JENA ECONOMIC RESEARCH PAPERS

2014 – 010

Experience in Public Goods Experiments

by

Anna Conte
M. Vittoria Levati
Natalia Montinari

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

Experience in Public Goods Experiments

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, P.O. Box 7082, 22007 Lund, Sweden*

Abstract

We use information on students' past participation in economic experiments, as stored in our database, to analyze whether behavior in public goods games is affected by experience (i.e., previous participation in social dilemma-type experiments) and history (i.e., participation in experiments of a different class than the social dilemma). We have three main results. First, at the aggregate level, the amount subjects contribute and expect others to contribute decrease with experience. Second, a mixture model reveals that the proportion of unconditional cooperators decreases with experience, while that of selfish individuals increases. Finally, history also influences behavior, although to a lesser extent than experience. Our findings have important methodological implications for researchers, who are urged to control for subjects' experience and history in their experiments if they want to improve the external validity and replicability of their results.

JEL classification: C35; C51; C72; H41

Keywords: Public goods experiments; Social preferences; Mixture models; Experience; History

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

1 Introduction

Participants in laboratory economic experiments are often recruited repeatedly. This gives them an opportunity to reflect on their past choices (and outcomes) before revisiting the laboratory and, consequently, to learn across different experimental sessions. In 1984, Isaac, Walker, and Thomas, in an article which is one of the most reputed and cited in the area of public goods experiments, raised the issue of whether this form of learning affects subjects' contribution behavior. Since then, no other research has specifically investigated this matter. The present study addresses this noteworthy question. In particular, we investigate whether and how contribution choices and their dynamics in public goods experiments are affected by i) previous participation in social dilemma-type experiments, which will be referred to as *experience*; and ii) previous participation in experiments different from the social dilemmas, which will be referred to as *history*.[1]

This learning-from-previous-participation process 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 reexperiencing a specific environment, it is likely that individuals fasten on their past experience when making decisions in the new situation. Therefore, a direct study of the effect of experience and history on subjects' behavior allows us to tackle, in a public goods setting, the issue raised by Smith (2010) concerning the experimentalist's interpretation of single play observations as isolated and with no precedents. According to Smith (2010), it is unwarranted to assume that the play of a specific game in the laboratory is unaffected by past experience accumulated in the world or in the laboratory. Accounting for participants' experience and history has both a direct and indirect positive effect. As a direct effect, it permits a more comprehensive interpretation of the data generated in a certain experiment; as an indirect effect, it leads to an improvement of experimental results in terms of external validity and replicability.

To delineate the present study's perspective on the significance of experience, consider two samples: one drawn from a population of students who have never faced a similar choice situation before (the inexperienced), and another drawn from a population of students who have already experienced such a situation (the experienced). If our analysis demonstrates that the two populations share similar contribution behavior, there is no reason to forcefully select out one or the other type of subjects from the subject pool of the experiment. However, if—as we believe—inexperienced and experienced subjects behave differently, drawing experimental samples jointly from the two populations can give rise to disruptive interaction effects which are especially relevant in small samples, particularly when the analysis aims to test treatment effects. Moreover, even when the samples to be compared are drawn from the same population, an additional issue raises concerns. In fact, the experienced subject pool may have been confronted with a higher number of other experiments, and probably more variegated than the inexperienced subject pool. History may matter as well as subjects' experience and determine behavioral differences even though the samples to be compared are drawn from the same population. For these reasons, being able to disentangle the effects of these two factors and to assess how they influence subjects' contribution behavior is essential. The unique point of our paper is, indeed, in the attempt to capture this

[1] We acknowledge that we use the term "history" with a meaning different from the one common in the public goods literature. History is, indeed, usually used to signify the decisions a player observes during the game (e.g., Gunnthorsdottir et al. 2007).

aspect in a laboratory experiment.

Following a common approach in the public goods literature (see, e.g., Fischbacher et al. 2001), we identify different types of player, defined on the basis of their cooperative preferences. To do so, we use a finite mixture model (see Bardsley and Moffatt 2007, and Conte and Levati 2014) considering three types of player: unconditional, conditional, and selfish contributors. The mixture approach together with our data set—which contains information about subjects' lab background—enable us to separate the effect of experience from that of history on behavior, controlling for first-order beliefs.

The ultimate scope of using this approach is to assess the existence of behavioral changes due to the effect of experience and history. In particular, compared to existing studies, the novelty of our contribution consists in establishing whether these behavioral changes (if any): i) are limited to the sphere of subjects' beliefs about the others' contribution; ii) are due to a variation in the composition of the population in terms of behavioral types; and iii) are explicable by a combination of the previous two points. Fischbacher and Gächter (2010) show that the decline in contributions in repeated public goods games can be essentially explained by a mismatch between contributions and beliefs about others' contributions. However, not much is known about how beliefs are affected by history and experience and whether groups that differ in the level of history and experience also differ in the composition of types. Our work aims at filling this gap, exploring the behavioral differences between experienced and inexperienced subjects with a special focus on the role of beliefs and player types.

Despite the early call by Isaac et al. (1984, p. 141) for additional research on "the factor of experience", little attention has been paid so far to the impact of previous participation in other experiments (both similar and dissimilar to the public goods environment) on contribution decisions. However, we are not alone in our pursuit of this question. Other experimental fields have already recognized, and thoroughly assessed, the relevance of experience. In industrial organization, it is worth mentioning the studies by Harrison et al. (1987) and Benson and Faminov (1988). Specifically, Harrison et al. (1987) 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. Moreover, experienced subjects seem to achieve tacit cooperation (i.e., collusion) more often than inexperienced subjects. In a *threshold* public goods game, Marwell and Ames (1980) and Isaac et al. (1989) do not observe significant differences when comparing subjects who have previously taken part in similar experiments and subjects who have 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 (i.e., dictator and ultimatum games), Matthey and Regner (2013)'s analysis reveals that previously participation in experiments tends to increase the amount subjects reserve for themselves, especially if they already have knowledge of that particular sort of experiments.

The rest of the paper is organized as follows. Section 2 discusses the literature that most closely relates to the present paper. Section 3 describes the experimental design, discusses the treatments implemented, and presents the hypotheses about subjects' behavior. Section 4 sets out some descriptive statistics of the samples and draws some conclusions from the data at the aggregate level. Section 5 develops the econometric model, and presents and discusses its results and relative implications. Section 6 describes an econometric model that enables to disentangle the effect of experience from that of history on the relative composition of the two samples. Section 7 reports an aggregate analysis per type of player. Section 8 concludes.

2 Related literature

In this section we briefly outline the contribution of our paper to the existing literature, discussing its novelties with respect to the studies close to the issue we are addressing.

In his review of public goods experiments, Ledyard (1995) emphasizes that much more attention has been devoted to learning within an experimental session than to learning from previous participation in similar games. To the best of our knowledge, there are only two works studying the effect of experience on contribution decisions, namely Isaac et al. (1984) and Zellmer (2003). Isaac et al. (1984) detect an increase in free riding due to subjects' previous participation in similar experiments.[2] Compared to Isaac et al. (1984), we are able to disentangle experience and history, and to study how these two factors affect cooperative preferences. In fact, while Isaac et al. (1984) control only for the participants' experience, we draw on a richer data set containing detailed information on the exact dates, the numbers, and the types of experiments previously attended by the subjects in both the experienced and the inexperienced groups. This enables us to obtain a precise estimation not only of the effect of history and experience *per se* but also of the effect associated with participation in each additional (similar and dissimilar) experiment. Moreover, unlike Isaac et al. (1984) who just look at contribution decisions, we analyze the differences between the experienced and the inexperienced participants, considering both contributions and beliefs.

