Published in final edited form as:
*Nat Hum Behav*. 2017 September ; 1(9): 650–656. doi:10.1038/s41562-017-0191-5.

# Reciprocity and the Tragedies of Maintaining and Providing the Commons

**Simon Gächter**[1,4,5,*], **Felix Kölle**[2], and **Simone Quercia**[3]

[1]School of Economics, University of Nottingham, UK

[2]Faculty of Management, Economics and Social Sciences, University of Cologne, Germany

[3]Institute for Applied Microeconomics, University of Bonn, Germany

[4]CESifo, Munich, Germany

[5]IZA, Bonn, Germany

## Abstract

Social cooperation often requires collectively beneficial but individually costly restraint to maintain a public good1–4, or it needs costly generosity to create one1,5. Status quo effects6 predict that maintaining a public good is easier than providing a new one. Here we show experimentally and with simulations that even under identical incentives, low levels of cooperation (the 'tragedy of the commons'2) are systematically more likely in Maintenance than Provision. Across three series of experiments, we find that strong and weak positive reciprocity, known to be fundamental tendencies underpinning human cooperation7–10, are substantially diminished under Maintenance compared to Provision. As we show in a fourth experiment, the opposite holds for negative reciprocity ('punishment'). Our findings suggest that incentives to avoid the 'tragedy of the commons' need to contend with dilemma-specific reciprocity.

### Keywords

Tragedy of the Commons; public goods; strong and weak reciprocity; evolution of human cooperation



[*]**Correspondence** should be addressed to S.G. simon.gaechter@nottingham.ac.uk.

**Data availability.** The data for the statistical analyses are stored in Dryad Data package title: Reciprocity in Maintaining and Providing Public Goods; http://dx.doi.org/10.5061/dryad.8d9t2

**Code availability.** We used STATA 14.2 for data analysis. The codes are stored in Dryad Data package title: Reciprocity in Maintaining and Providing Public Goods; http://dx.doi.org/10.5061/dryad.8d9t2

**Author Contributions**
SG, FK, and SQ developed the research ideas and designed the study; FK and SQ conducted the experiments, and analysed data. SG, FK, and SQ wrote the manuscript.



**Competing interest**
The authors declare no competing interests.


Europe PMC Funders Author Manuscripts

Humans are an exceptionally cooperative species able to collaborate for the creation of common benefit9,11–13. Collective actions such as voting, participating in political movements, the provision of the welfare state, charity, volunteering and teamwork are examples of public goods that come into existence by the generosity of many people that puts the greater good before self-interest1,5. Cooperation is, however, not always about providing collectively valuable resources, but often about maintaining existing ones1–4. Limiting $CO_2$ emissions, sustaining natural resources, or maintaining common pastures and biodiversity are important examples of cooperation problems that require restraint in exploiting existing socially beneficial public goods.

In this paper, we show experimentally and with simulations that cooperation for maintaining an initially existing public good is substantially and systematically weaker than cooperation for creating a new public good even if they are otherwise identical social dilemmas. This is unexpected, given that many people are biased towards the status quo and defaults6, which should ease cooperation when the public good already exists compared to when it needs to be provided.

We show that the reason for lower cooperation in the maintenance dilemma is that reciprocity, a fundamental force behind the evolution of cooperation and human sociality7–10, is substantially diminished in maintaining compared to providing a public good. Simulations show that, despite some variability, lower cooperation in Maintenance than Provision is a systematic effect to be expected with a likelihood of 70%. The simulation results also provide an explanation for the mixed findings in some related literature.14–23

In our experiments, we focus sharply on the behavioural differences between initially existing and inexistent public goods (Fig. 1) and abstract from technological complexities, loss aversion, time discounting and institutional details relevant in real world social dilemmas1,24–29. In ‘Maintenance’, a group of four people possesses a common pool of 80 tokens and each member can withdraw up to 20 tokens. Upholding the status quo by withdrawing nothing earns each group member 32 money units (MU); if all withdraw maximally, everyone earns 20 MU. In ‘Provision’, the common pool is initially empty and 80 tokens are distributed equally among group members who decide simultaneously how many tokens (up to 20) to contribute to the pool. In the status quo all earn 20 MU, and all contributing maximally earns each member 32 MU.

Using the setup described in Fig. 1, we run three series of experiments with 704 participants who interact anonymously in three generic settings of social interaction (see Methods). All experiments involve a between-subjects comparison of cooperation in Maintenance and Provision. We also elicit beliefs about group members’ contributions to (or withdrawals from) the public good. Participants need to successfully complete a comprehension test before the experiment starts.

In the first experiment, called One-shot, participants ($n$ = 288) take a single decision only. This experiment is a basic measure of people’s cooperativeness in the absence of strategic incentives to cooperate. In a second experiment, called Strangers, participants ($n$ = 256) play the games of Fig. 1 for 27 iterations with randomly changing group composition in each

round. This experiment is a sequence of one-shot interactions that permit learning about cooperativeness in the population30–32. The third experiment ($n$ = 160), called Partners, keeps group composition constant across the 27 iterations, which creates strategic incentives for cooperation32,33.

The effective size of the public good (after withdrawals or contributions) is smaller in Maintenance than Provision in all experiments (Fig. 2, Supplementary Table 1). In One-shot, the public good in Maintenance is on average 27% smaller than in Provision (Fig. 2, Panel 1; 23.8 vs. 32.6; two-sided t-test, $t$ = -2.51, $P$ = 0.014). Low levels of the public good (less than 10% of the optimal size of 80), are more likely in Maintenance than Provision (23% vs. 0%; $\chi^2(1) = 9.51$, $P$ = 0.002).

In Strangers, the public good starts out 23% lower in Maintenance than Provision (22.7 vs. 29.5; two-sided t-test, $t$ = -1.92, $P$ = 0.059) and decays on average to about 5% of the socially efficient level in both problems (Fig. 2, Panel 2). Thus, the tragedy of the commons is almost maximal in both Maintenance and Provision.

In Partners, the public good starts 33% smaller in Maintenance than Provision (27.7 vs. 41.3; two-sided t-test, $t$ = -2.96, $P$ = 0.005) and drops over time (Fig. 2, Panel 3). On average the public good is 37.3% smaller in Maintenance than Provision (10.6 vs.16.9; linear mixed effects model, $P$ = 0.035).

Comparing Partners and Strangers reveals the extent to which strategic incentives help the provision of the public good. We find that in Maintenance the average size of the public good is only 3.6 units higher in Partners than Strangers (10.6 vs. 7.0; linear mixed-effects model, $P$ = 0.346, Supplementary Table 2), while in Provision the public good is on average twice as large in Partners than Strangers (16.9 vs. 8.2; linear mixed effects model, $P$ = 0.004). Thus, strategic incentives to increase cooperation are substantially weaker in Maintenance than Provision.

