

May The Forcing Be With You: Experimental Evidence on Mandatory Contributions to Public Goods*

Pietro Battiston[†], Lorán Chollete[‡], and Sharon Harrison[§]

April 2023

Declarations of interest: none

Abstract

We run a repeated public good experiment where subjects are randomly forced to contribute. All group members are made aware when a forcing event occurs. We predict that forced individuals should be more likely to decrease subsequent contributions than the other members of the group, but we show that the sign of the effect is *ex ante* ambiguous, depending on the propensity towards reciprocity. We find that when random forcing is present, intended contributions are significantly larger in absolute terms. However, contributions decrease significantly after being forced to contribute, while they tend to increase after another group member is forced to contribute. Moreover, we find that doubling the probability of forcing does not result in a further increase in contributions. Our results indicate that forcing mechanisms have indirect effects that must be taken into account when assessing the overall impact of policies enforcing contributions to public goods.

Keywords: unintended consequences, public good game, laboratory experiment, reciprocity.

JEL classification: C92, D04, H41.

*We thank Caterina Giannetti, Riccardo Ghidoni, Pietro Guarnieri, Jeffrey Guo, Marco Mantovani, and Silvio Ravaioli, together with seminar participants at University of Pisa and at RegulationResearch Conference 2022 in Regensburg, for valuable feedback. We thank Jillian Harrison for exceptional research assistance. This research was approved by the Barnard College and Columbia University IRBs, and pre-registered at <https://osf.io/59j78>. We acknowledge support from Barnard College, Sharon Harrison, PI.

[†]University of Parma. ORCID 0000-0001-5343-7861 Corresponding author; present address: University of Pisa, Via Cosimo Ridolfi, 56124 Pisa, Italy; email me@pietrobattiston.it

[‡]Welch College of Business and Technology, ORCID 0000-0003-2130-2684, email chollete@sacredheart.edu

[§]Department of Economics, Barnard College, Columbia University, ORCID 0000-0003-0155-6062, email sh411@columbia.edu.

1 Introduction

In a standard public good game, positive externalities are known to motivate agents to contribute, even though they may be able to free ride and still reap the benefits of the good (see Ledyard, 1995; Chaudhuri, 2011 for comprehensive surveys of the literature). In this paper, we study the effects of a policy that forces agents to contribute to the public good. We manipulate the policy assignment over a repeated public good game.

In our experiment, subjects play a standard public good game for 30 rounds, but, in the treated groups, each round, have an exogenous probability of being forced to contribute their entire account. Our experimental design brings to mind the literature on noisy repeated games (Fudenberg and Maskin, 1990; Bereby-Meyer and Roth, 2006; Rand et al., 2015). For instance, Ambrus and Greiner (2012) run a public good game where the choices of participants are occasionally reversed, and Arechar et al. (2017) run a similar study on a repeated prisoners dilemma. Our setup crucially differs in that i) other group members are informed about the manipulation taking place, and ii) the manipulation only goes in one direction: that of increasing the contribution; so its structure is purposely different from that of a random *noise*, and meant to be perceived as such.

Forced contributions in public good games have mostly been studied in the context of threshold public good games (Dawes et al., 1986; Cartwright and Stepanova, 2017). Our study differs in two ways. First, we are interested in forced contributions which are not a consequence of other participants' choices. In a threshold public good game with forced contributions, the decision to contribute is akin to a (costly) vote: hence, the provision of the public good, and the forced contributions that may result from it, can be seen as the result of a collective decision process. In our experiment, by contrast, whether a contribution is forced does not depend in any way on *intended* contributions. Instead, forced contributions are a purely exogenous shock. Second, since our forced contributions are randomly assigned, we can estimate their causal effects. Compared to the literature on mandatory minimum contributions (Andreoni, 1993; Keser et al., 2017), our setup allows us to investigate dynamic effects, and to discriminate between effects on own and other group members' subsequent contributions. Brekke et al. (2011) and Bruns and Perino (2021) study public good games where, in some cases, part of the payoff is subtracted and given to charities. Such contributions are fixed and might resemble our policy intervention, but they do not influence the payoff of group members. Most importantly, the forced contributions implemented by these authors are not exogenously imposed, as participants self-select into such groups. A family of experiments in which individuals are randomly forced to give more than intended is that studying the effect of tax auditing (Kogler et al., 2016; Mittone et al., 2017); however such experiments involve both a specific framing (tax declarations) and a peculiar game structure, where evasion is endogenous and payments do not affect other subjects; similarly, the literature studying climate change in the lab (Ghidoni et al., 2017) shares with our design the presence of random shocks within a PGG, but with different effects on payoffs.

Conceptually, our work is closely related to the issue of reciprocity and conditional cooperation. In the literature on voluntary provision of public goods, reciprocity has long been a central subject of study (Sugden, 1984). Several scholars (including Keser and Van Winden,

2000; Fischbacher et al., 2001; Chaudhuri, 2011 among others) have observed conditional cooperation in public good games, providing evidence that many individuals are more willing to contribute conditional on their peers doing so. One standard explanation for this phenomenon is that of *fairness*: if subjects value negatively large differences in payoff between group members, they have an incentive to contribute more than the Nash equilibrium of zero (Geanakoplos et al., 1989; Rabin, 1993). Other behavioral explanations, however, have been put forward: Bigoni et al. (2018) study the issue of *betrayal aversion* (Cubitt et al., 2017; Bohnet et al., 2008), relating contributions in a public good game to those in a trust game.¹ This allows them to study how individual differences in betrayal aversion explain differences in contributions. However, their experiment does not consider the general question of whether betrayal aversion explains much of the observed conditional cooperation in the general population. Discriminating between betrayal aversion and fairness motives in contributions to public good games requires a distinction between *intended* contributions and *effective* contributions.² In our experiment, when subjects contribute more than they intended to, other group members are made aware of this. Hence, we can observe the extent to which conditional contributors react to actual contributions, as opposed to intentions to contribute.

We are also concerned about the welfare consequences of a policy forcing contributions to a public good; we explicitly look for aggregate effects of the forcing treatment in Section 2.1. While a policymaker’s goal is ideally to increase contributions to the public good, we are interested in the potential *unintended consequences* of this policy. The concept of unintended consequences denotes policy outcomes that differ from the goals of the policymaker, and its applications span the social sciences.³ The definition builds on that of Merton (1936), who applied the concept to sociological matters, and in our design it relates to the question of whether the policy will backfire, crowding out individual motivations and ultimately decreasing contributions.

2 Theory

We analyze a standard public good game where the payoff, π , of player i at time t is given by

¹Betrayal aversion is elicited by comparing the contributions in a trust game to those in an analogous game where the other player’s choice is replaced by a random draw.

²Gächter et al. (2017) show that the tendency to reciprocate varies significantly between different framings of the same game; but even in their design, contributions are, to other participants, indistinguishable from intentions to contribute.

³In Economics, these can be due to a number of different rational and behavioral explanations, which include moral hazard, time inconsistency, and externalities. For example, a well-intentioned policy with negative spillovers may be implemented (Elster, 2017). Such policies may or may not achieve their desired effect, and may also produce unintended outcomes — see Davis (2008), who also studies a random policy, and Antecol et al. (2018). Examples of recent policies with unintended consequences may be found at <http://sharongharrison.blogspot.com/>.

$$\pi_i^t = w_i - \bar{x}_i^t + c\bar{G}^t \quad \text{with } \bar{G}^t = \sum_{j=1}^N \bar{x}_j^t$$

where w_i is player i 's endowment, \bar{x}_i is her contribution, and the marginal per capita return (MPCR) $c > \frac{1}{N}$ determines the importance of the externality effect. In the experiment described later, we will adopt $\omega = 10$, and a multiplier of 1.5, with $c = \frac{1.5}{N} = \frac{1.5}{3} = 0.5$. At each round, a subject can be randomly selected for a forced contribution: when this happens, $\bar{x}_i^t = \omega$. We denote with $x_i^t \leq \bar{x}_i^t$ the *intended* contribution of player i at time t (with $G^t \leq \bar{G}^t$ denoting total intended contributions).