The meta-analysis conducted by Zellmer (2003), based on data from 27 public goods experiments, documents a negative effect of participants' experience on average contributions.

While little research has specifically addressed the issue of experience and history in public goods experiments, a few studies, distinguishable into two groups, have investigated related issues. A first group of studies, comprising Volk et al. (2012) and Sass and Weimann (2012), examines the stability of preferences over time. Evidence in this research area comes from repeated observations on pools of subjects who are asked to participate repeatedly in an identical public goods experiment (spaced out or not with other games) within a certain time lapse. Volk et al. (2012) reinvite subjects to the lab up to four times in one-week intervals; Sass and Weimann (2012) reinvite subjects up to three times in two-and-a-half month intervals. In both studies, subjects are classified into types following Fischbacher et al. (2001)'s method. Results are mixed. Volk et al. (2012) observe that cooperation preferences are rather invariant at the aggregate level, although they are not stable at the individual level. Sass and Weimann (2012) find that other-regarding preferences fade away eventually as about one third of initially conditional cooperators turn into free riders over the course of the experiment, determining a decline in contributions over time.[3] Compared to this

[2] In this study, subjects participate in a 10-period public goods game under a partner matching protocol. Besides controlling for experience, the authors vary, in a between-subject design, the number of people in a group (4 vs. 10) and the marginal rate of return for the public good (0.3 vs. 0.75).

[3] These contradictory findings may be due to the different feedback provided to the participants: in Volk et al. (2012) feedback is given after each repetition, while in Sass and Weimann (2012) it is given only at the end, ruling out, by design, any effect of learning on the elicitation of preferences.

first group of studies, we have a different focus: our main interest is not in examining the stability of preferences but in establishing whether and how contribution behavior is affected by previous participation, not only in social dilemma experiments but also in experiments different from the social dilemmas. We are motivated to provide the experimenter with an answer about the necessity to take experience and history into account when recruiting participants.

A second group of studies focus on repetitions within an experimental session and discuss the role of learning and confusion in explaining subjects' dominated contribution choices. The literature on this issue is huge and goes back to Andreoni (1988, 1995) and Palfrey and Prisbrey (1996 1997). More recent contributions are due to Houser and Kurzban (2002), Ferraro and Vossler (2010), and Bayer et al. (2013), among others. The results from these studies are twofold. On the one hand, they seem to converge on the idea that learning within a session only partially explains the increasing choice of the dominant free-riding strategy over repetitions. On the other hand, they draw attention to the relevance of heterogeneous social preferences as a complementary explanation for the incomplete decay of contribution choices.

Among the numerous studies that have dealt with learning, confusion, and other-regarding preferences, Andreoni (1988) is especially worthy of notice. In his 1988 paper, Andreoni presents a public goods experiment where, after 10 periods of play (showing the usual decline in average contributions), participants were given a surprise announcement by the experimenter, namely that they would play some additional periods. In both the partners and the strangers conditions, Andreoni observes what is now known as restart effect: average contributions increase after the restart and then begin to decline again. Andreoni (1988)'s work is relevant to us because (i) it has a simple and straightforward design (comparable to ours) and (ii) our experiment allows us to investigate the presence of a restart effect across sessions rather than within an experimental session.

3 The experiment

3.1 The public goods game

The basic decision situation is a linear public goods game. Let $N = \{1, \dots, 30\}$ stand for a population of 30 individuals who interact in pairs for $t = 1, \dots, 15$ periods according to a perfect stranger matching design ensuring that nobody meets the same person more than once.[4] 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: $\mathcal{A} \in \{(0, 100), (10, 90), \dots, (50, 50), \dots, (90, 10), (100, 0)\}$, 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 ($a = 0, \dots, 10$) 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.[5] In each period $t = 1, \dots, 15$, the monetary payoff of

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

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

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 the relationship between contributions and beliefs is crucial to the identification of a subject's type.[6]

Beliefs are elicited by endowing participants with 100 tokens and asking them to allocate these to the 11 alternatives available to their partner. Participants are asked to allocate 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.[7] 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 attaches to the event that his partner in period t chooses alternative a, i.e., $c_{j,t} = a \times 10$. 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. Subject i's payoff for accuracy of predictions then is:

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

where $\mathbb{1}\,(\cdot)$ is an indicator function taking on the value 1 if the statement in brackets is true and 0 otherwise.[8] Note that since beliefs are elicited in percentage points, they have to be divided by 100 to obtain probabilities.[9]

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

3.2 Treatments and hypotheses

We compare two treatments which are defined on the basis of subjects' experience. Depending on whether or not they 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

[6] 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 belief elicitation procedures is far from being conclusive (e.g., Wilcox and Feltovich, 2000) and does not concern stranger matching protocols.

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

[8] A similar rule has been used, e.g., by 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 former yields more accurate beliefs.

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

database, we invite two different groups of participants to the lab: the experienced group (E) and the inexperienced group (I). The experiment is administered to the groups in separate sessions (between-subject design). Details on the experimental procedures can be found in the Appendix.

Except for the participants' experience (and history), the two treatments are identical: subjects are faced with the same basic decision situation described in Section 3.1 and, at the end of the experiment, 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; 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 the inexperienced, the experienced 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 the free-riding behavior of others, and if those who participated in sessions with repeated interactions recall the dynamics of contributions (and, in particular, their frequently observed decay), then the experienced, compared to the inexperienced, should contribute and expect the others to contribute smaller amounts. Moreover, if previous participation in similar experiments improves subjects' understanding of the environment and of the others' behavior, the experienced should hold more accurate beliefs than the inexperienced. 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 in which subjects took part or whether, alternatively, it is determined by the mere fact of their having been exposed once again to the social dilemma environment.

Several experiments have documented the existence of heterogeneity in cooperative preferences. In the context of social dilemmas, these studies have essentially identified three types of player: selfish, unconditional cooperators, and conditional cooperators (see, e.g., Burlando and Guala 2005; Kurzban and Houser 2005; Arifovic and Ledyard 2012; Conte and Levati 2014 and references therein). Along similar lines, we recognize the importance of dealing with individual heterogeneity and, therefore, formulate specific hypotheses for each type of cooperative preferences.

Hypothesis 4a: **The Selfish**
The proportion of selfishly-behaving subjects is larger among the experienced than among the inexperienced.

Selfish subjects simply maximize their own payoff and, as a consequence, choose the free-riding (dominant) action. If the contribution observed in many public goods experiments is the result of either subjects' confusion or mistakes (Andreoni 1995; Palfrey and Prisbrey 1996, 1997), and if learning from previous experience plays a role, we should observe an increase in the number of free-riding actions chosen by experienced subjects.

Hypothesis 4b: **Unconditional Cooperators**
The proportion of unconditional cooperators is smaller among the experienced than the inexperienced.

Unconditional contributors choose to contribute to the public good irrespective of the others' contributions. This attitude has been frequently attributed to a lack of understanding of the decision situation, to altruism, or efficiency (e.g., Andreoni and Miller 2002; Burlando and Guala 2005). Whatever the underlying motive, unconditional contributors are exploited by free riders and, compared to the latter, are more likely to earn a lower payoff. 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 them to revise their choices. Alternatively, if unconditional contributors are motivated by altruism or efficiency, then experience should not induce any change in their behavior.

Hypothesis 4c: **Conditional Cooperators**
Experienced conditional cooperators expect lower contribution from others' and, consequently, contribute smaller amounts than inexperienced conditional cooperators.