Taken together, these results show that high levels of the public good are harder to achieve in Maintenance than Provision in One-shot and in the first period of Partners and Strangers. This is surprising given that in Maintenance the public good enjoys a head start because it is already provided at the outset. Furthermore, while in Strangers the size of the public good converges to similar long-run equilibrium levels, in Partners the initial differences are persistent and lead to different long-run outcomes between Provision and Maintenance. The aim of our further analysis is to understand the differences in cooperation outcomes by investigating whether initial resource allocation affects reciprocity in response to restraint and generosity, respectively.

Studying reciprocity is particularly interesting due to its fundamental role for human sociality7–10. In our settings, reciprocity takes the form of conditional cooperation: the willingness to cooperate provided others do the same30,32,34,35. Here, we distinguish between two forms of conditional cooperation, which are inspired by the concepts of weak and strong reciprocity9,10,36. Weak reciprocity can occur in stable relationships and means behaving conditionally cooperative for self-regarding strategic reasons. By contrast, strong reciprocity entails non-selfish conditional cooperation not only in repeated interactions but

also in one-shot games. Strong reciprocity is a preference for conditional cooperation, whereas weak reciprocity is a behavioural strategy deployed for self-regarding reasons.

Studying reciprocity as a preference requires looking beyond cooperation outcomes and to measure attitudes to cooperation separately from outcomes. The reason why this is important is that people who differ in their *ex ante* attitudes can *ex post* make the same cooperation decision. To see why, consider that a conditional cooperator's ex ante attitude is to cooperate only if they believe their group members do so too. But there may also be 'free riders', who never want to contribute to the public good irrespective of their beliefs how much others contribute. A conditional cooperator who believes that others do not contribute and a person with a free rider attitude both contribute nothing: their ex post behavior is observationally equivalent despite different ex ante attitudes. Thus, if cooperation is a function of attitudes and beliefs, the challenge is to separate them empirically. Our approach, which we call the 'ABC of cooperation', achieves this separation. This also allows us to compare strong reciprocity as measured by the ABC approach with reciprocity estimated from observed behaviour.

The ABC approach measures individual attitudes ($a_i$), beliefs ($b_i$), and effective contributions ($c_i$) separately and explains cooperation as $a_i(b_i) \rightarrow c_i$. It is inspired by30 and implemented as follows. All three experiments start with an incentive-compatible elicitation of attitudes without feedback in a one-shot version of either the Maintenance or the Provision dilemma. The elicited attitudes are our main measure of strong reciprocity. Eliciting attitudes involves specifying a vector $a_i$ of contributions or withdrawals as a function of all possible average contributions or withdrawals of other group members. We classify participants as conditional cooperators (that is, strong reciprocators) if the entries in the vector $a_i$ are increasing in others' contributions or withdrawals, or as a free rider if a participant's $a_i$ consists of only zero contributions or maximal withdrawals. We refer to the remaining participants as 'others'. After attitude elicitation, the three experiments proceed as described above. In all experiments, we elicit incentivized beliefs ($b_i$) about other group members' average withdrawal or contribution and we observe effective contributions ($c_i$) to the public good (see Methods).

In the repeated direct interactions of Strangers and Partners we measure conditional cooperation in linear mixed-effects models by regressing individual contributions or withdrawals on the average contribution or withdrawals of other group members in the previous period (Supplementary Information). The relation between these two variables, the coefficient $\beta_1$, is our measure of conditional cooperation. We will call $\beta_1$ 'estimated reciprocity'.

In Strangers, $\beta_1$ is an estimate of strong reciprocity because there are no strategic incentives to pretend being a reciprocator. Because $\beta_1$ is estimated from behavior only, it is a proxy for strong reciprocity. But we expect that participants with attitudes that classify them as conditional cooperators will have $\beta_1 > 0$, whereas people with a free rider attitude will display $\beta_1 \approx 0$.

In Partners, conditional cooperators will also have $\beta_1 > 0$, which may be larger than in Strangers due to added incentives for weak reciprocity. Free riders may therefore also display $\beta_1 > 0$. Furthermore, we will use the attitudes $a_i$ and $\beta_1$ to study the link between strong and weak reciprocity.

Elicited attitudes are significantly different in Maintenance and Provision ($\chi^2(2) = 31.03$, $P < 0.001$; Fig. 3a). In Maintenance, participants are significantly less likely to be conditional cooperators than in Provision (42% vs. 64%; $\chi^2(1) = 31.03$, $P < 0.001$); are significantly more likely to be free riders (28% vs. 17%; $\chi^2(1) = 10.46$, $P = 0.001$) and are also significantly more likely to display an unclassified attitude ('others'; 30% vs. 19%; $\chi^2(1) = 11.08$, $P = 0.001$). Thus, in Maintenance 58% of participants do not reciprocate their group member's effective contributions, which is almost the mirror image of the 64% in Provision who do reciprocate.

Estimated reciprocity $\beta_1$ in the repeated games is also significantly lower in Maintenance than Provision in both Strangers and Partners (Fig. 3b, panel 1; multilevel mixed-effects models, $P < 0.001$; Supplementary Table 3). The added strategic incentives for weak reciprocity significantly increase estimated reciprocity in both Maintenance and Provision (Fig. 3b, panel 1; multilevel mixed-effects models, $P < 0.001$; Supplementary Table 4).

Estimated reciprocity is also consistent with attitude types elicited prior to the repeated games (Supplementary Tables 5-6). Participants classified as conditional cooperators show high degrees of estimated reciprocity in Strangers and Partners, significantly above that of free riders in both Maintenance and Provision (Fig. 3b, panels 2 and 3; multilevel mixed-effects models, $P < 0.001$). Conditional cooperators also display significantly higher $\beta_1$ in Partners than Strangers (multilevel mixed-effects models, $P < 0.001$). As predicted, free riders in Strangers display low estimated reciprocity but show increased $\beta_1$ in Partners compared to Strangers. Participants classified as 'others' do display a substantial $\beta_1$ but do not react to strategic incentives (Fig. 3b, panel 4; multilevel mixed-effects models, $P > 0.166$).

Our next step is to investigate whether the differences in reciprocity across Maintenance and Provision can explain the observed differences in cooperation outcomes (Fig. 2). We do this by applying our ABC framework that uses attitudes and beliefs to explain effective contributions. We calculate predicted effective contributions $[a_i\,(b_i) \rightarrow c_i^*]$ and compare them with actual effective contributions $c_i$ from One-shot as well as with the effective first-period contributions in the repeated experiments (Methods). Predicted and actual effective contributions are highly significantly positively correlated in One-shot as well as in all repeated games (all Spearman's $\rho > 0.59$; $P < 0.001$).

We also calculate individual-level deviations from the predicted effective contribution, $c_i^* - c_i$. In One-shot, this measure lies within ± 2 tokens in 63% and 62% of the cases in Maintenance and Provision, respectively, with no differences between treatments ($\chi^2(1) = 0.01$, $P = 0.903$). We obtain similar results for first-period effective contributions in Strangers (66% and 63%; $\chi^2(1) = 0.43$, $P = 0.514$) and Partners (74% and 64%; $\chi^2(1) = 1.86$, $P = 0.172$). Finally, effective contributions differ significantly between attitude types:

free riders contribute significantly less than conditional cooperators and ‘others’ in all conditions (Supplementary Fig. 1).

