Preferences and Perceptions in Provision and Maintenance Public Goods*

Simon Gächter\textsuperscript{a,b,c}, Felix Kölle\textsuperscript{d}, and Simone Quercia\textsuperscript{e}

15 January 2021

Abstract
We document a robust asymmetry in preferences and perceptions in incentive-equivalent Provision and Maintenance laboratory public good games. We find fewer conditional cooperators and more free riders in Maintenance than Provision, a difference which is replicable, stable, and reflected in perceptions of kindness. Incentivized control questions administered before game play reveal dilemma-specific misperceptions but controlling for them neither eliminates game-dependent conditional cooperation, nor differences in perceived kindness of others' cooperation. Thus, even when sharing the same game form, Maintenance and Provision are different social dilemmas that require separate behavioural analyses. A theory of revealed altruism can explain our results.

Keywords: social preference elicitation, misperceptions, maintenance and provision social dilemmas, framing, conditional cooperation, kindness, experiments, revealed altruism.

JEL: C92, H41

* This work was supported by the European Research Council [grant number ERC-AdG 295707 COOPERATION] and the Economic and Social Research Council [grant number ES/K002201/1]. The research reported in this paper was approved by the Research Ethics Committee of the Nottingham School of Economics. The authors declare they have no relevant or material financial interests that relate to the research described in this paper. We thank Ben Beranek for excellent research support and Abigail Barr, Gary Charness, Jim Cox, Robin Cubitt, Martin Dufwenberg, Urs Fischbacher, Maria Garcia-Vega, Werner Güth, Friederike Mengel, Elena Manzoni, Vjollca Sadiraj, Maroš Servátka, Chris Starmer, and participants from various seminars and conferences for helpful comments.

\textsuperscript{a} School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK, e-mail: simon.gaechter@nottingham.ac.uk (corresponding author); \textsuperscript{b} CESifo Munich; \textsuperscript{c} IZA Bonn. \textsuperscript{d} University of Cologne, Albertus Magnus Platz, 50923 Cologne, e-mail: felix.koelle@uni-koeln.de. \textsuperscript{e} Economics Department, University of Verona, Via Cantarane 24, 37129 Verona, e-mail: simone.quercia@univr.it. The authors declare no relevant or material financial interests that relate to the research described in this paper.
1. Introduction

In this paper, we analyse the behavioural structure—preferences and perceptions—of two generic forms of voluntary cooperation: providing initially inexistent public goods and maintaining existing ones. Contributing to charities, volunteering, being a team player, or participating in collective action, are examples of voluntary cooperation that provides public goods. Shared natural resources, known as “common-pool resources” (e.g., Ostrom (1990)), but also biodiversity and a stable climate, are important public goods that nature has provided but people need to limit extraction or environmentally damaging emissions if they want to maintain them. Similarly, public goods that previous generations created, such as democracy and the rule of law, only continue existing if people limit rent-seeking and corruption.

As the examples above illustrate, “provision” and “maintenance” public goods differ along many dimensions. Crucially, however, they are all social dilemmas: providing or maintaining the public good is often collectively beneficial, but individual incentives are to hold back on provision and to exploit rather than to maintain the public good – the “tragedy of the commons” (Hardin (1968)). In contrast to this prediction, literature across the behavioural sciences, using field and lab research, has shown that people often cooperate voluntarily and limit the tragedy of the commons. However, until recently, most studies only looked at cooperation behaviour, and much less at the psychological mechanisms – cooperative preferences and perceptions – that produce cooperation. Here, we ask whether from the perspective of social preferences and perceptions, maintenance and provision dilemmas are psychologically different social dilemmas, if, to make them comparable, the only difference between them is whether the public good initially exists or not.

Understanding preferences and perceptions as drivers of cooperation is theoretically interesting because the cited evidence shows that people often cooperate even in anonymous one-shot games without communication, where mechanisms that can support cooperation, such as reputation or repeated interactions (e.g., Dal Bó and Fréchette (2018); Rand and Nowak (2013)) do not apply. Do people cooperate in one-shot games because they misperceive their incentives in a social dilemma, or do they have a ‘genuine’ preference for cooperation? Are maintenance and provision dilemmas distinctive in this regard? The

---

1 The behavioural evidence about cooperation in maintenance and provision dilemmas comes from largely separate literatures. For cooperation in maintenance (common-pool resource) problems, see, e.g., the surveys by Ostrom (1990) and Ostrom (2006). Evidence on cooperation in public goods provision problems is surveyed in, e.g., Ledyard (1995); Gächter and Herrmann (2009); Chaudhuri (2011) and Fehr and Schurtenberger (2018).
theoretical relevance of this question is whether cooperative choices reveal a preference for cooperation, when it is also possible that cooperation results from a misperception, which might also be dilemma specific. Furthermore, if after controlling for misperceptions, elicited preferences for cooperation are also dilemma specific, provision and maintenance dilemmas are different games, not just different “frames”, even when the dilemmas are incentive equivalent. The practical relevance is that any policy implication must rest on accurate behavioural mechanisms, which might be dilemma specific.