It is well established in the literature that although the only Nash equilibrium (for $c < 1$) envisages zero contributions, experimental subjects frequently choose to contribute non-zero amounts. In the present analysis, we are specifically interested in the possibility that the choice of x_i^t includes an element of *reciprocity* — that is, the desire to imitate other player's kindness, be it observed or expected. As observed by (Falk and Fischbacher, 2006), however, the kindness of an action can be evaluated either according to its *consequences*, or to its author's *intentions*.⁴

While in most economic experiments, including typical public good games, the two dimensions are indistinguishable, a growing literature is both theoretically and empirically disentangling them. For example, (Falk et al., 2003, 2008) compare reactions in sequential two-player games depending on the choice set of the first mover. The general finding of this literature is that intentions matter, although reciprocity also emerges to some extent in reaction to involuntary actions — consistent with early theoretical analyses of the topic (Rabin, 1993).

Two crucial features of our setting are that, as in typical public good games, more than two subjects are involved, and they make their decisions simultaneously. Both aspects enrich significantly the analysis of intentions and consequences. First, when we investigate the extent to which intention or consequences are rewarded, the latter go beyond changes in *own* payoffs, as a peer's contribution has an effect not just on own payoffs, but also on those of other peers. Second, each decision to contribute (except at the first and last round) is *at the same time* a possible reaction to contributions of peers (previous, and expected), and a possible way to trigger peers' reciprocity. So while conceptually the distinction between intentions and consequences precisely replicates that drawn by (Falk and Fischbacher, 2006), the implications for a repeated public good game are non-trivial.

In the literature on public good games, reciprocity has been extensively investigated by eliciting *conditional cooperation* depending on other peers' contributions (Fischbacher et al., 2001). But results are consistent with both reciprocating intentions, and consequences. In this model, as in the experiment later presented, we do not employ strategy methods, and

⁴Reciprocity to intentions is in some sense the symmetric effect to betrayal aversion (Bohnet et al., 2008): whereas the latter involves *distaste* for a payoff reduction due to another subject's choice (rather than pure luck), the former involves *appreciation* for a payoff increase due to another subject's choice (rather than luck). In another sense, it is the counterpart to the warm glow effect (Andreoni, 1995), seen from the point of view of the receiver, rather than of the giver.

participants are placed in the context of uncertainty on peers' contributions that is typical of PGGs. However, to the extent that they have a propensity to reciprocity, they will pick their contribution also based on what they *expect* their peers to contribute (Dufwenberg and Kirchsteiger, 2004). Note that in practice, in a repeated PGG such as the one we implement, such expectation is likely to depend on observed past behavior of one's peers. In what follows, we do not explicitly model such dependence as we are not interested in the level of contributions as such, but only in the difference in contributions due to our treatment, that is, the forcing.

We hence denote as follows the choice of contributions from round 2 onward:

$$x_i^t = x^t(\{G_{-i}^\tau\}_{\tau < t}, \mathbb{E}[G_{-i}^t], S_{-i}^{t-1}, S_i^{t-1}). \quad (1)$$

where $G_{-i}^t = G^t - x_i^t$ denotes aggregate intended contributions by peers, $S_i^{t-1} = \bar{x}_i^{t-1} - x_i^{t-1}$ denotes the extent of forcing on i , and S_{-i}^{t-1} the effect of forcing on i 's peers. Notice that we allow for the function x^t to be period-specific — that is, our formalization can nest any symmetric equilibria (made feasible for instance by altruism, or reciprocity) of repeated public good games (including without forcing, when $S_{-i}^{t-1} = S_i^{t-1} = 0$). As initial condition, we set a common value $x_i^1 = \alpha_0$.

Notice that in our experiment, a maximum of one peer per group and round is forced, so $\min(S_{-i}^{t-1}, S_i^{t-1}) = 0$.⁵

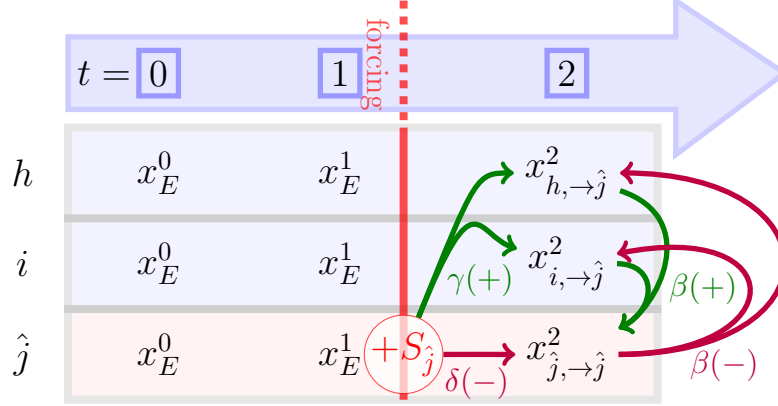
As already mentioned, both the dependence of x^t on its second and its third argument are related to reciprocity. However, the two arguments have markedly different roles with respect to the classification discussed above. Specifically, x^t can be increasing in S_{-i}^{t-1} only as a reaction to the (forced) peer's contribution. Instead, x^t could be increasing in $\mathbb{E}[G_{-i}^t]$ both because subjects reciprocate the peer's (expected) *willingness* to contribute, or because they reciprocate her consequent *actual* contribution (for instance, in our experiment, intended and actual contributions always coincide for at least two peers out of three). In other words, an individual who reciprocates *only* intentions should not reciprocate forced contributions. Vice-versa, an individual who is only interested in consequences should react in the exact same way to the intended component and to the forced component of peers' contributions.

Finally, x^t also depends on S_i^{t-1} because it is likely (for instance according to prospect theory, or even just fairness motives) that being forced to contribute *decreases* subsequent willingness to contribute, that is, that x^t is *decreasing* in its fourth argument. Note that both the reaction to own and peers' forcing could include a (positive) component related to the formation of a *habit* to contribute — which our experiment is not designed to disentangle.

Summing up, our model produces two counteracting effects, that our experiment allows us to compare. Whenever a peer $j \neq i$ is forced to contribute in period $t - 1$, then S_{-i}^{t-1} , the forcing shock affecting i 's peers, increases. At the same time, S_j^{t-1} also increases, and hence in period t individual j will want to decrease x_j which decreases $\mathbb{E}[G_{-i}^t]$. In other

⁵Implicit in our notation is that i knows S_{-i}^{t-1} : while strictly speaking this does not hold in our experiment, subjects observe the forcing event and can easily estimate any S_j^{t-1} from j 's contributions in previous rounds. Similarly, i may not exactly know G_{-i}^τ , but does have an unbiased estimate based on observed intended contributions.

Figure 1: Effects of forcing



Note: example of forcing on group member \hat{j} at period 1.

words, a subject who reciprocates consequences will react to the past — the peer’s forced contribution —, while a subject who reciprocates intentions will react to the expected future decrease in (intended) contribution.

To clarify this tension we consider, and illustrate graphically in Figure 1, a linear formulation of x_i^t that naturally reconnects to our later analysis of experimental data:

$$x_i^t = \alpha_t G_{-i}^{t-1} + \beta \mathbb{E}[G_{-i}^t] + \gamma S_{-i}^{t-1} + \delta S_i^{t-1}. \quad (2)$$

We allow the coefficient α_t to change over rounds in accordance with the abundant evidence that contributions decay over time in repeated PGGs (Andreoni, 1988). Indeed, from Equation (2) we immediately get that, before any forcing event occurs, in equilibrium

$$x_E^t = \alpha_t G_{-i}^{t-1} + \beta \mathbb{E}[G_{-i}^t] = \alpha_t G_{-i}^{t-1} + \beta(N-1)x_E^t \quad (3)$$

$$\implies x_E^t = \frac{\alpha_t G_{-i}^{t-1}}{1 - \beta(N-1)} = \frac{\alpha_t(N-1)x_E^{t-1}}{1 - \beta(N-1)} = \frac{(N-1)^t \prod_{\tau=0}^t \alpha_\tau}{1 - \beta(N-1)} \quad (4)$$

(notice that G_{-i}^{t-1} does not include forcing, and hence consists in $N-1$ individual contributions of x_E^t each), so that, in absence of forcing, x_E^t is uniquely determined by the parameters.