The fact that the ABC approach predicts equally well in Maintenance and Provision allows us to use the elicited attitudes and beliefs as a ‘population pool’ from which we can sample at random to run ’simulated experiments’ (Methods and Supplementary Information). The advantage of simulations is that we are not restricted to a specific laboratory sample we happen to draw at a given instance (with hitherto unobservable attitudes and beliefs); we can cost effectively perform a large number of identical experiments and therefore elicit a distribution of likely cooperation ratios of Maintenance relative to Provision. This also allows us to check how systematic the results are that we observe.

The results of 1000 simulated experiments (Fig. 4) show that effective cooperation levels in Maintenance are lower than in Provision in 70% of all simulated experiments. This result shows that our findings that cooperation in Maintenance is lower than in Provision are systematic.

Given that our results reveal important asymmetries in positive reciprocity between Maintenance and Provision, it is interesting to study whether initial resource allocation also affects negative reciprocity, which in our setting takes the form of punishment9,36. Furthermore, punishment is an expression of moral disapproval and social norms37 that are important in many real world public goods38. If the differences in positive reciprocity in Maintenance and Provision also translate into negative reciprocity, we should observe less punishment in Maintenance than Provision and, therefore, also a reduced effectiveness of punishment to stabilize cooperation in Maintenance compared to Provision.

We study punishment in a fourth experiment (‘Partners with Punishment’; $n = 172$), which is identical to Partners except for an added punishment stage in each period after group members have made their withdrawal or contribution decisions39. In the punishment stage, each group member can assign up to 5 punishment points to each other member, where each punishment point costs one MU and reduces the earnings of the punished group member by three MU (see Methods).

The attitudes elicited prior to the experiment replicate the results from Fig. 3a (Supplementary Fig. 2). Contrary to expectations, negative reciprocity, estimated as assigned punishment in reaction to negative deviations of others from own effective contribution, is substantially and significantly higher in Maintenance than in Provision. This effect is present both overall and for each attitude type (Fig. 5a, panels 1-4), and it is not driven by different frequencies of punishers (Supplementary Figure 3). There are no treatment differences for positive deviations (Fig. 5a, panel 1; Supplementary Table 7). Interestingly, in contrast to estimated positive reciprocity, estimated negative reciprocity does not differ between conditional cooperators and free riders (Fig. 5a, panels 2-4; Supplementary Table 8).

As expected40, punishment increases the public goods to substantially higher levels compared to Partners (Fig. 5b; linear mixed effect models; Maintenance: 43.1 vs. 10.6, $P < 0.001$; Provision: 44.1 vs. 16.9, $P < 0.001$; Supplementary Tables 9-10). Remarkably, the sizes of public goods are now very similar in Maintenance and Provision (linear mixed effect

models; Maintenance: 43.1, Provision: 44.1, $P = 0.904$). Besides stronger negative reciprocity, a further reason for this result is that reactions to received punishment (in terms of change in effective contributions) are also stronger in Maintenance than Provision (Supplementary Table 11).

One way to reconcile the results on positive and negative reciprocity in Partners and Partners with Punishment, respectively, is to argue that also in Partners people engage in punishment by reducing their contributions in the current period as a reaction to previous negative deviations of others from own effective contributions. If such 'implicit' punishment is stronger in Maintenance than Provision, it could explain why the decay in effective contributions is stronger in Maintenance than Provision. However, this conjecture is not borne out by the data.

We find that participants in Partners significantly increase their contributions in round $t$ in response to positive deviations of others from own contributions in round $t\text{-}1$; the reverse holds for negative deviations. However, we find both of these reactions to be significantly more pronounced in Provision than in Maintenance (linear mixed effect models; both $P < 0.018$; Supplementary Table 12). This confirms once again stronger conditional cooperation in Provision compared to Maintenance in Partners. It also suggests another interpretation of the results of Partners with Punishment: because voluntary conditional cooperation is weaker in Maintenance than Provision, stronger extrinsic incentives are needed, here in the form of punishment, to stabilize cooperation in Maintenance at similar levels than in Provision.

Our analysis has revealed that the important principles of human cooperation of strong and weak reciprocity7–10 are substantially diminished when cooperation requires restraint in exploiting a public good as opposed to when cooperation calls for generosity to provide a public good. Our findings are consistent with the observation that failing to contribute to a public good is judged more morally blameworthy than exploiting an existing public good.41

Our results can also be explained by a model of revealed altruism42,43, according to which initial resource allocation affects perceptions of generosity of actions and hence subsequent reciprocity. Because in Provision cooperation is the result of an act of commission (contributing), while in Maintenance cooperation is achieved by omission (not withdrawing), cooperation in Provision is perceived as more generous than in Maintenance and thus Provision triggers stronger positive reciprocity than Maintenance. By contrast, our results suggest that negative reciprocity, as expressed by people's costly punishment, does not follow this logic because punishment is more severe in Maintenance than Provision, likely to compensate for weaker voluntary cooperation in Maintenance.

Our findings from the experiments without punishment and the simulations also help explaining the mixed evidence from previous related literature, which, with a few exceptions35,44,45, only compared cooperation outcomes, that is, the effective size of the public good after contribution or withdrawal decisions. Some of these studies find higher cooperation in so-called 'give-some' vs. 'take-some' games14–17, some find the reverse18 and some find no significant differences19–23. The simulations based on our ABC approach can explain these mixed results (Fig. 4) but they also show that on average cooperation in

Maintenance is generally expected to be lower than in Provision. The finding that Maintenance and Provision are systematically different also suggests that future research should choose the game (Maintenance or Provision) that comes closest to the social dilemma of interest.

Our results also have potential policy relevance.46 Recent policy proposals to foster cooperation build on the power of reciprocity in combination with economic incentives47,48. Policy makers who reckon with reciprocity should therefore consider that the extent of reciprocity that can be evoked is dilemma-specific. Moreover, a problem of incentives is that they might 'crowd out' strong reciprocity because incentives typically strengthen self-regarding motives to cooperate49. Our finding of higher reciprocity in Provision than Maintenance suggests that crowding out may be more problematic in provision problems than in maintenance problems, because in Maintenance more people display non-reciprocal attitudes. Future research will need to address these issues, including how reciprocity and incentives interact in non-linear settings with thresholds, resource rivalry, discounting, and hybrid social dilemmas where provision and exploitation can take place at the same time.

## Methods

### Isomorphism of Maintenance and Provision under monetary incentives

In *Maintenance*, each group of 4 members is initially endowed with 80 tokens placed in a "group project"; individual members have no endowment. Material incentives are described by equation (1):

$$\pi_i = w_i + 0.4\left(80 - \sum_{j=1}^{4} w_j\right) \quad (1)$$

where $0 \leq w_i \leq 20$ indicates the withdrawal of individual $i$ from the project.

