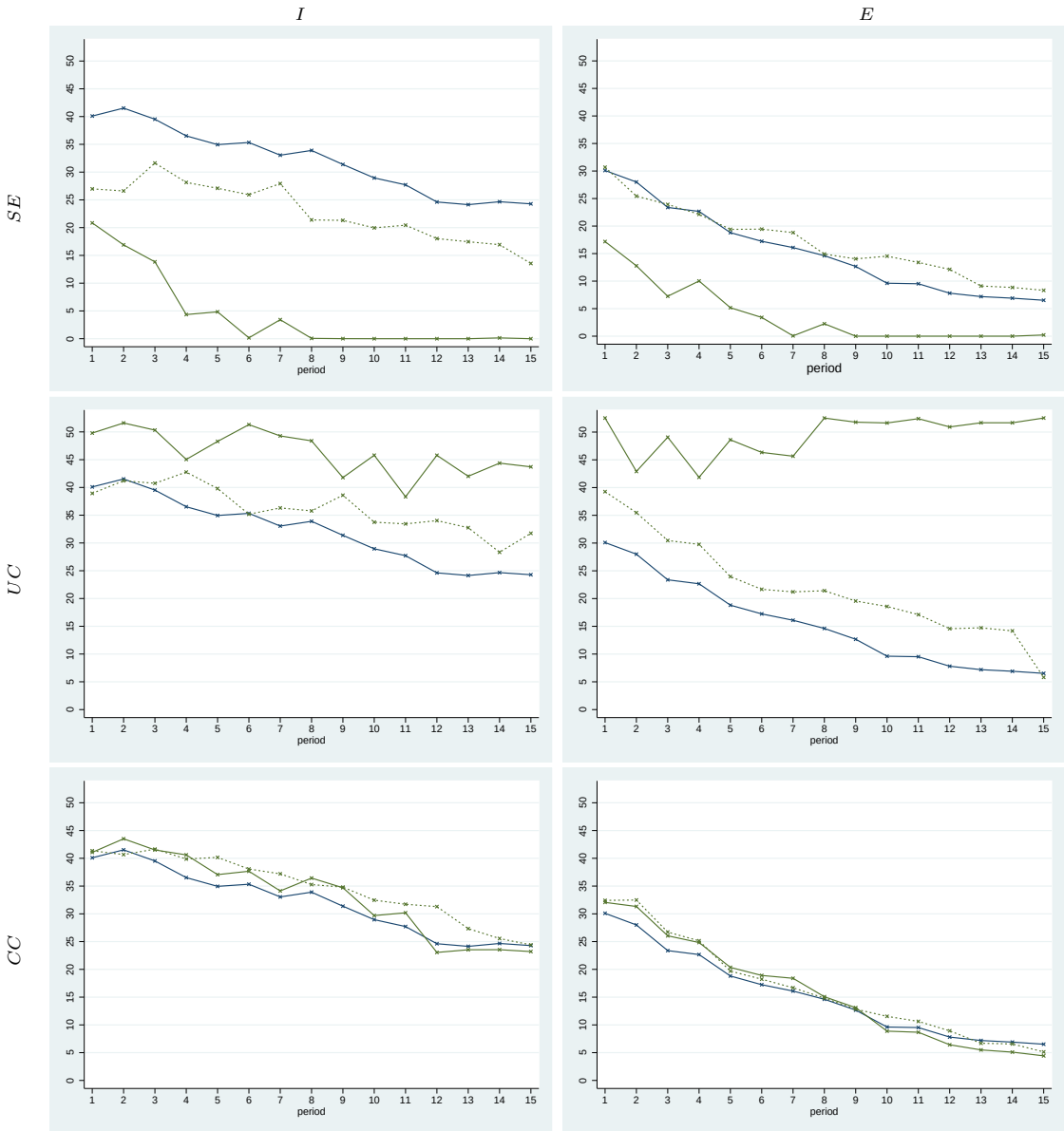

Figure 4: Average contribution (solid green line), $\overline{c_t}$, and average expected contribution (dotted green line), $\overline{\mathrm{E}_t(c_t)}$, by type and by treatment. A blue line representing the full-sample average contribution (see Fig. 1) is superimposed.