Equation (4) is undefined for $\beta = \frac{1}{N-1}$ and takes negative values for $\beta > \frac{1}{N-1}$. Indeed, for such parameter values each agent wants to contribute more than the average of peers, which is impossible. As a special case, we obtain an indeterminate form if $\alpha_t = 0$ and $\beta = \frac{1}{N-1}$, that is, if subjects are perfect conditional cooperators in expectations — Equation (3) becomes a tautology. In practice, the experimental literature suggests that β is typically distributed between 0 and $\frac{1}{N-1}$, including both extremes (Fischbacher et al., 2001): we can consider β to be on average lower than $\frac{1}{N-1}$.

Moving to the case in which one subject was forced at round $t - 1$, and denoting as $x_{i,\rightarrow j}^t$ the contribution at round t by agent i after agent j was forced, we find that for $i \neq \hat{j}$

$$\begin{aligned} x_{i,\rightarrow \hat{j}}^t &= \alpha_t G_{-i}^{t-1} + \beta \mathbb{E} \left[G_{-i}^t \mid \rightarrow \hat{j} \right] + \gamma S_{-i}^{t-1} \\ &= \alpha_t G_{-i}^{t-1} + \beta (x_{j,\rightarrow \hat{j}}^t + (N - 2)x_{i,\rightarrow \hat{j}}^t) + \gamma S_{-i}^{t-1} \end{aligned} \quad (5)$$

which holds because a subject whose peer was forced expects to interact with that peer plus $N - 2$ other (non-forced) peers. Similarly,

$$\begin{aligned} x_{\hat{j},\rightarrow \hat{j}}^t &= \alpha_t G_{-\hat{j}}^{t-1} + \beta \mathbb{E} \left[G_{-\hat{j}}^t \mid \rightarrow \hat{j} \right] + \delta S_{\hat{j}}^{t-1} \\ &= \alpha_t G_{-\hat{j}}^{t-1} + \beta (N - 1)x_{i,\rightarrow \hat{j}}^t + \delta S_{\hat{j}}^{t-1}. \end{aligned} \quad (6)$$

We define the effect of a peer being forced as $\Delta_p = x_{i,\rightarrow \hat{j}}^t - x_E^t$ and of being oneself forced as $\Delta_o = x_{\hat{j},\rightarrow \hat{j}}^t - x_E^t$. The assumption, discussed above, that being forced to contribute decreases, all else equal, the willingness to contribute in the future can be summarized as:

Assumption 1. $\delta < 0$

which allows us to conclude what follows:

Proposition 1. *Under Assumption 1, $\Delta_p < 0$ implies $\Delta_o < 0$. That is, the effect of forcing on a peer can be negative only if the effect on the forced individual is negative.*

Proof. See Appendix A. □

Clearly, a consequence of Proposition 1 is that $\Delta_o > 0$ implies $\Delta_p > 0$: if the effect of forcing is positive for the forced individual, then this is true also for peers.

The analysis outlined above rests on the assumption that x_E^t are equilibrium values of contributions before any forcing event occurs. However, the algebra developed applies straightforwardly to x_E^t being equilibrium values *conditional* on a history of forcing events. The assumption that these equilibrium values are equal for all group members is not even required: for instance we can abandon it by reformulating Equation (3) as $x_{Ei}^t = \alpha_t G_{-i}^{t-1} + \beta \sum_{j \neq i} x_{Ej}^t$, Equation (5) as $x_{i,\rightarrow \hat{j}}^t = \alpha_t G_{-i}^{t-1} + \beta \sum_{j \neq i} x_{j,\rightarrow \hat{j}}^t + \gamma S_{-i}^{t-1}$ and Equation (6) as $x_{\hat{j},\rightarrow \hat{j}}^t = \alpha_t G_{-\hat{j}}^{t-1} + \beta \sum_{j \neq \hat{j}} x_{j,\rightarrow \hat{j}}^t + \delta S_{\hat{j}}^{t-1}$: Proposition 1 and its proof are unchanged. The only relevant assumption is that the *effects* Δ_o and Δ_p are homogeneous across group members, regardless of the past history of forcing events; this is reasonable precisely because equilibrium contributions can internalize the effect of past forcing events.

2.1 Hypotheses

The first-order effect of forcing is to rise contributions to the public goods, but the above model suggests it can backfire, depending on the parameters values. We now outline hypotheses that express, at different levels of analysis, whether forcing contributions achieves the intended goal, or if instead there are unintended consequences to this policy.

- (H1) If a subject is forced, this affects subsequent contributions of other group members: $x_{j \rightarrow i}^t \neq x_E^t$.

With the linear specification provided by Equation (2), this hypothesis is equivalent to testing the sign of $\beta\delta + \gamma$.

- (H2) If a given round is forced, this affects subsequent contributions of the subject: $x_{\hat{j} \rightarrow \hat{j}}^t \neq x_E^t$.

In the linear specification, this amounts to testing the sign of $\beta\gamma(N-1) + \delta(1 - (N-2)\beta)$. Notice that Proposition 1 predicts that if $x_{\hat{j} \rightarrow \hat{j}}^t > x_E^t$ (as in (H1)), then necessarily $x_{i \rightarrow \hat{j}}^t > x_E^t$.

- (H3) When random forcing is present, average intended contributions differ from when forcing is absent.

This more general hypothesis is not only a combination of hypotheses (H1) and (H2). The average effect of a forcing event is a weighted sum of its effect on the forced subject and on the peers, but the values x_E^t could also be affected by the *expectation* of forcing, and contribute to the overall effect of the forcing scheme.

3 Experimental design

In our experiment, we bring this model into the lab. The sessions were run online, on Zoom, between June 22 and June 30, 2021 under the auspices of the CELSS lab at Columbia University. Subjects were recruited using ORSEE (Greiner, 2015). The experimental software was developed in Otree (Chen et al., 2016).

Each session consisted of 30 rounds. At the beginning of each session, subjects were randomly and anonymously matched in groups of 3, which remained stable for the entire session.

At the beginning of each round, each subject was asked to decide how much to contribute – out of the initial endowment of \$10 – to their group’s common pool. The contribution was restricted to be an integer amount between 0 and 10. Subjects were instructed that the total amount contributed was then multiplied by 1.5 and redistributed in equal shares to the group members (for a marginal per capita return of $1.5/3 = 0.5$). Before the experiment began, subjects answered control questions to verify that they understood the game. At the end of the experiment, each subject was paid a \$5 show up fee, and the earnings from a round drawn randomly by the computer.

Table 1: Summary of experimental sessions

Session	1	2	3	4	5	6	7	8	9	10	11	12	Total
Sample	15	15	12	12	9	12	12	9	18	15	15	12	156
Control	x				x				x			x	54
LOW		x		x		x				x			54
HIGH			x				x	x			x		48

Note: Excluded: three subjects from each of sessions 2 and 8, because of one group member leaving the session prematurely.

We ran 4 sessions with each of three treatments, for a total of 12 sessions, on 156 subjects.⁶ Table Our treatments were: a *control* with no forcing (C) and 2 *policy* treatments, one with a LOW (L) ($p=0.3$) probability of forcing, and one with a HIGH (H) ($p=0.6$) probability of forcing, the exact meanings of which are described below. The numbers of subjects in each session are summarized in Table 1.