In *Provision,* the "group project" is initially empty and each group member has an endowment of 20 tokens instead. The material incentives for each individual $i$ are described by equation (2):

$$\pi_i = 20 - c_i + 0.4\sum_{j=1}^{4} c_j \quad (2)$$

where $0 \leq c_i \leq 20$ denotes the contribution of individual $i$ to the project.

Hence, under rationality and money maximization, Maintenance and Provision are isomorphic social dilemmas. Using $c_j = 20 - w_j$ for $j = 1, \ldots, 4$ and substituting into eq. (2) we obtain (1). Analogously, using $w_j = 20 - c_j$ for $j = 1, \ldots, 4$ and substituting into (1) yields (2).

## Experimental design details

The experiments were approved by the Research Ethics Committee in the School of Economics at the University of Nottingham. We conducted four series of experiments using the two decision situations described above and in Fig. 1. Each experiment was composed of three parts that allow to elicit the three components of our ABC framework: an individual *i*'s attitude ($a_i$) towards cooperation (*i*'s 'type'), *i*'s beliefs ($b_i$) about others' contribution, and *i*'s contribution decision ($c_i$). Participants knew that the experiment consisted of several parts but only received information about the relevant part upon progression of the experiment. To avoid spillover effects between different parts, information about decisions and payoffs were given only at the very end of the experiment. Experimental instructions are in the Supplementary Information.

In Part 1, participants were introduced to either the Maintenance or Provision problem. Before continuing, participants answered a set of computerized control questions.

In Part 2, we elicited cooperation attitudes $a_i$ using a variant of the strategy method50, which allows eliciting an individual's willingness to cooperate as a function of the other group members' cooperation decisions. Participants were asked to make an unconditional and a conditional cooperation decision. In the unconditional decision, participants were simply asked how much they want to withdraw from (contribute to) the common pool. In the conditional contribution participants had to fill a withdrawal (contribution) table in which they had to state their withdrawal (contribution) decision for each possible (rounded) average withdrawal (contribution) of the other three group members. This gives us the vector $a_i$, our measure of strong reciprocity. To achieve incentive compatibility, in each group a random mechanism selected three members for which the unconditional decision was payoff-relevant and one member for whom the conditional decision for the (rounded) average unconditional withdrawal (contribution) of the three other group members was payoff-relevant.

Part 3 comprised a direct-response interaction that differed in its exact design protocol across the four experiments as described in the main text (One-shot, Strangers, Partners, and Partners with Punishment). This elicits component $c_i$ of the ABC framework.

In all repeated experiments (Strangers, Partners, and Partners with Punishment), participants were matched in groups of four and interacted for 27 consecutive rounds under payment rules (1) or (2). Participants were not told how many rounds the experiment would last.51 This avoids endgame effects and also seems realistic for many common resource problems, which do not have a known endpoint. In Strangers, participants were re-matched randomly in 16-participants matching groups after every round, while in Partners and Partners with Punishment group composition remained constant across all 27 rounds. At the end of each round, participants received aggregate feedback on choices and outcomes.

In all rounds of the direct-response interactions, we also elicited beliefs about average effective contributions of the other three group members. Participants were paid for the accuracy of their beliefs. They earned 3 points if their belief was exactly correct, and 2 (1) points when their belief deviated by 1 (2) point(s) from the true average effective

contribution. If their estimation was off by more than two points, they received no additional money. This elicits component $b_i$ of our framework.

## Data collection and subject-pool socio-demographics

A total of $n = 876$ students participated in our experiments (Maintenance: $n_M = 432$, Provision: $n_P = 444$). Participants were recruited with the help of ORSEE52 from the volunteer student subject pool at the University of Nottingham. Participation was upon informed consent. The average age was 20.1 years (s.d. 2.25 years); 57% were females. 59% were British, 22% Asian, 12% from other European countries and the rest from other countries. 20% were economics or business students, 18% other social sciences, 20% humanities, 14% sciences, 12% engineering, and 12% medical science, and 4% law. We conducted all experiments in the CeDEx lab at the University of Nottingham using z-Tree.53 The experiments lasted between 70 to 210 minutes depending on the experimental condition. Participants earned on average £20.60.

## Predicting effective contributions

In One-shot, Strangers, and Partners, the ABC approach allows us to predict contributions using elicited cooperation attitudes $a_i$ and beliefs $b_i$:$a_i(b_i) \rightarrow c_i^*$. By matching beliefs with the corresponding decision in the contribution (withdrawal) table, we predict a contribution (withdrawal) decision $c_i^*$ for each subject and compare $c_i^*$ with the actual contribution $c_i$ that we observe in the direct-response experiment.30,54

## Classification of attitudes

We analyse cooperation in the strategy-method experiment treating each participant's effective contribution schedule (the vector $a_i$) as an independent observation. We classify cooperation attitudes into three main behavioural types30: a participant is a *conditional cooperator* if either his/her effective contribution schedule exhibits a (weakly) monotonically increasing pattern, or if the Spearman correlation coefficient between his/her schedule and the others' average contribution is positive and significant at the 1% level; a *free rider* if he/she never contributes anything (always withdraws everything) irrespective of how much others contribute (withdraw); (iii) *other* if neither (i) nor (ii) applies. Attitudes are a proxy for cooperation preferences because they reflect a willingness to pay for cooperation as a function of other group members' cooperation.

## Simulations

For each simulated experiment, we randomly sample (with replacement) from the participant pool of Maintenance experiments attitudes and beliefs (n = 60, the median sample size in related studies also using linear public goods14–23) and calculate simulated effective contributions [$a_i(b_j \rightarrow \hat{c})$]. We do the same for $n = 60$ Provision attitudes and beliefs. This resembles an experiment where a researcher invites 60 participants per treatment and then observes their effective contribution. As a participation pool, we use all $n = 876$ attitudes from our four experiments (Maintenance: $n_M = 432$, Provision: $n_P = 444$), and $n = 544$ beliefs (Maintenance: $n_M = 268$, Provision: $n_P = 276$) from One-shot as well as the first period of Strangers. Details are in the Supplementary Information.