In both treatments, the average contributions from the subjects classified as unconditional cooperators (UC) lie mostly above their average expected contributions. The latter mimic quite well the full-sample average contributions in treatment I (with a mild positive bias). The experienced's beliefs, instead, appear to overestimate the average expected contributions systematically, but it has to be noted that the number of experienced unconditional cooperators is almost negligible. Note that, in treatment E, the average contributions seem more stable than in treatment I, outstandingly from mid-game on. These results again reflect the high noisiness of unconditional cooperators' decisions discussed in Section 4.1. If subjects, with experience, improve their

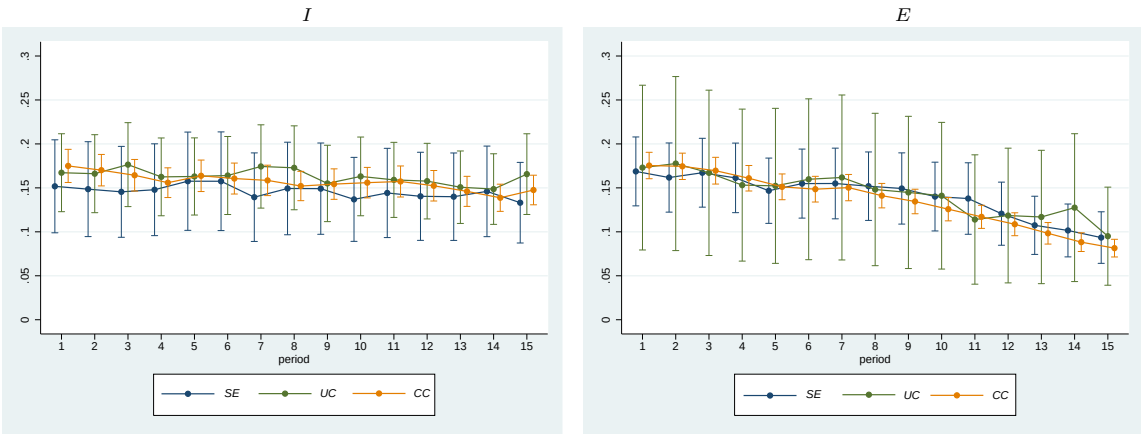

Figure 5: Average accuracy of individual belief distribution from all sample contribution distribution by period, by treatment and by type. The range plots represent 95% confidence intervals.

understanding of the dynamics of interaction, then in the E subject pool those who are classified as unconditional cooperators should not be regarded anymore as confused subjects but as truly altruists.

Finally, let us look at the subjects classified as conditional cooperators (CC). These subjects, which represent the most popular type in both subject pools, appear the most similar across treatments: irrespective of their experience, their contributions match pretty well both the expected contributions and the full-sample contributions, on average. The most noticeable between-treatment difference is in the sequences of contribution: compared to treatment I, treatment E starts lower and decline eventually following a steeper trend.

Fig. 5 represents the average accuracy (and relative 95% confidence intervals) of individual belief distribution from the all sample contribution distribution, by period, treatment and type,, calculated as follows

$$\overline{\delta_t}^k = \frac{1}{\sum_{i\in\mathcal{S}} pp_i^k} \sum_{i\in\mathcal{S}} pp_i^k \times \delta_{i,t}, \quad t = 1, \ldots, 15, \quad k \in \mathcal{K}. \tag{16}$$

In each treatment, there seem to be no remarkable difference in the accuracy of prediction among types. Selfish subjects appear the most accurate type in treatment I and conditional cooperators appear the most accurate in treatment E towards the end of the game. Anyhow, we want to stress once again that the differences among types as evidenced by our indicator are negligible. Concerning the differences between sample I and E, what noted about the accuracy of prediction from the entire samples (see Fig. (3)) still holds here at the individual level, so that all types in sample E are as accurate as all types in sample I at the beginning of the game, but happen to improve considerably their predictions after only a few rounds. The same occurs only marginally to all the types in sample I.

:-)

Appendix

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). Participants were undergraduate students from the University of Jena. Participants were recruited by the ORSEE (Greiner, 2004) software such that the samples composing treatment I were made of students who had never participated in public goods and prisoner dilemma experiments before; while participants in treatment E were recruited among those students who had previously participated in at least one public goods game experiment of sort. Upon entering the laboratory, participants were randomly assigned to visually isolated computer terminals. The instructions (reproduced in the supplement) were distributed and then read aloud to establish public knowledge. Before starting the experiment, subjects had to answer control questions which tested their comprehension of the payoff function in Eq. 1. The experiment did not start until participants had answered all the questions correctly. We can therefore safely assume that they understood the game.

