Preferences and strategic behavior in public goods games

Gilles Grandjean* Mathieu Lefebvre† Marco Mantovani ‡

Abstract

In finitely repeated public goods games, contributions are initially high, and gradually decrease over time. Two main explanations are consistent with this pattern: (i) the population is composed of free-riders, who never contribute, and conditional cooperators, who contribute if others do so as well; (ii) strategic players contribute to sustain mutually beneficial future cooperation, but reduce their contributions as the end of the game approaches. This paper analyzes experimentally these explanations, by manipulating group composition to form homogeneous groups on both the preference and the strategic ability dimensions. Our results highlight the role of strategic ability in sustaining contributions, and suggest that the interaction between the two dimensions also matters: we find that groups that sustain high levels of cooperation are composed of members who share a common inclination toward cooperation and also have the strategic abilities to recognize and reap the benefits of enduring cooperation.

Keywords: Voluntary contribution, public goods, conditional cooperation, free riding, strategic sophistication.
JEL codes: H41, C73, C92.

---

*CEREC, Université Saint-Louis - Bruxelles, Boulevard du Jardin Botanique 43, B-1000 Brussels, Belgium.

†Aix Marseille Univ, CNRS, AMSE, 5-9 Bvd Maurice Bourdet, 13205, Marseille, France.

‡Corresponding author. Dept. of Economics, University of Milan Bicocca, P.za dell'Ateneo Nuovo 1, 20126, Milan, Italy. marco.mantovani@unimib.it

Acknowledgements: The authors are grateful to the associate editor and two anonymous referees for their very helpful comments and suggestions. The authors gratefully acknowledge financial support by the project "IDex Attractivité" funded through the University of Strasbourg IDEX Unistra, by the project "PRIN 2017K8ANN4" funded by the Italian Ministry of University and Research, by the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX.

Replication package: the data, the software used in the analysis, and the experimental software are available in the following public repository: https://osf.io/dqye4/?view_only=43d38912295d46218fbbddc383a485f0.

1 Introduction

In finitely repeated public goods games, players are not expected to contribute according to standard equilibrium analysis. However, highly replicable experimental evidence shows that contributions are initially high, and gradually decrease over rounds (Chaudhuri, 2011; Ledyard, 1995). Research on the topic has typically focused on two families of explanations: the first one based on (non-standard) preferences, the second on strategic motivations. According to the 'preference-based' explanation, these empirical patterns are due to the interaction of conditional cooperators and free riders. Conditional cooperators are willing to contribute to the public good only if the other group members also contribute. Free-riders do not contribute to the public good. Contributions would therefore decrease over rounds because conditional cooperators reduce their contributions when they interact with free riders.[1] According to the 'strategy-based' explanation, subjects contribute larger amounts in the initial periods because current contributions may sustain mutually beneficial future cooperation. These incentives are higher at the beginning and vanish as the game approaches the end.[2]

The literature has typically focused on one of these dimensions. This paper, instead, studies both preferences and strategic ability in a finitely repeated public goods game within the same experimental design. We categorize subjects as free-riders, unconditional cooperators and conditional cooperators, based on their choices in a one-shot public goods game, using the so-called Strategy Method (Fischbacher et al., 2001; Selten, 1967). We order subjects by strategic ability ac-

[1] Among the papers that established the presence of these heterogeneous types in one-shot public goods games, see Brandts and Schram (2001) and Fischbacher et al. (2001). For the analysis of their role in explaining patterns in finitely repeated public goods games, see Ambrus and Pathak (2011) – who show theoretically and empirically how those patterns can be seen as an equilibrium of the game with heterogeneous players – and Fischbacher and Gachter (2010) – who show that a bias toward one's own payoff in the contributions of conditional cooperators may be sufficient to generate those patterns, even in the absence of free riders. A number of papers compare the contributions of groups composed of free-riders to those of groups composed of cooperators (Burlando and Guala, 2005; Gächter and Thöni, 2005; Gunnthorsdottir et al., 2007). We discuss these papers in the main text.

[2] The theoretical underpinnings of this argument date back to Kreps et al. (1982). There is evidence that forward-looking strategic considerations are relevant in finitely repeated social dilemmas. Sonnemans et al. (1999) study a finitely repeated public goods game with partial rematching at known stages. They show that participants strategically reduce their contributions when approaching a rematch of the group. Muller et al. (2008) investigate a finitely repeated public goods game with strategy method in all rounds. They find that contributions decline also after controlling for one's partners contributions, and that a longer horizon is sufficient to slow down this decline. Evidence of strategically-motivated cooperation can be found in the literature on partners-versus-strangers protocols (e.g., Keser and van Winden (2000)). Baader and Vostroknutov (2017) investigate the interaction between non-selfish preferences and strategic cooperation. They show that the former provide incentives for the latter in the traveler's dilemma, and find that only sophisticated subjects respond to these incentives.

cording to their average score in a cognitive reflection test (Frederick, 2005), a race game (Dufwenberg et al., 2010; Gneezy et al., 2010) and a beauty contest (Nagel, 1995). We then manipulate group formation and study groups that are homogenous either in terms of preferences or in terms of strategic ability. Subjects are not informed of the matching protocol. Therefore, we can identify the effect of group composition, disentangling it from the effect that information about group composition would have through beliefs. This also distinguishes our work from recent papers showing that providing information about subjectsâ€™ abilities (Lambrecht et al., 2021) or personality traits (Drouvelis and Georgantzis, 2019) affects behavior.

We hypothesize that strategically able individuals respond to repeated-game incentives by contributing more, even if the number of repetitions is finite, and that they respond more to group composition than individuals of lower ability. As a consequence, they should increase their contributions when matched in homogeneous groups of high-ability subjects. Preferences and ability provide distinct intrinsic and extrinsic motives to contribute and shape individuals' reaction to others' contributions. We investigate potential complementarities between the two, and hypothesize a positive interaction between preferences and ability. This implies that groups composed of high-ability conditional cooperators should contribute more than other groups.

Previous research that manipulates group composition in finitely repeated public goods games has focused on preferences. Burlando and Guala (2005) is the closest to our paper. They classify subjects according to their preference profile and compare their contributions in homogenous groups to those of the same subjects in randomly matched groups, without informing subjects of the matching protocol. They find that conditional cooperators contribute more when matched with other conditional cooperators, and that total contributions are higher when subjects are matched according to their preference profile than when they are matched randomly. Gächter and Thöni (2005) find similar results by forming homogenous groups based on choices in a one-shot public goods game, when subjects are informed about the matching protocol.[3] One of our treatments is conceptually similar to the exercise of Burlando and Guala (2005), to which we add a further treatment where the matching is based on strategic ability. For measuring ability, we adopt a procedure close to Bosch-Rosa et al. (2018): we use the same tasks as theirs and also form groups according to the resulting strategic-

[3]See also Gunnthorsdottir et al. (2007) – who rematch groups period-by-period according to their current contribution level, not informing subjects of the procedure – and Ones and Putterman (2007) – who classify subjects according to both their attitudes toward cooperation and their attitudes toward punishment.

ability ranking. They show that homogeneous groups of high-ability subjects do not generate bubbles in asset markets.

Proto et al. (2019) match subjects according to (cognitive) ability in a social dilemma. They show that homogeneous groups of high-ability subjects play more cooperatively in an indefinitely repeated prisoner's dilemma. Drouvelis and Pearce (2021) find similar results in a sequential public goods game where they elicit the participants strategies. Low-ability subjects are more likely to free-ride after a cooperative action by their partner, while social preferences are similar across ability groups.[4]

The role of strategic motivations is, however, different when a game is indefinitely, rather than finitely repeated. Under standard assumptions, mutual cooperation is part of a Pareto superior equilibrium in the former case, but not in the latter.[5] One may expect that more strategic groups play closer to equilibrium and, therefore, less cooperatively in the finitely repeated case. However, this deduction may be too naïve: minimal departures from common knowledge of rationality and preferences induce rational cooperation until the last rounds of the game (Kreps et al., 1982). This paper is the first to investigate empirically the role of strategic ability in a finitely repeated social dilemma.

Our results show that strategic ability is crucial to sustain cooperation even in a context where cooperation cannot be sustained under standard equilibrium notions. Comparing contributions in the first period of the repeated public goods game with those in the one-shot version that we use to classify them, we find that high-ability subjects respond to the incentives of repeated interaction by increasing their contributions more than low-ability ones, independently of their preferences. Matching high-ability subjects together boosts their contributions as compared to random matching. Perhaps surprisingly, we do not find that conditional cooperators contribute more when they are matched in homogeneous groups.[6] However, we observe a positive interaction between preferences and ability: groups that sustain high levels of contributions until the final rounds are

[4]These studies measure cognitive abilities using the Raven test. We use a composite measure of strategic ability, where the closest component to the Raven test is the CRT. For the relation between cognitive and strategic ability, see, e.g., Gill and Prowse (2016), Basteck and Mantovani (2018).

[5]Reuben and Suetens (2012), Dreber et al. (2014) and Cabral et al. (2014) all conclude that cooperation is mainly explained by instrumental motives in indefinitely repeated games.

[6]This result is at odds with Burlando and Guala (2005). As we will discuss (see Section 3 and Appendix D), this difference probably arises from differences in the samples and in the classification procedures. Burlando and Guala (2005) use a combination of different tasks for classification, while we follow Fischbacher et al. (2001) and use only one. As a result, their conditional cooperators have stronger cooperative attitudes than ours. The difference in the results vanishes when we focus on conditional cooperators that are ex-ante more similar. We are grateful to Roberto Burlando and Francesco Guala who have made their original data available to us.

made of high-ability conditional cooperators. Thus, the data suggest that both a shared inclination toward cooperation and the strategic ability to recognize and reap the benefits of enduring cooperation are relevant for a group to sustain high levels of cooperation in finitely repeated public goods games.

The rest of the paper proceeds as follows. Section 2 provides the details of the experimental design. Section 3 reports our results. Section 4 discusses some limitations of the paper and concludes.

2 Experimental design

2.1 Main task

The main task of the experiment is a finitely-repeated linear Public Goods Game (PGG).[7] The PGG is played in fixed groups of three subjects (i.e. 'partners matching') for 15 periods. At the start of each period, each subject receives an endowment of 20 tokens and decides how many tokens she wants to invest in a public account. Decisions are individual and simultaneous. Each token invested in the public account yields 0.6 token to each member of the group. Subjects keep for themselves the tokens they do not invest. Therefore, in each period, the earnings of individual i in a group with j and k, given the contribution decisions to the public account c_i, c_j and c_k, are given by:

$$\pi_i = 20 - c_i + 0.6(c_i + c_j + c_k)$$

At the end of every period, each group member is informed of her earnings in that period and a new period begins. We analyze the behavior of individuals in groups formed according to individuals' preferences and abilities. The classification of the subjects is based on their behavior in four independent tasks, which are played before the repeated PGG.

2.2 Classification: preferences

To classify subjects according to their preferences, we use a one-shot PGG in Strategy Method (Fischbacher et al., 2001; Selten, 1967). The framing of the problem is in all respects similar to the one described above for the finitely-repeated PGG, the only difference being that they play the game only once, rather than 15 times. Subjects are asked to make two decisions. First, they choose a contribution in a one-shot PGG – the 'unconditional contribution'. Then, they fill in

[7] See Appendix A for a complete transcript of the instructions.

a contribution table: they select a contribution for every possible average contribution of the other group members.[8] In each group, earnings are computed using the unconditional contributions of two randomly-selected subjects and the contribution table of the third one. Subjects are informed of their earnings in this task only at the end of the experiment.

We classify subjects into 'preference types' according to their contribution table, following a procedure that is similar to the one used by Thöni and Volk (2018) in their meta-analysis of preference types in one-shot PGG. If a subject's average entry in the contribution table is below 10 percent of the endowment, she is labeled a 'free rider'. If the average entry is higher and the standard deviation of the entries in the contribution table is below 5 percent of the endowment, the subject is labeled an 'unconditional cooperator'. If a subject is not a free-rider, nor an unconditional cooperator, and the correlation between her entries in the contribution table and the corresponding average contribution of others is above .7, the subject is labeled a 'conditional cooperator'. If none of these criteria is met, she is assigned to a residual category.

Two remarks are worth mentioning at this stage. First, our procedure, while quite established in this literature, is different from the one used by Burlando and Guala (2005), who combine four tasks and different criteria for classification. We thoroughly discuss those differences, which are likely to be relevant in interpreting our results, in Appendix D. Second, we are conceptually interested in conditionally cooperative 'preferences'. However, we do not observe preferences, and classify subjects according to a *behavioral trait* that we infer to be driven by the underlying preferences. We come back to this issue when discussing the interpretation of our results in Section 4.

2.3 Classification: strategic ability

To classify subjects according to their strategic ability, we use three different tasks.[9]

A **Cognitive reflection test (CRT)** (Frederick, 2005) measures the ability to switch one's reasoning mode from the routine system to the reflective system. The CRT consists of three algebraic items. Each item has an intuitive incorrect answer and

[8] The earnings of individual i in a group with j and k given the contribution decisions to the public account c_i, c_j and c_k are given by: $\pi_i = 200 - c_i + .6(c_i + c_j + c_k)$. The rescaling of incentives with respect to the repeated PGG was used to ensure appropriate incentives in all tasks while maintaining payments within the standard experimental oness.

[9] Our classification of the strategic ability of subjects follows closely Bosch-Rosa et al. (2018). They aggregate performances in the same three tasks that we use, although with slightly different procedures.

a non-intuitive correct one. The score of a subject in the CRT is her percentage of correct answers.[10]

A **Race to 26 game** (Dufwenberg et al., 2010; Gneezy et al., 2010) measures the subject's ability to plan strategic decisions ahead and perform backward induction. The subject and the computer sequentially choose numbers between 1 and 5. Those are added up, until the target of 26 is reached. The subject wins if she reaches 26 before the computer. Each player has a dominant strategy. It identifies an action between 1 and 5 for each position between 1 and 25. These actions are such that, added to the current position, they lead into the set of 'losing positions' $\{2,8,14,20,26\}$, whenever possible. The computer never leads the subject to a losing position, but picks the winning number in case it has to make a choice between 21 and 25. Therefore, the subject can secure the victory from the first move by playing according to her dominant strategy. We observe the moment where a subject switches to this strategy. The percentage of consecutive losing positions a subject reaches, starting from the last, represents her score in the task.

A **Beauty contest** (Nagel, 1995) measures the ability to perform iterative reasoning in a strategic environment. It is commonly used to classify subjects into levels of reasoning in normal form games. Subjects are asked to choose a number between 0 and 100. The subject whose choice is closest to 2/3 of the average of the numbers chosen in her group receives a prize. The score of a subject in this game is the normalized distance between her choice and 2/3 of the average in the session (as done in Bosch-Rosa et al. (2018)). Formally, let a subject's entry be b_i, and the average entry in the session be μ. The score of subject i in the Beauty contest, p_i, is given by:

$$p_i = 100 \cdot \max\left\{0, 1 - \left[\frac{|b_i - 2/3\mu|}{66 - 2/3\mu}\right]\right\}$$

Subjects whose entries are dominated are assigned a score of zero. A score of 100 is assigned to subjects for whom $b_i = (2/3) \cdot \mu$.

Subjects are informed of their earnings in these tasks only at the end of the experiment. The strategic ability score of a subject is obtained by averaging her score in the CRT, the race game and the beauty contest. It is therefore a number between 0 and 100, and higher numbers correspond to a higher strategic ability. We label subjects with a score above the median in the whole sample 'high-ability subjects', and we label other subjects 'low-ability subjects'. We test our hypotheses both around this median split – i.e., comparing all high-ability subjects to

[10]See Appendix B for a transcript of the cognitive reflection test.

all low-ability subjects – and by comparing the top 30 percent to the bottom 30 percent of the strategic ability distribution.

The three tasks capture skills that are relevant for the repeated PGG. Backward induction is the archetypical reasoning procedure in finitely repeated games. Performance in this type of reasoning is measured through the Race to 26. However, backward induction is useful in a finitely repeated PGG only when it is combined with a good model of other players' behavior. The Beauty contest captures how sophisticated and accurate is a subject's model of others' behavior. To sustain cooperation in a finitely repeated PGG, subjects need to resist the temptation to shirk in any given stage, and avoid the sequence of punishments and counter-punishments that would follow. The CRT measures the ability to avoid unsupervised instinctive responses.[11] We assume that we can aggregate these different abilities, and that this aggregation represents the latent construct labeled 'strategic ability'. Part of our analysis rests on this assumption, because treatments are based on this aggregate measure of strategic ability. However, the possibility to aggregate the three performances is debatable, and the interpretation of performance in each task is more straightforward than the aggregate measure. Therefore, when possible, we will also test for the predictive power of each single ability task.