Conditional cooperators condition their behavior on what others do or are believed to do (Fischbacher et al. 2001). With regard to the relative popularity of this type among inexperienced and experienced subjects, we are not able to formulate any prior hypothesis, nor do previous studies help in this respect. Nevertheless, the behavior which characterizes conditional cooperators as well as the large proportion of the population that they represent make their presence crucial to the decay in contributions observed in public goods experiments (as documented by, e.g., Fischbacher and Gächter 2010). For this reason, if experienced subjects have a better understanding of the heterogeneity of preferences in the population and already experienced free riding, then they should expect lower contributions. This should, in turn, induce them to contribute smaller amounts.

Concerning the accuracy of beliefs of the three types, we cannot formulate any provisional hypotheses different from those formulated at the aggregate level (see Hypotheses 2 and 3).

4 Description of data and aggregate results

4.1 Biographical information and previous participation in experiments

In this section, we compare our two groups of participants on the basis of additional information to be provided by them in the postexperimental questionnaire. Out of the 420 participants, only 3 subjects in treatment E refused to provide this additional information. As the refusal of just 3 subjects does not make a case for sample selection bias, the analysis that follows is based only on the 207 experienced subjects who disclosed their details. Therefore, without loss of generality,

No. of lab experiments		S	Mean	Std. dev.	Min	Max
I		210	1.871	2.689	0	14
E	*total*	207	7.415	5.342	1	36
	≠ 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

the sample size, denoted as S, corresponds to 210 subjects in the case of sample I, and to 207 subjects in the case of sample E.

In our samples, the inexperienced are aged 23.452 years (s.d. 3.887, min 18, max 65, S=210), on average, and the experienced 22.807 years (s.d. 2.981, min 18, max 36, S=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 participated, on average, in 2.2 social dilemma experiments.[10] The experienced reported more participation in other experiments than the inexperienced (overall: 7.4 vs. 1.9; per year: 6.44 vs. 1.66). Moreover, when they participated in our experimental session, 214.05 days had passed from the first experimental session for the experienced, and 128.93 for the inexperienced.

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 such as, for example, risk elicitation experiments. Group B includes experiments inspired by 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, bargaining, coordination games and some other experiments which do not fall into the previous three categories. It should be noted that, except for the experiments in Group A, the percentage of experienced subjects who participated in at least one experiment of the other categories is much larger than that of inexperienced subjects.

Given these between-treatment differences in subjects' history, if experience is shown to have an effect, then it must be purged from the possible influence that other type of experiments may have on individual behavior. Therefore, a conclusive evidence on the determinants of the observed differences in both contributions and beliefs across treatments needs to be subjected to a discriminating analysis of all plausible causes including the participation in non-public-goods-like experiments. This will be the object of investigation in Section 6.

4.2 Contributions and expected contributions

In this section, we introduce the main characteristics of the two treatments at the aggregate level. Sections 5 contains a structural analysis of contributions and beliefs at the individual level.

[10]Note that, on the basis of the information provided by the ORSEE system, we are only able to observe whether a subject has taken part in a social dilemma experiment or not, but we cannot observe other characteristics of the experiment, as, for example, matching protocol, money earned, length and repetitiveness of the interactions, etc.

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

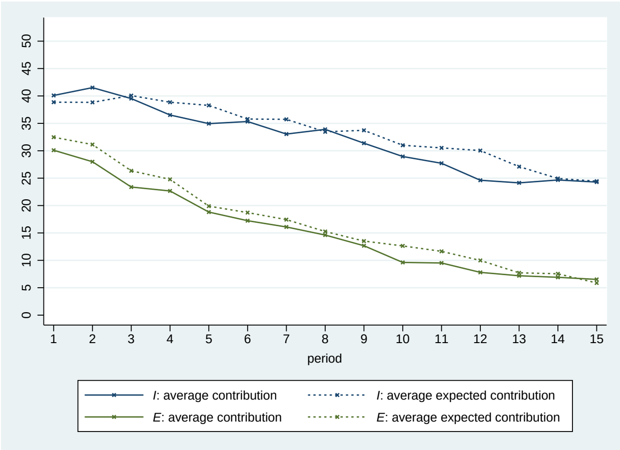

Figure 1: Average contribution $(\sum_{i=1}^{S} c_{i,t}/S)$ and average expected contribution $(\sum_{i=1}^{S} \mathrm{E}_{i,t}(c_{j,t})/S)$ against period, $t = 1, \ldots, 15$

The following descriptive analysis can be subsumed into three results which correspond to the first three hypotheses formulated in Section 3.2.

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 those of the inexperienced from the middle of the game.*

For each of the two treatments, Figure 1 displays the evolution of average contributions (solid lines) as well as average expected contributions (dashed lines).[11]

A first glance at the figure reveals several striking features of the data at hand: both time series of average contributions start from quite a high level in the first period and then steadily decrease; in each treatment, average expected contributions lie mostly above average contributions, even if they remain rather close and almost coincide in the last couple of periods; both average

[11] By "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 possible contributions, weighted for the corresponding beliefs. More exactly, expected contributions in period $t = 1, \ldots, 15$ are computed as

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

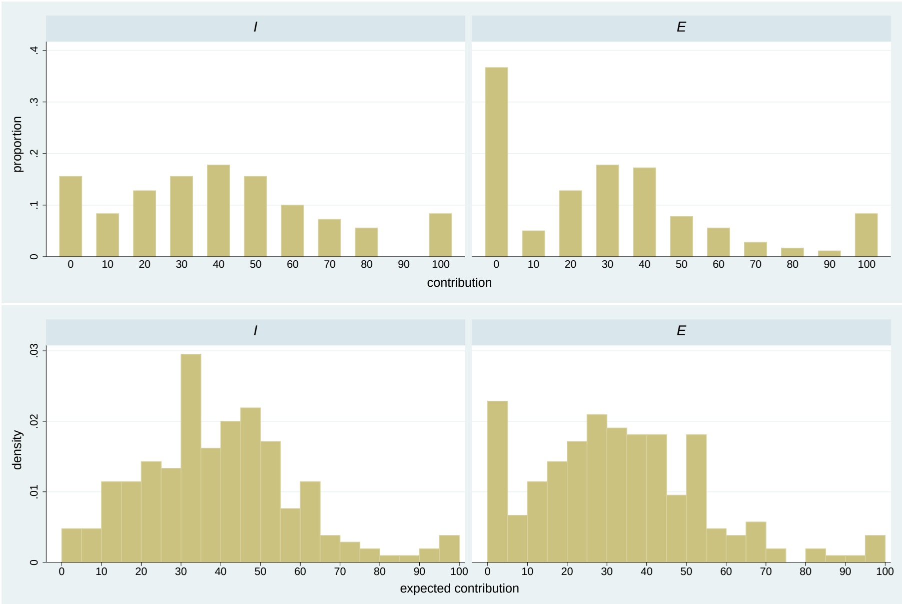

Figure 2: Top panes display bar diagrams of period-1 contributions, $c_{i,1}$, (bar heights indicate the proportion of times the corresponding contribution is chosen); bottom panes display histograms of period-1 expected contributions, $\mathrm{E}_{i,1}(c_{j,1})$

contributions and expected contributions in treatment E start 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 a Kolmogorov-Smirnov test, we cannot accept the null hypothesis that the distributions of contributions in period 1, $c_{i,1}$, from the two treatments are equal (p-value=0.001). The same result holds (p-value=0.001) for the distributions of expected contributions in period 1, $\mathrm{E}_{i,1}(c_{j,1})$. We reach exactly the same conclusions and statistical significance when we perform that test using session averages of contributions and expected contributions (aggregated over all 30 players and 15 periods) as independent observation units.[12] 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' contributions, based on reported beliefs, by period and by treatment, is displayed in Figure 3.[13] In the first four periods, both treatments share a similar pattern and a decreasing trend. Anyhow, starting from period 5, but more markedly from

[12]Given our rematching protocol, the number of statistically independent observations is 7 in both treatments.