Overall, we ran fourteen sessions: seven for treatment inexperienced, and seven for the experienced treatment. In each session we had 30 participants so that, in total, our analysis relies on 210 individuals observed in treatment I and 210 individuals observed in treatment E.[20]

Participants were paid according to their contributions in one randomly selected period t_1, at a rate of €0.15 per ECU, and according to the accuracy of their belief statements in another randomly selected (without replacement) period $t_2 \neq t_1$ on the basis of Eqq. 1 and 2, respectively. Sessions lasted, on average, one and a half hours with most of the time being used up for reading the instructions and answering the control questionnaire.Average earnings per subject were € (inclusive of a €2.50 show-up fee), ranging from € in treatments EI to € in treatment EI. Additional €3 were paid to those who accepted to disclose their biographical data and information about their previous participation(s) in experimental sessions.

To this end, for each participant, we prepared an envelope containing all the details about their previous pateicipations in lab experiments, as recorded in our database. At the end of the experimental session, we asked the participants who accepted to share their information to fill them in the post-experimental questionnaire. Thus, we are able to track the complete history of participation of our subjects: total number of experiments undertaken, dates, classes of experiments and some additional biographical information.

In both treatments, invited students were not told that they were going to participate in a public goods game experiment; neither, in the E treatment case, they were made aware that their partners had had previous experience of public goods game experiments or other experiments of sort.

Computation of the ratio α_i/β_i

Fehr and Smith (1999) suggest a possible range of values for α_i and β_i. With pooled data, we apparently succeed in estimating the distributions over the population of both parameters. Actually, we only succeed in estimating standard deviations and correlation coefficient of the

[20]The first six sessions of treatment I have been already analysed in Conte and Levati (forthcoming).

underlying bivariate normal distributions but not the two means.[21] In facts, we can only get an estimate of μ_α and μ_β minus the logarithm of the unknown standard deviation of the error term in the CC-type model (Eq. 9). Had these two lognormal distributions been independent, we could have obtained the distribution of the ratio α_i/β_i cleaned of unknown elements (still distributed lognormal with parameters made of a combination of the parameters of the distributions of α_i and β_i). Anyhow, we can still say something interesting about the CC model. For this purpose, we draw $100,000$ values for α_i and β_i from two (one for I and one for E) bivariate lognormal distributions having as parameters the estimates of μ_α, σ_α, μ_β, σ_β and ρ from Table (3), and calculated their ratio, α_i/β_i, getting a value of 0.503 from sample I and 0.583 from sample E.

Likewise, for each of the 200 bootstrapped samples per treatment, we calculate the ratio α_i/β_i using parameter estimates from that sample. The standard deviation of the values so-obtained constitutes the standard errors of the values of α_i/β_i obtained from the original samples. We use this procedure to calculate these standard errors, so that they reflect the sampling variation in α_i/β_i for each treatment.

A graphical representation of the posterior probabilities from Eq. 13 and estimates results in Tab. 3

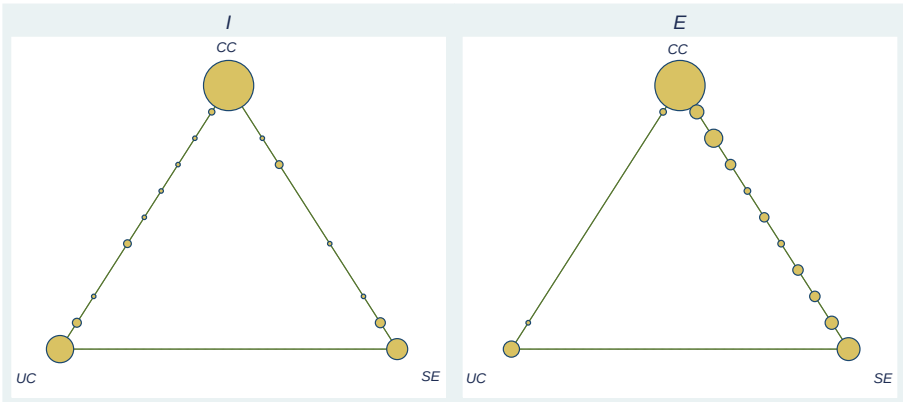

Figure 6: Posterior probabilities distribution of the three types from the models estimated in Table 3.