2.4 Treatments

The experiment has three treatments that differ in terms of the procedure used to form groups. In a baseline treatment, groups are matched at random (treatment RAND). In two other treatments, we manipulate the composition of groups to obtain groups that are homogeneous in terms of preferences (treatment PREF), and groups that are homogeneous in terms of strategic ability (treatment STRAT). In treatment PREF, the matching procedure maximizes the number of homogeneous groups. First, within each preference type, all possible homogeneous groups are formed at random. Second, all subjects that are not assigned to a homogeneous group are randomly matched. In treatment STRAT, the three subjects with the highest strategic ability score form one group, the three subjects with the highest score among the remaining ones form another group, and so on.[12] Treatments are ex-ante identical from the point of view of subjects: they are only told that they are matched with two other subjects in the session and that

[11] More broadly, the CRT also correlates with more time-consuming measures of cognitive abilities that predict cooperation in repeated games, such as the Raven score (Drouvelis and Pearce, 2021; Proto et al., 2019).

[12] Ties are broken at random.

they will play in the same group for fifteen rounds.

2.5 Procedures

The experiment was conducted in spring 2016 in the Experimental Economics Lab at the University of Strasbourg. It was programmed using Z-tree (Fischbacher, 2007). 192 subjects were recruited through ORSEE (Greiner, 2015), distributed over 8 experimental sessions, with 24 subjects in each session. Each subject participated only in one session.

All sessions followed an identical procedure. After their arrival, subjects were randomly assigned to cubicles in the laboratory. Instructions were read aloud before each task. To ensure that everybody understood the tasks, participants had to answer a set of control questions before the one-shot PGG, the Race to 26 game, the Beauty Contest and the repeated PGG. These tasks would start only after all subjects had cleared the control questions.[13] In every session, participants faced the classification tasks first, and then played the 15 repetitions of the PGG. Finally, they filled in a questionnaire which included qualitative information about their strategies and self-reported quantitative measures of risk preferences extracted from the SOEP German panel.[14] At the end of the repeated PGG, the computer selected at random one of the four classification tasks for the whole session. The monetary payment of the subjects was based on the tokens earned in this task and those earned in the repeated PGG. The tokens were paid according to the exchange rate: 40 tokens = 1€. Subjects could earn between 7.5 and 16.5€ in the repeated PGG, and between 0 and 11€ in the other tasks.[15] Participants earned 13.60€ on average and sessions lasted around 60 minutes.

2.6 Hypotheses

The main hypotheses target the role of strategic ability in the finitely repeated PGG. In theory, minimal departures from common knowledge of rationality or from universal selfish preferences allow cooperation to be sustained until the last rounds. Under these conditions, subjects have an extrinsic incentive to contribute in order to induce profitable higher levels of contributions in the future. We hypothesize that subjects of higher strategic ability will be better at understanding

[13]Subjects played two trial versions of the race game before playing the one which was relevant for classification and payment. Trials featured a slightly different game, to avoid mechanical learning of losing positions. See Appendix A.

[14]For the use of the risk questions to measure risk preferences, see Dohmen et al. (2011) and Vieider et al. (2015).

[15]In particular: 0€, 2.5€, 5€ or 7.5€ in the CRT, 0€ or 5€ in the race game, 0€ or 5€ in the beauty contest, and between 0€ and 11€ in the one-shot PGG.

these repeated-game incentives. The cleanest shot to this hypothesis comes from the comparison of contributions in the one-shot game with first-period contributions in the finitely repeated PGG, because in the first round contributions are not affected by the observation of partners' behavior.

Hypothesis 1. *The difference between the contribution in the first period of the repeated PGG and the one-shot unconditional contribution is larger for high-ability subjects than for low-ability ones.*

Understanding that one's contribution today leads to higher contributions of group members tomorrow bears consequences also after the first period. The typical pattern of contributions in lab experiments features decay over time, as players progressively undercut past group contributions. We expect high-ability subjects to sustain cooperation levels longer in the game.[16]

Hypothesis 2. *For given contribution of their partners in the previous round, high-ability subjects contribute more than low-ability ones.*

Hypotheses 1 and 2 characterize how we expect ability to affect individual behavior, and constitute the foundation of our hypotheses regarding treatment effects. High-ability subjects contribute in a given period in order to foster cooperation in future periods. A high-ability subject should, therefore, sustain higher levels of cooperation when she observes that her strategy is effective. We expect this to be the case when she is matched with high-ability partners who share the same view about the benefits of long-run cooperation.

Hypothesis 3. *High-ability subjects contribute more in treatment STRAT than in treatment RAND.*

In treatment PREF groups are matched according to the subjects' preferences. Conditional cooperators are willing to contribute only if others do so as well. Matching them in homogeneous groups of conditional cooperators should lead them to contribute more than when they are matched at random.

Hypothesis 4. *Conditional cooperators contribute more in treatment PREF than in treatment RAND.*

[16]As the last period approaches, strategic motives to contribute fade away. In the last period, subjects should not contribute, unless intrinsically motivated. While one may conjecture that this is particularly true for high-ability subjects, we cannot make hypotheses about the comparative statics between low- and high- ability subjects in the last period(s). There are multiple reasons for not contributing in these repetitions, including not seeing repeated-game incentives from the beginning, and, thus, not sustaining cooperation also in previous rounds – what we expect from low-ability subjects.

Finally, we investigate the interaction between preferences and ability. Running a treatment where we match groups according to preferences and ability was practically unfeasible. Nevertheless, we can test for this interaction at the individual level and at the group level, exploiting the ex-post realized matches.

Conditionally cooperative preferences and ability provide distinct incentives to contribute: an intrinsic incentive to contribute when others do so, and an extrinsic incentive to contribute to promote profitable future cooperation. Both imply a similar conditionality on (beliefs about) others' contributions. We ask if and how the preference-based and the strategic motives interact. We hypothesize that a subject that has an intrinsic preference for conditional cooperation will contribute more when she also sees a strategic reason to do so. Therefore, high-ability conditional cooperators should contribute more than other subjects, and we expect a positive interaction between the preferences and the ability of the individual. Since their contributions reinforce each other, at the group level, homogeneous groups of high ability conditional cooperators should sustain the highest levels of contributions of all groups.

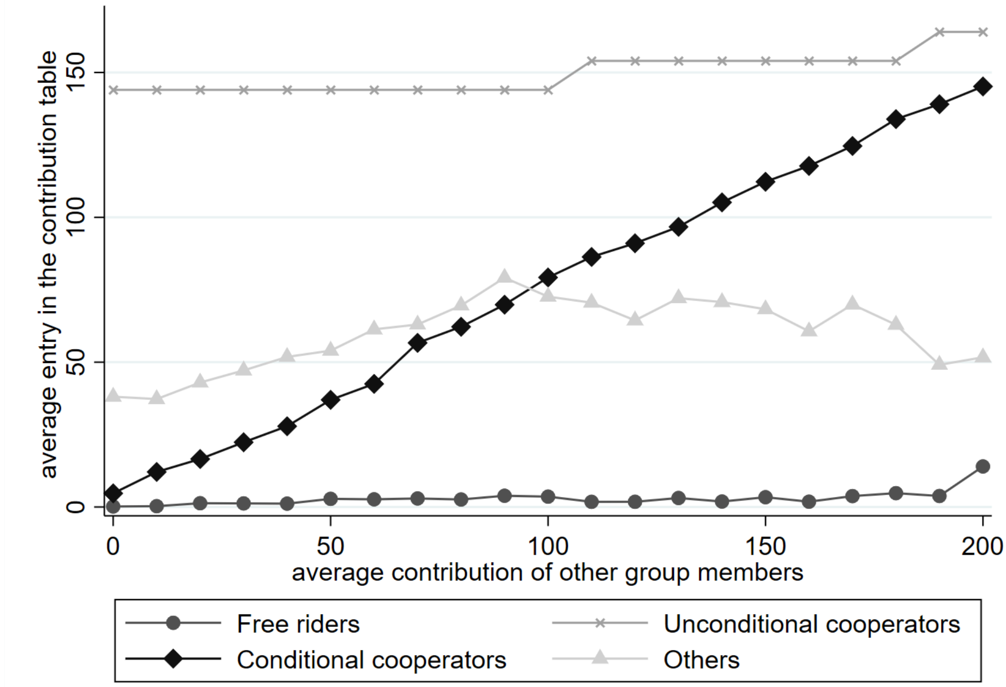
Hypothesis 5. *a) For given contribution of their partners in the previous round, high-ability conditional cooperators contribute more than both low-ability ones and than other high-ability subjects; b) homogeneous groups of high-ability conditional cooperators contribute more than other groups composed by either conditional cooperators or high-ability subjects.*

3 Results

3.1 Classification

We analyze in this section the results obtained in the classification tasks. Two features of the data from the classification tasks can affect the tests of the hypotheses. First, we need to assess the balance across treatments over the preference and the ability dimensions. Second, we need to scrutinize potential correlations between preferences and ability. We find that (i) the proportion of preference types is consistent with previous studies, (ii) the types of subjects are evenly distributed among treatments, (iii) there is no correlation between measured strategic ability and preferences. We are therefore confident about the robustness of our results. Subjects receive no feedback on others' behavior during the classification tasks. Therefore, the analysis in this section is always based on one independent observation per subject.

Figure 1: Contribution patterns of each preference type (strategy method)



Notes: the Figure reports, for each preference type, the average entry in the contribution table (strategy method) against the corresponding average contribution of the other group members.

Table 1 reports summary statistics on the classification tasks. Overall, 30 percent of the subjects are classified as free riders and 50 percent as conditional cooperators. Figure 1 reports the average entries in the contribution table for each preference type (see also the individual contribution tables in Figures C.1 - C.3 in Appendix C). These percentages are in line with previous studies (see the meta-analysis in Thöni and Volk, 2018). Given that we only have 5 unconditional cooperators (3 of low and 2 of high ability), and given the unspecific nature of the residual category, the analysis of preference types focuses on conditional cooperators and free riders. The proportion of free riders and conditional cooperators does not differ significantly in any two treatments.[17] In addition, the unconditional and average conditional contributions of the subjects are not statistically different in any pair of treatments.[18] The average strategic ability of subjects in the whole sample is 44.7 (see Figure C.4 in Appendix C for the full distribution).

[17] Two-sample proportion tests, conditional cooperators: RAND vs PREF, P-val = .41; RAND vs STRAT, P-val = .82; PREF vs STRAT, P-val = .24; free riders: RAND vs PREF, P-val = .24; RAND vs STRAT, P-val = .67; PREF vs STRAT, P-val = .13.

[18] Wilcoxon rank-sum tests, unconditional contribution: RAND vs PREF, P-val = .31; RAND vs STRAT, P-val = .95; PREF vs STRAT, P-val = .30; average conditional contribution: RAND vs PREF, P-val = .12; RAND vs STRAT, P-val = .43; PREF vs STRAT, P-val = .18.

Table 1: Summary statistics by treatment and overall

	RAND	**PREF**	**STRAT**	**Overall**
Number of subjects	48	72	72	192
Free riders	13 (27%)	27 (37%)	17 (24%)	57 (30%)
Conditional cooperators	25 (52%)	32 (44%)	39 (54%)	96 (50%)
Unconditional cooperators	1 (2%)	1 (1%)	3 (4%)	5 (3%)
One-shot Contribution	85.73 (43%)	71.81 (36%)	84.18 (42%)	79.93 (40%)
Contribution table (average)	63.54 (32%)	46.70 (23%)	54.49 (27%)	53.83 (27%)
CRT	34.67	42.67	46.33	42.0
Race game	36.2	50.2	42.8	44.0
Beauty contest	49.8	48.5	50.00	49.4
Average ability	40.27	47.11	45.37	44.75
# of High-ability	23 (48%)	37 (52%)	36 (50%)	96 (50%)

Notes: 'One-shot contribution' is the average unconditional contribution in the one-shot game (endowment = 200). 'Contribution table' is the average entry in the strategy-method version of the one-shot game. 'CRT', 'Race game' and 'Beauty contest' report the average normalized (i.e., on a 0-100 scale) performance in the three strategic-ability tasks. 'Average ability' is the average score obtained in the three ability tasks (on a 0-100 scale). The last rows report the number of high-ability subjects, i.e., those above the median strategic-ability score.

In the various treatments, the average ability of subjects ranges from 40.3 to 47.1 and the proportion of high-ability subjects (those with an ability higher than the median, i.e., 43.4) ranges from 48% to 52%. These differences are not statistically significant.[19]

Table 2 reports statistics on strategic ability for each preference type. On average our measure of strategic ability is stable across types (Wilcoxon rank-sum test: 96 conditional cooperators vs 57 free riders, P-val $= .95$). There are slightly more high-ability subjects among free riders (56%) than among conditional cooperators (47%) but the difference is not significant (Proportion test: 96 conditional cooperators vs 57 free riders, P-val $= .44$). In addition, there is no correlation between the strategic ability of subjects and either one-shot contributions, or average contributions in the contribution table.[20] Finally, scores in the the CRT

[19] Strategic ability, Wilcoxon rank-sum test: RAND vs PREF, P-val $= .20$; RAND vs STRAT, P-val $= .27$; PREF vs STRAT, P-val $= .82$; Number of high-ability subjects, proportion test: RAND vs PREF, P-val $= .71$; RAND vs STRAT, P-val $= .82$; PREF vs STRAT, P-val $= .87$.

[20] On aggregate, we do not find evidence of a clear relation between strategic ability and preferences. We run a number of additional tests looking at each ability task. Among correlations across tasks, the only significant one shows that a better performance in the race game is associated with lower average numbers in the contribution table. Comparing each ability task across preference types, a Kruskall-Wallis test can never reject the null of equality across the four populations. Most pairwise comparisons of types for each ability task also fail to reject the null. Even the two that do not – conditional cooperators perform significantly better than free riders in the CRT, free riders perform significantly better than conditional cooperators in the race game – do

Table 2: Ability of each preference type and correlation across tasks

	Abilities by preference type			
	Free riders	Conditional cooperators	Others	
Average ability	45.62	44.90	43.12	
# of High-ability	32 (56%)	45 (47%)	19 (49%)	
	Correlations across classification tasks			
	Race game	Beauty contest	Unconditional contribution	Conditional contribution
CRT	.32***	.03	.13*	.06
Race game		−.03	−.11	−.20***
Beauty contest			.01	.08
Unconditional contrib.				.46***

Notes: 'Average ability' is the average score obtained in the three ability tasks (on a 0-100 scale). '# of High' report the number of high-ability subjects, i.e., those above the median strategic-ability score. 'CRT' / 'Race game' / 'Beauty contest' refer to the score obtained in these tasks. 'Unconditionl (cond.) contrib.' is the unconditional (average conditional) contribution in the one-shot public good game with strategy method. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively.

and in the Race to 26 are significantly correlated, while their correlation with the score in the beauty contest is not significant. The relatively low correlations between the scores in the three ability measures challenges the consistency of our aggregate measure of strategic ability. The correlations among the scores in the various ability tasks are larger than in Table 2 when restricting the sample to the top and bottom 30 percent of the distribution of strategic ability (CRT-Race game: $\rho = .48$, P-value= .000; CRT-Beauty contest: $\rho = .41$, P-value= .000; Race game-Beauty contest: $\rho = .20$, P-value= .032). This motivates testing our hypothesis about strategic ability also on this restricted sample, given that our measure of strategic ability has a higher consistency there. No correlation between ability tasks and either the unconditional or the conditional contributions remains significant in the restricted sample. We do not find preference types to differ in terms of risk preferences (Kruskall-Wallis: P-value = .718). The same holds for low- and high-ability subjects (Wilcoxon rank-sum test: P-value = .740).

3.2 Behavior in the repeated public goods game

We present in this section our findings on the repeated PGG. We show that (i) high-ability subjects increase more their contribution in the first period with respect to their unconditional contribution in the one-shot PGG (Hypothesis 1); (ii)

not survive correction for multiple hypothesis testing. These findings are consistent with the literature on cognitive ability and preferences. There, results typically depend on how ability and preferences are measured, and fail at establishing a clear relation between the two (see Ben-Ner and Halldorsson, 2010; Burks et al., 2009; Chen et al., 2013; Lohse, 2016; van den Bos et al., 2010).

in later periods, high ability subjects contribute more given the level of others' contributions in the previous period (Hypothesis 2) and (iii) high-ability subjects respond more to group composition relative to low-ability ones (Hypothesis 3). On the preferences side, (iv) matching conditional cooperators with other conditional cooperators is not sufficient to increase their contributions relative to random matching (Hypothesis 4). Nevertheless, the data indicate a significant interaction between preferences and ability: (v) high-ability conditional cooperators contribute more than other high-ability subjects and than low-ability conditional cooperators given the level of others' contributions in the previous period; groups composed by high-ability conditional cooperators contribute more than any other type of group (Hypothesis 5).