Studying people’s preferences for cooperation is important because it can explain why people contribute or not to public goods in lab (e.g., Fischbacher and Gächter (2010); Gächter, et al. (2017)) and field contexts (e.g., Frey and Meier (2004); Alpizar et al., (2008); Rustagi et al., (2010); Fehr and Leibbrandt (2011)). Studying perceptions is important because of the way they interact with people’s preferences. The reason is that the description of a decision situation can affect people’s perceptions of their incentives as well as their evaluation of others’ actions. If people misunderstand their incentives, they may implement choices they otherwise wouldn’t. Thus, observing cooperation without controlling for perception of incentives is not yet a conclusive revelation of a preference for cooperation (for related arguments see, e.g., Koszegi and Rabin (2008) and Cason and Plott (2014)). Perceptions of other players’ actions are important because in social dilemmas, perceptions of other’s kindness can explain why some people cooperate in the first place (e.g., Falk and Fischbacher (2006)). Thus, to understand the fundamental nature of cooperation in maintenance and provision dilemmas, we study preferences and perceptions in conjunction.

We focus sharply on the social dilemma dimension of provision and maintenance public goods, and abstract from technological features (e.g., resource rivalry in many common-pool resources vs. non-rivalrous public goods) and institutional details (rules and regulations) that define the specifics of real-world social dilemmas (e.g., Ostrom (1990); Cornes and Sandler (1996); Apesteguía and Maier-Rigaud (2006); Levin (2014)). Thus, for our purposes, the only relevant difference between maintenance and provision dilemmas is whether the public good initially exists or not. We also abstract from the fact that real-world public goods are sometimes hybrids of provision and maintenance dilemmas.

We compare two versions of a linear public good game that share the same game form: A Provision game and a Maintenance game. In Provision, the public good initially does not exist; four players in a group are endowed with 20 tokens each and decide simultaneously how many of them to contribute to the public good. In Maintenance, players have no
endowment, but the public good already exists because 80 tokens are invested at the outset in the public good. Players decide simultaneously how many (up to 20) tokens to withdraw from the public good. Any token contributed to the public good (in Provision) or not withdrawn from the public good (in Maintenance) is worth 1.6 money units to the group, which is then shared equally between group members; any token not contributed to the public good or withdrawn from the public good is worth 1 money unit.

Maintenance and Provision share the same game form because the strategy sets are the same and material consequences are incentive equivalent. This also implies that the social dilemmas are identical in terms of their strategic complexity: from a material incentives point of view, contributing nothing or withdrawing everything are dominant strategies.

From the viewpoints of standard economics and theories of social preferences (Section 4), behaviour should not be affected by the default position of the public good, because of equivalent game forms. Given game-form equivalence, the two dilemmas might be considered as two different frames of the same game (Salant and Rubinstein (2008)). The framing in terms of different default positions of the public good in Maintenance and Provision might put people into different psychological mindsets: The literature on the behavioural consequences of defaults suggests that people tend to stick to them (see the meta-analysis in Jachimowicz et al., (2019)). One might therefore expect that the resulting size of the public good after contributions or withdrawal decisions are made, will be higher in Maintenance than Provision: the default in Maintenance is an already existing public good, which is absent in the default situation of Provision.

In contrast to the prediction that cooperation is higher in Maintenance than Provision, we found in a related paper (Gächter et al., (2017)) that cooperation is significantly higher in Provision than in Maintenance. For example, in one-shot experiments people on average contributed 41% of their endowment in Provision and 30% in Maintenance. We also found higher cooperation in Provision than Maintenance in repeated one-shot games (“Strangers”) and, like Shanley and Grossman (2007), in repeated games with fixed groups (“Partners”).

Here, we start by replicating these findings in a one-shot experiment using a diverse online subject pool (MTurk). In this new experiment, people contribute 52% of their endowment in Provision compared to 39% in Maintenance. Our anonymous one-shot experiments leave no room for incentives coming from repeated play or from reputation effects that can influence behaviour. In addition, all experiments only started after people
had completed a quiz that tested their understanding of the respective public goods payoff function. Given such robust results and the clean conditions under which we observe them, positive effective contributions might be seen as a revealed preference for cooperation. This would suggest that Maintenance and Provision are different games, not just different frames, because despite identical game forms the two dilemmas would be different with respect to the utility players derive from own and others’ actions.

Although we will conclude our paper by showing that Maintenance and Provision are different games because they trigger different preferences, at the outset it is premature to conclude that contribution choices reveal contribution preferences. The reason is that perceptions of the game and beliefs about other’s actions might also be dilemma specific.