The posterior probabilities based on Eq. 13 and the mixture model results in Tab. 3 are displayed in Fig. 6 by means of 2-simplexes. Each vertex of the simplex represents one type (bottom-left: unconditional cooperator; bottom-right: selfish; top: conditional cooperator). Subjects are points in the simplex, with their closeness to each vertex representing their posterior type-probabilities. Small circles represent individual subjects. Larger circles represent concentrations of subjects in the same location; the larger the circle, the higher the concentration of subjects in that area of the simplex. In order to create the graphs, all posterior probabilities have been rounded to the nearest 0.05.

In the simplex from both samples, our mixture model appears to be very successful at assigning subjects to types: the vast majority of subjects, with just few exceptions, are at the vertices. In the graph representing the posterior probabilities from sample E, the fact that the vast majority of subjects are close to the top-right edge of the simplex is consistent with the very low estimate of

[21] More on this sort of identification problems can be found in Train (2003, pag. 45).

the proportion of unconditional cooperators (0.063). Segregation between the CC and SE types appears to be slightly less successful in sample E, due to the larger (than in sample I) amount of subjects who, especially from mid-game on, choose to contribute 0 and expect their partner to contribute 0 as well. Similar circumstances make a CC hardly distinguishable from a SE (see footnote 9 for limit cases of this sort). Despite these few cases, the power of our mixtures at assigning subjects to types remains very high, with 88% and 72% of subjects in I and E, respectively, assigned to types with a posterior probability larger than 0.95.

Three simultaneous equation model of posterior probabilities from Eq. 13 and computation of the marginal effects, $\frac{\Delta E(pp_i^k)}{\Delta x_h}$, and relative standard errors reported in Tab. 4

Given $\mathcal{K} \in \{SE, UC, CC\}$, let us assume that:

$$pp_i^k = \frac{\exp(\gamma_k' X_i + \eta_{ki})}{\sum_{k' \in \mathcal{K}} \exp(\gamma_{k'}' X_i + \eta_{k'i})} = \frac{\lambda_i^k \exp(\eta_{ki})}{\sum_{k' \in \mathcal{K}} \lambda_i^{k'} \exp(\eta_{k'i})}, \quad \forall k \in \mathcal{K}, \tag{17}$$

where $\lambda_i^k = \exp(\gamma_k' X_i)$. With this specification we are allowing the λ's to depend linearly on a vector of individual characteristics (X_i), including subjects' experimental background. The η_{ki}'s, $\forall k \in \mathcal{K}$, are i.i.d. individual random effects, which we assume to be distributed gamma, so that $\lambda_i^k \exp(\eta_{ki})$, $\forall k \in \mathcal{K}$, also follow a gamma distribution with parameters (λ_i^k, 1). Given these hypotheses, it happens that the vector of posterior probabilities (pp_i^{SE}, pp_i^{UC}, pp_i^{CC}), $\forall i \in S$, is distributed according to a Dirichlet (Multivariate Beta) distribution with parameters (λ_i^{SE}, λ_i^{UC}, λ_i^{CC}) and probability density function:[22]

$$g(pp_i^{SE}, pp_i^{UC}; \lambda_i^{SE}, \lambda_i^{UC}, \lambda_i^{CC}) = \frac{\Gamma\left(\sum_{k \in \mathcal{K}} \lambda_i^k\right)}{\left(\prod_{k \in \mathcal{K}} \Gamma\left(\lambda_i^k\right)\right)} \prod_{k \in \mathcal{K}} \left(pp_i^k\right)^{\lambda_i^k - 1}, \tag{18}$$

where $\Gamma(\cdot)$ is the gamma function and $pp_i^{CC} = 1 - (pp_i^{SE} + pp_i^{UC})$. We estimated this model of posterior type-probabilities, calculated as explained above, jointly from both samples by maximising the logarithmic sum of Eq. (18). Only merging the two samples, we can distinguish the effect on subjects' behaviour of the participation in public goods experiments from other background experiences and characteristics. The reason for this approach is simple. Both inexperienced and experienced subjects may have faced other type of experiments (see Table 2). We can only resolve the joint effect of experience in social dilemma games and experience in other type of experiments by an analysis conditional on the background of the subjects from both treatments.