The statistical tests related to Hypothesis 1 are based on one independent observation per subject. After the first period, subjects observe the behavior of the other members of their group. Therefore, there is no statistical independence within groups. For Hypotheses 2-5, non-parametric tests are run at the group level, using one independent observation per group. In regressions, non-independence within groups is addressed by clustering errors at the group level.

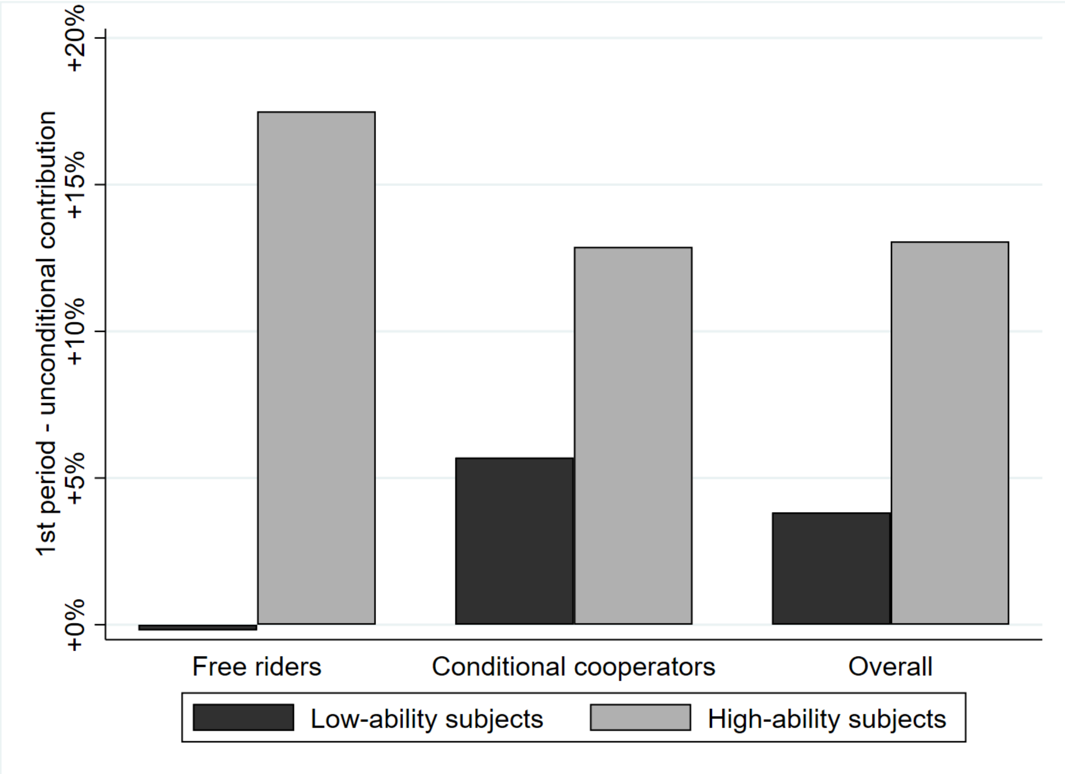
3.2.1 Strategic ability and contributions: Hypotheses 1 and 2

Figure 2 shows the difference between the first-period contribution in the repeated PGG and the unconditional contribution in the one-shot PGG. On average, subjects contribute 49 percent of their endowment in the first period of the repeated PGG, and 40 percent in the one-shot. Subjects' contributions are 22 percent higher when their current behavior may be rewarded in the future. According to Hypothesis 1, this increase should be larger for high-ability subjects than for low-ability ones.

The average contribution in the one-shot game does not depend on the ability: low-ability and high-ability subjects contribute respectively 39.14 and 40.79 percent of their endowment in this task.[21] However, high-ability subjects contribute more than low-ability subjects in the first period of the PGG. The contributions of high-ability subjects in the first period of the PGG are 54.8 percent of their endowment on average, 32 percent higher than in the one-shot game. Low-ability subjects contribute 44.5 percent of their endowment, 14 percent higher than in the one-shot game. A test for equality in the difference-in-differences rejects the null against the two-sided alternative at the 5 percent level both on the whole sample (Wilcoxon rank-sum test: 192 independent observations, P-val= 0.048)

[21] The difference is not statistically significant (Wilcoxon rank-sum test: 192 independent observations, P-val= 0.961).

Figure 2: Difference between repeated (first-period) and one-shot game

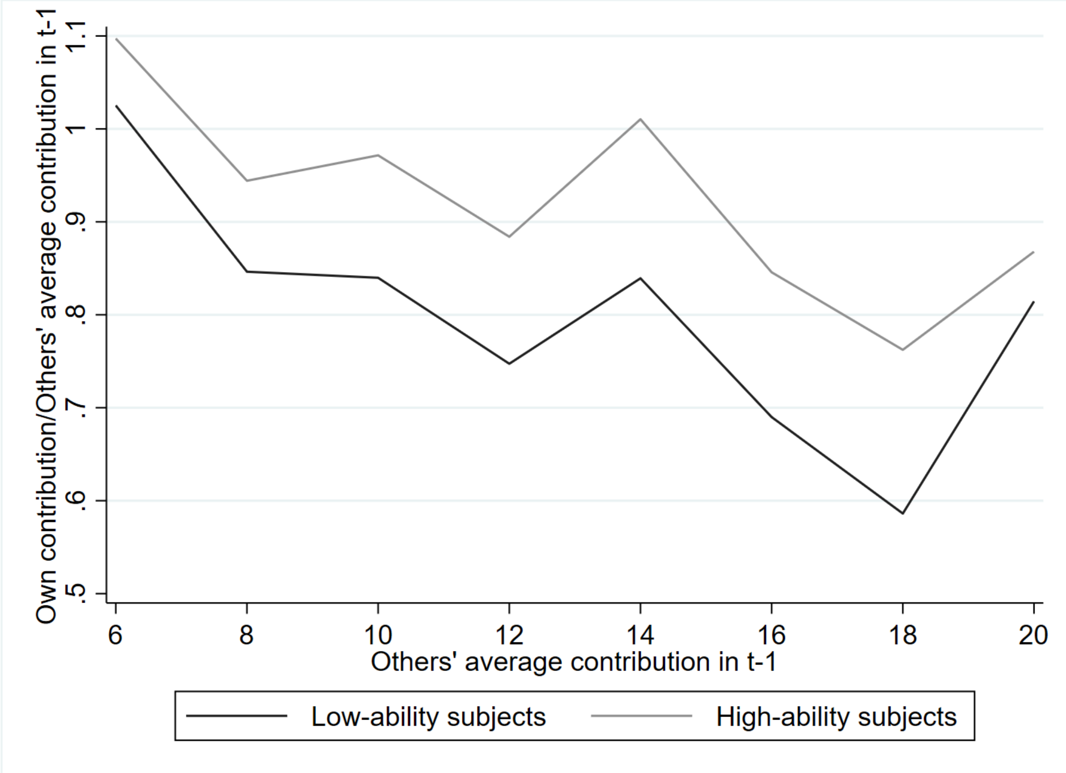


Notes: the Figure shows the average differences between the first-period contributions in the finitely repeated PGG (as a percentage of the endowment) and the unconditional contribution in the one-shot game (as a percentage of the endowment), for the subsamples of conditional cooperators, free riders and for the whole sample.

and when comparing top and bottom 30 percent of the distribution of strategic ability (Wilcoxon rank-sum test: 116 independent observations, P-val$= 0.039$; see also Figure C.5 in Appendix C). Looking at differences between subjects with different preferences, a Kruskall-Wallis test cannot reject the null that increases in contributions in the repeated PGG do not differ across preference types (192 independent observations, P-val$= 0.444$).

We therefore observe that high-ability subjects respond more to repeated-game incentives in the first period of the repeated game. After the first period, and before the last one, there are both a shadow of the future and a shadow of the past: participants can condition their choice on past contributions. Undercutting relative to past contributions would typically trigger further reductions by others, driving contributions down. We hypothesize that high-ability subjects are less undercutting than low-ability ones (Hypothesis 2). Figure 3 shows individuals' contributions as a fraction of the average partners' contributions in the previous period. It is in line with the hypothesis: it shows that, in the first 12 periods, the contribution of a low-ability subject is 24 percent lower than the

Figure 3: Contributions relative to partners' previous-period contributions



Notes: the Figure shows subjects' contributions, as a fraction of their partners' last-period average contribution (y-variable) against the partners' last-period average contribution (x-variable, rounded at the next even number). The Figure shows separate averages for low- and high-ability subjects over the first 12 rounds of the repeated PGG.

contribution of her group members in the past period, while high-ability subjects contribute only 8 percent less. Columns (1)-(2) of Table 3 confirm these findings and support Hypothesis 2. High-ability subjects contribute significantly more (or reduce significantly less their contribution relative to their partners' past contributions) also after controlling for the period, for partners' past contribution and for the subject's preferences.[22] As shown in Columns (1)-(2) of Table 4, we obtain similar conclusions on the restricted sample of top and bottom 30 percent of the distribution of strategic ability (see also Figure C.6 in Appendix C).

We perform the tests of Hypotheses 1 and 2 on each single task, on both the

[22]At the end of the experiment, we asked subjects the following question: "Can you shortly describe your strategy during Task 5?". Subjects stated in words what had determined their decision to contribute or not. We coded all subjects' responses and applied text mining tools for lemmatization and word counting. This allows us to compare the (relative) frequency of each keyword in different treatments and between types of subjects. The analysis shows that the words "betting", "contributing" and "trust" are positively correlated with being of high ability. As an example, high-ability subjects express more often their willingness to contribute to encourage the group members to do so (e.g., "My strategy was to encourage my group to contribute.." or "...induce a dynamics by contributing 20 tokens...").

Table 3: Individual contributions: whole sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var:	Contribution (Fraction)			Contribution (Absolute value)			
Conditional cooperator	0.232**	2.326***	4.215***		2.659***	1.516**	1.327*
	(0.110)	(0.653)	(1.327)		(1.000)	(0.663)	(0.783)
High ability	0.269***	1.495***		-0.791	-0.0182	0.186	0.105
	(0.101)	(0.554)		(1.095)	(1.052)	(0.743)	(0.851)
PREF			1.648	-0.453	0.137		
			(2.066)	(1.784)	(1.614)		
STRAT			1.733	-1.028	-0.608		
			(1.651)	(1.605)	(1.539)		
Conditional cooperator*PREF			-1.984				
			(2.596)				
Conditional cooperator*STRAT			-1.316				
			(1.557)				
High ability*PREF				1.891	1.380		
				(1.785)	(1.705)		
High ability*STRAT				4.546**	3.664**		
				(1.882)	(1.747)		
Conditional cooperator*High ability						1.741**	1.887**
						(0.886)	(0.961)
Others' contribution$_{t-1}$	-0.0458***	0.452***				0.452***	0.448***
	(0.0130)	(0.0430)				(0.0431)	(0.0435)
period	-0.0142	-0.176***	-0.425***	-0.425***	-0.425***	-0.176***	-0.177***
	(0.0185)	(0.0414)	(0.0448)	(0.0448)	(0.0449)	(0.0414)	(0.0417)
_cons	1.327***	3.531***	7.872***	10.430***	3.369	4.137***	-1.773
	(0.226)	(0.635)	(1.365)	(1.266)	(7.287)	(0.672)	(6.108)
N	1899	2112	2880	2880	2880	2112	2112
Controls	NO	NO	NO	NO	YES	NO	YES

Notes: The table reports panel regressions with random effects. The dependent variable is the subject's contribution as a fraction of her partners' last-period average contribution in Column (1). It is the subject's contribution in Columns (2)-(7). 'High ability' and 'Conditional cooperator' are dummies for high-ability subjects and conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. Additional controls include gender, age, field of study, nationality and self-reported risk preferences. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

whole sample and the restricted samples of top and bottom performers.[23] Overall, we find similar support for both hypotheses when tested on ability in the CRT and in the Race to 26. Performance in the Beauty contest does not seem to explain contributions. Detailed results can be found in Tables C.1, C.2 and C.3 in Appendix C.

Investigating endgame effects is useful to understand if the higher contributions by high-ability subjects are due, as we hypothesize, to strategic motives related to the shadow of the future. In the last period there are no repeated-game

[23] The proportion of subjects included in these restricted sample is constrained by the distribution of ability in each task. We end up comparing the top 20% (3 correct answers) and the bottom 35% (0 correct answers) in the CRT, the top 20% (4 or 5 iterations of backward induction) and the bottom 45% (0 or 1 iteration) in the Race game, the top and bottom 30% in the Beauty contest.

Table 4: Individual contributions: top vs bottom 30 percent

Dep. Var:	(1) Contribution (Fraction)	(2)	(3)	(4) Contribution (Absolute value)	(5)	(6)
Conditional cooperator	0.221	2.561***		2.770***	2.487***	2.170**
	(0.135)	(0.784)		(1.054)	(0.830)	(0.959)
High ability	0.214*	1.654**	-0.925	0.493	1.574	2.465*
	(0.123)	(0.724)	(1.461)	(1.951)	(1.063)	(1.335)
PREF			0.179	0.226		
			(1.900)	(1.677)		
STRAT			-0.587	-0.714		
			(1.732)	(1.604)		
High ability*PREF			2.260	2.800		
			(2.188)	(2.464)		
High ability*STRAT			5.107**	4.522*		
			(2.403)	(2.616)		
Conditional cooperator*High ability					0.151	0.205
					(1.322)	(1.473)
Others' contribution$_{t-1}$	-0.0527***	0.448***			0.447***	0.437***
	(0.0190)	(0.0503)			(0.0503)	(0.0506)
period	-0.0291	-0.213***	-0.417***	-0.417***	-0.213***	-0.215***
	(0.0225)	(0.0574)	(0.0603)	(0.0605)	(0.0574)	(0.0581)
_cons	1.539***	3.759***	9.979***	-6.896	3.805***	-10.46*
	(0.358)	(0.865)	(1.288)	(6.781)	(0.860)	(5.958)
N	1146	1276	1740	1740	1276	1276
Controls	NO	NO	NO	YES	NO	YES

Notes: The table reports panel regressions with random effects. The dependent variable is the subject's contribution as a fraction of her partners' last-period average contribution in Column (1). It is the subject's contribution in Columns (2)-(7). 'High ability' is a dummy for subjects in the top 30 percent of the distribution of strategic ability (baseline = bottom 30 percent). 'Conditional cooperator' is a dummy for conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. Additional controls include gender, age, field of study, nationality and self-reported risk preferences. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

incentives, and only intrinsic motives to contribute are present. Our classification of preference types assumes all subjects – i.e., both of low and of high ability – understand the one-shot monetary incentives. Last period contributions are not significantly different between low and high-ability subjects (see Table C.4 in Appendix C, Columns(2)-(4)). Instead, preferences matter (joint F-test on types from Column (1): P-val=0.012) and, in particular, conditional cooperators contribute more than free riders. However, even conditional cooperators contribute less in the last period than they do in the one-shot game. Indeed, 66 percent of conditional cooperators contribute zero in the last period (low ability: 64%; high-ability: 67%). The same figure is 86 percent for free-riders (low ability: 84%; high-ability: 89%).

Even before the last period, the difference between low- and high-ability participants vanishes as the shadow of the future shortens. The results presented in Figure 3 do not hold in the last three periods of the game: low- and high-ability subjects undercut past contribution at a similar rate (39 and 35 percent, respectively). Similarly, when we add an interaction between the high-ability dummy and a dummy variable for the last three periods, we find that being in the last periods has no significant effect above the linear negative time trend for low-ability subjects, while it has a negative and significant effect for high-ability subjects. The results are presented in Table tab:endgame in Appendix C for the whole sample and the restricted sample of top and bottom performers. In the same table we also present regressions on contributions, including a quadratic term for the period. We do so separately for top and bottom performers in the aggregate strategic ability measure and in the race game only, which is expected to capture backward induction ability. We find a negative linear and a negative quadratic term for the bottom performers, both of which are non-significant, resulting in a predicted pattern of almost-linear decay over time. We find a positive linear term and a negative and significant quadratic term for the top performers, resulting in a concave pattern, non-decreasing in the early periods and then decreasing at increasing speed (see Figure C.7 in Appendix C). This evidence on endgame effects, while only exploratory, supports our interpretation of the strategic motives driving the contributions of high-ability subjects.