We separate preferences from beliefs and perceptions as follows. First, we recognize that simple contribution choices are not unambiguous with regard to the preference for cooperation they reflect. In particular, pessimistic conditional cooperators who believe others will not contribute might contribute nothing, which makes their behaviour observationally equivalent to a free rider who does not want to cooperate regardless of what others do. To circumvent this problem, we adopt the Fischbacher et al., (2001) strategy-method experiment to separate beliefs from preferences. This incentive-compatible design allows measuring a proxy for cooperation preferences (in the sense of a willingness to pay for cooperation) by controlling for beliefs. We describe this method in Section 3. We find that there are systematically fewer conditional cooperators and more free riders in Maintenance than Provision. In separate experiments, described in Section 5, we show that, within-participants, our measure of cooperation preferences is stable over time and, together with beliefs, predicts contribution or withdrawal decisions in one-shot games played immediately or five months after preferences were elicited.

Second, we measure people’s perceptions about others’ behaviour. In Section 6 we show that in two different subject pools people perceive failing to contribute in Provision to be unkind than withdrawing everything in Maintenance. In contrast, a full effective contribution in Maintenance (withdrawing nothing) is considered less kind than contributing everything in Provision. While this finding supports a preference interpretation, it is possible that some people misperceive the game form because they do not understand the material incentives of the public good game. Such “confusion” is likely because considerable levels of it have been found previously (e.g., Andreoni (1995a); Houser and Kurzban (2002); Ferraro and Vossler (2010); Bayer et al., (2013)). Moreover, irrespective of the game form,
cooperation preferences as measured by the strategy method might be influenced by misperceptions (Burton-Chellew et al., 2016). Conditional cooperation might be a consequence of mistaking the public goods game for a coordination game and thereby failing to see the dominant strategies of withdrawing everything or contributing nothing.

To test for misperception of incentives, we designed a set of eight incentivized control questions that people answered after they had correctly solved ten standard understanding questions covering payoffs in the public good game. We administered the incentivized control questions before we measured participants’ cooperation preferences using the strategy method by Fischbacher et al. (2001). As we explain in detail in Section 7, the eight incentivized questions comprise four payoff and four goal questions. The payoff questions test for the understanding of payoff consequences of various decisions in the public goods game. The goal questions ask what a subject with a particular goal in mind, e.g., maximizing group payoff or maximizing individual payoff, would need to contribute or withdraw. Across all eight questions, we find low rates of misunderstandings: 8.1% of answers in Maintenance and 10.6% of answers in Provision are wrong. The treatment difference is small, but significant ($p = 0.03$), suggesting dilemma-specific misunderstandings.

Our main analysis with regard to misperceptions starts by showing that without controlling for misperceptions, the distribution of types is highly significantly different between Maintenance and Provision with more conditional cooperators (34% vs. 51%) and fewer free riders (39% vs. 27%) in the latter than the former. We then control for misperceptions by applying three measures of misperceptions by excluding: (1) all participants who made at least one mistake in the payoff questions ($M$: 13%; $P$: 22%); (2) all participants who made at least one mistake in the goal questions ($M$: 29%; $P$: 37%); and (3) all mistaken conditional cooperators ($M$: 5%; $P$: 13%). These rates of misperception are on average smaller than the ones found in earlier studies, which (using different methodologies) reported misperception rates of about 50% (Andreoni 1995a; Houser and Kurzban 2002; Ferraro and Vossler 2010; Fosgaard et al. 2017). Interestingly, all three measures of misperceptions are significantly more likely in Provision than Maintenance, which reinforces the observation that misperceptions are frame specific. Applying the exclusion criteria when testing for different type distributions slightly reduces the gaps in

---

2 To measure confusion, previous studies used various experimental designs, like changed incentive structures (Andreoni 1995a), information conditions (Bayer, et al. 2013), or computerized players (Houser and Kurzban 2002; Ferraro and Vossler 2010; Burton-Chellew, et al. 2016). Fosgaard, et al. (2017) used an incentivised post-experimental questionnaire.
the distribution of conditional cooperators and free riders in *Maintenance vs. Provision*, but none of these gaps turns insignificant by controlling for any of the three measures of misperceptions. Perceptions of kindness are also unaffected by misperceptions.

Our results have two theoretical implications: First, while some conditional cooperators are mistaken (e.g., Burton-Chellew *et al.* (2016)), the typical one is not: their conditional choices reveal a properly understood preference for conditional cooperation. All theories of social preferences we consider here can explain this preference. Second, even cooperation preferences free of misperceptions are dilemma specific, which implies that *Maintenance* and *Provision* are different games, not just different “frames”. As we show in Section 4, this contradicts consequentialist theories of social preferences, such as theories of distributional preferences (Fehr and Schmidt (1999); Bolton and Ockenfels (2000)), as well as theories of reciprocity (Rabin (1993); Dufwenberg and Kirchsteiger (2004); Falk and Fischbacher (2006)), but is potentially consistent with guilt aversion (Battigalli and Dufwenberg (2007); Dufwenberg *et al.*, (2011)) and a theory of revealed altruism (Cox *et al.*, (2008)) in which reciprocity is status-quo dependent. The empirical tests we report in Section 8 show that our results are qualitatively consistent only with the theory of revealed altruism.