[13]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:

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

It represents the square root of a quadratic deviation of subject i's beliefs from the empirical distribution of contributions: the lower $\delta_{i,t}$ is, the closer the subject's beliefs are to such a distribution.

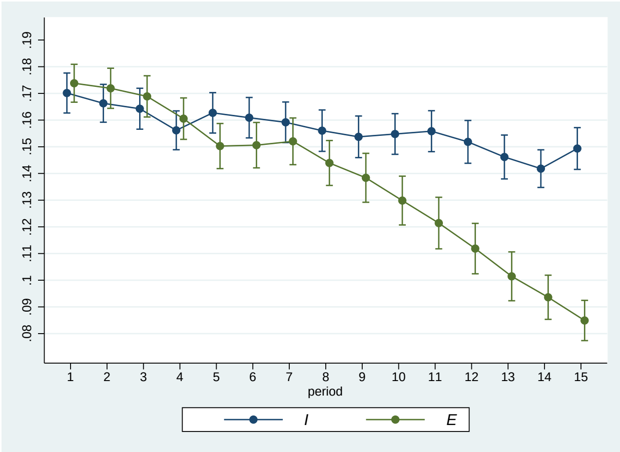

Figure 3: Average accuracy of individual beliefs distribution from the all-sample distribution of contributions by period and by treatment ($\sum_{i=1}^{S} \delta_{i,t}/S$). The range plots represent 95% confidence intervals

period 8 on, the experienced's beliefs become more and more accurate so that, in the last period, the average accuracy index halves with respect to the beginning of the game. The same does not occur to the inexperienced, whose average measure of belief accuracy reduces in the end, though only marginally.

Players who are already acquainted with a repeated social dilemma situation should have apprehended that, in the long run, contributions tend to converge toward the dominant strategy. In principle, then, on 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, where the lines representing average contributions and expected contributions from sample E seem a continuation of the lines from sample I, except for a little jump in period 1 in E with respect to period 15 in I. There are at least three conjectures (which are *not* mutually exclusive) we can make about this finding. A first key to understanding lies in the so-called restart effect, documented by Andreoni (1988) during a public goods experiment as a result of an unannounced call to start afresh during the experimental session. Another possible explanation may be associated to the fact that participants in treatment E are not facing exactly the same environment as experienced in previous experiments so that they need some time to connect the new situation back to the already experienced one(s) and recognize the similarities. Finally, it is plausible that other-regarding preferences matter so that, regardless of their awareness of 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 results mostly as a consequence of people's imperfect cooperative propensity. In these authors' view, this is sufficient to enact the decreasing reciprocal contributions/beliefs dynamics. This hypothesis is supported by Fig. 1 in that it shows the experienced as willing to cooperate, although with amounts slightly smaller than those expected from the other participants, exactly as the inexperienced seem to do.

However, all these results and conjectures leave several issues open to interpretation. They do not provide sufficient clues to 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 the relative proportion of types in the two samples, or both? Moreover, they do not say anything about the effect of history on subjects' behavior. These are still to be determined and will be the object of investigation in the following sections.

5 The mixture assumption

Factors as confusion, heterogeneity of social preferences as well as subjects' beliefs about others' contributions seem to be able to explain the pattern of decay in contributions observed in almost all finitely repeated public goods experiments (see Houser and Kurzban 2002, and Fischbacher and Gächter 2010), ours included. In this section, we introduce a mixture model to deal with subjects' heterogeneity and noisiness while controlling for their beliefs.

The finite mixture approach adopted here pools data over subjects and allows the latter to be of different types. This aspect 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 result of an analysis of our data per type of player, conditional on first-order beliefs, should clarify: i) whether the causes of the observed differences in contributions between treatments can be solely imputed to changes in beliefs (in that case, we should obtain a similar distribution of types from the two samples); ii) whether the relative proportions of types in the two populations has changed (in that case, we should obtain a dissimilar distribution of types from the two samples). In the latter case, we will try to extricate the different effects experience has on beliefs and on the popularity of a certain type within the two populations, and to distinguish the effect of subjects' experience from that of history.

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 and its specific content in terms of preferences and beliefs.

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

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

As argued, e.g., by 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, taking 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 suboptimal behavior by a tremble representing the probability that a selfish player – for whatever 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:

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

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

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

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

Following Bardsley and Moffatt (2007), we take m_i to equal the median of i's 15 contributions observed during the experiment. Similar to the SE type, we allow for the possibility of suboptimal 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} lead to the same interpretation here regarding 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 behavioral 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 others' actions in the way explained in the next paragraphs. We assume that a *conditional cooperator* (CC) dislikes inequitable outcomes. To characterize 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):

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

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 logistically. 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. In combination with the expected utility function defined in (8), these assumptions give rise to the model:

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

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

Given that subject i is of type CC, the likelihood contribution of subject i, choosing alternative $c_{i,t}$, $t = 1, \ldots, 15$, is:[14]

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

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. Similar 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 suboptimal 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:

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

where π_{SE}, π_{UC}, and π_{CC} are the mixing proportions of type SE, UC, and CC, respectively,

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

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:

$$\log L\left(\pi_{SE}, \pi_{UC}, \pi_{CC}, \theta^{SE}, \tau^{SE}, \theta^{UC}, \tau^{UC}, \theta^{CC}, \tau^{CC}, \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \rho\right) = \sum_{i=1}^{S} \log L_i. \tag{12}$$

The model is estimated using data (choices and beliefs) from each treatment separately. Our samples consist of 210 subjects (S) for treatments I and 207 for treatment E; each subject's contribution and vector of beliefs are observed $T = 15$ 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 (100 draws per subject).[15]

5.1 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 from unconditional cooperators and selfish players and to estimate the proportion of the population who are of each type.[16]

	I	E
π_{SE}	0.1306 (0.0233)***	0.2315 (0.0339)***
π_{UC}	0.1997 (0.0315)***	0.0638 (0.0173)***
π_{CC}	0.6697 (0.0335)***	0.7047 (0.0340)***
θ^{SE}	0.4257 (0.1562)***	0.2320 (0.1558)
τ^{SE}	-0.4738 (0.1010)***	-0.3374 (0.0913)***
θ^{UC}	0.6325 (0.1271)***	0.6650 (0.1502)***
τ^{UC}	-0.1134 (0.0339)***	-0.2022 (0.0496)***
θ^{CC}	0.1301 (0.0541)**	0.1455 (0.0851)*
τ^{CC}	-0.0837 (0.0526)	-0.2999 (0.0643)***
μ_α	-0.0053 (0.2114)	0.7765 (0.1725)***
σ_α	1.7515 (0.1904)***	1.2513 (0.1667)***
μ_β	0.2185 (0.1475)	0.5032 (0.1539)***
σ_β	1.2822 (0.1759)***	1.1110 (0.1652)***
ρ	0.5802 (0.1153)***	0.3717 (0.1465)**
S (no. of subjects)	210	207
T (observations per subject)	15	15
log-likelihood	-4766.74	-3049.80

Table 3: Maximum simulated likelihood estimates of the mixture model's parameters (the log-likelihoods are maximized using two sequences of 100 Halton draws). Standard errors, in parentheses, are bootstrapped (200 replications). ***, ** and * denote a p-value < 0.01, < 0.05 and < 0.10, respectively

Our findings are summarized in Result 4:

Result 4 *The distribution of types varies between treatments. The selfish type is more popular*

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

[16] 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.

among the experienced than the inexperienced, as opposed to the unconditional cooperator type. The conditional cooperator is the most popular type among both the inexperienced and the experienced.