3.2.2 Treatment effects: Hypotheses 3 and 4

We now turn to the effect of group matching. Treatments PREF and STRAT match together individuals that are like-minded along different dimensions. There are 10 homogeneous groups of conditional cooperators and 8 homogeneous groups of free riders in treatment PREF. We form 10 homogenous groups of high-ability subjects, and 10 of low ability, in treatment STRAT; the 4 remaining groups are mixed. For comparison, in treatment RAND, we form 3 groups of conditional cooperators, 0 groups of free riders, 2 groups of high-ability subjects, 2 of low ability. Within-group variance in contributions decreases and across group variance increases in both treatments relative to the baseline RAND.[24]

Columns (3)-(5) of Table 3 test for the effect of the matching protocol on the contributions of participants with different abilities and preferences. Individuals of High ability respond to treatment STRAT: they contribute more when they are matched with other high-ability individuals than when they are matched at random (Post-estimation test from Column (4); H_0: STRAT + High*STRAT = 0, P-val= 0.039). There is no treatment effect for low-ability subjects. Columns (3)-(4) of Table 4 display a similar analysis comparing the top and bottom 30 percent of the distribution of strategic ability, and come to similar conclusions

[24] This effect is larger in treatment PREF. For instance, the ratio between within and across-group variance is .66 in treatment RAND, .55 in treatment STRAT and .37 in treatment PREF.

(Post-estimation test from Column (3); H_0: STRAT + High*STRAT = 0, P-val= 0.027). We find the same treatment effect when considering top and bottom 30 percent of performers in the CRT, but not those in the other ability tasks (Table C.3 in Appendix C). This result is unsurprising: the treatment is only partially assortative when considering performance in each single task, since it matches subjects according the aggregate ability measure. Indeed, the task that enjoys the highest correlation with the aggregate measure is the CRT ($\rho = 0.73$ in the whole sample, $\rho = 0.85$ in the restricted sample), followed by the Race game ($\rho = 0.63$ and 0.73) and by the Beauty contest ($\rho = 0.52$ and 0.69), and treatment effects for the three tasks are ranked in the same order. For the same reason, we do not report in the Tables treatment effects around the median split for each ability task (all not significant): around one third of subjects above the median in each task is classified as low-ability and matched accordingly, and vice versa, making the interaction with the matching protocol meaningless.

Conditional cooperators contribute more than others in all treatments. Perhaps surprisingly, however, we fail to detect treatment effects for them: they do not contribute more when matched with other conditional cooperators than when they are matched at random (Post-estimation test from Column (3); H_0: PREF + Conditional cooperator*PREF = 0, P-val= 0.859).[25] These results are consistent with what can be visualized in Figures 4 and 5.[26] Overall, we find support for Hypothesis 3, but not for Hypothesis 4: high-ability subjects contribute more in treatment STRAT than in treatment RAND, but conditional cooperators do not contribute more in treatment PREF than in treatment RAND.

Treatment PREF is a conceptual replication of Burlando and Guala (2005). Contrary to their results, our data do not show that conditional cooperators contribute more when matched with other conditional cooperators than when they are matched at random. We conducted a supplemental analysis comparing our data to theirs in order to investigate potential reasons behind these different results. We report it in full in Appendix D, and only mention the two main findings here. First, conditional cooperators appear to be partly different in the two studies in terms of their contribution tables in the one-shot PGG. This is probably driven by the different classification procedures used. In our sample, conditional cooperators show the typical bias toward their own payoff: while they tie their contributions to those of their partners, on average they choose to contribute less

[25]We obtain similar results by estimating all models with 2-limit Tobit regressions (see Table C.6). The results are robust to using the (continuous) ability score or the deciles of the distribution of the ability score rather than the dummy for high-ability subjects. They are also robust to restricting to the first twelve periods. These additional robustness checks are available upon request.

[26]See Figure C.8 in Appendix C for a visualization of the aggregate behavior across treatments.

Figure 4: Average contributions by preference type and treatment

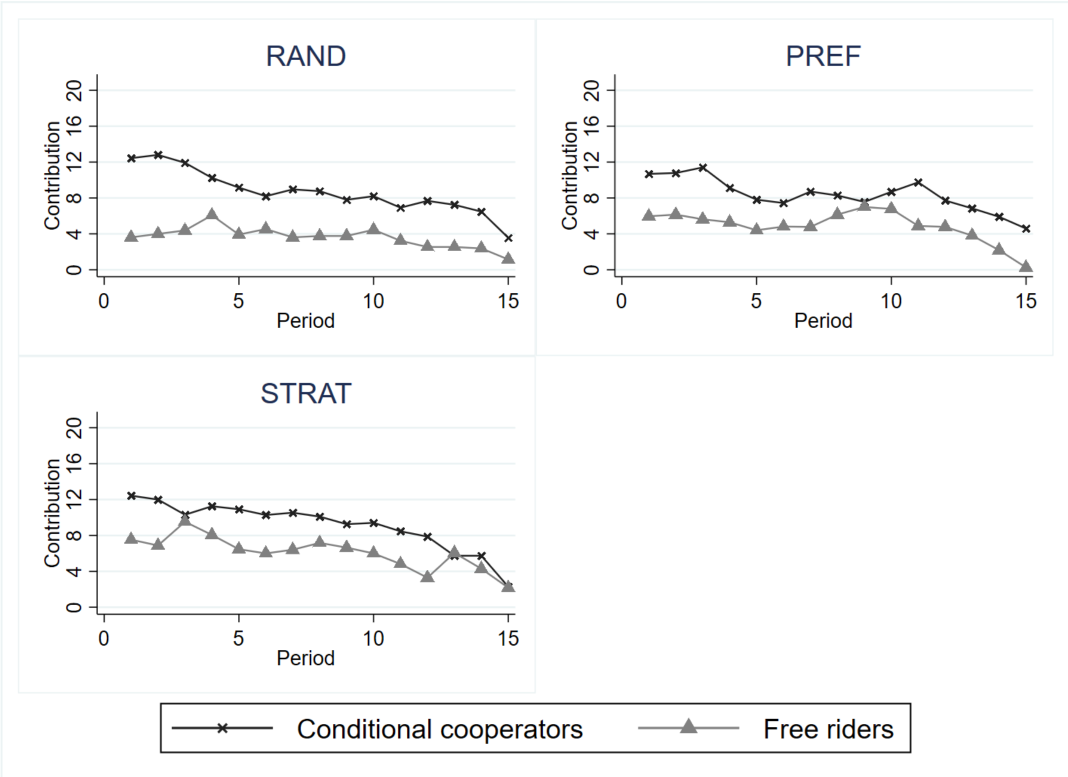


Figure 5: Average contributions by strategic ability and treatment

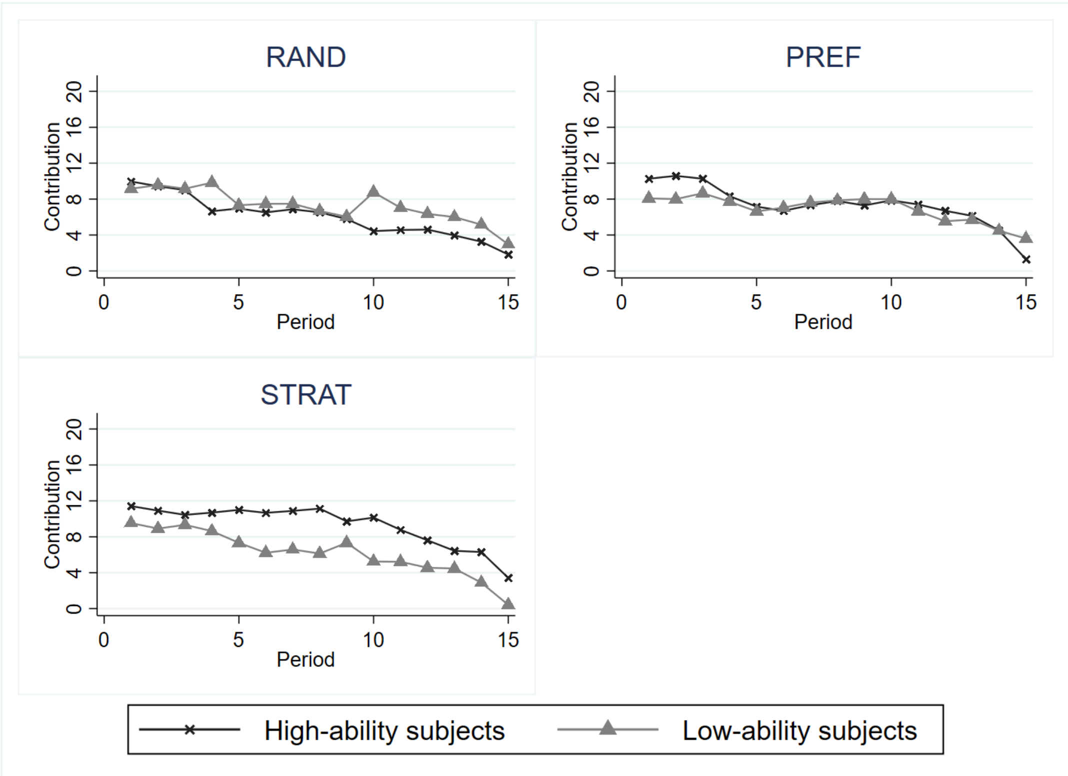
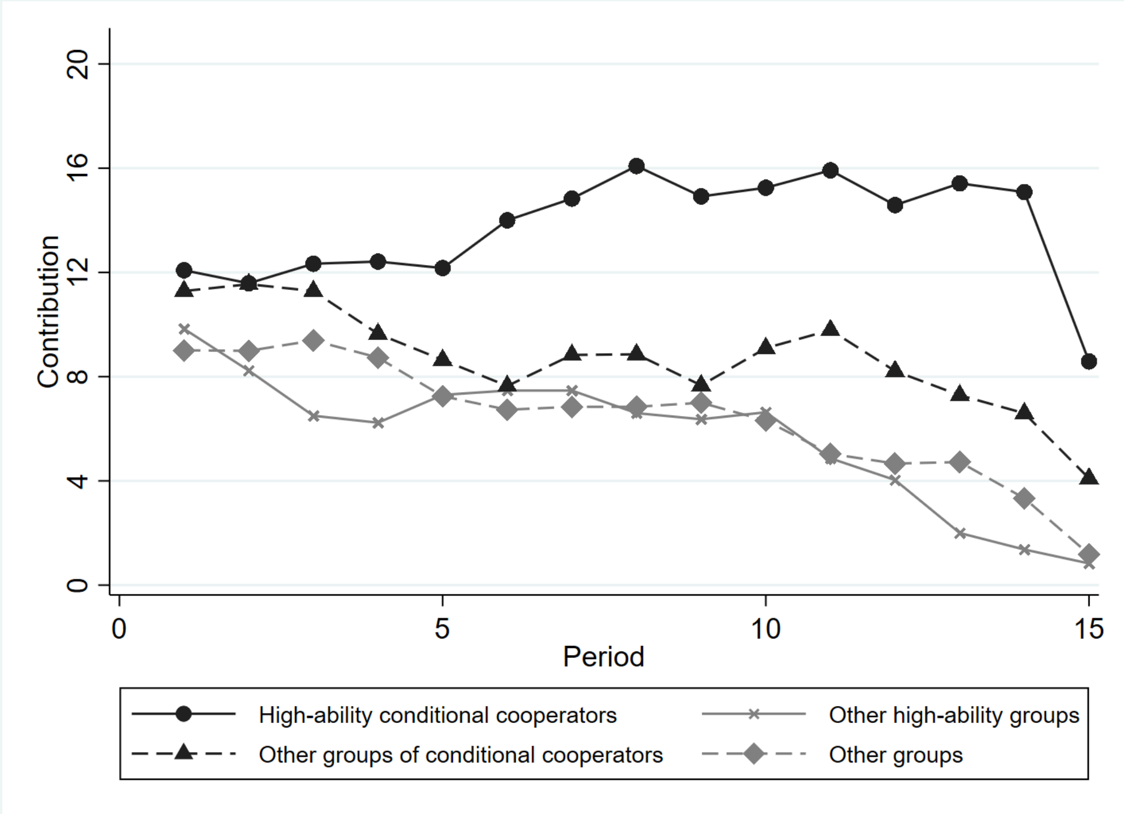


Figure 6: Average contributions in homogenous and heterogeneous groups



Notes: the Figure reports the average contribution over periods in groups of different composition: homogeneous groups of high-ability conditional cooperators, other homogeneous groups of high-ability subjects, other homogeneous groups of conditional cooperators and, finally, all other groups.

than their partners' past contribution. Conditional cooperators in Burlando and Guala (2005)'s sample show no such bias on average. This suggests that they may be more cooperative and perhaps more responsive to others' cooperative traits than our conditional cooperators. Second, in line with this interpretation, we recover results similar to those in Burlando and Guala (2005) when we restrict the analysis to those conditional cooperators in our sample that are similar to those in their sample. These subjects also contribute more when matched in groups that share similar cooperative attitudes. Therefore, one can read our data as qualifying their results, rather than failing to replicate them: matching together conditional cooperators increases contributions only when subjects have sufficiently strong cooperative attitudes. Otherwise, as already noted by Fischbacher and Gachter (2010), imperfect conditional cooperation is sufficient to drive contributions down.

3.2.3 Interaction between preferences and ability: Hypothesis 5

We report regressions on the interaction between preferences and ability (Hypothesis 5) in Column (6) and (7) of Table 3. Conditional cooperators contribute more than other types when comparing subjects of low ability, although the effect is weaker when we add the individual controls in Column (7). On the other hand, high-ability subjects do not contribute more than low-ability ones when they are not conditional cooperators. Both regressions show a positive and significant interaction between preferences and ability. Linear tests of hypotheses on the coefficients show that high-ability conditional cooperators contribute more than both conditional cooperators of low ability (Post-estimation test from Column (6); H_0: High + Conditional cooperator*High = 0, P-val= 0.003) and other high-ability subjects (Post-estimation test from Column (6); H_0: Conditional cooperator + Conditional cooperator*High = 0, P-val= .000). This supports the first part of Hypothesis 5.

The evidence goes in the same direction, though it is weaker, when we look at the restricted sample of top and bottom 30 percent of the distribution of strategic ability. The interaction terms in Columns (5) and (6) of Table 4 are not significant. Nevertheless, it is still true that high-ability conditional cooperators contribute more than both conditional cooperators of low ability (Post-estimation test from Column (5); H_0: High + Conditional cooperator*High = 0, P-val= .050) and other high-ability subjects (Post-estimation test from Column (5); H_0: Conditional cooperator + Conditional cooperator*High = 0, P-val= .026). When we focus on each single ability task (see Table C.3 in Appendix C), we find similar coefficients to those in Column (5) of Table 4 for top performers in both the CRT and the Race game. Tests of hypotheses also give similar results, except that conditional cooperators that are top performers in the Race game do not contribute significantly more than conditional cooperators with a low performance in that game (Post-estimation test from Column (6); H_0: 'Backward Induction $\geq$ 4' + Conditional cooperator*'Backward Induction $\geq$ 4' = 0, P-val= .127; see the Table notes for all post-estimation tests results).

The interaction between preferences and ability at the level of group composition also appears in Figure 6, showing the contributions over time of groups of different composition. It compares groups of high-ability conditional cooperators, groups of high-ability subjects that are not all conditional cooperators, groups of conditional cooperators that are not all of high ability and all other groups (that are not homogeneous in terms of high ability or conditional cooperation). Between period 6 and 14, groups of high-ability conditional cooperators contribute on average 64 percent more than other groups of conditional cooperators, and more than twice as much as the other groups. We test for these differences through non-parametric tests, using one independent observation per group (small-sample exact statistics are reported). A Kruskall-Wallis test rejects the null that group composition does not matter (P-value= .032). Wilcoxon rank-sum tests confirm that homogeneous groups of high-ability conditional cooperators contribute more than a) groups of high-ability subjects that are not all conditional cooperators (P-value= .014) b) groups of conditional cooperators that are not all of high ability (P-value= .047) and c) all other groups (P-value= .011). This result, which supports the second part of Hypothesis 5, is limited by the small number of groups of high-ability conditional cooperators: we only have four such groups.[27] Focusing on top and bottom 30 percent of the distribution of strategic ability further reduces the number of observations.[28] Figure C.10 in Appendix C reports similar results based on a different perspective: we analyze how the subjects' contributions depend on their type (preference and ability) and on the types of their partners.

4 Concluding remarks

Previous studies have highlighted the relevance of matching together people sharing similar cooperative attitudes or preferences to sustain cooperation in finitely repeated interactions. In this paper, we extend these findings and show how like-mindedness of group members matters also in a different sense, related to the common understanding of the strategic features of the game. Our results indicate that groups that sustain high levels of cooperation until the end of the game are formed by subjects that share both a high strategic ability and a positive attitude toward conditional cooperation.