2. Related literature and our contribution

Our research question relates us to the literature on framing and other-regarding behaviour. For example, previous evidence from dictator games reveals that people are less willing to give if the choice set also includes the option to take away money (List (2007); Bardsley (2008); Cappelen *et al.*, (2013); Dreber *et al.*, (2013); Korenok *et al.*, (2014)). In social dilemmas, early papers on framing typically focused on cooperative behaviour in simultaneous game play without measuring perceptions or preferences. This holds in particular for research in psychology, which compared so-called Take-some vs. Give-some dilemmas (e.g., Brewer and Kramer (1986); McCusker and Carnevale (1995); De Dreu and McCusker (1997); Sell and Son (1997); van Dijk and Wilke (2000)) and also early research in economics (see the seminal paper by Andreoni (1995b); and, e.g., Park (2000); Shanley and Grossman (2007)), with mixed results on which frame induces higher levels of cooperation.

More recent studies separately elicited the effects of framing on beliefs about others’ cooperation (Cubitt *et al.*, (2011a); Dufwenberg *et al.* (2011); Ellingsen *et al.*, (2012)) and found that frames affect beliefs. These studies also did not control for perceptions and
preferences. Finally, recent papers (Frackenpohl et al., (2016); Gächter et al. (2017); Fosgaard et al. (2017)), elicited preferences for conditional cooperation and detected dilemma-specific conditional cooperation, but only Fosgaard et al. (2017) provided some ex post measure of misperception. Measuring perception of incentives before game play heeds arguments by Koszegi and Rabin (2008) and Cason and Plott (2014) that establishing framing effects requires controlling for perception of incentives. Doing so, allows us to show that some subjects do misperceive their incentives, but controlling for such misperceptions neither eliminates conditional cooperation, nor dilemma-specific differences of it.

In light of this previous literature, our main contribution is to provide comprehensive evidence on the interplay between cooperation preferences and (mis)perceptions of two generic social dilemma problems. Our combination of lab and online experiments, replications, between and within-subject stability tests, measurement of perceptions of kindness as well as of the incentives in the games and embedding our comparative analysis in theories of social preferences, make our study the most comprehensive of its kind.

3. The basic setup and the proxy for cooperation preferences

Our setup consists of the two social dilemmas described above, Provision and Maintenance. In both conditions, participants are randomly assigned to groups of $n = 4$. In Provision, each group member $i$ is endowed with 20 tokens, which they can either keep or (partly or fully) contribute ($c_i$) to a “group project”. Contributions to the group project are summed up, multiplied by a factor of 1.6, and distributed equally among the four group members. Equation (1) describes the material incentives of individual $i$:

$$\pi_i = 20 - c_i + \frac{1.6}{4} \sum_{j=1}^{4} c_j.$$  

In Maintenance, 80 tokens are initially placed in a “group project”. Each group member $i$ decides about the allocation of 20 tokens, which they can either leave or (partly or fully) withdraw ($w_i$) from the project. Material incentives are described by equation (2):

$$\pi_i = w_i + \frac{1.6}{4} (80 - \sum_{j=1}^{4} w_j).$$

---

3 Our focus on social preferences and (mis-)perception does not deny the possibility that cognitive ability, risk preferences and loss aversion might matter too, but we leave this for future research, not least to keep this paper manageable.
From a material incentives point of view, (1) and (2) are incentive-equivalent social dilemmas because $c_i = 20 - w_i$. Furthermore, because the material costs of cooperation outweigh its benefits, both the Maintenance and Provision dilemma have full free riding ($c_i = 0; w_i = 20$) as the unique Nash equilibrium in dominant strategies, no matter what other members of their group (are believed to) do.

All experiments were based on these two incentive-equivalent social dilemmas and consisted of three parts. In the first part, participants were introduced to the basic decision situation explaining either the Maintenance or the Provision dilemma and its incentive structure, that is, each participant only faced one of the two social dilemmas (between-participants design). To ensure understanding, participants then had to complete a set of ten computerized control questions. Only after correctly answering all of them, participants could proceed with the experiment. In the second part, which is the main focus of this paper, we implemented a strategy-method public goods game (described below) through which we measure cooperation attitudes, our main proxy for cooperation preferences. After all participants had finished the strategy-method experiment, they received instructions for the third part of the experiment comprising a direct-response interaction that varied in its exact design protocol across different experiments. At the beginning of the experiment, participants were told that the experiment consists of several parts, but that the details about later parts would be disclosed only after they had completed the respective parts. The different designs of the third part could therefore not affect behaviour in previous parts. To avoid spillover effects between the different parts, we randomly re-matched groups after the second part and we informed group members about their decisions only at the very end of all experiments. All instructions and control questions are in Online Appendix A.