Support for Result 4 is shown by Table 3, which reports the estimation results of the mixture model described in the previous section. In view of the validation of our hypotheses, the mixing proportions deserve a particular attention. In both treatments, CC seems to be the most common type, representing 67% and 70% of the population estimated from I and E, respectively. The estimated mixing proportion of SE is 13% from sample I and 23% from sample E. Finally, the proportion of UC is estimated to be 20% and 6% from I and E, respectively. The predominance of conditional cooperators in our samples reflects the results of Fischbacher et al. (2001), which are obtained with a different classification method based on the slope of the individual contribution schedule revealed via the strategy method.

Considering each mixing proportion singularly, we observe that the proportion of SE in E is significantly larger than the proportion of SE in I (z=2.89, p-value< 0.01, one-sided test). The proportion of the population who are UC, estimated from E, is strongly significantly smaller compared to the estimate we obtain from sample I (z=-7.82, p-value< 0.01, one-sided test). These results confirm our *Hypotheses 4a* and *4b*. Finally, we cannot reject the hypothesis that the proportion of CC in E is equal to that in I (z=0.95, p-value=0.30, two-sided test). A Wald test for the joint null hypothesis, stating that the mixing proportions from E are not significantly different from the mixing proportions estimated from I, strongly rejects the null ($\chi^2(2)$=65.24, p-value< 0.01). Therefore, we can rule out that sample E is drawn from a population with characteristics similar to those exhibited by the population from which sample I is drawn.[17]

In both treatments, selfish agents' choices are quite noisy in early periods but markedly less so in treatment E (w_1^{SE} is about 43% in I and 23% in E) with a rather high rate of decay (signalled by the negative sign of τ^{SE}) in the following periods so that the probability of choosing at random already approaches zero at mid-game. At the beginning of the experiment, unconditional cooperators 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%), while 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 smaller than 15% in the first period 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 significant in E. These findings imply that, in I, CC subjects keep the initial noisiness throughout the game, while, in E, such a noisiness completely disappears after only a few periods.

The last five rows of Table 3 report estimation results from the CC-type model. To characterize the behavior 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 attaches to distances (positive or negative) between his own and his partner's payoff (contribution). According to our estimation results, the inequality

[17] All the tests reported here are bootstrapped (200 replications) with asymptotic refinement (see Cameron and Trivedi 2005, among others).

$\alpha_i > \beta_i$ (predicted by Fehr and Smith 1999) holds for the proportion 0.438 (s.e. 0.050) of the population of I subjects and for the proportion 0.586 (s.e. 0.054) of the population of E subjects.[18] 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 proportion of the population that is larger among the experienced, than among the inexperienced.[19] We can also add that, as postulated by Fehr and Smith (1999), we get a positive and statistically significant correlation between α_i and β_i from both samples. 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.

A noteworthy feature of these results is that noisiness characterizes 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 unconditional cooperators happen to be the most noisy type, as they keep trembling until the very end of the game. We observe that even experienced subjects take time to understand how to play the optimal strategy dictated by their type's behavioral rule. As discussed in Section 4.2, we are inclined to attribute this to the so-called restart effect and/or to having already faced a *similar* decision framing which does not mirror the current situation perfectly.

The smaller proportion of unconditional cooperators among the experienced may be interpreted as a form of learning deriving from the participation in previous experimental sessions involving social dilemma games. Inspired by Palfrey and Prisbrey (1996), it may be inferred from this finding that subjects' confusion accounts for a large portion of the positive contributions observed in sample I. Since we acknowledge a larger proportion of selfish subjects among the experienced, we might conclude that, with experience, subjects are simply learning how to play the optimal strategy. As an alternative to this interpretation, one might assume that unconditional cooperators have selected themselves out of the experienced sample, after their first participation in a public goods experiment. If so, we should have expected a proportional increase in the mixing proportions of both the other two types, not only in the proportion of selfish players.[20] Hence, we are inclined towards the first explanation. Anyhow, since we cannot establish which of the two explanations holds, we will refrain from causal interpretations of the effect of experience (and history) on subjects' preferences and we will only refer to variations in the composition of types in the population due to these and other factors.

Understanding the reasons of the differences in the two populations is behind the main scope of this paper. From *our* methodological point of view, this is not relevant. What really matters is the altered composition of the population of the experienced with respect to that of the inexperienced, which deserves particular attention regardless of its causes.

[18] For details on the identification and calculation of these proportions, the reader is referred to the Appendix.

[19] Blanco et al. (2011) also find that, at an individual level, the inequality $\alpha_i > \beta_i$ is often violated.

[20] The estimated relative proportion of selfish subjects with respect to conditional cooperators is 0.195 (s.e. 0.041) and 0.328 (s.e. 0.062) from sample I and E, respectively, the latter being significantly different from the former (z=2.12, p-value= 0.03, two-sided test).

6 Understanding the differences in the mixing proportions: The role of experience and history

In this section, an econometric model is developed and estimated with the purpose of isolating the effect of experience from that of history and other background characteristics of subjects on the proportions of the population who are of each type.

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

$$pp_i^k = \Pr[i = \text{type } k \mid \text{obs}_i] = \frac{\Pr[i = \text{type } k] \times \Pr[\text{obs}_i \mid i = \text{type } k]}{\Pr[\text{obs}_i]} \quad (13)$$
$$= \frac{\pi_k \times \Pr[\text{obs}_i \mid i = \text{type } k]}{\Pr[\text{obs}_i]} = \frac{\pi_k \times l_i^k}{L_i}, \quad \forall k,$$

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 of both samples is reported in the Appendix.

As mentioned, we want to disentangle the effect of experience of public goods games from the effect of background experience of other kinds of experiment on the mixing proportions of types. With this in mind, we deal with the posterior probabilities obtained from maximizing Eq. (12) as the dependent variables in a three simultaneous equations (one for each type) model, as explained in 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), that reflect the way in which the mixing proportions are affected by this factor. Details on the computation of marginal effects and their standard errors can be found in Appendix. Here we just need to know that x_h^b is the *base* value and x_h^f the *final* value of the variable of interest, with respect to which we calculate the marginal effects. The table reports three different specifications of Eq. 18. The p-value of a Wald test of the joint significance of a certain coefficient in the three equations of model (18) is reported in the third column of each sub-table.

Specification 1 includes two binary indicators, one that captures experience and another that captures subjects' history. The results reveal that experience of public goods games affects positively the proportions of SE and CC, and negatively the proportion of UC.[21] Having taken part in other experiments seems to have a mild effect only on the proportion of SE subjects, which increases of 3 percentage points but has no clear effect on the other two types.

[21] The numbers in Table 4 have to be interpreted as follows. Specification 1 attests, for example, that, if we draw two samples which are equal in everything but in subjects' experience, then we have to expect to find in the experienced sample, with respect to the inexperienced sample, that the proportion of selfish (conditional cooperator) is 6.37 (6.81) percentage points larger and the proportion of unconditional cooperators is 13.18 percentage point smaller.