We believe our results represent an important step towards a better understanding of cooperation in finite dynamic interactions, one that incorporates forward-looking strategic thinking and anticipation of others' choices. A number of questions remain to be answered. Some of these stem from limitations of our study. We would like to highlight the most important ones, for the sake of a correct interpretation of our paper.

The construct of strategic ability, and the composite measure we adopt for it,

[27]There are three groups of high-ability free riders. Contributions in these groups are 3.9 on average and are similar to those in other groups of free riders, in line with the non-significant coefficient for high ability in Columns (6) and (7) of Table 3.

[28]Yet, the difference between groups of high-ability conditional cooperators and a) groups of high-ability subjects that are not all conditional cooperators and b) groups where subjects are not all conditional cooperators and not all of high ability are significant at the .1 level. See also Figure C.9 in Appendix C.

is partly challenged by our data, since we observe low correlations across the three tasks, which questions its reliability. We address the issue running a set of robustness checks. We perform our tests both on the whole sample and on a restricted sample of top and bottom performers in the distribution of strategic ability, where the degree of consistency between the three tasks is higher. Results are consistent across the two samples on Hyposeses 1, 2, 3 and 5, although the evidence on the interaction between preferences and ability is weaker.

We also test our hypotheses studying separately the three tasks. The performance in the Beauty contest has no predictive power on our Hypotheses across the board. Ability in the CRT and in the Race game predicts both responding to repeated game incentives by contributing more, and sustaining higher contributions for given contributions of one's partners (Hypotheses 1 and 2). The assortative matching treatment interacts with performance in the CRT (Hypothesis 3). Estimated coefficients have the same sign and order of magnitude for the Race game, but the treatment effect is not significant. This is relatively unsurprising, because the treatment matches according to the aggregate strategic-ability measure. Finally, we find support for Hypothesis 5 when considering ability in the CRT, only partially when considering ability in the Race game.

Taken together, our results on the role of strategic ability in finitely repeated public goods game appear robust. The question of which type of ability is key to sustaining cooperation in those games remains open. The evidence also suggests there exists a relevant interaction between preferences and ability. These latter results are less robust, and are sometimes based on a limited number of observations. A treatment where subjects are matched according to both their preferences and their ability would be a welcome development to validate our finding on groups of high-ability conditional cooperators.

A question parallel to the one for strategic ability exists for preferences. We follow a standard procedure for classification and follow a consolidated practice in inferring 'preferences' from behavior in the contribution table. However, we should consider the possibility that this inference may not always be valid. Through careful instructions and control questions we try to avoid macroscopic confounding factors, such as confusion and misunderstandings (Burton-Chellew et al., 2016). Indeed, conditional cooperators do not differ systematically from others in the ability tasks. Nevertheless, we cannot deny that the questions of who are the 'real' reciprocal players and what is their impact on cooperation in finitely repeated games remain, at least partly, open.

References

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

Baader M., Vostroknutov A., 2017. Interaction of reasoning ability and distributional preferences in a social dilemma. Journal of Economic Behavior & Organization 142, 79–91.

Basteck C., Mantovani M., 2018. Cognitive ability and games of school choice. Games and Economic Behavior 109, 156–183.

Ben-Ner A., Halldorsson F., 2010. Trusting and trusworthinesss: what are they, how to measure them, and what aaffect them. Journal of Economic Psychology 31, 64–79.

Bosch-Rosa C., Meissner T., Bosch-Domènech A., 2018. Cognitive bubbles. Experimental Economics 21, 132–153.

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

Burks S.V., Carpenter J.P., Goette L., Rustichini A., 2009. Cognitive skills affect economic preferences, strategic behavior, and job attachment. Proceedings of the National Academy of Sciences 106, 7745–7750.

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

Burton-Chellew M.N., El Mouden C., West S.A., 2016. Conditional cooperation and confusion in public-goods experiments. Proceedings of the National Academy of Sciences 113, 1291–1296.

Cabral L., Ozbay E.Y., Schotter A., 2014. Intrinsic and instrumental reciprocity: An experimental study. Games and Economic Behavior 87, 100–121.

Chaudhuri A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics 14, 47–83.

Chen C.C., Chiu I., Smith J., Yamada T., et al., 2013. Too smart to be selfish? measures of cognitive ability, social preferences, and consistency. Journal of Economic Behavior & Organization 90, 112–122.

Dohmen T., Falk A., Huffman D., Sunde U., Schupp J., Wagner G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. Journal of the European Economic Association 9, 522–550.

Dreber A., Fudenberg D., Rand D.G., 2014. Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. Journal of Economic Behavior & Organization 98, 41–55.

Drouvelis M., Georgantzis N., 2019. Does revealing personality data affect prosocial behaviour? Journal of Economic Behavior & Organization 159, 409–420.

Drouvelis M., Pearce G., 2021. Leadership under the shadow of future: intelligence and strategy choice in infinitely repeated games. mimeo .

Dufwenberg M., Sundaram R., Butler D.J., 2010. Epiphany in the game of 21. Journal of Economic Behavior & Organization 75, 132–143.

Fallucchi F., Luccasen III R.A., Turocy T.L., 2019. Identifying discrete behavioural types: A re-analysis of public goods game contributions by hierarchical clustering. Journal of the Economic Science Association 5, 238–254.

Fischbacher U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171–178.

Fischbacher U., Gachter S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. American Economic Review 100, 541–56.

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

Frederick S., 2005. Cognitive reflection and decision making. The Journal of Economic Perspectives 19, 25–42.

Gächter S., Thöni C., 2005. Social learning and voluntary cooperation among like-minded people. Journal of the European Economic Association 3, 303–314.

Gill D., Prowse V., 2016. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. Journal of Political Economy 124, 1619–1676.

Gneezy U., Rustichini A., Vostroknutov A., 2010. Experience and insight in the race game. Journal of Economic Behavior & Organization 75, 144–155.

Greiner B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. Journal of the Economic Science Association 1, 114–125.

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

Keser C., van Winden F., 2000. Conditional cooperation and voluntary contributions to public goods. Scandinavian Journal of Economics 102, 23–39.

Kreps D., Milgrom P., Roberts J., Wilson R., 1982. Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic Theory 27, 245–252.

Lambrecht M., Proto E., Rustichini A., Sofianos A., 2021. Intelligence disclosure and cooperation in repeated interactions .

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

Lohse J., 2016. Smart or selfish—when smart guys finish nice. Journal of Behavioral and Experimental Economics 64, 28–40.

Muller L., Sefton M., Steinberg R., Vesterlund L., 2008. Strategic behavior and learning in repeated voluntary contribution experiments. Journal of Economic Behavior & Organization 67, 782–793.

Nagel R., 1995. Unraveling in guessing games: An experimental study. The American Economic Review 85, 1313–1326.

Ones U., Putterman L., 2007. The ecology of collective action: A public goods and sanctions experiment with controlled group formation. Journal of Economic Behavior & Organization 62, 495–521.

Proto E., Rustichini A., Sofianos A., 2019. Intelligence, personality, and gains from cooperation in repeated interactions. Journal of Political Economy 127, 1351–1390.

Reuben E., Suetens S., 2012. Revisiting strategic versus non-strategic cooperation. Experimental Economics 15, 24–43.

Selten R., 1967. Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperimentes. In: H. Sauermann (Ed.), Beiträge Zur Experimentellen Wirtschaftsforschung. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie, Mohr Siebeck GmbH & Co.

Sonnemans J., Schram A., Offerman T., 1999. Strategic behavior in public good games: when partners drift apart. Economics Letters 62, 35–41.

Thöni C., Volk S., 2018. Conditional cooperation: Review and refinement. Economics Letters 171, 37–40.

van den Bos W., Westenberg M., van Dijk E., Crone E.A., 2010. Development of trust and reciprocity in adolescence. Cognitive Development 25, 90–102.

Vieider F.M., Lefebvre M., Bouchouicha R., Chmura T., Hakimov R., Krawczyk M., Martinsson P., 2015. Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. Journal of the European Economic Association 13, 421–452.

A Experimental instructions

Thank you for participating in this experiment in decision-making. In the experiment, your earnings will depend **on both your decisions and those of the other participants**. It is important that you read these instructions carefully. They will help you understand the experiment. All your decisions will remain anonymous. The computer will never ask you to enter your name. It is forbidden to communicate with the other participants during the experiment. If you have questions, please raise your hand, and one of us will come to answer your question at your seat. If you violate this rule, you will be expelled from the experiment and will receive no payment.

The experiment consists of five different tasks. The instructions describe in details the five tasks and will be read aloud. Before taking each task, you will answer a series of comprehension questions to make sure that you fully understand the task. During the experiment, your earnings are calculated in tokens. Among the first four tasks, one will be randomly chosen and your gains in this task will be added to your total gains. Your gains in the fifth task will be automatically added to your total gains.

At the end of the experiment, your total earnings in tokens will be converted at the following rate:

40 tokens = 1 €

Your earnings will then be paid in cash.

TASK 1

During the first task, you have five minutes to answer correctly to three problems that will be displayed on screen.

The problems are of varying difficulty, but for each correct answer, you will earn 100 tokens. Therefore, if you answer correctly to three problems, your earnings will be 300 tokens in this task.

Your earnings in this task will be communicated to you once the whole experiment is finished.

TASK 2

In this task, participants are divided in groups of 3, so that you are in a group with two other participants chosen at random. You will be a member of this group only for this task. You will not know the identity of the other members of your group nor will the other members know your identity. You do not know the identity of the members of the other groups either.

At the beginning of the task, you receive 200 tokens and you must decide how to use this endowment. More precisely, you must decide how many tokens you want to contribute in a common project of the group and how many tokens you want to keep for yourself. Choosing your contribution in the project automatically determines the number of tokens you keep for yourself (200 minus your contribution). Each member of your group makes the same decision, and the total contributions to the common project entitles you to an income. Each token contributed to the project gives 0.6 token to each member of the group.

> *For example, if the total amount contributed by the three members of the group is 600 tokens, each group member receives an income of* $0.6 * 600 = 360$ *tokens. If the total amount contributed by the three members of the group is 150 tokens, each group member receives an income of* $0.6 * 150 = 90$ *tokens from the project.*

Your gains are then the sum of two amounts:

1. The tokens you keep from your endowment.
2. The tokens you earn from the common project.

Your payoff = (200 - contribution to the project) + 0.6 * (total contributions)

When you make your decision, a calculator is available on the screen. It may help you calculate the potential gains from yours and the others' contributions to the project.

In this task, you have to make two types of decision, both regarding your contribution to the project:

A) You have to decide how many of the 200 tokens you want to contribute to the common project.

B) You have to fill out a contribution table. In the table you have to indicate how many tokens you want to contribute to the project for each possible average contribution of the other group members (rounded to the next multiple of ten). You must enter a number between 0 and 200, representing your contribution to the project if the others contribute 0 token, 10 tokens, 20 tokens, etc.

Average others' contribution	0	10	20	30	40	50	60	70	80	90	100
Your contribution											
Average others' contribution	110	120	130	140	150	160	170	180	190	200	
Your contribution											

After all participants made their decisions A and B, in each group a member will be randomly selected. For the randomly chosen subject only, the contribution table (decision B) will be payoff-relevant. For the other group members, their payoffs will be determined according to decision A. Since you do not know whether you are going to be selected, please be careful when making both your decision A and B. Both decisions can be relevant for your payoffs.

The following example should clarify:

> *Example: If you are selected by the computer, payoffs will be determined by your contribution table. For the other two group members, the decision A is relevant. Assume that they have contributed 30 and 90 tokens, respectively. The average contribution of these two subjects is 60 tokens= (30+90)/2. If you indicated in your contribution table that you will contribute 130 tokens when the others contribute 60 tokens on average, then the total contribution to the project is 30 + 90 + 130 = 250 tokens. All group members therefore earn 0.6*250=150 tokens plus the tokens (off the initial endowment of 200 tokens) that they have not contributed to the project. The total payoff for the first member is 200-30+150=320 tokens. The second member earns 200-90+150=260 tokens. And your payoff is 200-130+150=220 tokens. If, instead of 130 tokens, you indicated in the contribution table that you will contribute 180 tokens when the others contribute 60 tokens on average, then the total contribution to the project is 30 + 90 + 180 = 300 tokens. All group members therefore earn 0.6*300=180 tokens from the project. The total payoff for the first member is 200-30+180=350 tokens. The second member earns 200-90+150=260 tokens. And your payoff is 200-130+180=250 tokens.*

The random choice between decision A and B, as well as your payoff will be calculated at the end of the experiment.

TASK 3

In this task, you will be playing a game against a computer opponent. The computer is programmed to play in response to your decision.

The figure hereafter shows the situation you are going to face in this task. The cross indicates the initial position. In this game, you and your opponent start at position 1 (the cross in the grid). You will move first. You will choose a number between 0 and 5 included. This number adds to your current position to determine a new position for both of you. Then the computer will choose a number between 0 and 5 included. This number adds to your current position to determine a new position for both of you. It is then your turn and the game continues with each player taking turns incrementing the current position.

Example: You start in position 1. You choose 4 and you reach the position 5. If the computer then selects 2, you are now both in position 7 (5+2) and it is you turn to decide, etc.

The game continues until one player (you or the computer) reaches position 26. If you reach position 26 first, you earn 200 tokens. If the computer reaches first the final position, you earn nothing.

Before the game starts, you have 90 second to think about your choice. If you find it helpful, you can use the grid below as you like.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13
	X												
Position	14	15	16	17	18	19	20	21	22	23	24	25	26

TASK 4

In this task you are randomly assigned to a group with two other participants. You will be a member of this group only for this task. You will not know the identity of the other members of your group nor will the other members know your identity.

Each group member, as well as you, has to choose a number between 0 and 100. The winner is the group member whose number is closest to 2/3 times the average of all the numbers chosen by the group members. The winner earns 200 tokens.

If there are more than one winner, the 200 tokens are equally split among the winners.

TASK 5

You now participate to the last task. This task includes 15 successive periods and you are assigned to a group of 3 players, so that you are in a group with two other participants. You will not know the identity of the other members of your group nor will the other members know your identity. Your group will remain the same throughout the 15 periods. You are a member of this group only for this task.

The rules of the game are similar to task 2; the only change is the value of your endowment in each period. At the beginning of each period, you receive 20 tokens and you must decide how many tokens you want to contribute to a common project and how many tokens you want to keep for yourself.

Each member of your group makes the same decision and the total contributions to the common project entitles you to an income. Each token contributed to the project gives 0.6 token to each member of the group.

After each member of your group made its decision, you are informed of the total amount invested in the project (i.e., your contribution and that of the others). You are also informed of your earnings in that period. Your earnings in each period are the sum of two amounts:

1. The tokens you keep from your endowment.

2. The tokens you earn from the common project.

Your payoff = (200 - contribution to the project) + 0.6 * (total contributions)

When you make your decision, a calculator is available on the screen. It may help you calculate the potential gains from yours and the others' contributions to the project.

At the end of the period, your earnings for the period will be announced and another period will begin.

Your total earnings for this task are the sum of the tokens you earn throughout the 15 periods.

B Cognitive reflection test questions

1. A bat and a ball cost 1.10 dollars in total. The bat costs 1.00 dollar more than the ball. How much does the ball cost?