To elicit a proxy for cooperation preferences we used the design introduced by Fischbacher et al. (2001), which employs a variant of the strategy method (Selten (1967)). This design elicits an individual’s willingness to cooperate as a function of other group members’ cooperation. Participants played a one-shot version of the game and were asked to make an unconditional and a conditional contribution (withdrawal) decision. In the unconditional decision, participants chose one contribution or withdrawal level. In the conditional decision, participants were asked to fill in a table in which they had to indicate their contribution (withdrawal) decision for each possible (rounded) average contribution (withdrawal) of the other three group members. To guarantee incentive compatibility, in
each group a random mechanism selected three members for whom the unconditional decision was payoff-relevant and one member for whom the conditional decision was payoff-relevant. For this participant, the conditional decision was calculated according to the (rounded) average unconditional decision of the other three group members. The incentive-compatibly elicited attitudes are a proxy for cooperation preferences in the sense that they measure people’s willingness to pay for conditional cooperation.

Following Fischbacher et al. (2001), we classify a participant as a (i) conditional cooperator if their contribution/withdrawal schedule exhibits a (weakly) monotonically increasing pattern, or if the Spearman correlation coefficient between their schedule and the others’ average contribution (withdrawal) is positive and significant at $p < 0.01$; (ii) a free rider if they never contribute anything or withdraw everything irrespective of how much the others contribute (withdraw); and (iii) as other if none of the criteria in (i) & (ii) apply.⁴

Our data come from six experiments and three main sources: the CeDEx laboratory at the University of Nottingham; the online platform Amazon Mechanical Turk (MTurk); and online experiments conducted with students at the University of Nottingham (see Table A1 in the appendix for an overview of our experiments). A total of 2,854 people participated in our experiments. We used z-Tree (Fischbacher (2007)) for conducting the laboratory sessions. For the online experiments on MTurk and the University of Nottingham, we used the survey software Qualtrics. For the lab and online experiments at Nottingham, we recruited student participants from various disciplines at the University of Nottingham using the software ORSEE (Greiner (2015)). Students were only allowed to participate in one lab or online session. On MTurk, we recruited participants from various age and demographic backgrounds.⁵ Average payments were £20.60 for lab sessions, and $2.60 for MTurk sessions (corresponding to an average hourly wage of about $13.00).

4. Theoretical considerations

We start with the observation that all models of social preferences we consider here can explain conditional cooperation. Conditional cooperation is the expression of the

---

⁴ As a robustness check, we used an alternative classification method by Thöni and Volk (2018), who proposed a refinement of the criteria of Fischbacher, et al. (2001). All results are qualitatively and quantitatively in line with those reported below.

⁵ See Horton, et al. (2011) and Arechar, et al. (2018) for a detailed description of MTurk, and a comparison of MTurk versus lab experiments. Both studies as well as Snowberg and Yariv (2020) demonstrate that behaviour in a variety of games is similar on MTurk and the lab.
fundamental principle of reciprocity long deemed relevant for the voluntary supply of public goods (Sugden (1984)) and the evolution of cooperation (e.g., Bowles and Gintis (2011); Guttman (2013)). Our main theoretical question is whether theories of social preferences can explain the observation of a higher share of conditional cooperators and fewer free riders in Provision than in Maintenance as found in previous research (Fosgaard et al., (2014); Frackenpohl et al. (2016); Gächter et al. (2017)) – a finding we expect to replicate here.

Here we only sketch the arguments why particular theories of social preferences can or cannot explain game-dependent conditional cooperation. All details are in Online Appendix C. We first note that theories of distributional preferences (Fehr and Schmidt (1999); Bolton and Ockenfels (2000); Charness and Rabin (2002)) do not predict game-dependent conditional cooperation because material payoff consequences are identical across the two incentive-equivalent social dilemmas. A similar argument holds for models of altruism and fairness such as those by Levine (1998) and Cox et al., (2007).

In theories of reciprocity (Rabin (1993); Dufwenberg and Kirchsteiger (2004); Falk and Fischbacher (2006)), agents’ motivations derive from their material payoff as well as a psychological payoff that depends on their first- or second-order beliefs about others’ actions. These theories can explain conditional cooperation (Dufwenberg et al. (2011)) because agents want to reward kind actions with kindness and to punish hostile actions with unkindness. Kindness is evaluated with a reference point based on material payoffs, which is the same in Maintenance and Provision. Therefore, beliefs are the only channel through which simultaneous game play may differ across Maintenance and Provision. Since in our strategy-method experiment first-order beliefs are fixed because participants condition their contributions on all possible average contributions of others, these models do not predict game-dependent conditional cooperation.