Specification 1

1(experience> 0) $x^b = 0; x^f = 1$	SE	0.0637(0.0300)**	
	UC	-0.1318 (0.0271)***	p-value< 0.01
	CC	0.0681 (0.0345)**	
1(other experiments> 0) $x^b = 0; x^f = 1$	SE	0.0325 (0.0182)*	
	UC	-0.0132 (0.0220)	p-value=0.12
	CC	-0.0193 (0.0337)	

Specification 2

		Δx $x^b = 0; x^f = 1$	$x^b = 1; x^f = 2$	$x^b = 2; x^f = 3$	
# Public Goods Games	SE	0.0237 (0.0083)***	0.0259 (0.0105)***	0.0279 (0.0127)**	
	UC	-0.0496 (0.0113)***	-0.0405 (0.0074)***	-0.0322 (0.0043)***	p-value< 0.01
	CC	0.0259 (0.0130)**	0.0146 (0.0121)	0.0042 (0.0128)	
# other experiments	SE	-0.0004 (0.0024)	-0.0004 (0.0024)	-0.0004 (0.0024)	
	UC	0.0016 (0.0026)	0.0017 (0.0026)	0.0017 (0.0026)	p-value=0.215
	CC	-0.0012 (0.0037)	-0.0012 (0.0037)	-0.0012 (0.0038)	

Specification 3

1(experience> 0) $x^b = 0; x^f = 1$	SE	0.0566 (0.0296)**	
	UC	-0.1086 (0.0296)***	p-value= 0.03
	CC	0.0520 (0.0385)	
1(Group A> 0) $x^b = 0; x^f = 1$	SE	0.0118 (0.0167)	
	UC	0.0194 (0.0182)	p-value= 0.42
	CC	-0.0312 (0.0275)	
1(Group B> 0) $x^b = 0; x^f = 1$	SE	0.0002 (0.0197)	
	UC	-0.0347 (0.0282)	p-value= 0.21
	CC	0.0345 (0.0420)	
1(Group C> 0) $x^b = 0; x^f = 1$	SE	-0.0079 (0.0181)	
	UC	-0.0080 (0.0239)	p-value= 0.94
	CC	0.0160 (0.0349)	
1(Group D> 0) $x^b = 0; x^f = 1$	SE	0.0381 (0.0188)**	
	UC	-0.0150 (0.0244)	p-value= 0.07
	CC	-0.0231 (0.0350)	
male $x^b = 0; x^f = 1$	SE	0.0809 (0.0166)***	
	UC	0.0523 (0.0196)***	p-value< 0.01
	CC	-0.1332 (0.0298)***	
age $x^b = 23; x^f = 24$	SE	-0.0016 (0.0023)	
	UC	0.0010 (0.0031)	p-value= 0.58
	CC	0.0007 (0.0049)	

Table 4: The tables report marginal effects of a change in one of the regressors at a time on the expected posterior probabilities, based on the estimation results of three specifications of model (18). The p-values on the right of each group of marginal effects refer to a bootstrapped Wald test of the joint significance of that particular regressor. To characterize experience we use: a binary indicator in spec. 1 and 3; the number of participations in a public goods experiment in spec. 2. To characterise history we use: a binary indicator in spec. 1; the number of participations in a public goods experiment in spec. 2; and a binary indicator of the participation in at least one of the experiments included in the groups of experiments listed in Tab. 2 in spec. 3. A binary indicator of gender and participants' age are also used in spec. 3. The marginal effects of age are computed with respect to the approximate mean age of the participants in both samples. Standard errors, in parentheses, are bootstrapped (200 replications). ***, ** and * denote p-values < 0.01, < 0.05, and < 0.10, respectively. The p-values of the significance tests of the marginal effect of experience on the proportion of SE and UC refer to a one-sided test, compatibly with *Hypotheses 4a,b*. All the other tests are two-sided.

Specification 2 is similar to specification 1 but uses the number of participations in public goods games and other types of experiments instead of binary indicators. The positive (negative) and significant marginal effect of an additional participation in a public goods experiments seems to be persistent on the proportion of selfish (unconditional cooperator). Its positive effect on the proportion of conditional cooperators, instead, fades away after the first participation in a public goods game. The number of participations in experiments different from the public goods game does not seem to have any significant effect on the popularity of the three types.

In specification 3, subjects' history is decomposed into binary indicators capturing the participation in at least one of the experiments as classified in Table 2. A dummy variable, taking the value 1 if the subjects is a male, 0 otherwise, is added as well as participants' age. The positive and negative effects on the proportions of selfish and unconditional cooperators, respectively, are still significant in this specification. There seems to be no effect of experience on the proportion of conditional cooperators. Participation in experiments included in Groups A, B and C do not determine any statistically significant effect on the mixing proportions of types. However, there is a positive and statistically significant effect of the participation in experiments included in Group D (auctions, bargaining, coordination and other experiments) on the proportion of selfish subjects in the population. We also tried to disaggregate these groups, but we did not obtain any significant effect of the single experiment.[22] 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 emerges only when those experiments are pooled together.

Gender has a very strong effect on all the types. The proportion of selfish and unconditional cooperators increases of 8 and 5 percentage point, respectively, and the proportion of conditional cooperators decreases of 13 percentage points if a male is sampled. Subjects' age does not seem to contribute significantly to the proportion of the different types in the populations.

In summary, our hypotheses on the effect of experience on the proportion of selfish and unconditional cooperators are confirmed by these results. The effect (positive) of experience on the proportion of conditional cooperators is almost absent. There seems to be a positive effect on the proportion of selfish players of the participation in experiments included in the aggregate labelled as Group D. Gender is a strong predictor of type.

7 Aggregate analysis by type

This section essentially retraces most of the aggregate analysis of our data described in Section 4.2. This time, though, we use posterior probabilities (see Eq. (13)) and the maximization results of Eq. (12) as weights to calculate average contributions and average expected contributions *by type*. These are computed, respectively, as:

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

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

Figure 4 displays average contributions and average expected contributions by type and 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: SE; center: UC; bottom: CC). For convenience, in each pane a line representing the full-sample

[22] Table 4 only reports the relevant results. We used several different controls such as, among others, the number of participations in a particular type of experiment, the time of the first and/or last participation in an experiment both for public goods and other experiments, 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.

average contribution (see Fig. 1) is superimposed.