### Statistical analysis

In the One-shot direct interaction, we treat ($b_i$, $c_i$) as independent observations. In the repeated interactions, we treat beliefs and effective contributions at the matching group level as an independent observation. Matching groups are composed by 16 participants in Strangers and by 4 participants in Partners and Partners with Punishment, respectively. For the repeated experiments, all estimations are performed using linear mixed models with random intercepts at the matching group and the individual level (see Statistical Analysis in SI for details on model specifications).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank A. Alonso Arechar, B. Beranek, T. Cason, J. Cox, F. Fallucchi, U. Fischbacher, M. García-Vega, R. Hernán-González, D. Houser, L. Molleman, D. van Dolder, T. Weber, O. Weisel, and various seminar audiences for helpful comments. B. Beranek provided valuable research assistance. This work was supported by the ERC-AdG 295707 COOPERATION and the ESRC Network for Integrated Behavioural Sciences (NIBS, ES/K002201/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

1. Van Lange, PAM., Balliet, D., Parks, CD., Van Vugt, M. Social dilemmas The psychology of human cooperation. Oxford University Press; Oxford: 2014.
2. Hardin G. The tragedy of the commons. Science. 1968; 162:1243.
3. Ostrom, E. Governing the commons The evolution of institutions for collective action. Cambridge University Press; Cambridge: 1990.
4. Collier, P. The plundered planet. Penguin Books; London: 2011.
5. Gintis, H.Bowles, S.Boyd, R., Fehr, E., editors. Moral sentiments and material interests The foundations of cooperation in economic life. MIT Press; Cambridge: 2005.
6. Samuelson W, Zeckhauser R. Status quo bias in decision making. J Risk Uncertain. 1988; 1:7.
7. Trivers RL. Evolution of reciprocal altruism. Quarterly Review of Biology. 1971; 46:35.
8. Axelrod R, Hamilton W. The evolution of cooperation. Science. 1981; 211:1390. [PubMed: 7466396]
9. Bowles, S., Gintis, H. A cooperative species: Human reciprocity and its evolution. Princeton University Press; Princeton: 2011.
10. Rand DG, Nowak MA. Human cooperation. Trends in Cognitive Sciences. 2013; 17:413. [PubMed: 23856025]
11. Nowak, MA., Highfield, R. Supercooperators: Altruism, evolution, and why we need each other to succeed. Free Press; New York: 2011.
12. Turchin, P. Ultrasociety: How 10,000 years of war made humans the greatest cooperators on earth. Beresta Books; LLC Chaplin, CT: 2016.
13. Tomasello, M. A natural history of human morality. Harvard University Press; Cambridge, MA: 2016.
14. Andreoni J. Warm glow versus cold prickle - the effects of positive and negative framing on cooperation in experiments. Quarterly Journal of Economics. 1995; 110:1.
15. Park E-S. Warm-glow versus cold-prickle: A further experimental study of framing effects on free-riding. Journal of Economic Behavior and Organization. 2000; 43:405.
16. Khadjavi M, Lange A. Doing good or doing harm: Experimental evidence on giving and taking in public good games. Experimental Economics. 2015; 18:432.

17. Cox CA. Decomposing the effects of negative framing in linear public goods games. Economics Letters. 2015; 126:63.
18. Sell J, Son Y. Comparing public goods and common pool resources: Three experiments. Social Psychology Quarterly. 1997; 60:118.
19. Van Dijk E, Wilke H. Is it mine or is it ours? Framing property rights and decision making in social dilemmas. Organizational Behavior and Human Decision Processes. 1997; 71:195.
20. Cubitt R, Drouvelis M, Gächter S. Framing and free riding: Emotional responses and punishment in social dilemma games. Experimental Economics. 2011; 14:254.
21. Dufwenberg M, Gächter S, Hennig-Schmidt H. The framing of games and the psychology of play. Games and Economic Behavior. 2011; 73:459.
22. Messer KD, Zarghamee H, Kaiser HM, Schulze WD. New hope for the voluntary contributions mechanism: The effects of context. Journal of Public Economics. 2007; 91:1783.
23. Cox JC, Ostrom E, Sadiraj V, Walker JM. Provision versus appropriation in symmetric and asymmetric social dilemmas. Southern Economic Journal. 2013; 79:496.
24. Apesteguía J, Maier-Rigaud FP. The role of rivalry. Public goods versus common-pool resources. Journal of Conflict Resolution. 2006; 50:646.
25. Levin SA. Public goods in relation to competition, cooperation, and spite. Proceedings of the National Academy of Sciences. 2014; 111:10838.
26. Brewer MB, Kramer RM. Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. Journal of Personality and Social Psychology. 1986; 50:543.
27. De Dreu CK, McCusker C. Gain–loss frames and cooperation in two-person social dilemmas: A transformational analysis. Journal of Personality and Social Psychology. 1997; 72:1093.
28. Van Dijk E, Wilke H, Wilke M, Metman L. What information do we use in social dilemmas? Environmental uncertainty and the employment of coordination rules. Journal of Experimental Social Psychology. 1999; 35:109.
29. McCusker C, Carnevale PJ. Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. Organizational Behavior and Human Decision Processes. 1995; 61:190.
30. Fischbacher U, Gächter S. Social preferences, beliefs, and the dynamics of free riding in public good experiments. American Economic Review. 2010; 100:541.
31. Janssen MA, Holahan R, Lee A, Ostrom E. Lab experiments for the study of social-ecological systems. Science. 2010; 328:613. [PubMed: 20431012]
32. Keser C, van Winden F. Conditional cooperation and voluntary contributions to public goods. Scandinavian Journal of Economics. 2000; 102:23.
33. Andreoni J. Why free ride - strategies and learning in public-goods experiments. Journal of Public Economics. 1988; 37:291.
34. Rustagi D, Engel S, Kosfeld M. Conditional cooperation and costly monitoring explain success in forest commons management. Science. 2010; 330:961. [PubMed: 21071668]
35. Fosgaard TR, Hansen LG, Wengström E. Understanding the nature of cooperation variability. Journal of Public Economics. 2014; 120:134.
36. Gintis H. Strong reciprocity and human sociality. Journal of Theoretical Biology. 2000; 206:169. [PubMed: 10966755]
37. Fehr E, Fischbacher U. Social norms and human cooperation. Trends in Cognitive Sciences. 2004; 8:185. [PubMed: 15050515]
38. Nyborg K, et al. Social norms as solutions. Science. 2016; 354:42. [PubMed: 27846488]
39. Fehr E, Gächter S. Altruistic punishment in humans. Nature. 2002; 415:137. [PubMed: 11805825]
40. Sigmund K. Punish or perish? Retaliation and collaboration among humans. Trends in Ecology and Evolution. 2007; 22:593. [PubMed: 17963994]
41. Cubitt R, Drouvelis M, Gächter S, Kabalin R. Moral judgments in social dilemmas: How bad is free riding? Journal of Public Economics. 2011; 95:253.
42. Cox JC, Friedman D, Sadiraj V. Revealed altruism. Econometrica. 2008; 76:31.
43. Cox JC, Servátka M, Vadovič R. Status quo effects in fairness games: Reciprocal responses to acts of commission versus acts of omission. Experimental Economics. 2017; 20:1.