We begin by describing the control treatment. In these sessions, there was no forcing. Subjects simply played the game for 30 rounds, deciding how much to contribute each round. At the end of each round, each subject was told their groupmates’ contributions,⁷ how their own payoff was calculated from that round, and the value of their payoff. See the screenshots in Appendix B. The *policy* treatments were the same as the control treatment, except that in every round, each group was, with a probability of either $p=0.3$ (LOW) or $p=0.6$ (HIGH), subject to forcing. When this happened, one randomly chosen group member was forced to contribute their entire endowment, \$10.⁸ Subjects were informed of this before starting the game, and when they were forced, the computer automatically contributed for them. In these treatments, the information provided at the end of every round included whether or not a subject’s groupmate was forced to contribute. Another way to describe the intervention, also included in the experimental instructions, is that each subject was selected with a probability of 0.1 and 0.2, and that whenever a subject was forced, neither of their groupmates was forced in that round. It is worth emphasizing that if p had been even larger than 0.6, we would have observed few rounds unaffected by forcing, making it difficult to study the effects of a forcing event in isolation — that is, on subsequent non-forced rounds.

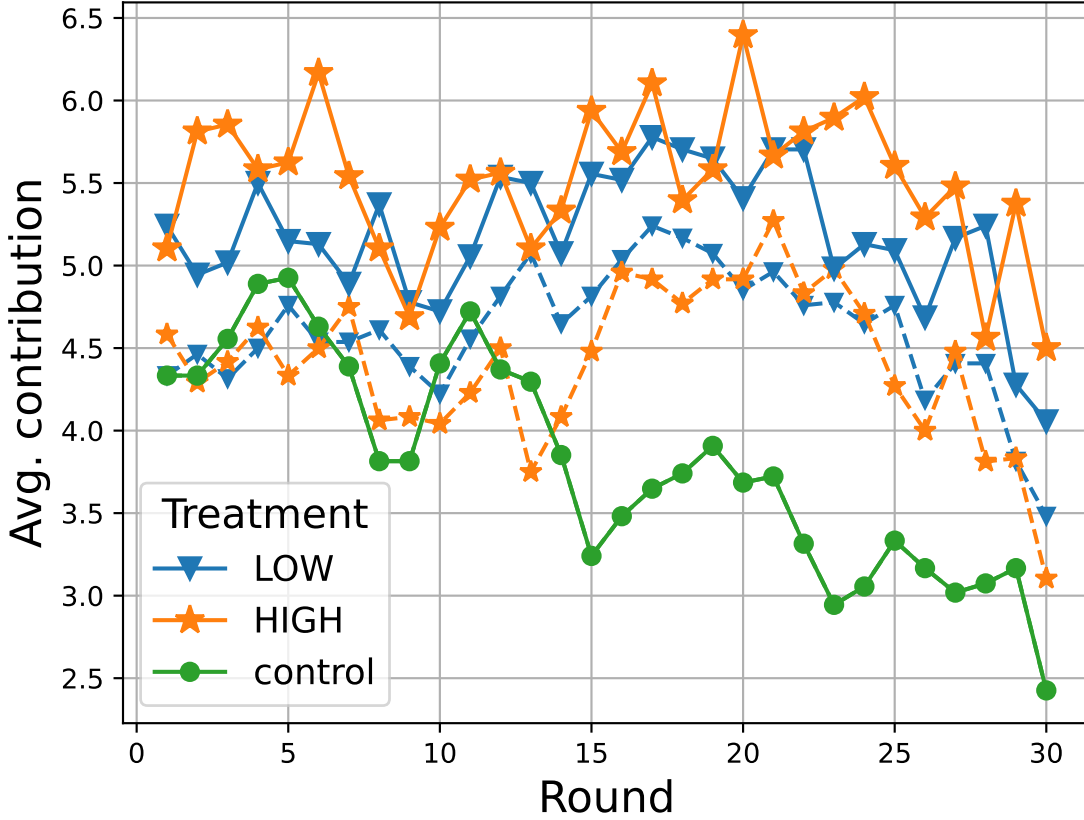
After the experiment, participants completed a short demographics questionnaire, including a question asking to explain choices made. The average payoff was \$16.70, including

⁶Observations from six more subjects were dropped because two of them left the experiment early, invalidating their entire groups’ observations.

⁷It was necessary to provide individual, rather than average, past contributions because, when a peer is forced, the other two peers know it — hence, knowing the average of contributions is equivalent to knowing the individual contribution of the non-forced peer. Since the literature has shown that information on individual contributions can affect behavior differently from information on average contributions (Cason and Khan (1999), footnote 1), this would have represented a confounding effect to forcing.

⁸We purposely avoided framing the intervention in a negative way — e.g. by referring to it as a “punishment” — because our aim was for forced contributions to be closely comparable to intended contributions; in the same spirit, we did not attach to them any sunk costs that are typical of the experimental literature on punishment and tax audits.

Figure 2: Contributions over time



Note: Average contributions per round, disaggregated by treatment. Solid lines: actual contributions (including forced); dashed lines: intended contributions.

the \$5 show up fee; sessions lasted between 43 and 71 minutes. Instructions can be found in Appendix C.

4 Results

Figure 2 displays average contributions over time, depending on treatment. Both LOW and HIGH sessions result in larger contributions than CONTROL: this holds true also for intended contributions (that is, disregarding forcing).

Contributions in LOW and HIGH sessions are relatively similar, and the increase in the forcing probability from 0.3 to 0.6 does not seem to result in a further increase in intended contributions (if anything, the opposite).

Looking at the evolution of contributions over repetitions, treated sessions do not display

any different propensity to contribute in the very first rounds, but they seem more immune to the progressive decrease of contributions that, consistent with the literature on repeated PGGs (Fehr and Gächter, 2000), characterizes CONTROL sessions. In what follows, we confirm and further investigate these observations by employing OLS regressions.

Table 2: Cross-treatment results: actual contributions

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.154*** (0.881)	2.082*** (0.662)	5.131*** (0.874)	1.999*** (0.659)	1.047*** (0.323)	5.741*** (0.876)
Treatment	1.543*** (0.356)	0.640 (0.397)	1.398*** (0.419)	0.251 (0.436)	0.688*** (0.126)	0.646 (0.408)
Treatment \times Peer contr.		0.007 (0.048)		0.038 (0.050)		
Treatment \times Round						0.058*** (0.020)
Peer contr.		0.291*** (0.043)		0.292*** (0.043)		
HIGH			0.311 (0.448)	0.938** (0.438)		
Peer contr. \times HIGH				-0.073* (0.041)		
Own contribution					0.548*** (0.026)	
Peer int. contr.					0.124*** (0.014)	
Shock size					0.127*** (0.013)	
Round	-0.028*** (0.010)	-0.017** (0.008)	-0.028*** (0.010)	-0.017** (0.008)	-0.013*** (0.004)	-0.065*** (0.016)
Observations	4,680	4,524	4,680	4,524	4,524	4,680
R ²	0.055	0.269	0.056	0.272	0.437	0.059

Note: Dependent variable: actual (including forced) contributions; full sample. All featured variables except *Treatment*, *Round* and *H* are lagged. All estimates are run controlling for age, a dummy variable taking value 1 for females and a dummy variable indicating reported ethnicity different from “white”. Clustered standard errors at the subject level in parentheses. Two-sided p -values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We begin by comparing overall contributions between treatment and CONTROL sessions. This is done in Table 2, where the dependent variable, actual contributions, includes forced contributions.⁹ The dummy variables *Treatment* is equal to 1 for both treatments with forcing, while *H* is equal to 1 only for the HIGH treatment. *Peer contr.* denotes average

⁹Note that, while Table 2 includes lagged variables, specifications in columns (1), (3) and (6) are only based on exogenous variables.

contributions of peers in the previous round, *Peer int. contr.* the average of intended contributions, *Shock size* the average difference between actual and intended contributions, and *Own contribution*, own contribution in the previous period.