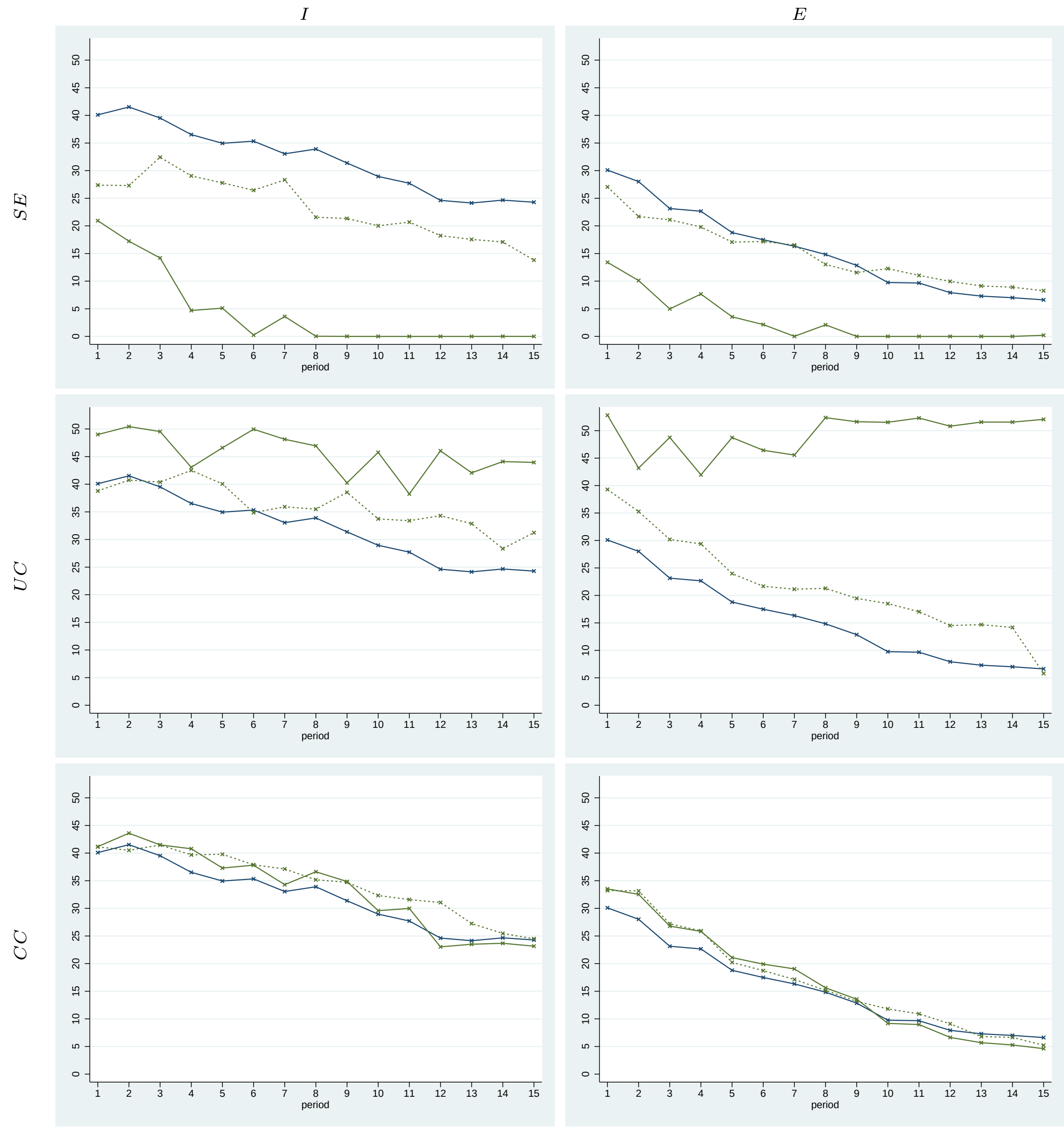

Figure 4: Average contribution (solid green line), $\overline{c_t}$, and average expected contribution (dotted green line), $\overline{\mathrm{E}_t(c_t)}$, by type and treatment. A blue line representing the full-sample average contribution (see Fig. 1) is superimposed

Consider first the subjects classified as selfish. The experienced's and the inexperienced's average contributions follow a similar decreasing trend. As noted in Section 5.1, 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 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 they might be motivated by pessimistic expectations about the others' cooperative behavior, rather they seem to be motivated by their own payoff maximization. In the opposite case, we should have observed the average expected contributions from selfish subjects to match their average contribution pattern more closely.

In both treatments, the average contributions from the subjects classified as unconditional cooperators lie above their average expected contributions. The latter mimic quite well the full-sample average contributions in treatment I (with a mildly positive bias). The experienced's beliefs, instead, appear to overestimate the average expected contributions systematically even if their declining trend is followed, but it has to be noted that the number of experienced unconditional cooperators is almost negligible. 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 5.1. If, with experience, subjects improve their understanding of the dynamics of interaction, then in the E subject pool those who are classified as unconditional cooperators should no longer be regarded as confused subjects but as true altruists.

Finally, let us look at the subjects classified as conditional cooperators. These subjects, who represent the most popular type in both subject pools, appear the most similar across treatments: irrespective of experience, their contributions match rather well both the expected contributions and the full-sample contributions, on average. The most noticeable between-treatment difference lies in the sequences of contribution: compared to treatment I, treatment E starts at a lower level and declines, eventually following a steeper trend. This evidence can be summarized in Result 5, which supports our *Hypothesis 4c*:

Result 5 *Conditional cooperators expect the others to contribute smaller amounts and contribute smaller amounts when they are experienced than when they are inexperienced.*

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:

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

No remarkable difference seems to exist in any of the treatments regarding the accuracy of prediction among types. Selfish subjects appear to be most accurate in treatment I, and conditional cooperators appear to be most accurate in treatment E toward the end of the game. Yet we want to stress once again that the differences among types as evidenced by our indicator are negligible. Concerning the differences between samples I and E, what noted about the accuracy of prediction from the entire samples (see Fig. (3)) still holds here at the individual level. Thus, all types in sample E are as accurate as all types in sample I at the beginning of the game but considerably improve their predictions after only a few periods. This occurs only marginally with respect to all types in sample I.

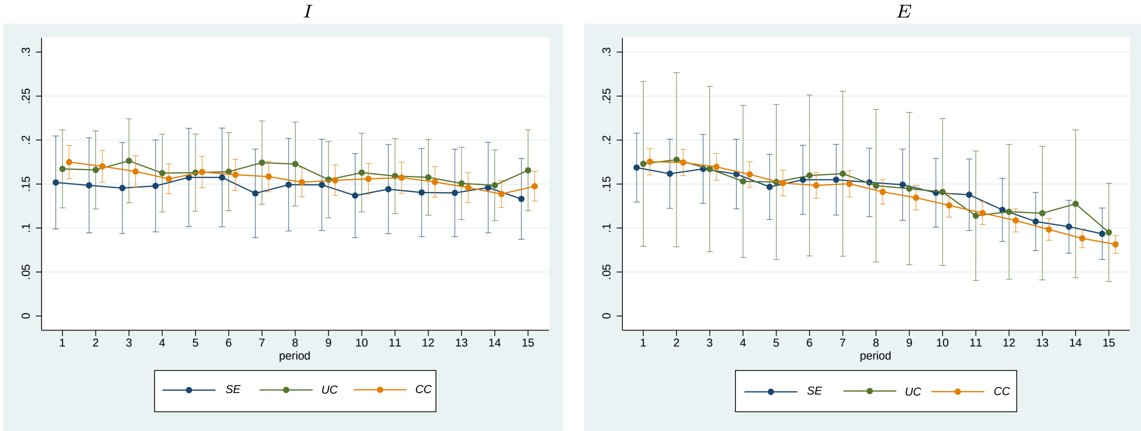

Figure 5: Average accuracy of individual belief distribution from all-sample contribution distribution of contributions by period, treatment and type. The range plots represent 95% confidence intervals

8 Conclusion

Thirty years ago, Isaac et al. (1984) acknowledged the importance of subjects' experience for behavior in public goods experiments and, given the inconclusive evidence, called for a more systematic study of the matter. In this paper, we have welcomed this call and studied whether and how past participation in economic experiments affects subjects' behavior in a sequence of one-shot two-person linear public goods games where subjects must decide on their contributions and report their first-order beliefs about their partners' contributions.

Relying on the information stored in our database, we have allocated the effect of previous participation into two components referred to as: i) experience, which designates previous participation in social dilemma-type experiments, and ii) history, which designates participation in experiments of a different class than social dilemma.

The results of our analysis on the impact of *experience* on contribution behavior are crystal clear. At the aggregate level, the experienced—compared to the inexperienced—contribute smaller amounts, expect the others to contribute smaller amounts, and hold more accurate beliefs. This evidence already indicates important differences between the two populations. Yet, it does not provide sufficient clues to their reasons. With the help of a finite mixture model, distinguishing selfish, conditional, and unconditional contributors, we find that the proportion of unconditional cooperators in the experienced subject pool decreases, prominently in favor of the selfish type. Our data reveals that *history* also influences the proportions of the population who are of each type, but such an effect is less trenchant than in the case of *experience*. Interestingly, the number of participations in public goods experiments seems to have a cumulative effect on the proportion of both selfish and unconditional cooperators. This does not seem to be the case for the conditional cooperators, whose prominent proportion of the population is only mildly affected by the simple fact of having taken part in public goods experiments, no matter the frequency.