Two theories that can potentially reconcile more conditional cooperation in Provision than Maintenance are guilt aversion and revealed altruism. In modelling guilt aversion, we rely on a model of simple guilt by Battigalli and Dufwenberg (2007) and applied to public goods games in Dufwenberg et al. (2011). An agent’s utility depends on her material payoff as well as her second-order beliefs, that is, what she believes the other players believe she will do. Guilt aversion predicts that player $i$ will suffer guilt if $i$ contributes less than what $i$ thinks the other three group members expect $i$ to contribute (on average). If the disutility from guilt becomes large enough, player $i$ has an incentive to contribute whatever she thinks
others expect her to contribute. Adopting the same specification as in Dufwenberg et al. (2011), i’s utility in the Provision dilemma is defined by:

\[
u_i(c_i, c_j, c_k, c_l, b_{iji}, b_{iki}, b_{ili}) = 20 - c_i + 0.4 \sum_{j=1}^{4} c_j - \gamma_i \max \left(\frac{b_{iji} + b_{iki} + b_{ili}}{3} - c_i, 0\right),\]

where j, k and l denote the other players; b_{iji}, b_{iki}, b_{ili} denote the second-order beliefs of player i and \(\gamma_i\) measures player i’s game-independent degree of guilt aversion. An analogous specification describes utilities in the Maintenance dilemma.

Since the strategy-method experiment fixes only first-order beliefs, differences in second-order beliefs can predict differences in cooperation preferences: If more participants in Provision than in Maintenance have second-order beliefs that others expect them to reciprocate their contributions, the perceived guilt from not matching others’ contributions in Provision may be higher than in Maintenance, which, in turn, can lead to a higher fraction of conditional cooperators in Provision than Maintenance. Similarly, if more participants in Maintenance than in Provision believe that others expect them to free ride, the perceived guilt from actual free riding may be lower in Maintenance than Provision, which would predict more free riders in Maintenance than Provision.

The second theory that can explain more conditional cooperation in Provision compared to Maintenance is an axiomatic theory of revealed altruism by Cox et al. (2008). Revealed altruism assumes that an agent’s reciprocity is triggered by the generosity of others’ actions. Applying Cox et al.’s reciprocity axiom, Axiom R, to our context, the model predicts that participants will perceive higher contributions by the other group members as more generous towards them, and, as a consequence, they will be more altruistic towards the others. In our strategy-method experiment, this will be manifested in a positive slope of the contribution schedule (see Cox et al., (2013) for a related analysis, and Online Appendix C for formal definitions of the axioms and derivation of these predictions).

In a second axiom, Axiom S, Cox et al. (2008) assume that, based on the psychological asymmetry behind omission and commission (see, e.g., Spranca et al., (1991)), generous actions that change the status quo trigger stronger reciprocity than generous actions that just uphold the status quo. That is, Axiom S strengthens or weakens the effect of Axiom R depending on the status quo of where resources are allocated initially. Applied to our strategy-method experiment in which there are three first movers and one second mover, Axiom S predicts that the same average effective contribution by the first movers strengthens
reciprocity and therefore triggers higher contributions by the second mover in *Provision* than in *Maintenance*. The reason is that in *Provision* any positive contributions by the other three group members increase the payoff opportunities of the second mover compared to the status quo where nothing is contributed to the public good. In *Maintenance*, in contrast, where all resources are initially allocated to the public good, any withdrawal by the other three group members reduces the payoff opportunities for the second mover and, hence, will weaken reciprocity (see Proposition 3 in Online Appendix C).

To illustrate the effects of Axiom R and Axiom S, we introduce a utility function that represents preferences as in Cox *et al.* (2008). Consider the following utility function:

\[ u_i = \pi_i - \gamma_i^F \max(\bar{c}_{-i} - c_i, 0), \]

where \( \pi_i \) is the material payoff for player \( i \) and \( \gamma_i^F \) represents player \( i \)'s degree of reciprocity under \( F \in (M, P) \) *Maintenance* or *Provision*. The effective contribution of player \( i \) is \( c_i \) and the average effective contribution of the other group members is denoted by \( \bar{c}_{-i} \).

Taking the average contribution \( \bar{c}_{-i} \) of the others as given (as it is the case in our strategy-method experiment), the best response of individual \( i \) depends on \( i \)'s degree of reciprocity, \( \gamma_i^F \). Given the parameters of our experiment, any token contributed to (not withdrawn from) the public good generates a material cost of \( 0.4 - 1 = -0.6 \). Hence if \( \gamma_i^F > 0.6 \) the best reply of individual \( i \) will be to match others’ contributions, while if \( \gamma_i^F < 0.6 \) her best reply is to free ride and contribute nothing for every possible average contribution of the others. In terms of two axioms described above, people with \( \gamma_i^F > 0.6 \) satisfy Axiom R. Hence, under this parametrization, depending on the distribution of the reciprocity parameter, \( \gamma_i^F \), some people will be conditional cooperators, and some will be free riders.