The coefficient on *Treatment* is always positive. In particular, it is significant in column (1), indicating that absolute contributions are larger with forcing and allowing us to conclude the following.

Result 1. *The presence of forced contributions raises overall contributions.*

In particular, Hypothesis 1 is confirmed, with forced contributions *increasing* intended contributions.

It might seem unsurprising that forcing increases contributions — it has an obvious mechanical effect, given that in general, average contributions are much lower than 10 (see Figure 2). However, this result is relevant in terms of welfare analysis, as it clarifies that the possible unintended effects of the treatment are not strong enough to overturn this mechanical effect. In fact, the treatment coefficient significantly decreases from 1.543 to 0.640, and loses significance, when we control for lagged contributions of peers, and interact this variable with the treatment dummy (column (2)). The reduction in the size of the coefficient on the treatment variable, along with the positive and highly significant coefficient on *Peer contr.* suggests that the higher contributions observed in treatment sessions are at least in part a reaction to higher contributions by peers. Moreover, the interaction coefficient, though it is close to zero and not significant, hints at a bit of a reaction in intended contributions. Indeed, if participants were choosing contributions comparable to those in the control sessions, we would observe a *negative* coefficient here — signaling the tendency to not reciprocate the forced component of a peer’s past contribution. In other words, reciprocity implies a positive relation between the contribution of a subject and the subsequent contributions of her peers: forcing is a random shock which, all else equal, should weaken this relation. However, overall, the degree to which participants reciprocate seems comparable between control and treatment sessions, since the coefficient on this interaction variable is small and not significant whenever it is included.

When disaggregating along the dimension of forcing probability in column (3), the treatment effect is mildly (0.311, non-significant) larger in HIGH sessions, again consistent with participants being forced to give more (regardless of peers’ past contributions – see column (4)).

In column (5), we look in more detail at the reaction to contributions in the previous round, disaggregating them between *Own contribution* and peers’ intended contributions (*Peer int. contr.*) on one side, and the size of the forcing (if any, *Shock size*) on the other. We hence find a strong significant autocorrelation between subjects’ choices, and a comparatively small (but still strongly significant) reaction to peers’ past intended contributions.¹⁰ The size of the forcing also has a relatively small but strongly significant effect. We will analyze

¹⁰It is worth mentioning that subjects do not always observe *Peer int. contr.* (intended contribution of forced subjects remain hidden); but since forcing is random, what subjects observe does *on average* correspond to the intended contributions of a given group at a given round.

this later in more detail, as this variable conflates forcing that affects a participant and forcing that affects her peers.

Finally, column (6) confirms the observation made in Figure 2 that the decay of contributions is attenuated under the forcing scheme — in fact, the sum of coefficients for “Round” and “*Treatment* × *Round*” is close to zero and non-significant ($p = 0.316$). That is, in treated groups, subjects do not contribute less over time.

Table 3: Cross-treatment results: intended contributions

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.156*** (0.952)	2.035*** (0.694)	5.172*** (0.963)	2.024*** (0.708)	0.525* (0.305)	5.755*** (0.946)
Treatment	0.726* (0.377)	-0.691* (0.406)	0.824* (0.446)	-0.704 (0.455)	-0.232** (0.110)	-0.190 (0.419)
Treatment × Peer contr.		0.054 (0.049)		0.076 (0.052)		
Treatment × Round						0.059*** (0.021)
Peer contr.		0.293*** (0.043)		0.294*** (0.043)		
HIGH			-0.210 (0.518)	0.035 (0.492)		
Peer contr. × HIGH				-0.044 (0.047)		
Own contribution					0.615*** (0.027)	
Peer int. contr.					0.144*** (0.013)	
Shock size					0.143*** (0.013)	
Round	-0.027** (0.011)	-0.016** (0.008)	-0.027** (0.011)	-0.016** (0.008)	-0.010*** (0.003)	-0.065*** (0.016)
Observations	4,680	4,524	4,680	4,524	4,524	4,680
R ²	0.027	0.334	0.027	0.338	0.567	0.032

Note: Dependent variable: intended contributions; full sample, except where indicated. All featured variables except *Treatment*, *Round* and *H* are lagged. All estimates are run controlling for age, a dummy variable taking value 1 for females and a dummy variable indicating reported ethnicity different from “white”. Clustered standard errors at the subject level in parentheses. Two-sided p -values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We next look at effects on *intended* contributions in Table 3, which closely mirrors Table 2 except for the different dependent variable. We still find a positive and marginally significant ($p = 0.055$) treatment effect in column (1), this time providing evidence that forcing increases participants’ overall *willingness* to contribute, and allowing us to conclude what follows, and

confirm hypothesis (H3).

Result 2. *The presence of forced contributions raises intended contributions.*

Column (2) reveals that when we condition on peers' previous contributions, interacted with the treatment, the coefficient of the treatment variable becomes negative. Although it is only marginally significant ($p = 0.089$), it represents suggestive evidence that forcing can backfire, in the sense of decreasing *conditional* willingness to contribute. Results are consistent when we disaggregate between the LOW and the HIGH treatment (columns (3) and (4)). Column (5) again disaggregates past contributions into own and peers' intended contributions and the size of the forcing, featuring results analogous to those in Table 2 for the new variables, but a negative and significant sign instead for the treatment variable.

Results from column (6) are in line with those featured in Table 3: specifically, the sum of coefficients for "Round" and "*Treatment* \times *Round*" is close to zero and non-significant ($p = 0.377$), while the interaction coefficient itself is positive and strongly significant, allowing us to state the following.¹¹

Result 3. *The presence of forced contributions neutralizes the decay of contributions over repetitions.*

In order to get deeper insights on the mechanisms at play, and to test the two disaggregated hypotheses (H1) and (H2), we next focus on separately investigating the effect of forcing on a given subject, or on her peers: this is done in Table 4. We again see the positive autocorrelation in contributions, in every column. In columns (1) and (2), which feature a dummy for forced contributions and the size of the forcing effect, respectively,¹² we see that when a subject is forced to contribute, subsequent contributions strongly decrease. The effect of a peer being forced (columns (3) and (4)) is positive and significant, albeit only marginally so for the dummy variable ($p = 0.068$).¹³ Results are consistent, where significant, when we check for own and peer effects simultaneously (columns (5) and (6)).

Result 4. *Forcing results in an immediate decrease in contributions on behalf of the forced subject, which is larger the larger the extent of the forcing.*

This confirms hypothesis (H2); in particular, subsequent contributions of a subject are *negatively* affected by a forcing event. Evidence on hypothesis (H1) also suggests that a forcing event on a peer will actually *increase* subsequent contributions.

Result 5. *Forcing of a peer results in an immediate increase in contributions which is only marginally significant, but significantly increases with the extent of the forcing.*

¹¹Figure 2 suggests that there might be a decay in the very final rounds also for the treatment sessions, apparently compensated by an increase in middle rounds.

¹²While *Forced self* is a random unexpected shock, *Shock size self* is not: the extent of forcing is larger if the intended contribution was lower. Thus, it is important to include own lagged contribution among controls. Analogously, controlling for *Peer int. contr.* compensates for the non-randomness of *Shock size peer*. Notice that since groups have three members, there are twice as many observations with *Shock size peer*=1 than there are with *Shock size self*=1.

¹³Results do not change if we interact the forcing dummy with the forcing frequency (HIGH or LOW): the coefficient for the interaction variable being close to 0. These results are available upon request.