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

C Additional tables and figures

Figure C.1: Contribution tables of free riders

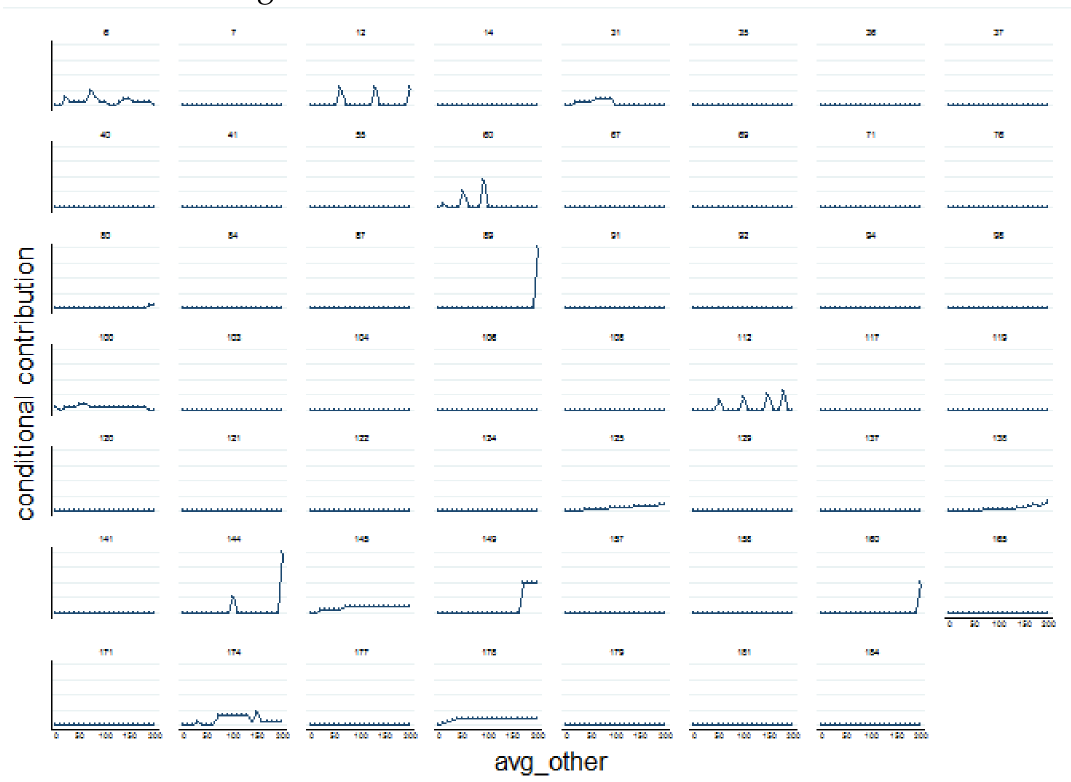


Figure C.2: Contribution tables of conditional cooperators

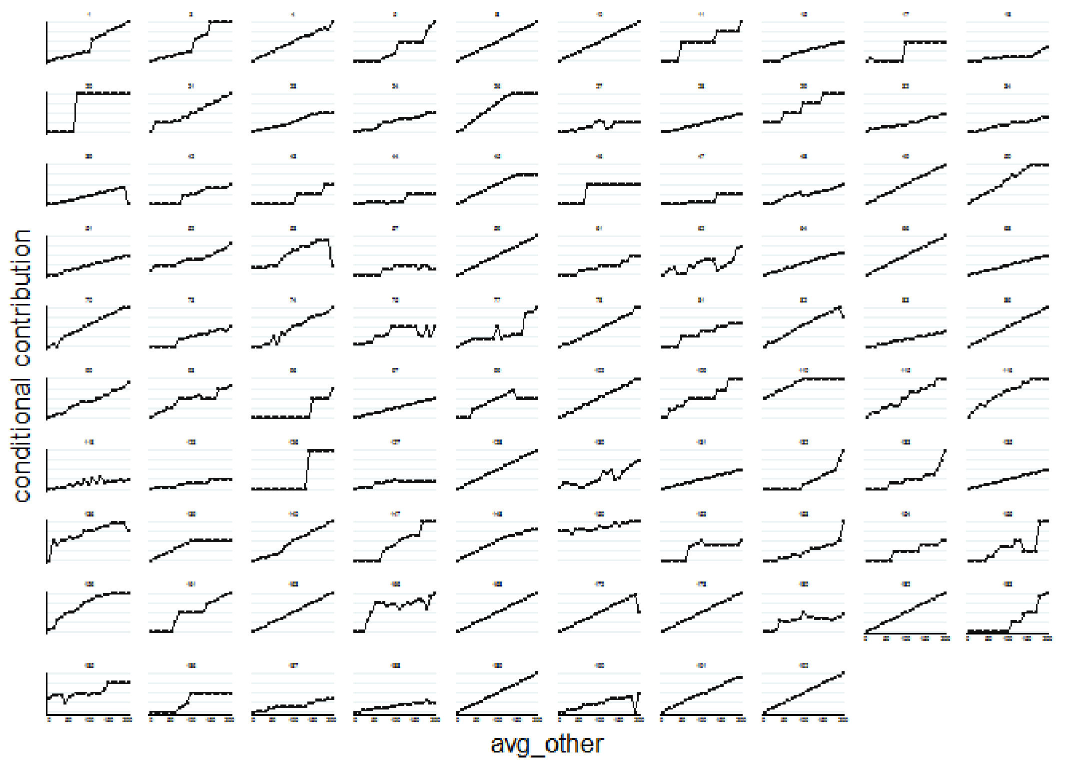


Figure C.3: Contribution tables of the residual category

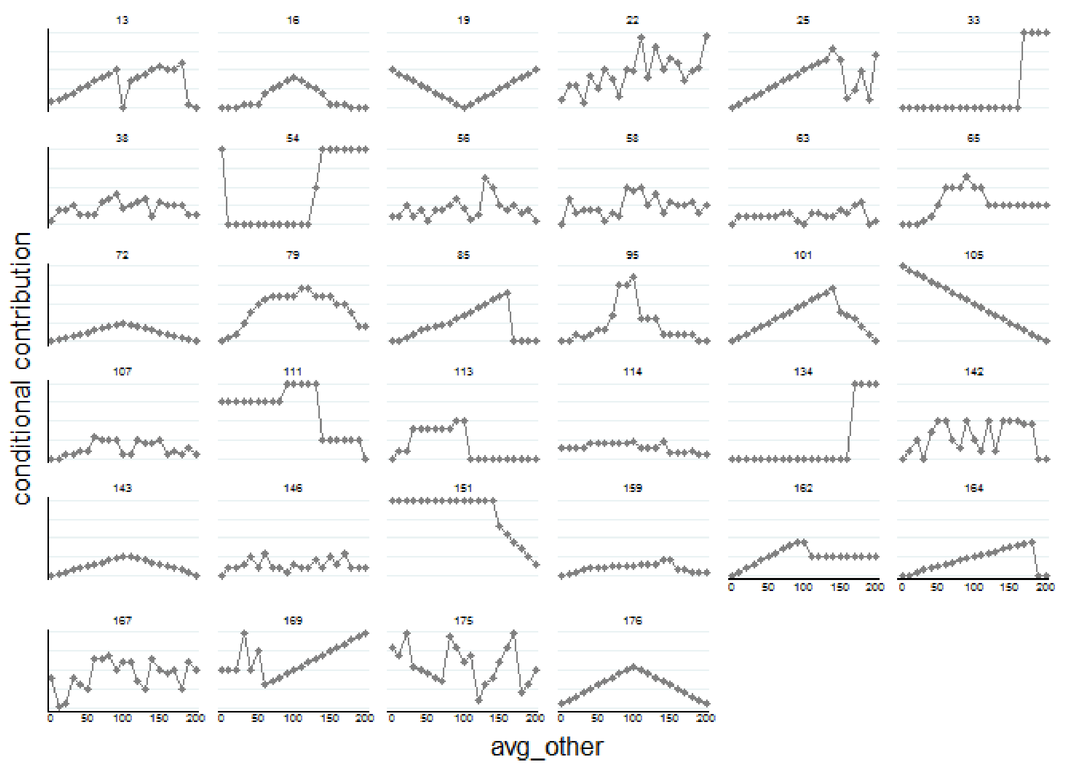


Figure C.4: Distribution of strategic ability

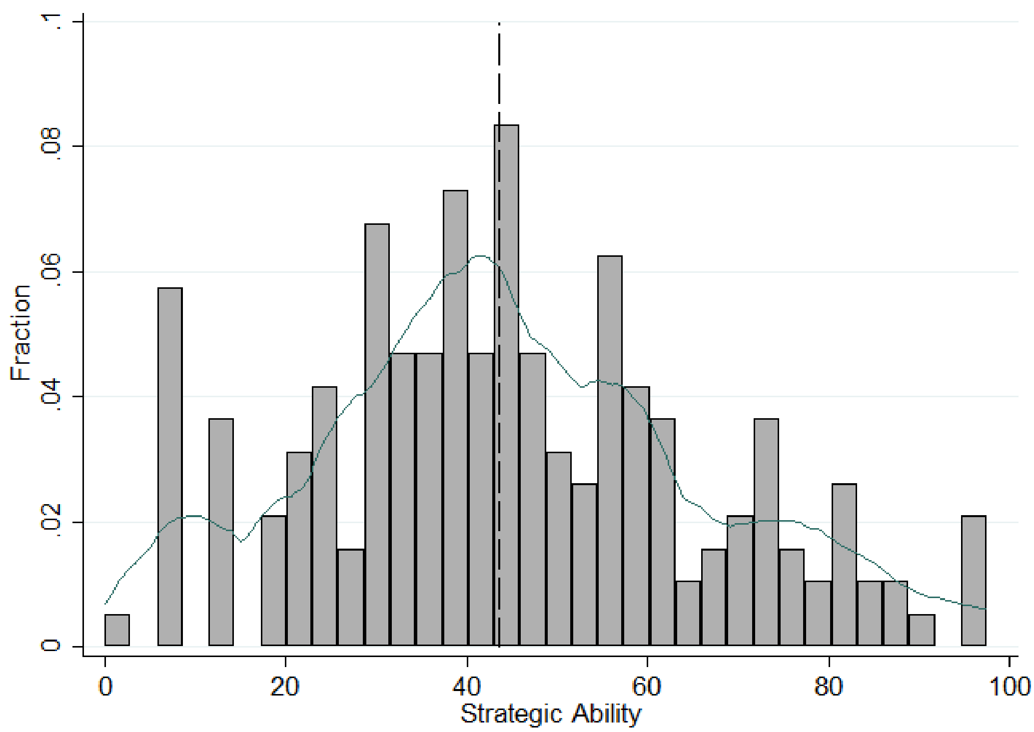
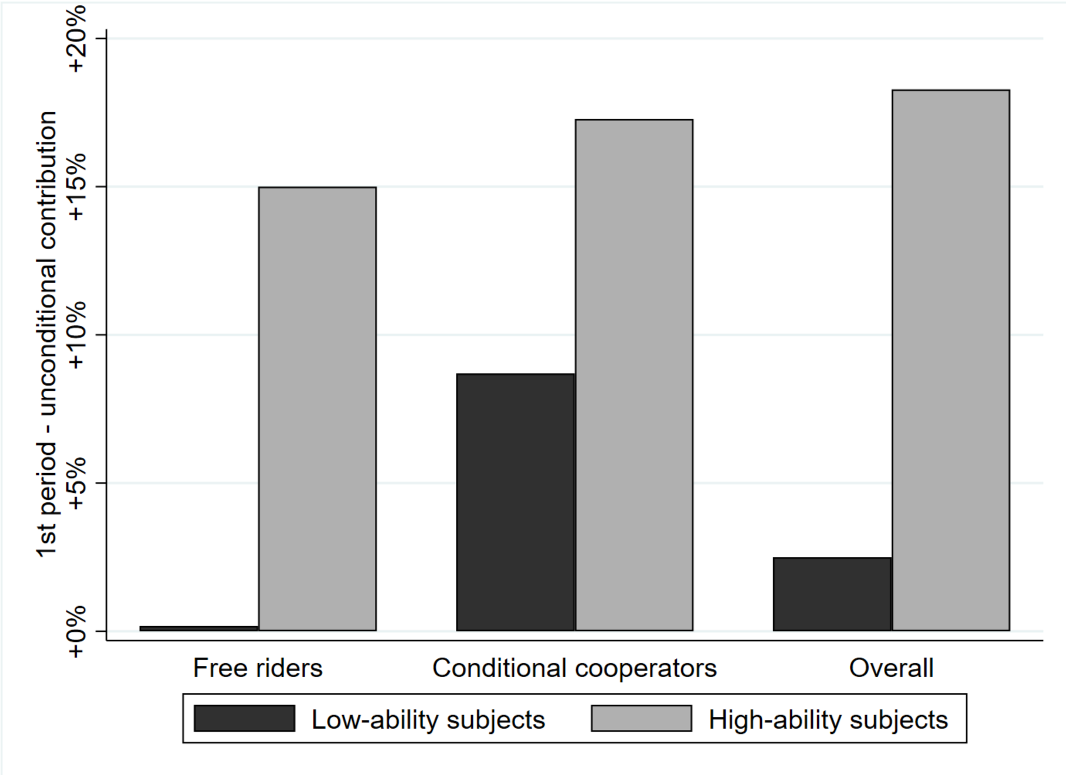


Table C.1: Difference between repeated (first-period) and one-shot game: statistics and tests for various ability measures and samples

		Strategic ability	CRT	Race game	Beauty contest
Whole sample (median split)	Difference low ability	5.4	2.0	3.9	11.4
	Difference high ability	13.0	12.9	14.0	6.9
	WRS test (P-value)	0.048	0.006	0.031	0.206
Restricted sample (top vs bottom performers)	Difference low ability	5.0	1.9	3.9	13.1
	Difference high ability	16.1	17.7	19.7	10.8
	WRS test (P-value)	0.040	0.005	0.004	0.520

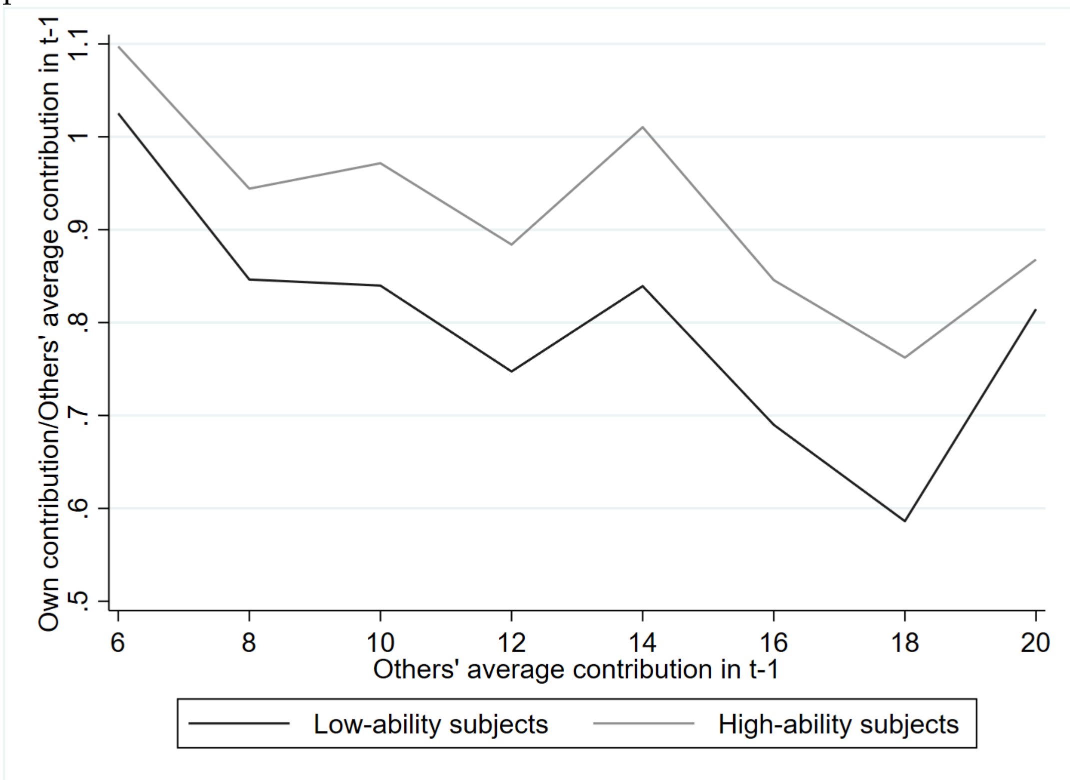
Notes: the table shows the average difference, in percentage points, between the contribution in the first period of the repeated game and the contribution in the one-shot game (both expressed as percentage of the endowment), for subjects of low and high ability. Ability is defined, depending on the column, as the aggregate strategic-ability measure or as the performance in each ability task. For each pair of differences we report the P-value of the Wilcoxon rank-sum test for the difference-in-differences between low and high-ability subjects. The tests use one independent observation per subject. We report the statistics for the whole sample – with low and high ability being defined around the median split – and comparing only bottom and top performers. Those are, respectively: the bottom and top 30% for the aggregate ability measure and for the Beauty contest; subjects with a score of 0/3 and 3/3 in the CRT; subjects with a level of backward induction ≤ 1 and ≥ 4 in the Race game.

Figure C.5: Difference between repeated (first-period) and one-shot game: top and bottom 30%



Notes: the Figure shows the average differences between the first-period contributions in the finitely repeated PGG (as a percentage of the endowment) and the unconditional contribution in the one-shot game (as a percentage of the endowment), for the subsamples of conditional cooperators, free riders and for the whole sample, comparing top and bottom 30 percent of the distribution of strategic ability.

Figure C.6: Contributions relative to partners' previous-period contributions: top and bottom 30%



Notes: the Figure shows subjects' contributions, as a fraction of their partners' last-period average contribution (y-variable) against the partners' last-period average contribution (x-variable, rounded at the next even number). The Figure shows separate averages for the bottom 30 percent (low-ability) and the top 30 percent (high-ability) of the distribution of strategic ability over the first 12 rounds of the repeated PGG.