The estimation results from Eq. (18) do not provide immediate information on the effects the change of a particular regressor, say x_h, has on the posterior type-probabilities, since x_h appears both in the numerator and in all the λ's in the denominator of Eq. (17), and it is therefore not easy to predict. For these reasons, these results are omitted, but are available from the authors on request. Nevertheless, we can use the property of the Dirichlet distribution that the expected posterior probability of each type is derivable as

[22] See Mosimann (1962) and Guimarães and Lindrooth (2007).

$$\text{(19)} \qquad \mathrm{E}\left(pp_i^k\right) = \frac{\lambda_i^k(X_i)}{\sum_{k'\in\mathcal{K}} \lambda_i^{k'}(X_i)}, \qquad \forall k \in \mathcal{K}$$

together with the estimation results from Eq. (18) to calculate the effect of a change in a regressor of interest on each expected posterior type-probability, or marginal effect, in the following way. Let us consider the effect on the posterior type-probabilities of a change in the variable x_h for subject i:

$$\text{(20)} \qquad \begin{aligned} \frac{\Delta \mathrm{E}\left(pp_i^k\right)}{\Delta x_h} &= \frac{\lambda_i^k\left(x_{1i},\ldots,x_h^f,\ldots,x_{Hi};\hat{\gamma}\right)}{\sum_{k'\in\mathcal{K}} \lambda_i^{k'}\left(x_{1i},\ldots,x_h^f,\ldots,x_{Hi};\hat{\gamma}\right)} - \frac{\lambda_i^k\left(x_{1i},\ldots,x_h^b,\ldots,x_{Hi};\hat{\gamma}\right)}{\sum_{k'\in\mathcal{K}} \lambda_i^{k'}\left(x_{1i},\ldots,x_h^b,\ldots,x_{Hi};\hat{\gamma}\right)}, \\ & \forall k \in \mathcal{K}, \end{aligned}$$

where $\hat{\gamma}$ are parameters' estimates from the maximisation across all subjects in the two samples of the logarithm of Eq. (18), x_h^b and x_h^f are the base value and the final value of the variable of interest, respectively, so that $\Delta x_h = \left(x_h^f - x_h^b\right)$ is the change in x_h with respect to which we want to calculate the change in $\mathrm{E}\left(pp_i^k\right)$, $\Delta\mathrm{E}\left(pp_i^k\right)$.

Based on this formula, we can calculate the marginal effect on the posterior type-probabilities of a change in x_h by averaging $\frac{\Delta\mathrm{E}\left(pp_i^k\right)}{\Delta x_h}$ across all i in the two samples, I and E.

To calculate the standard errors of the marginal effects so-obtained, we have to take into account the sample variation in the posterior probabilities (Eq. (17)), which have to reflect the uncertainty embodied in the estimates from the maximisation of model (12). For this reason, we follow the procedure described here for each bootstrapped sample. In other words, from each bootstrapped sample, we maximise Eq. (12) and use such parameters' estimates to calculated the posterior probabilities according to Eq. (17); we then maximise Eq. (18) and calculate the marginal effects according to Eq. (20). The standard deviations of the marginal effects so-calculated are used as standard errors of the marginal effects obtained from the original samples and reported in Table 4. Such standard errors are then used to perform the hypothesis tests in Section 5.

References

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(2), 297–323.

Andreoni, J., 1988. Why free ride? : Strategies and learning in public goods experiments, *Journal of Public Economics* 37(3), 291–304.

Andreoni, J., 1993. An Experimental Test of the Public-Goods Crowding-Out Hypothesis, *American Economic Review* 83(5), 1317–27.

Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion, *American Economic Review* 85(4), 891–904.

Andreoni, J., 1998. Toward a Theory of Charitable Fund-Raising, *Journal of Political Economy*, 106(6), 1186–1213.

Bardsley, N., Moffatt, P.G., 2007. The experimetrics of public goods: inferring motivations from contributions, *Theory and Decision* 62(2), 161–193.

Bayer, R.-C., Renner, E., Sausbruber, R., 2012. Confusion and Learning in the Voluntary Contributions Game, *Discussion Papers 2012-18*, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Benson and Faminov, 1988.

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

Bolton, 1991. A comparative model of bargaining: Theory and evidence, *American Economic Review*, 81, 1096–1136.

Brandts, J., Schram, A., 2001. Cooperation and noise in public goods experiments: applying the contribution function approach, *Journal of Public Economics* 79, 399-427.