Table 4: Within-treatment results

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	1.622*** (0.478)	1.633*** (0.476)	1.512*** (0.467)	0.366 (0.404)	1.583*** (0.476)	0.342 (0.411)
Forced self	-0.325*** (0.118)				-0.279** (0.125)	
Forced peer			0.192* (0.105)		0.132 (0.110)	
Shock size self		-0.042** (0.018)				0.010 (0.018)
Shock size peer				0.159*** (0.016)		0.159*** (0.016)
Own contribution	0.750*** (0.031)	0.744*** (0.032)	0.751*** (0.032)	0.604*** (0.035)	0.750*** (0.031)	0.605*** (0.035)
Peer int. contr.	-0.002 (0.004)	-0.003 (0.004)	-0.002 (0.004)	0.156*** (0.016)	-0.003 (0.004)	0.157*** (0.017)
Round	-0.009** (0.004)	-0.009** (0.004)	-0.009** (0.004)	-0.009** (0.004)	-0.009** (0.004)	-0.009** (0.004)
Observations	2,958	2,958	2,958	2,958	2,958	2,958
R ²	0.554	0.554	0.554	0.604	0.555	0.604

Note: Dependent variable: intended contributions; sample restricted to sessions with forcing (both LOW and HIGH). All featured variables except *Treatment* and *Round* are lagged. All estimates are run controlling for age, a dummy variable taking value 1 for females and a dummy variable indicating reported ethnicity different from “white”. Clustered standard errors at the subject level in parentheses. Two-sided p -values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.1 Dynamic analysis

Results 2, 4, and 5 might seem contradictory, as forcing is found to increase overall intended contributions, despite individual forcing events strongly *decreasing* forced subjects' subsequent contributions and having a positive but small effect on peers' subsequent contributions. These observations, however, can be easily reconciled in two ways.

First, by trivially observing that in groups of size three, a subject's peer is forced twice as often as the subject is. Second, we run an exploratory analysis of how contributions evolve in rounds that *do not* immediately follow a forcing event. In order to do so, we create a new variable "distance", defined for all observations that follow at least one forcing event at the group level, and recording the distance, in rounds, from the last event; for instance, "distance=1" means that a forcing event took place in the previous round. Looking at values of "distance" (from 0 to 8) for which at least two groups are observed, we find that between two forcing events, and controlling for usual covariates (round, age, female, ethnicity), contributions increase by 0.116 per round on average. When interacting "distance" with another variable "self" that takes value 1 if the last forcing event affected oneself and 0 if it affected a peer, we find that indeed contributions increase on average by 0.124 per round after a peer is forced. These results however are not significant ($p = 0.109$ without interaction with "self", $p = 0.130$ with) when we include clustered standard errors.¹⁴ No similar trend is found for the forced subject's contributions.

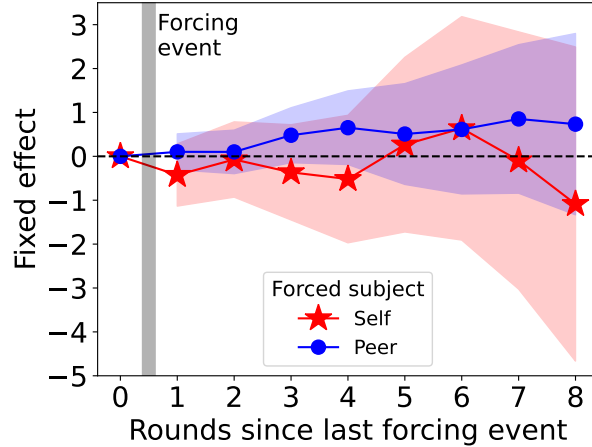
Figure 3 visually displays the same analysis where the "distance" variable was replaced with fixed effects for each value: while individual fixed effects are not significant, the overall increasing trend for "Peer" (referring to the non-interacted "distance" variable) is visible.

Overall, while this evidence must be considered suggestive, as coefficients are not significant to conventional levels with clustered standard errors, it helps to reconcile empirically the apparent inconsistency of results 2, 4 and 5, by showing that the immediate negative effect for the forced subject is more than compensated not just by the immediate positive effect for peers, but also by the following dynamics of contributions. Beyond statistical significance, one word of warning is required concerning the interpretation: what appears to be a delayed, indirect effect of forcing, might also be driven by the *absence* of new forcing events in the following rounds. For instance, subjects may be driven to contribute by the fact that a peer was forced to contribute *and they were not*, afterwards.

Another approach to analyzing the dynamic effects of forcing is to look at how its effectiveness changes over time, that is, to enrich specifications in columns (1) and (3) of Table 4 with the interaction of the forcing dummy and the round. If we do so, we obtain in both cases (with forcing on self and on peers) a negative but small and non-significant coefficient, again not providing any conclusive result.

¹⁴ $p = 0.005$, $p = 0.012$ without clustered standard errors. These results are otherwise robust to the selection of any maximum distance between 5 (observed 60 times) and 8 (observed 18 times).

Figure 3: Contributions over time after a forcing event



Note: Estimated coefficients from a model with fixed effects for the distance (number of periods) from the last forcing event, interacted with a dummy indicating whether “self” or a “peer” was forced. For ease of interpretation, both series are shifted so they take value 0 for distance=0. In other words, the coefficient of the “self” variable interacted with distance=0 (-0.030 , $p = 0.812$) — which is purely random, given the random assignment of forcing — is disregarded.

5 Conclusions

Our experiment shows that interventions that enforce higher contributions to public goods can foster not just overall contributions, but also *intended* contributions. Forced contributions appear to reinforce the trust- and reputation-building effect of intended contributions, as they neutralize the contribution decay over time typically observed in public good games.

In line with predictions of our model, the effect of forcing is more negative on the forced subject than on peers. Specifically, subjects react to being forced to contribute by *decreasing* subsequent contributions; instead, the larger the forcing of a subject, the higher the subsequent contributions of peers. The results also shed new light on the phenomenon of conditional cooperation. Subjects react to *actual* contributions in the same direction but to a much lower extent than if they were intended, which implies that a propensity to reciprocate intentions, and not just consequences, seems to be a key driver for conditional cooperation.

Increasing the frequency of forcing events does not result in a proportional effect on intended contributions: in fact, if any, doubling the probability of forcing reduces this effect. This suggests that experimenting with different probability of forcing could reveal the optimal policy from the point of view of social welfare.

Examining the potential unintended consequences of a forced contribution in a repeated public good game provides an understanding of the behavioral effects of a policy of this kind. Clearly, having purely random forced contributions is a design choice made for ease of interpretation rather than for realism; however, it mimics a large number of real world

instantances, from tax audits to eligibility thresholds, in which the obligation to contribute to a public good is heterogeneous in the population. Our results can help policymakers evaluate the net effects of policies concerning public goods.

References

- Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review* 102(7), 3317–32.
- Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics* 37(3), 291–304.
- Andreoni, J. (1993). An experimental test of the public-goods crowding-out hypothesis. *The American economic review*, 1317–1327.
- Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics* 110(1), 1–21.
- Antecol, H., K. Bedard, and J. Stearns (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies? *American Economic Review* 108(9), 2420–2441.
- Arechar, A. A., A. Dreber, D. Fudenberg, and D. G. Rand (2017). “I’m just a soul whose intentions are good”: The role of communication in noisy repeated games. *Games and Economic Behavior* 104, 726–743.
- Bereby-Meyer, Y. and A. E. Roth (2006). The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *American Economic Review* 96(4), 1029–1042.
- Bigoni, M., S. Bortolotti, M. Casari, and D. Gambetta (2018). At the Root of the North–South Cooperation Gap in Italy: Preferences or Beliefs? *The Economic Journal* 129(619), 1139–1152.
- Bohnet, I., F. Greig, B. Herrmann, and R. Zeckhauser (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review* 98(1), 294–310.
- Brekke, K. A., K. E. Hauge, J. T. Lind, and K. Nyborg (2011). Playing with the good guys. a public good game with endogenous group formation. *Journal of Public Economics* 95(9–10), 1111–1118.
- Bruns, H. and G. Perino (2021). Point at, nudge, or push private provision of a public good? *Economic Inquiry* 59(3), 996–1007.
- Cartwright, E. and A. Stepanova (2017). Efficiency in a forced contribution threshold public good game. *International Journal of Game Theory* 46(4), 1163–1191.