Table C.2: Individual contributions and performance in ability tasks

	CRT		Race game			Beauty contest		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var:	Contribution (absolute)							
Conditional cooperator	2.186***	2.124***	2.436***	2.487***	2.433***	2.344***	2.343***	2.356***
	(0.673)	(0.684)	(0.654)	(0.644)	(0.649)	(0.666)	(0.664)	(0.670)
Others' contribution$_{t-1}$	0.446***	0.445***	0.451***	0.447***	0.450***	0.450***	0.450***	0.449***
	(0.0427)	(0.0426)	(0.0432)	(0.0425)	(0.0431)	(0.0434)	(0.0435)	(0.0432)
period	-0.178***	-0.178***	-0.176***	-0.177***	-0.176***	-0.177***	-0.176***	-0.177***
	(0.0417)	(0.0418)	(0.0414)	(0.0415)	(0.0415)	(0.0415)	(0.0415)	(0.0416)
CRT (top 50%)	1.433***							
	(0.541)							
CRT score = 1		1.078						
		(0.673)						
CRT score = 2		1.487						
		(0.869)						
CRT score = 3		1.877**						
		(0.813)						
Race (Top 50%)			0.564					
			(0.543)					
Bacward induction =2 steps				0.772				
				(0.707)				
Backward Induction = 3 steps				-0.820				
				(0.735)				
Backward Induction $\geq$ 4 steps				1.451**				
				(0.720)				
Backward Induction (continuous)					0.250*			
					(0.173)			
Beauty (Top 50%)						-0.117		
						(0.603)		
Beauty (continuous score)							-0.0692	
							(0.889)	
Beauty (number chosen)								0.00994
								(0.0115)
_cons	3.418***	3.453***	3.896***	3.910***	3.654***	4.307***	4.281***	3.833***
	(0.618)	(0.619)	(0.630)	(0.634)	(0.674)	(0.667)	(0.752)	(0.769)
N	2112	2112	2112	2112	2112	2112	2112	2112

Notes: The Table reports panel regressions with random effects, similar to Column (2) of Table 3. The only difference is that, rather than a dummy for high-ability subjects (that uses the aggregate measure of ability), it adopts different measures of performance in each of the three ability tasks. The dependent variable is the subject's contribution in periods 2-12. 'Conditional cooperator' is a dummy for subjects that are conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. 'CRT (top 50%)' is a dummy for subjects above the median in the CRT. 'CRT score' is a categorical variable representing the number of correct answers in the CRT. 'Race (top 50%)' is a dummy for subjects above the median in the Race to 26. 'Backward induction' is a categorical variable representing the number of steps of backward induction performed in the Race to 26. 'Backward Induction (continuous)' is the same variable used as continuous. 'Beauty (top 50%)' is a dummy for subjects above the median in the Beauty contest. 'Beauty (continuous score)' is a continuous variable representing the subjects's score in the Beauty contest. 'Beauty (number chosen)' is a continuous variable representing the subjects's choice in the Beauty contest. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

Table C.3: Individual contributions and performance in ability tasks: top versus bottom performers

	CRT			Race game			Beauty contest		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var:	Contribution (absolute)								
Conditional cooperator	2.577***		2.385***	3.007***		2.956***	1.692**		1.972*
	(0.829)		(0.883)	(0.708)		(0.726)	(0.771)		(1.125)
Others' contribution$_{t-1}$	0.433***		0.432***	0.474***		0.473***	0.481***		0.479***
	(0.0597)		(0.0598)	(0.0508)		(0.0506)	(0.0572)		(0.0566)
period	-0.183***	-0.422***	-0.183***	-0.171***	-0.415***	-0.171***	-0.165***	-0.436***	-0.166***
	(0.0649)	(0.0663)	(0.0650)	(0.0504)	(0.0537)	(0.0504)	(0.0577)	(0.0601)	(0.0577)
CRT score = 3	1.785**	1.215	1.430						
	(0.835)	(1.236)	(1.292)						
Backward Induction $\geq$ 4 steps				1.521**	-1.133	1.435			
				(0.702)	(1.821)	(1.153)			
Beauty (Top 30%)							0.320	-0.245	0.594
							(0.706)	(1.477)	(1.095)
PREF		0.351			-0.958			0.602	
		(1.508)			(1.801)			(2.255)	
STRAT		0.477			1.197			1.054	
		(1.240)			(1.810)			(1.832)	
Ability dummy*PREF		1.119			3.294			-0.550	
		(2.484)			(2.601)			(2.408)	
Ability dummy*STRAT		3.658*			2.526			0.866	
		(2.153)			(2.831)			(2.104)	
Conditional coop.*Ability dummy			0.598			0.187			-0.544
			(1.476)			(1.563)			(1.567)
_cons	3.386***	8.731***	3.477***	3.334***	10.48***	3.375***	4.211***	10.45***	4.082***
	(0.792)	(0.999)	(0.804)	(0.698)	(1.286)	(0.676)	(0.874)	(1.461)	(1.003)
N	1100	1500	1100	1408	1920	1408	1265	1725	1265

Notes: The Table reports panel regressions with random effects, similar to Columns (2), (3) and (5) of Table 4. The only difference is that, rather than comparing top and bottom performers according to the aggregate measure of strategic ability, it compares top and bottom performers in each of the three ability tasks. The dependent variable is the subject's contribution. 'Conditional cooperator' is a dummy for subjects that are conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. 'CRT score =3' is a dummy that (approximately) identifies the top 20% of the distribution of CRT scores. The baseline is 'CRT score = 0' and identifies the bottom 33% of the same distribution. 'Backward induction $\geq$ 4' is a dummy that (approximately) identifies the top 20% of the distribution of performance in the Race to 26. The baseline is 'Backward induction $\leq$ 1' and identifies the bottom 47% of the same distribution. 'Beauty (top 30%)' is a dummy for subjects in the top 30% of ability in the Beauty contest. The baseline is 'Beauty (bottom 30%)' and identifies the bottom 30% of the same distribution. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level. **Post-estimation tests relevant for the test of Hypothesis 3** (H_0: STRAT + Ability dummy*STRAT = 0): CRT (Column (2)), P-val= 0.027; Race game (Column (5)), P-val= 0.163; Beauty contest (Column (7)), P-val= 0.307. **Post-estimation tests relevant for the test of Hypothesis 5a**: H_0: High + Conditional cooperator*High = 0: CRT (Column (3)), P-val= 0.032; Race game (Column (6)), P-val= 0.127; Beauty contest (Column (8)), P-val= 0.960; H_0: Conditional cooperator + Conditional cooperator*High = 0: CRT (Column (3)), P-val= 0.033; Race game (Column (6)), P-val= 0.033; Beauty contest (Column (8)), P-val= 0.185.

Table C.4: Individual contributions in the last period

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var:	Contribution in the last period					
Unonditional cooperator	-0.035					
	(1.033)					
Conditional cooperator	2.350***		2.316***	1.308**	3.215***	2.450***
	(0.726)		(0.668)	(0.645)	(0.967)	(0.770)
Other type	0.112					
	(0.785)					
High ability		-0.0208	0.0757	-0.105	0.975	1.050
		(0.774)	(0.710)	(0.526)	(0.699)	(0.633)
Conditional cooperator*High ability					-1.798	-2.318*
					(1.462)	(1.246)
Others' contribution$_{t-1}$				0.361***		0.367***
				(0.0800)		(0.0810)
_cons	1.035***	2.240***	1.034**	0.0135	0.565***	-0.608*
	(0.521)	(0.570)	(0.422)	(0.374)	(0.198)	(0.321)
N	192	192	192	192	192	192

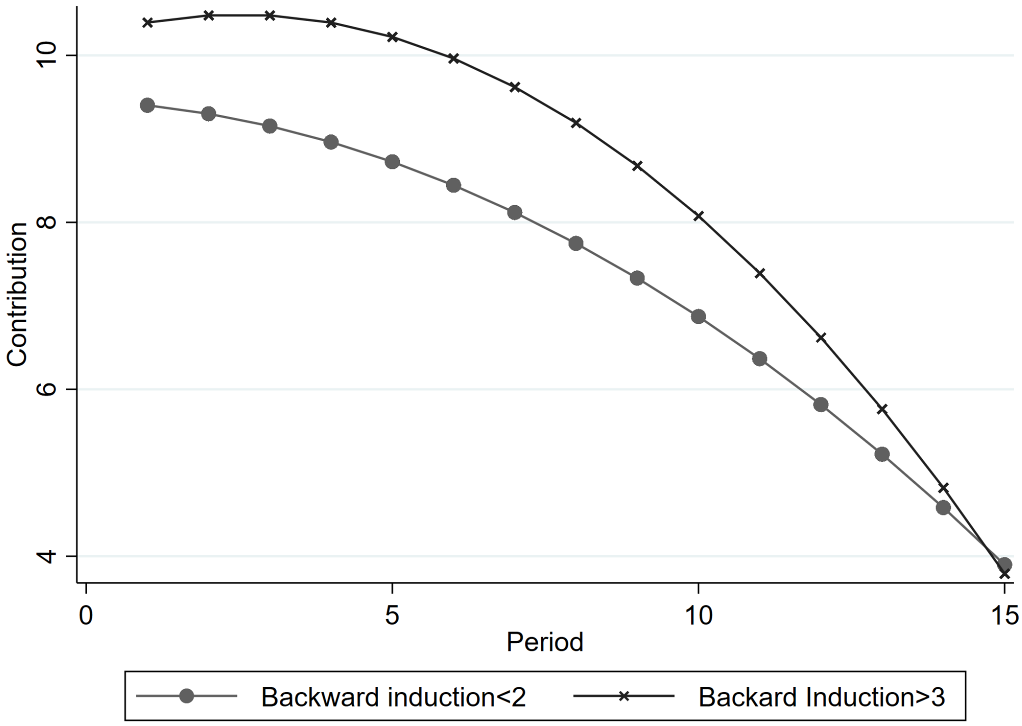
Notes: The table reports regressions on the subjects' contributions in the last-period of the repeated game. 'Unconditional cooperator', 'Conditional cooperator' and 'Other type' are dummies for the various types of preferences (baseline: free rider). 'High ability' is a dummy for high-ability subjects. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

Table C.5: Endgame effects: non linear effects of period for low and high ability subjects

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Contribution (Fraction)		Contribution (Absolute value)			
Conditional cooperator	0.213**	0.172				
	(0.103)	(0.134)				
High ability	0.218**	0.215*				
	(0.101)	(0.122)				
Period $\geq$ 13	-0.0858	0.00914				
	(0.158)	(0.178)				
High ability*Period $\geq$ 13	-0.299**	-0.282*				
	(0.140)	(0.192)				
Others' contribution$_{t-1}$	-0.0435***	-0.0491***				
	(0.0119)	(0.0171)				
period	-0.0179	-0.0331	-0.0349	0.214	-0.174	0.111
	(0.0183)	(0.0223)	(0.251)	(0.334)	(0.276)	(0.286)
period^2			-0.0224	-0.0428**	-0.0140	-0.0342**
			(0.0150)	(0.0201)	(0.0148)	(0.0160)
_cons	1.358***	1.559***	9.460***	10.22***	9.024***	10.61***
	(0.224)	(0.346)	(0.804)	(1.333)	(1.149)	(0.983)
N	2316	1401	1380	540	855	885
Sample	Whole	Restricted	Backward induction<2	Backward Induction>3	Bottom 30%	Top 30%

Notes: The Table reports panel regressions with random effects. The first column refers to the whole sample, the second to the restricted sample of top and bottom performers. The third and fourth columns report separate regressions for bottom (Backward induction<2) and top (Backward induction>3) performers in the Race game. The fifth and sixth columns report separate regressions for bottom 30% and top 30% performers in the strategic ability measure. The dependent variable is the subject's contribution in periods 2-15 as a fraction of her partners' last-period average contribution in Columns (1) and (2). It is the subject's contribution in Columns (3)-(6). 'Conditional cooperator' is a dummy for subjects that are conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. 'Period $\geq$ 13' is a dummy that takes value one in the last three periods. 'Period^2' is a quadratic term for the period.*,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

Figure C.7: Predicted contributions over time of top and bottom performers in the Race game



Notes: the Figure depicts the predicted contributions of subjects that performed at most 1 step (Bakward induction <2) and at least 4 steps (Backward induction >3) of backward induction in the Race game, representing the bottmom 47% and the top 20% of the distribution of performance in that task. The predictions come from the regressions (5) and (6) in Table C.5, which include a quadratic term for the period.

Table C.6: Tobit regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var:	Contribution (Fraction)			Contribution (Absolute value)			
Conditional cooperator	0.446***	3.639***	8.307***	5.710***	5.526***	1.167**	0.918
	(0.136)	(0.897)	(2.847)	(1.808)	(1.910)	(0.497)	(0.596)
High ability	0.287***	1.737**		-1.292	-0.780	0.103	-0.036
	(0.107)	(0.769)		(1.826)	(2.070)	(0.566)	(0.693)
PREF			2.413	-0.212	0.393		
			(4.463)	(3.148)	(3.130)		
STRAT			3.984	-0.597	-0.526		
			(3.529)	(2.934)	(2.921)		
Conditional cooperator*PREF			-2.708				
			(5.100)				
Conditional cooperator*STRAT			-3.500				
			(3.203)				
High ability*PREF				2.316	1.779		
				(3.019)	(3.224)		
High ability*STRAT				5.324*	5.764*		
				(3.263)	(3.361)		
Conditional cooperator*High ability						1.676**	1.905**
						(0.742)	(0.830)
Others' contribution$_{t-1}$	-0.001	1.135***				0.693***	0.694***
	(0.009)	(0.113)				(0.048)	(0.049)
period	-0.027*	-0.210***	-0.884***	-0.880***	-0.881***	-0.108***	-0.108***
	(0.015)	(0.056)	(0.102)	(0.102)	(0.102)	(0.032)	(0.032)
_cons	0.521***	-3.983***	6.573**	8.612***	-0.127	1.909***	-3.935
	(0.185)	(1.198)	(2.896)	(2.522)	(14.332)	(0.517)	(5.250)
N	1899	2112	2880	2880	2880	2112	2112
Controls	NO	NO	NO	NO	YES	NO	YES

Notes: The table reports Tobit regressions similar to the panel regressions reporten in Table 3. The dependent variable is the subject's contribution as a fraction of her partners' last-period average contribution in Column (1). It is the subject's contribution in Columns (2)-(6). 'High ability' and 'Conditional cooperator' are dummies for high-ability subjects and conditional cooperators. 'Others' contribution$_{t-1}$' is the average contribution of one's partners in the last period. Additional controls include gender, age, field of study, nationality and self-reported risk preferences. *,**,***: statistically significant at the 10%, 5% and 1% level, respectively. Between parentheses, we report robust standard errors, clustered at the group level.