44. Fleishman JA. The effects of decision framing and others' behavior on cooperation in a social dilemma. The Journal of Conflict Resolution. 1988; 32:162.
45. Frackenpohl G, Hillenbrand A, Kube S. Leadership effectiveness and institutional frames. Experimental Economics. 2016; 19:842.
46. Fehr-Duda H, Fehr E. Game human nature. Nature. 2016; 530:413. [PubMed: 26911767]
47. MacKay DJC, Cramton P, Ockenfels A, Stoft S. Price carbon — I will if you will. Nature. 2015; 326:315.
48. Rand DG, Yoeli E, Hoffman M. Harnessing reciprocity to promote cooperation and the provisioning of public goods. Policy Insights from the Behavioral and Brain Sciences. 2014; 1:263.
49. Bowles S. Policies designed for self-interested citizens may undermine “the moral sentiments”: Evidence from economic experiments. Science. 2008; 320:1605. [PubMed: 18566278]
50. Fischbacher U, Gächter S, Fehr E. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters. 2001; 71:397.
51. Rand DG, et al. Positive interactions promote public cooperation. Science. 2009; 325:1272. [PubMed: 19729661]
52. Greiner B. Subject pool recruitment procedures: Organizing experiments with ORSEE. J Econ Sci Assoc. 2015; 1:114.
53. Fischbacher U. Z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics. 2007; 10:171.
54. Fischbacher U, Gächter S, Quercia S. The behavioral validity of the strategy method in public good experiments. Journal of Economic Psychology. 2012; 33:897.
55. Bardsley N, Moffatt PG. The experimetrics of public goods: Inferring motivations from contributions. Theory and Decision. 2007; 62:161.


Europe PMC Funders Author Manuscripts

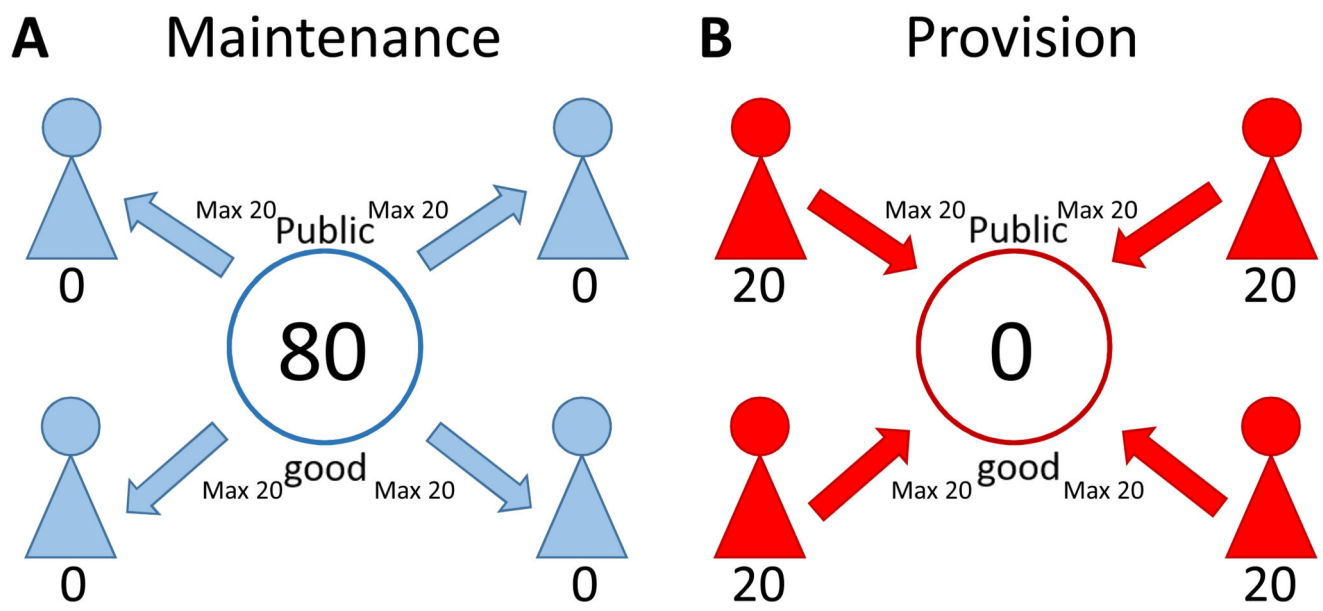


**Figure 1. The isomorphic social dilemmas of maintaining and providing a public good.**

The figure illustrates initial resource allocation in Maintenance and Provision prior to decision-making. **a,** Maintenance: Initially, group members have 0 tokens and 80 tokens are provided in the public good. Group members can simultaneously withdraw up to 20 tokens. **b,** Provision: Initially, each group member has 20 tokens and the public good is empty. Group members can simultaneously contribute up to 20 tokens. Each token withdrawn or not contributed is worth 1 MU to the respective group member alone. Each token in the common resource is worth 0.4 MU for each group member. Material incentives therefore are to withdraw 20 tokens in Maintenance and to contribute 0 tokens in Provision, yielding 20 MU for each group member. The socially beneficial decisions of withdrawing nothing and contributing everything earn each group member 32 MU. Further details are in Methods.