Applying Axiom S to our parametrization implies that the reciprocity parameter \( \gamma_i^F \) is not the same across *Provision* and *Maintenance*, but that (on average) \( \gamma_i^P \geq \gamma_i^M \). While for participants satisfying either \( 0.6 > \gamma_i^P \geq \gamma_i^M \) or \( \gamma_i^P \geq 0.6 \), behaviour is predicted to be identical across dilemma types, participants with \( \gamma_i^P > 0.6 > \gamma_i^M \) are predicted to be conditional cooperators in *Provision* and free riders in *Maintenance*. Hence, this model is consistent with more conditional cooperators in *Provision* compared to *Maintenance*.

In sum, all theories of social preferences considered here can explain conditional cooperation but, with the exception of guilt aversion and revealed altruism, predict that the frequency of conditional cooperators and free riders is the same in *Maintenance* and
Provision dilemmas. All theories also assume that individuals have well-formed and stable preferences. Thus, before we test for potential game-form misperceptions, we first report analyses that test the robustness and stability of game-dependent preferences both in different samples and over time. We further provide evidence for the validity of our preference elicitation method by testing whether our proxy for cooperation preferences together with beliefs predict cooperation in a direct response public good game.

5. Robustness and stability of preferences in Provision and Maintenance dilemmas

We start by summarizing the findings from our previous study (Gächter et al. (2017)). We then compare these results with an online replication study conducted on MTurk. Being able to replicate the basic phenomenon we want to study is an important first step in our analysis. After that, we show that cooperation preferences are not only stable between different subject pools but are also stable over time within participants. Finally, we investigate the predictive power of the elicited cooperation preferences for simultaneous game play and compare it across Maintenance and Provision. All procedural details and further supporting evidence are in Online Appendix B1.

5.1. Gächter et al. (2017) and a replication on MTurk

The left panel of Figure 1 summarizes the main finding from Gächter et al. (2017), which was based on a one-shot strategy method experiment as described in Section 3. Participants were significantly more likely to be conditional cooperators ($\chi^2(1) = 31.03; p < 0.001$) and significantly less likely to be free riders ($\chi^2(1) = 10.46; p < 0.001$) and others ($\chi^2(1) = 11.08; p = 0.001$) in Provision than in Maintenance. In a one-shot direct response game played after the type elicitation, Gächter et al. (2017), further found that cooperation rates were higher in Provision than in Maintenance (41% vs. 30%; two-sided t-test: $p = 0.007$).

---


7 The category ‘others’ contains “unconditional cooperators” who contribute a constant positive amount irrespective of what other group members contribute (3% in P, 3% in M, $p = 0.473$); “anti-conditional cooperators” whose cooperation depends negatively on the cooperation of other group members (3% in P, and 7% in M, $p = 0.006$), “triangle cooperators” who are conditionally cooperative up to a certain level when they turn into anti-conditional cooperators (9% in P, 11% in M, $p = 0.387$) and the rest (5% in P, 9% in M, $p = 0.065$). Further details are in Online Appendix Table B1 (Panel A).
Our replication for the purposes of this paper was conducted on the online labor market platform MTurk with $n = 703$ US participants (instructions are in Online Appendix A).\footnote{We decided to replicate the findings of Gächter, et al. (2017) with a planned sample size of $n = 700$ because we were interested in the robustness of our lab results with undergraduates in a much more diverse subject pool. Based on the differences in the type distributions in the left panel of Fig. 1, a sample size of $n = 215$ would have sufficed to detect the same effect size with a power of 0.99 at $\alpha = 0.001$ (calculations based on G*Power 3.1, Faul, et al. (2007)). However, given the different socio-demographic characteristics of the subject pool and the online nature of the experiment in MTurk, we decided to increase the sample to $n = 700$.} The results match our previous findings. MTurkers contribute 52\% of their endowment in Provision compared to 39\% in Maintenance. While levels in the frequency of types are different, treatment differences are highly significant. As shown in the right panel of Figure 1, we find a larger fraction of conditional cooperators (80\% vs. 67\%, $\chi^2(1) = 14.75; p < 0.001$), and a lower fraction of free riders (8\% vs. 17\%, $\chi^2(1) = 10.75; p = 0.001$) and others (12\% vs. 16\%, $\chi^2(1) = 3.07; p = 0.080$) in Provision compared to Maintenance. Overall, the distribution of types is significantly different between dilemmas ($\chi^2(2) = 15.96, p < 0.001$).\footnote{Our MTurk results are similar to Kocher, et al. (2008): When comparing their US sample with our MTurk results, we find a remarkably similar distribution of types ($\chi^2(2) = 0.02; p = 0.992$): 81\% vs. 80\% conditional cooperators, 8\% vs. 8\% free riders, and 11\% vs. 12\% others. We thank M. Kocher for providing the data. Disaggregating the category ‘others’ shows similar results than in Gächter, et al. (2017). See Table B1 (Panel B) for the details.} Furthermore, like in our previous study, we find that cooperation rates, measured in a one-shot direct-response game played after the type elicitation, are significantly higher in Provision than in Maintenance (52\% vs. 39\%, two-sided t-test: $p < 0.001$). In both samples, we also find unconditional contributions in the strategy-method to differ significantly across
both dilemmas (Lab: \textit{Provision}: 42\%, \textit{Maintenance}: 34\%, two-sided t-test: \( p = 0.003 \); MTurk: \textit{Provision}: 53\%, \textit{Maintenance}: 38\%, two-sided t-test: \( p < 0.001 \)).