Figure C.8: Average contributions by treatment

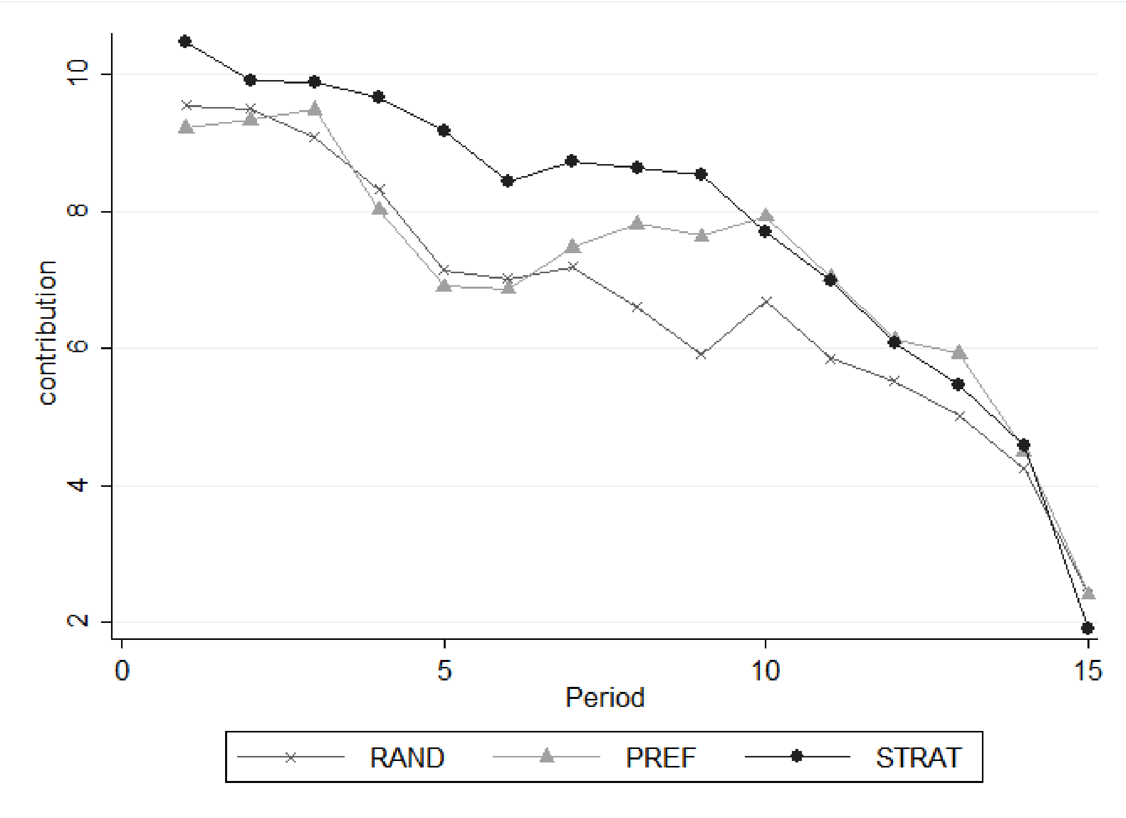
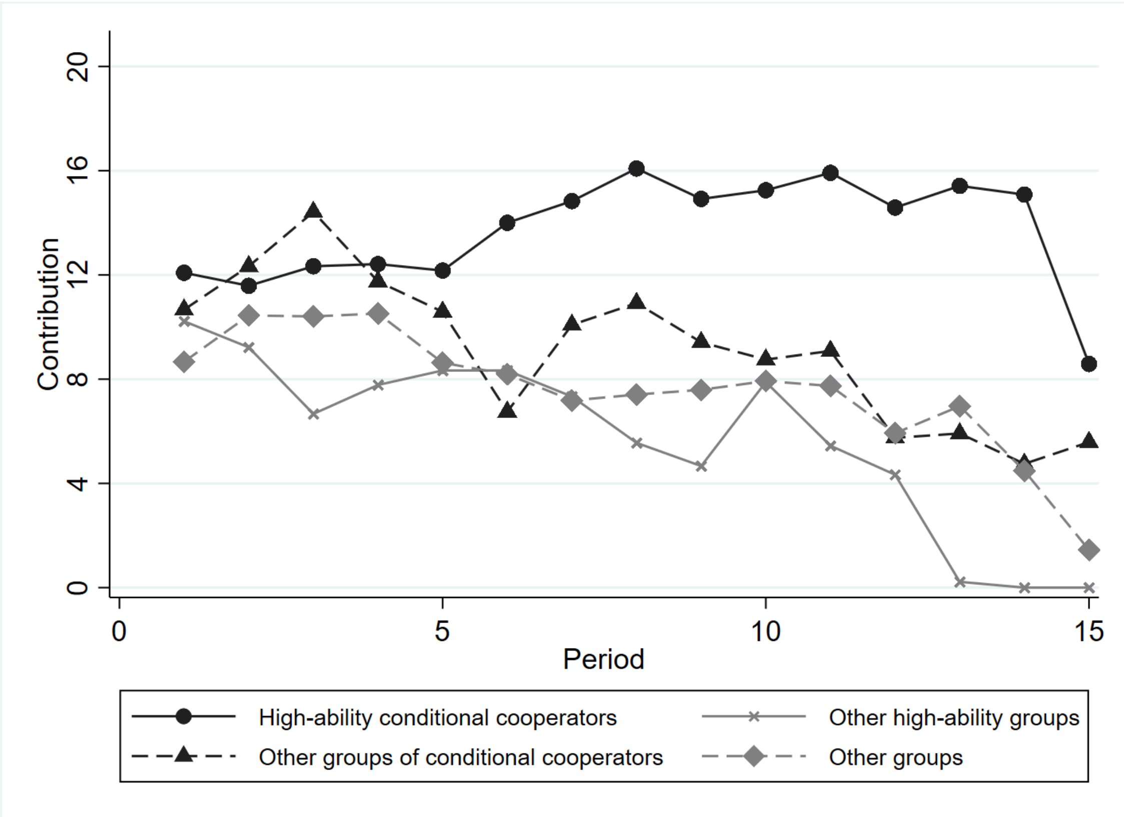
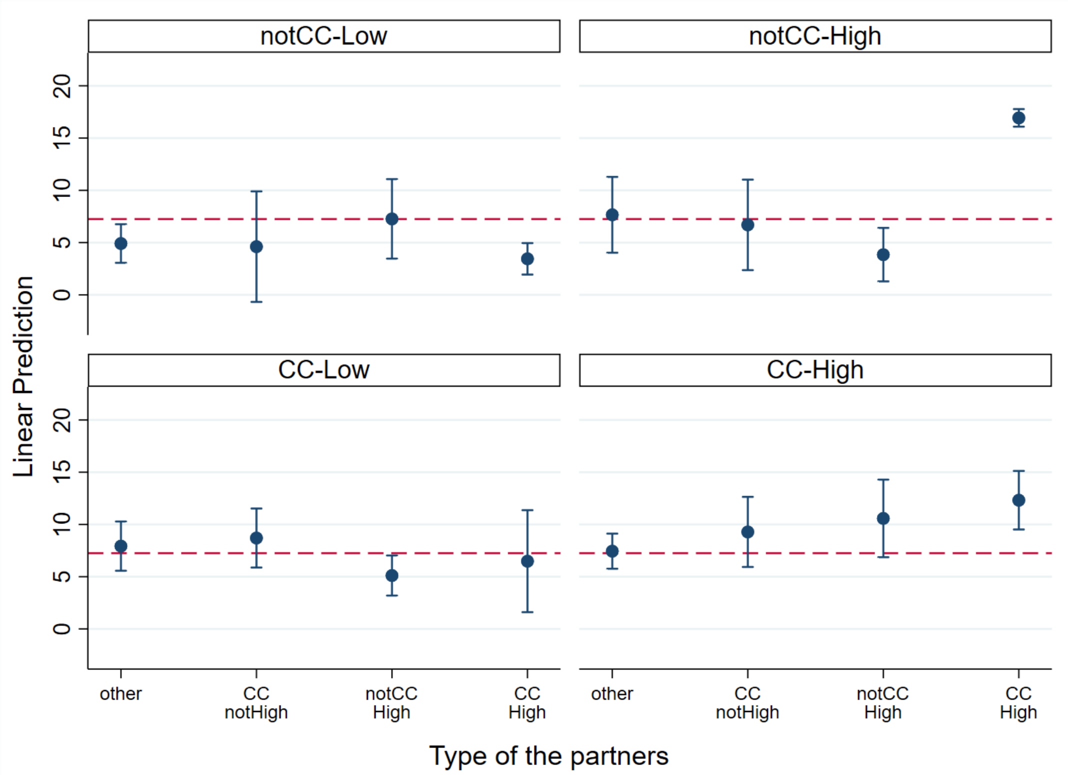


Figure C.9: Average contributions in homogenous and heterogeneous groups: top and bottom 30%



Notes: the Figure reports, for the subsample of bottom and top 30 percent of the distribution of strategic ability, the average contribution over periods in groups of different composition: homogeneous groups of high-ability conditional cooperators, other homogeneous groups of high-ability subjects, other homogeneous groups of conditional cooperators and, finally, all other groups.

Figure C.10: Marginal effects of group composition



Notes: the Figure depicts the linear predictions of the contributions of different types of subjects depending on the type of their partners. The predictions are based on a linear regression of contributions over a triple interaction *preferences* × *ability* × *type of the partners*. The first two terms are the own type of a subject. It can be conditional cooperator of low ability (CC-Low), not conditional cooperators of high-ability (notCC-High), conditional cooperator of high-ability (CC-High) or neither conditional cooperator nor high ability (notCC-Low). We distinguish between partners that are both conditional cooperators but not both of high ability (CC-notHigh), partners that are both of high ability but not both conditional cooperators (notCC-High), partners that are both conditional cooperators and of high ability (CC-High), and a residual category with the partners that are neither both conditional cooperators nor both of high ability (Other)

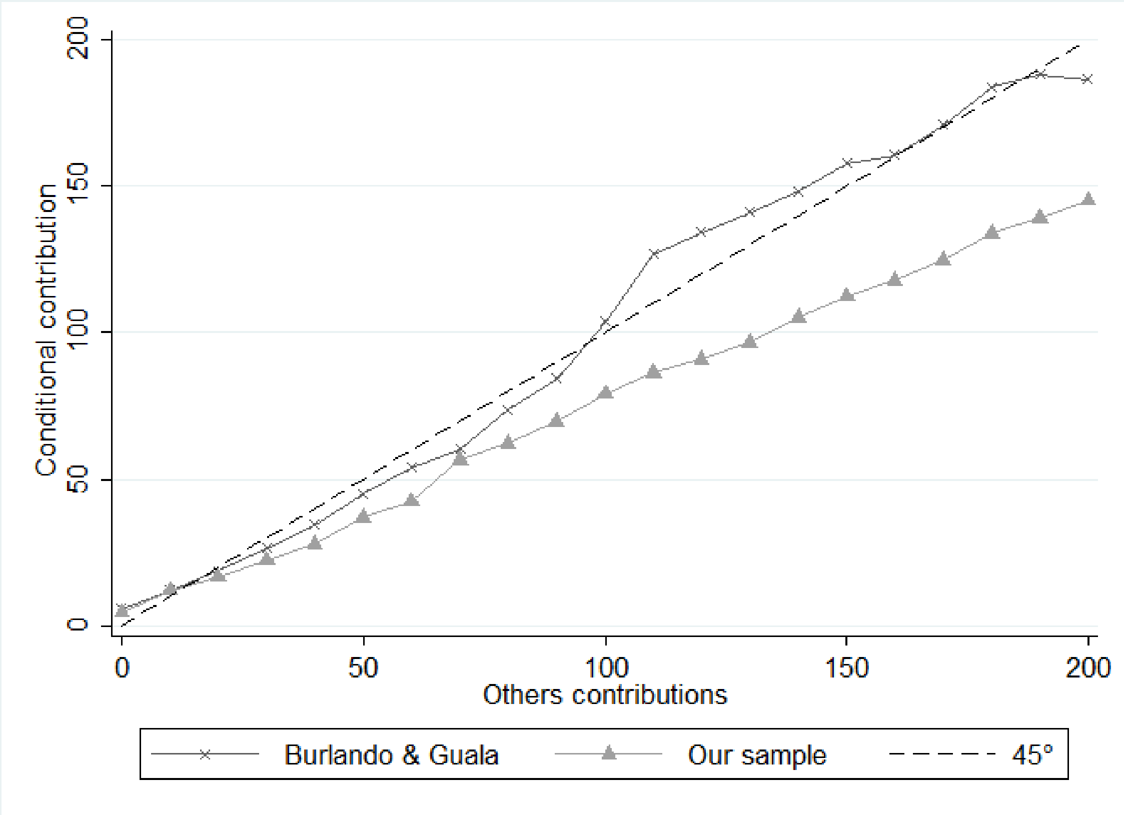
D Discussion: a comparison with Burlando and Guala (2005)

One striking result of our analysis is that, contrary to Burlando and Guala (2005) (hereafter BG), we do not find that conditional cooperators contribute significantly more in homogenous groups of conditional cooperators than in groups formed randomly. We argue in this section that these differences come from the different definitions of conditional cooperation used in the two papers. BG classify subjects according to 4 tasks: a one-shot public goods game in strategy method, a finitely repeated public goods game, a decomposed game and a questionnaire. They classify subjects within each task and attribute weights to the different tasks (20, 20, 40, and 20 percent, respectively). A subject is a conditional cooperator if she is classified as such in tasks that account for at least 50 percent of the weights. In the one-shot public goods game, subjects are classified as conditional cooperators if their conditional contribution pattern lies within a bandwidth of $\pm 10\%$ from the average contribution of their partners. In the finitely repeated public goods game, subjects are classified as conditional cooperators if their average contribution over rounds is within a bandwidth of $\pm 5\%$ of the endowment from the average group contribution. We classify subjects according to a one-shot public goods game in strategy method. Subjects are classified as conditional cooperators if the correlation between their contribution and that of their partners is above 0.7, if they contribute on average more than 10% (otherwise they are classified as free-riders) and if the standard deviation of their contribution pattern is greater than 5% of their endowment (otherwise they are classified as unconditional cooperators).

We argue that their classification selects subjects with a stronger cooperative attitude than ours. Furthermore, if we restrict our attention to groups composed of the conditional cooperators with the highest cooperative attitude in our sample, we do in fact have results that are consistent with theirs.

Their classification selects subjects with a stronger cooperative attitude than ours for two main reasons. First, they select subjects that are willing to match closely others' contributions in the repeated public goods game – that is, in the same game where the homogeneous matching is then applied. Second, they are likely to select subjects who show a low bias toward their own payoff in the strategy method. Figure D.1 shows the average contribution in the strategy method of subjects classified as conditional cooperators in the two samples. The difference between the two groups is apparent, and the difference between the two slopes is significant, as documented in Table D.1 (column (1)). A Wilcoxon rank-sum

Figure D.1: Conditional contributions of conditional cooperators in Burlando and Guala (2005) and in our study



tests also shows that individual average entry in the contribution table of conditional cooperators in BG is significantly higher than in our sample ($z = -3.87$, P-val< 0.01). The median average contribution of a conditional cooperator is significantly higher in BG than it is in our study. These observations confirm that conditional cooperators in BG are more cooperative than in our study.[29]

Finally, we show that homogenous partnerships increase the contributions of conditional cooperators also in our sample, provided that other members of the group have sufficiently strong cooperative attitudes. To do so, we split our sample of conditional cooperators in two categories: the strong and the weak conditional cooperators. *Strong* conditional cooperators are the conditional cooperators whose average entry in the conditional contribution table is above

[29]Our design differs from BG on many aspects, other than the classification method highlighted in the text: we adopt a 'between' design, they adopt a 'within' design; we have groups of three subjects, they have groups of four subjects; the public goods technology is not identical; they have 20 repetitions plus three trial periods, we have 15 repetitions with no trial. While each of these differences could in principle contribute to the difference in results, we believe that the different nature of conditional cooperators in the two experiments is the main explanation of the differences that we observe.

Table D.1: Behavior of conditional cooperators

	(1) Cond. contrib.	(2) Ind. contrib.	(3) Ind. contrib.	(4) Ind. contrib.	(5) Ind. contrib.
others_cont	0.731***				
	(0.009)				
BG	-4.228				
	(7.013)				
BG*other_cont	0.284***				
	(0.020)				
period		-0.468***	-0.583***	-0.468***	-0.583***
		(0.070)	(0.094)	(0.070)	(0.088)
others CC		1.257	1.594		
		(1.413)	(1.653)		
others_CC+				0.697	4.214**
				(1.663)	(1.985)
_cons	3.404	11.825***	13.413***	10.467***	13.009***
	(3.378)	(0.934)	(1.144)	(1.319)	(1.016)
Sample	CC's (BG + GLM)	CC's (GLM)	Strong CC's (GLM)	CC's (GLM)	Strong CC's (GLM)
N	2625	1440	735	1440	735

Notes: the dependent variable is the conditional contribution in the strategy method in model (1). It is the individual contribution in the repeated PGG in models (2) – (5). 'others_cont' is the average contribution of the other group members in the contribution table. 'BG' is a dummy for observations from the Burlando and Guala (2005) dataset. 'others_CC' is a dummy that takes value 1 when one's group members are both conditional cooperators; 'others_CC+' also implies that their average conditional contribution is above the median of the averages of conditional cooperators. In the row 'Sample', 'BG' stands for Burlando and Guala (2005); 'GLM' stands for our sample;'CC's' ('strong CC's') means we are restricting the analysis to subjects classified as conditional cooperators (strong conditional cooperators). In parentheses we report robust standard errors, clustered at the individual level in model (1), at the group level in models (2) – (5). **,***: statistically significant at the 5% and 1% level, respectively.

the median average contribution of the whole set of conditional cooperators.[30] Columns (2) and (3) of Table D.1 show that both (at-large) conditional coopera-

[30]The clustering analysis in Fallucchi et al. (2019) detects the existence of two separate groups of conditional cooperators, the strong and the weak, which are distinguished by how much their conditional contributions are biased toward their own payoff.

tors and strong conditional cooperators do not contribute more when matched in groups of (at-large) conditional cooperators. Since, we do not have homogeneous groups of strong conditional cooperators, we propose a measure of the cooperative attitude of subjects' group members. The group members of a subject are said to be *strongly cooperative* if the average of their mean entries in the conditional contribution table is above the individual median. In column (5) of Table D.1, we see that strong conditional cooperators contribute significantly more when they are matched in strongly cooperative groups than when they are matched in other groups, while we do not observe such difference for (at-large) conditional cooperators (column (4)). We believe that this qualification is relevant, since narrow variations of our definition of conditional cooperation are common in the literature (e.g., Fischbacher et al., 2001).