Brosig et al., 2007.

Cameron, A.C., Trivedi, P.K., 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Conte, A., Levati, M.V., forthcoming. Use of data on planned contributions and stated beliefs in the measurement of social preferences, *Theory and Decision*.

Costa-Gomes, M.A., Weizsäcker, G., 2008. Stated beliefs and play in normal-form games, *Review of Economic Studies* 75(3), 729–762.

Croson, R., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play, *Journal of Economic Behavior & Organization* 41(3), 299–314.

Eckel, Grossman and Johnston, 2005.

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

Ferraro, P.J., Vossler, C.A., 2010. The Source and Significance of Confusion in Public Goods Experiments, *The B.E. Journal of Economic Analysis & Policy* 10(1), 1–42.

Fischbacher, U., 2007. Zurich toolbox for readymade economic experiments, *Experimental Economics* 10(2), 171–178.

Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments, *American Economic Review* 100(1), 541–556.

Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment, *Economics Letters* 71(3), 397–404.

Gächter, S., Renner, E., 2010. The effects of (incentivized) belief elicitation in public good experiments, *Experimental Economics* 13(3), 364–377.

Goeree, J., Holt, C., Laury, S., 2002. Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior, *Journal of Public Economics* 83(2), 257–278.

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, 79–93.

Guala, F., 1999. The Problem of External Validity (Or 'Parallelism') in Experimental Economics, *Social Science Information* 38, 555–73. Reprinted in *New Developments in Experimental Economics*, Vol. II, edited by E. Carbone and C. Starmer. Cheltenham: Elgar, 2007.

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(2), 439–452.

Harrison et al. (1985).

Holt 1993.

Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments, *American Economic Review* 92(4), 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(1), 71–85.

Isaac, M., Schmidtz, D., andWalker, J. (1989). The Assurance Problem in a Laboratory Market, *Public Choice* 62, 217-236.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas, 1984. Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations, *Public Choice* 43(2), 113–49.

Ledyard, J.O., 1995. Public Goods: A Survey of Experimental Research, In *Handbook of Experimental Economics*, edited by J. Kagel and A. Roth, Princeton University Press.

Levitt, S.D., List, J.A., 2007. What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives* 21(2), 153–174.

Levitt, S.D., List, J.A., 2008.

Loomes, G., Moffatt, P.G., Sugden, R., 2002. A microeconometric test of alternative stochastic theories of risky choice, *Journal of Risk and Uncertainty* 24, 103-130.

Marwell, G. and R.E. Ames, 1980, Experiments on the provision of public goods II: Provision points, stakes, experience and the free rider problem, *American Journal of Sociology* 85, 926–937.

Moffatt, P.G., Peters, S.A., 2001. Testing for the presence of a tremble in economic experiments, *Experimental Economics* 4(3), 221–228.

Mosimann, J., 1962. On the compound multinomial distribution, the multivariate beta-distribution and correlations among proportions, *Biometrika* 49, 65-82.

Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations, and voluntary contributions in public goods, *The Economic Journal* 106(437), 817–845.

Offerman, T., Sonnemans, J., Van de Kuilen, G., Wakker, P.P., 2009. A truth serum for non-Bayesians: correcting proper scoring rules for risk attitudes, *Review of Economic Studies* 76(4), 1461–1489.

Palfrey, T.R., Prisbrey, J.E., 1993.

Palfrey, T.R., Prisbrey, J.E., 1996. Altruism, reputation and noise in linear public goods experiments, *Journal of Public Economics* 61(3), 409–427.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: how much and why? *American Economic Review* 87(5), 829–846.

Matthey, A. ,Regner, T., 2012. On the independence of history: experience spill-overs between experiments, *Theory and Decision.*

Rey-Biel, P., 2009. Equilibrium play and best response to (stated) beliefs in normal form games, *Games and Economic Behavior* 65(2), 572–585.

Selten, R.,1998. Axiomatic characterization of the quadratic scoring rule, *Experimental Economics* 1(1), 43–62.

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

Volk, S., Thöni, C., Ruigrok, W., 2012. Temporal stability and psychological foundations of cooperation preferences, *Journal of Economic Behavior & Organization* 81 (2), 664-676.

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

Zelmer, J. (2003). Linear Public Goods Experiments: A Meta-Analysis, *Experimental Economics*, 6(3), 299–310.