- Cason, T. N. and F. U. Khan (1999). A laboratory study of voluntary public goods provision with imperfect monitoring and communication. *Journal of Development Economics* 58(2), 533–552.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* 14(1), 47–83.
- Chen, D. L., M. Schonger, and C. Wickens (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88–97.
- Cubitt, R., S. Gächter, and S. Quercia (2017). Conditional cooperation and betrayal aversion. *Journal of Economic Behavior & Organization* 141, 110–121.
- Davis, L. (2008). The Effect of Driving Restrictions on Air Quality in Mexico City. *Journal of Political Economy* 116(1), 38–81.
- Dawes, R. M., J. M. Orbell, R. T. Simmons, and A. J. Van De Kragt (1986). Organizing groups for collective action. *American Political Science Review* 80(4), 1171–1185.
- Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and economic behavior* 47(2), 268–298.
- Elster, J. (2017). Emotions in Constitution-Making: The Johan Skytte 2016 Prize Lecture. *Scandinavian Political Studies* 40(2), 133–246.
- Falk, A., E. Fehr, and U. Fischbacher (2003). On the nature of fair behavior. *Economic Inquiry* 41(1), 20–26.
- Falk, A., E. Fehr, and U. Fischbacher (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior* 62(1), 287–303.
- Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and economic behavior* 54(2), 293–315.
- Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *The American Economic Review* 90(4), 980–994.
- Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71(3), 397–404.
- Fudenberg, D. and E. Maskin (1990). Evolution and cooperation in noisy repeated games. *The American Economic Review* 80(2), 274–279.
- Gächter, S., F. Kölle, and S. Quercia (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour* 1(9), 650.

- Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and Economic Behavior* 1(1), 60–79.
- Ghidoni, R., G. Calzolari, and M. Casari (2017). Climate change: Behavioral responses from extreme events and delayed damages. *Energy Economics* 68, 103–115.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* 1(1), 114–125.
- Keser, C., A. Markstädter, and M. Schmidt (2017). Mandatory minimum contributions, heterogeneous endowments and voluntary public-good provision. *Games and Economic Behavior* 101, 291–310.
- Keser, C. and F. Van Winden (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics* 102(1), 23–39.
- Kogler, C., L. Mittone, and E. Kirchler (2016). Delayed feedback on tax audits affects compliance and fairness perceptions. *Journal of Economic Behavior & Organization* 124, 81–87.
- Ledyard, J. (1995). Public good: a survey of experimental results. In J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*, pp. 111–194. Princeton, New Jersey: Princeton University Press.
- Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review* 1(6), 894–904.
- Mittone, L., F. Panebianco, and A. Santoro (2017). The bomb-crater effect of tax audits: Beyond the misperception of chance. *Journal of Economic Psychology* 61, 225–243.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 1281–1302.
- Rand, D. G., D. Fudenberg, and A. Dreber (2015). It’s the thought that counts: The role of intentions in noisy repeated games. *Journal of Economic Behavior & Organization* 116, 481–499.
- Sugden, R. (1984). Reciprocity: the supply of public goods through voluntary contributions. *The Economic Journal* 94(376), 772–787.

A Proofs

Proof of Result 1. We combine Equation (5) and (6), respectively, with Equation (3):

$$\begin{aligned}\Delta_p &= \alpha_t G_{-i}^{t-1} + \beta((N-2)x_{i,\rightarrow\hat{j}}^t + x_{\hat{j},\rightarrow\hat{j}}^t) + \gamma S_{\hat{j}}^{t-1} - \alpha_t G_{-i}^{t-1} + \beta(N-1)x_E^t \\ &= \beta((N-2)\Delta_p + \Delta_o) + \gamma S_{\hat{j}}^{t-1}\end{aligned}\tag{7}$$

$$\begin{aligned}\Delta_o &= \alpha_t G_{-i}^{t-1} + \beta(N-1)x_{i,\rightarrow\hat{j}}^t + \delta S_{\hat{j}}^{t-1} - \alpha_t G_{-i}^{t-1} + \beta(N-1)x_E^t \\ &= \beta(N-1)\Delta_p + \delta S_{\hat{j}}^{t-1}\end{aligned}\tag{8}$$

then replace Δ_o inside Δ_p :

$$\begin{aligned}\Delta_p &= \beta \left((N-2)\Delta_p + \beta(N-1)\Delta_p + \delta S_{\hat{j}}^{t-1} \right) + \gamma S_{\hat{j}}^{t-1} \\ &= \beta((N-2) + \beta(N-1))\Delta_p + (\beta\delta + \gamma)S_{\hat{j}}^{t-1} \\ &= \frac{\beta\delta + \gamma}{-(N-1)\beta^2 - (N-2)\beta + 1} S_{\hat{j}}^{t-1}\end{aligned}\tag{9}$$

and this last expression inside Δ_o :

$$\begin{aligned}\Delta_o &= \beta(N-1) \frac{\beta\delta + \gamma}{-(N-1)\beta^2 - (N-2)\beta + 1} S_{\hat{j}}^{t-1} + \delta S_{\hat{j}}^{t-1} \\ &= \frac{\beta\gamma(N-1) + \delta(1 - (N-2)\beta)}{-(N-1)\beta^2 - (N-2)\beta + 1} S_{\hat{j}}^{t-1}.\end{aligned}\tag{10}$$

The denominator in these two equations has one root in $\beta = \frac{1}{N-1}$, which we had already excluded based on Equation (4), and one in $\beta = -1$, which is also out of the admissible range $(0, \frac{1}{N-1})$: on such range, the polynomial is always strictly positive. As for the numerator of Δ_p , Assumption 1 guarantees $\delta < 0$, and since γ is likely non-negative (reflecting reciprocity towards a peer's forced contribution), the overall sign is the result of two contrasting forces: on one hand, when \hat{j} is forced other peers may be willing to reciprocate the (unintended) extra contribution (via γ); on the other hand, they may anticipate a lower subsequent contribution (via δ) and react to it (via β).

In principle, the sign of Δ_o is also undetermined, since $\beta < \frac{1}{N-1}$ implies $1 - (N-2)\beta > 0$: despite having being forced, a subject can react positively to the expected positive reaction of her peers. However, it can be easily observed from Equation (8) that if $\Delta_p < 0$, then since $\delta < 0$ (by Assumption 1), necessarily $\Delta_o < 0$. \square

B Screenshots

In the following, we provide screenshots for the *policy* treatment: the *control* treatment is simpler, in that all references to forcing are omitted.

Figure 4: Screenshot of the contribution phase

Player A: Contribution

[Click here for a reminder of instructions](#)

This is round 4 of 30.

Each of you is endowed with \$10.00 initially, and can decide to contribute any integer amount between 0 and 10. The contributions of all 3 group members will be summed, multiplied by 1.5, and then split in equal parts between you and the other group members.

Here is a summary of contributions in your group from previous rounds:

Round	Your contributions	Player B contributions	Player C contributions	Forced contributions	Total Contributions	Share received back	You kept	Your payoff
1	\$8.00	\$4.00	\$10.00 (forced)	1	\$22.00	\$11.00	\$2.00	\$13.00
2	\$7.00	\$3.00	\$5.00	0	\$15.00	\$7.50	\$3.00	\$10.50
3	\$9.00	\$10.00 (forced)	\$6.00	1	\$25.00	\$12.50	\$1.00	\$13.50

How much do you want to contribute?

0 1 2 3 4 5 6 7 8 9 10

Figure 4 displays the step in which a participant is asked to determine their contribution; Figure 5 displays the step in which the participant sees the results of the round; Figure 6 displays the same step in a case in which the participant was forced to contribute.