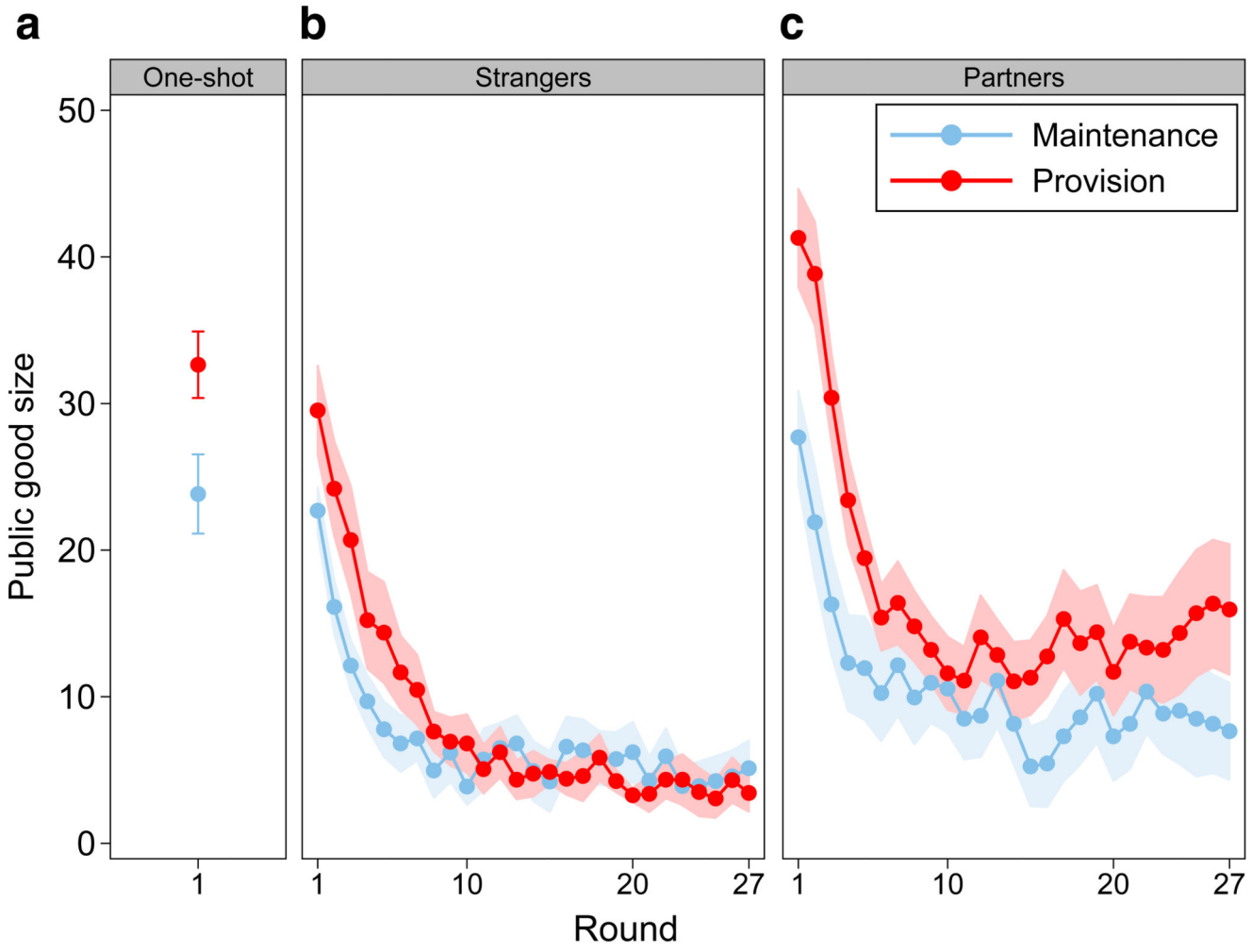


**Figure 2. Public good levels.**

Shown are the effective sizes of the public good per round after contribution or withdrawal decisions (± 1 s.e.m.). **a,** One-shot game, $n_M = 140$, $n_P = 148$. **b, c,** Effective public goods over the 27 rounds of interactions in randomly changing groups (Strangers, $n_M = 128$, $n_P = 128$) and fixed groups (Partners, $n_M = 80$, $n_P = 80$), in Maintenance and Provision, respectively. Supplementary Table 1 reports further summary statistics, including on beliefs about other group members' average effective contributions (which mirror the effective contributions).




Europe PMC Funders Author Manuscripts

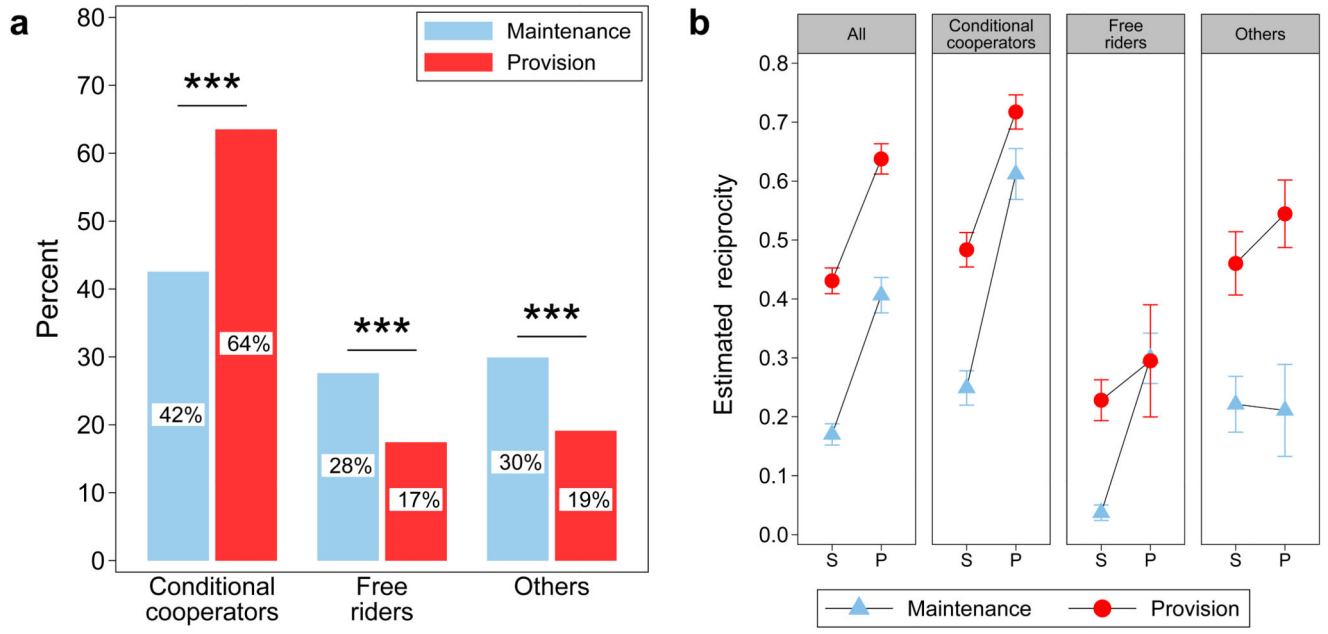


**Figure 3. Reciprocity in Maintenance and Provision.**

**a,** Strong reciprocity as measured by cooperation attitudes; type classification as in30, $n_M$ = 348, $n_P$ = 356. $\chi^2$-tests, *** $P < 0.01$. Results are robust to alternative classification methods (Supplementary Information, Section 1.2). **b,** Estimated reciprocity in repeated interactions (± 1 s.e.m.); Strangers (S), $n$ = 256; Partners (P), $n$ = 160 by treatment and attitude category (conditional cooperators, free riders, others). Positive reciprocity is estimated as the coefficient of lagged average contributions of the other group members ($\bar{C}_{-ij,t-1}$) from multilevel mixed-effects linear regressions (Supplementary Information, Section 1.1; Supplementary Table 3). An alternative estimation approach using finite mixture models55 confirms these results (Supplementary Table 14).




Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

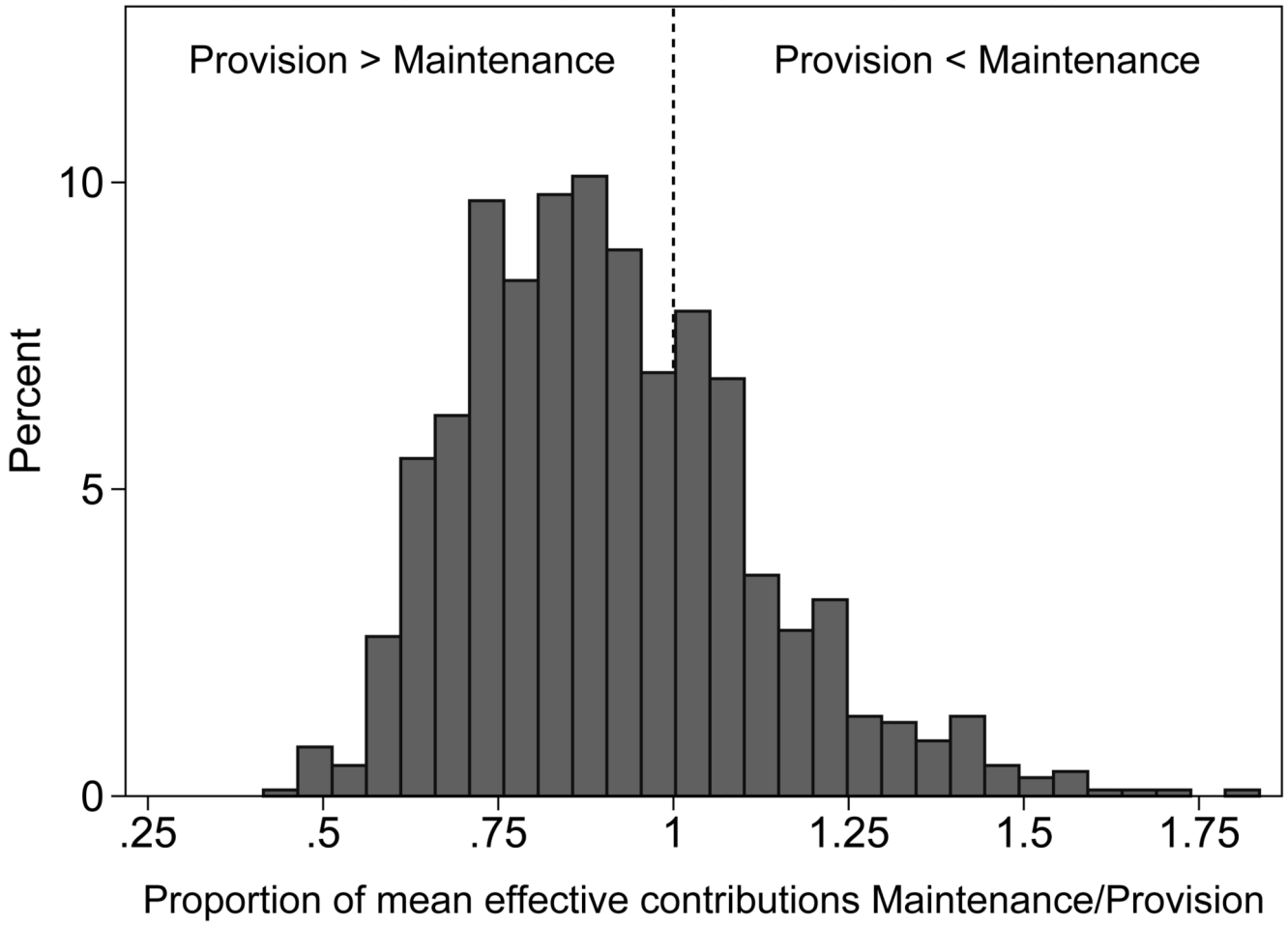


**Figure 4. Simulated effective contribution ratios.**

Distribution of 1000 simulated effective contribution ratios between Maintenance and Provision ($\bar{c}^M / \bar{c}^P$) using a sample of $n = 60$ per treatment and simulated experiment. The sample size reflects the median sample size in related literature.14–23 The mean is 0.91, median is 0.89, and IQR = 0.76 to 1.03. Further details are in Supplementary Information, Section 1.3. As a robustness check, we ran a simulation with $n = 100$, which returns a mean of 0.90, a median of 0.89, and an IQR of 0.79 to 1.00 (see also Supplementary Figure 4).

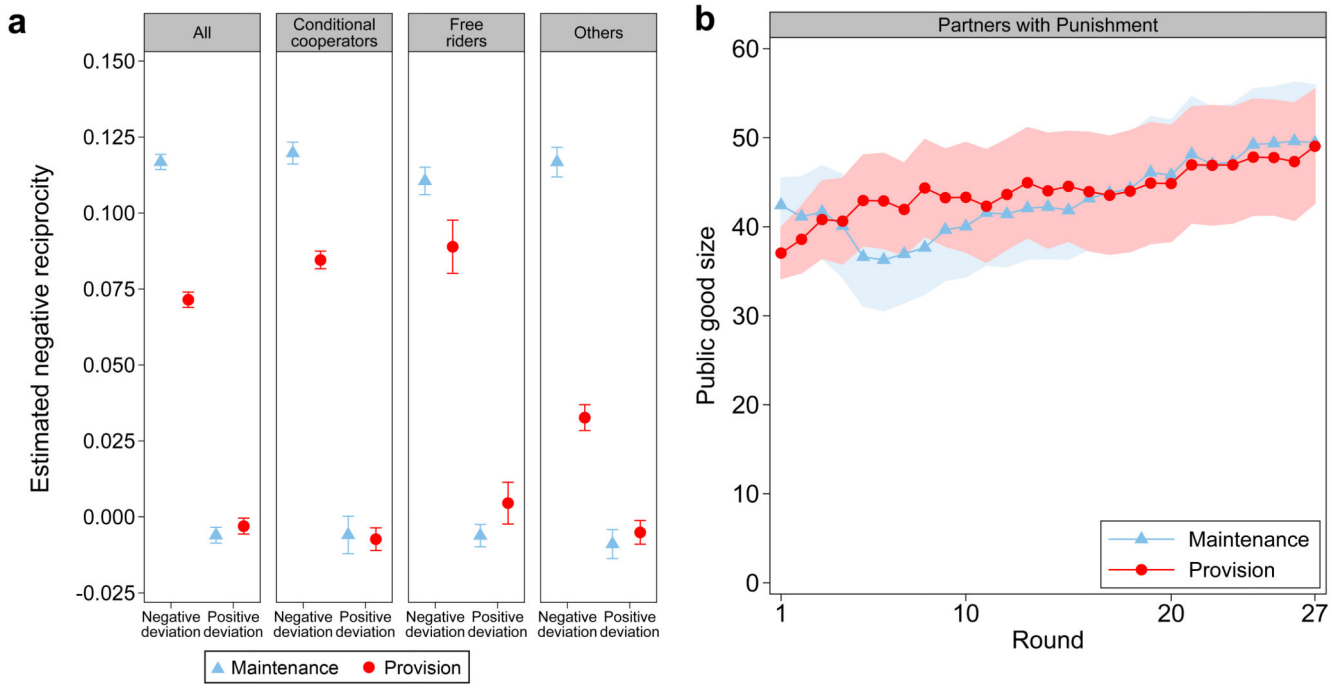


**Figure 5. Partners with Punishment.**

**a,** Estimated negative reciprocity (± 1 s.e.m.); by treatment and attitude category (conditional cooperators, free riders, others). We estimate negative reciprocity in multilevel mixed-effects linear regression as the number of punishment points assigned to effective contributions that deviate negatively from own contribution (Supplementary Table 7; Supplementary Information). **b,** Shown are the effective levels of the public goods (± 1 s.e.m.) over the 27 rounds of interactions ($n_M = 84$, $n_P = 88$).




Europe PMC Funders Author Manuscripts