The individual analysis demonstrates that the decisions of all types of player are noisy in early periods, but less so in the experienced population. 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 optimum use of the information about the others received in each period since their predictions only mildly refine in the end. Differently, after only a few periods the experienced's accuracy ameliorates dramatically. There appears to be no notable difference in belief accuracy among the three types regardless of the treatment.

The present study has important implications for the experimental methodology in the context of public goods games. In fact, both *experience* and *history* can, to some degree, be controlled in the recruitment of experimental participants. In the light of our results, both these factors should be properly accounted for when conducting economic experiments, and they should be documented in the experimental procedures. We postulate that this may lead to an improvement in terms of both external validity and replicability of the experimental results. Guala (1999) defines the set of properties (participants' characteristics, rules used for their recruitment, etc.) that a given experimental system requires to ensure generalization of the results it produces. Our proposal to enrich such a set with *experience* and *history* is meant to introduce elements typical of the real word—where experience matters—into the experimental background conditions. Controlling for 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 (that may or may not be located in different laboratories).

Therefore, 30 years after Isaac et al. (1984)'s contribution, we respond to their call and show in a rigorous manner that experience and history are important factors to be controlled and considered in selecting participants for public goods experiments.

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 Friedrich Schiller University Jena. They were recruited by the ORSEE (Greiner 2004) software such that the samples for treatment I were made up of students who had never participated in public goods and prisoner's 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. 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 14 sessions: 7 for treatment inexperienced, and 7 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.[23]

Participants were paid according to their contributions in one randomly selected period, t_1, at a rate of €0.15 per ECU and, depending on 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 spent reading the instructions and answering the control questionnaire. Average earnings per subject were €28.95 (inclusive of a €2.50 show-up fee), ranging from €16.8 to €47.5 in treatments E and from €15.00 to €45.3 in treatment I. Additional €3 were paid to those who agreed to disclose their biographical data and information about their previous participation(s) in experimental sessions.

To this end, we had prepared an envelope for each participant, containing all the details about their previous participations in lab experiments, as recorded in our database. At the end of the experimental session, we asked the participants who had agreed to share their information to enter them into the postexperimental questionnaire. Thus, we were able to track the complete history of participations 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; nor were they, in the E treatment case, made aware of the fact that their partners had had previous experience of public goods game experiments or other experiments.

Identification of α_i and β_i and 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 the correlation coefficient of the underlying bivariate normal distributions but not the two means.[24] In fact, we can only obtain an

[23] The first 6 sessions of treatment I have already been analyzed in Conte and Levati (2014).
[24] More on this sort of identification problems can be found in Train (2003, pag. 45).

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 $1,000,000$ 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 calculate their ratio, α_i/β_i, obtaining a value of 0.438 from sample I and 0.586 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

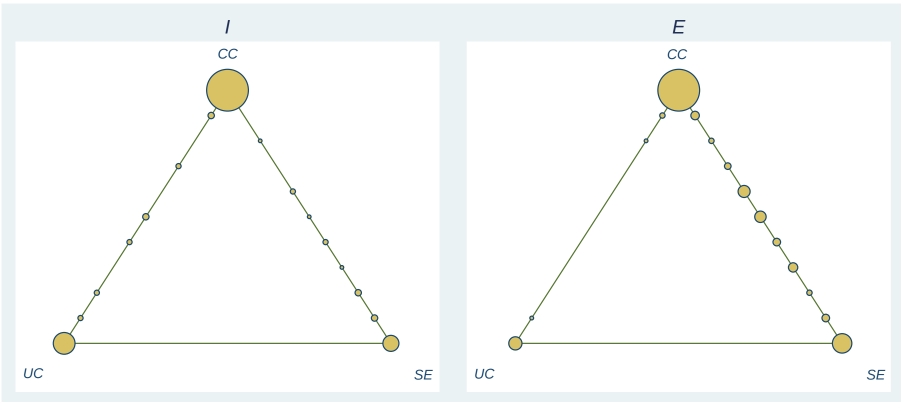

Figure 6: Posterior probabilities distribution of the three types from the models estimated in Table 3.

The posterior probabilities based on Eq. 13 and the mixture model results in Tab. 3 are displayed in Figure 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: with just few exceptions, the vast majority of subjects 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 the proportion of unconditional cooperators (0.064). 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 16 for limit cases of this sort). Despite these few cases, the power of our mixtures in 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. error terms, 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:[25]

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

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 maximizing the logarithmic sum of Eq. (18). Only by merging the two samples, can we distinguish the effect on subjects' behavior of participation in public goods experiments from other background experience 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

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

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

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:

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

where $\hat{\gamma}$ are parameters' estimates from the maximization 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 maximization of model (12). For this reason, we follow the procedure described here for each bootstrapped sample. In other words, from each bootstrapped sample, we maximize Eq. (12) and use such parameters' estimates to calculate the posterior probabilities according to Eq. (17); we then maximize 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 6.

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), 1317–1327.

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

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

Arifovic, J., Ledyard. J., 2012. Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism, *Journal of Public Economics* 96, 808–823.

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., Sausgruber, R., 2013. Confusion and learning in the voluntary contributions game, *Experimental Economics* 16, 478–496.

Benson, B.L., Faminow, M.D., 1988. The impact of experience on prices and profits in experimental duopoly markets, *Journal of Economic Behavior & Organization* 9(4), 345–365.

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(5), 1096–1136.

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

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

Conte, A., Levati, M.V., 2014. Use of data on planned contributions and stated beliefs in the measurement of social preferences, *Theory and Decision* 76(2), 201–223.

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.

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 goods 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 goods 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, 555–573.

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.

Gunnthorsdottir, A., Houser, D., McCabe, K., 2007. Disposition, history and contributions in public goods experiments, *Journal of Economic Behavior & Organization* 62, 304–315.

Harrison, G.W., McKee, M., Rutström, E.E., 1987. Experimental evaluations of institutions of monopoly restraints. In: Green, L., Kagel, J.H. (Eds.), *Advances in Behavioral Economics.* Ablex Press, Norwood, N.J., 54–94.

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., Walker, J., 1989. The assurance problem in a laboratory market, *Public Choice* 62, 217–236.

Isaac, R.M., Walker, J.M., Thomas, S.H., 1984. Divergent evidence on free riding: an experimental examination of possible explanations, *Public Choice* 43, 113–149.

Kurzban, R., Houser, D., 2005. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations, *PNAS* 102(5), 1803–1807.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J., Roth, A.E. (Eds.), *The Handbook of Experimental Economics.* Princeton University Press, 111–194.

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., Ames, R.E., 1980. Experiments on the provision of public goods. II. Provision points, stakes, experience and the free rider problem, *American Journal of Sociology* 85(4), 926–937.

Matthey, A., Regner, T., 2013. On the independence of history: experience spill-overs between experiments, *Theory and Decision* 75(3), 403–419.

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., 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.

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

Sass, M., Weimann, J., 2012. The dynamics of individual preferences in repeated public good experiments, FEMM Working Papers 120002, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.

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

Smith, V.L., 2010. Theory and experiment: what are the questions?, *Journal of Economic Behavior & Organization* 73, 3–15.

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, 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.