Figure 5: Screenshot of the results phase

Results for round 4

[Click here for a reminder of instructions](#)

Your payoff from this round is \$16.50.

You decided to contribute \$2.00.

Your fellow group members contributed \$10.00 (**forced**) and \$5.00.

Your group's total contribution was hence **\$17.00**, including the forced contribution.

Your group's total contribution was multiplied by **1.5** and then split in 3 equal shares of **\$8.50** each.

Hence, your payoff of \$16.50 is obtained as your initial endowment of \$10.00 minus your contribution of \$2.00 plus your \$8.50 share.

[Next](#)

Figure 6: Screenshot of the results phase, with forced contribution

Results for round 4

[Click here for a reminder of instructions](#)

Your payoff from this round is \$8.50.

You decided to contribute \$8.00. **You were randomly selected for a forced contribution of \$10.00. Your intended contribution was not taken into account for this round.**

Your fellow group members contributed \$2.00 and \$5.00.

Your group's total contribution was hence **\$17.00**, including the forced contribution.

Your group's total contribution was multiplied by **1.5** and then split in 3 equal shares of **\$8.50** each.

Hence, your payoff of \$8.50 is obtained as your initial endowment of \$10.00 minus your (forced) contribution of \$10.00 plus your \$8.50 share.

[Next](#)

C Instructions

The experimental instructions are reported below, for the LOW version. The “Forced contribution” panel was absent in the *control* version. Parts in **bold** were appropriately adapted for the HIGH version. Parts in *italic* were part of the script but they were only read aloud, not included in the on-screen instructions.

General instructions

Welcome, and thanks for your participation in this experiment. If you have a question at any time, please “raise your hand” or chat the experimenter.

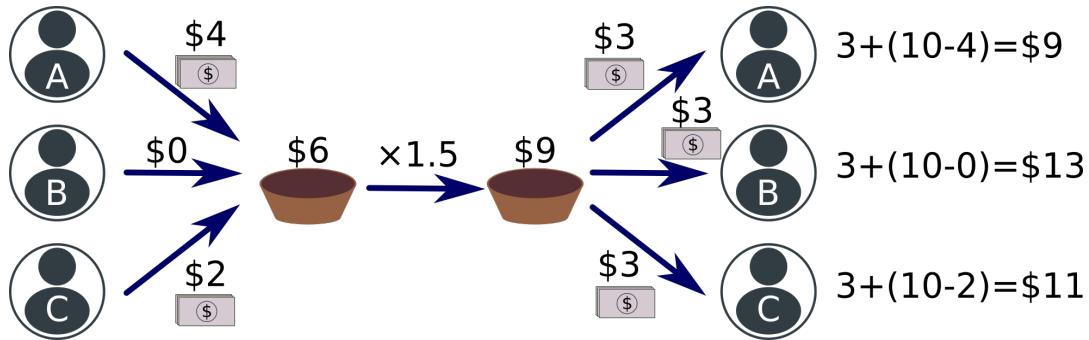
- Please silence your phone, and be sure that your environment is free of all of other distractions.
- At the beginning of the experiment, you will be randomly assigned to a group of 3 people.
- During the entire experiment, you will interact only with the other 2 members of your group.
- The composition of your group will never be revealed.
- All members in your group are subject to the same rules.

Structure of the task

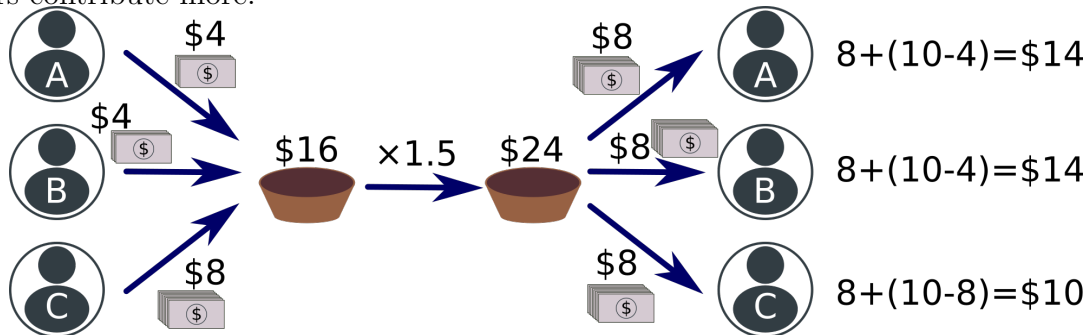
- There will be 30 rounds of the experiment.
- At the beginning of each round you will be given an account with \$10.
- You will then be asked to decide how much of your account you want to put into the pot. You must put in an integer dollar amount, with no cents. For example, you can put in \$0 or \$1 or \$2 or \$3 . . . up to \$10.
- The contributions of all 3 of your group members will be summed up, multiplied by 1.5, and split equally between the 3 group members.
- Hence, your payoff from the round is your share of the pot plus the remaining amount that you didn’t contribute out of your \$10 account.

Below are a couple of examples.

In this first example, player A contributes \$4, the other two group members contribute \$0 and \$2, respectively.



In this second example, player A still contributes \$4, but the other two group members contribute more.

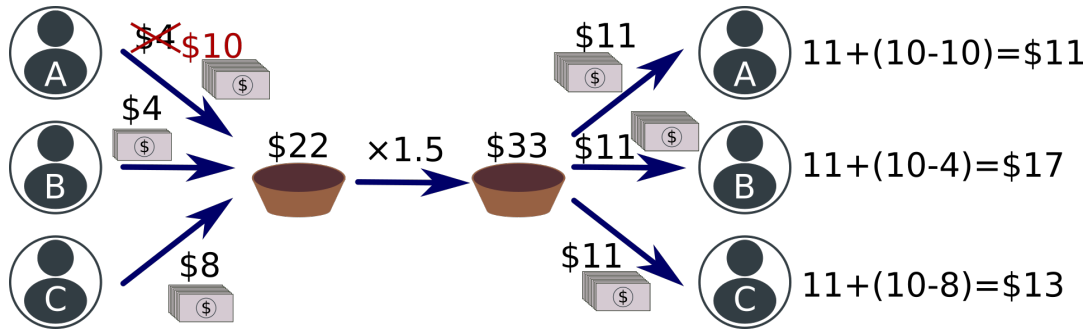


So as you can see, in general keeping money for one self is a guarantee to have that money at the end, but at the same time if participants contribute more, they get more as a group, because contributions are augmented.

Forced contributions

- Every round, after all of your group members have selected their contributions, with some probability one of the members of your group will be randomly selected to contribute their entire account. The probability that *you* will be selected is **0.1** or **10%**. In other words, you can expect to be forced to contribute the full \$10, your entire account, in **3** of the 30 rounds on average. The same applies to the two other members of your group.
- The computer will tell you if and when this happens to you, and when it does, it will contribute your entire account for you. Notice that in any given round, if you are forced, neither of your groupmates will be forced.
- If at a given round one of your groupmates was forced to contribute, you will be informed immediately after, and similarly they will be informed when you are forced.

Below is an example similar to the second previously shown, but with the difference that player A is being forced to contribute.



So if you prefer another way to state this is that your group is selected on average three rounds every ten, and when it is, one of the members will be randomly picked to select the entire contribution.

Notice that you might in principle never, or always, be forced: the numbers we give are only an average, and the fact that you were or not forced in the early rounds doesn't tell you anything about whether you will in the following ones.

Payoffs

- At the end of each round, you will be provided with a summary of the results of the round, including your payoff.
- At every step of the experiment, you will find in the top right of the page a link allowing you to access these instructions again in case you want a reminder.
- Everyone gets a \$5 show-up fee. To calculate your earnings for the experiment, at the end of the session, one round will be randomly drawn by the computer. We will add your earnings from this round to the \$5 show-up fee.

One final advice before we start: if you happen to get an error screen in your browser, please just try to reload or reopen the same address, and you should be back precisely where you left.