5.2 Temporal stability

To test whether the observed difference in the distribution of types is also stable \textit{within} participants, we ran an additional experiment in which we re-invited a subset of participants of our first experiment (left panel of Figure 1) four months earlier. Without knowing in advance, participants took part in sessions which were identical to the ones in which they participated before. We report results from \( n = 119 \) participants, \( n = 65 \) in \textit{Provision} and \( n = 54 \) in \textit{Maintenance}, who showed up in both waves.

At the aggregate level, cooperation preferences are remarkably stable over a period of four months. The distribution of types \textit{within treatments} is very similar and does not significantly change between waves, neither in \textit{Maintenance} (\( \chi^2(2) = 0.51, p = 0.776 \)) nor in \textit{Provision} (\( \chi^2(2) = 1.57, p = 0.456 \)). When comparing the distribution of types \textit{across treatments}, we find a significantly different distribution across \textit{Maintenance} and \textit{Provision} for both Wave 1 and Wave 2 (\( \chi^2(2) = 10.32, p = 0.006 \) and \( \chi^2(2) = 11.87, p = 0.003 \), respectively; see also Table B2). With regard to \textit{individual-level stability} of cooperation preferences, we find that 63\% of participants are classified as the same type in both waves. This is significantly higher than chance (Table B3) and very similar to a similar analysis by Volk \textit{et al.}, (2012) who, in a \textit{Provision} experiment, found a stability rate of 64\%. We also find no significant differences across treatments; the percentage of ‘stable’ participants is 59\% and 66\% in \textit{Maintenance} and \textit{Provision}, respectively (\( \chi^2(1) = 0.60, p = 0.438 \)).

5.3. Predictive power of cooperation preferences

If our proxy for cooperation preferences measures something fundamental about people’s attitude towards cooperation, it should be predictive of actual behaviour in another comparable environment. To test this, we rely on the third part of our experiment. After the strategy-method experiment in the second part, a subset of subjects participated in a one-shot direct-response public goods game in which they made a single contribution decision. We also elicited incentivized beliefs about the average contribution of the other group members. Following Fischbacher \textit{et al.}, (2012), by combining elicited cooperation preferences with stated beliefs we can make a point prediction about the contribution
decision, $\hat{c}_i$. For each individual, we then compare $\hat{c}_i$ with their actual contribution in the direct-response game, $c_i$, delivering an individual-level measure of consistency.

In total, we have (1) $n = 288$ observations from participants in our lab experiment, and (2) $n = 703$ observations from participants in our MTurk experiment. We further report data from (3) a set of $n = 116$ participants, for which the elicitation of cooperation preferences and the direct-response game took place in two separate sessions that lay five months apart. To our knowledge, this is the first paper that undertakes such a test.

The median deviation $(c_i - \hat{c}_i)$ in both treatments is zero for all samples, that is, participants’ contribution in the direct-response game is perfectly consistent with their predicted contribution from the strategy-method, even after a delay of 5 months. While not all participants are completely consistent, for none of the three samples the distribution of deviations is significantly different across Maintenance and Provision (Kolmogorov Smirnov tests; Lab: $p = 0.195$; MTurk: $p = 0.532$; Lab (5 months): $p = 0.472$). See Figure B1 in Online Appendix B for an illustration of this result.

5.4. Discussion

In a series of experiments, we have shown that participants are more likely to be conditional cooperators in Provision compared to Maintenance. This finding is consistent with similar previous results (Frackenpohl et al. (2016); Fosgaard et al. (2017); and Gächter et al. (2017)). The temporal stability we found is also consistent with a previous related lab study (Volk et al. (2012) and with stable social preferences found in field social dilemmas (Carlsson et al., (2014)). Given this wealth of replicable evidence, we consider the different distribution of types between Maintenance and Provision a stylised fact.

While replicability, stability, and predictive power are necessary conditions to interpret the effects as difference in underlying social preferences, they are not sufficient. An alternative interpretation of the observed differences is that they are due to stable and systematic misperceptions of the game form. In the next two sections, we disentangle the relative importance of these two explanations. We start in the next section by investigating whether the differences between Provision and Maintenance can be related to different social perceptions across the two contexts.
6. Perceptions of kindness differ between Maintenance and Provision

A prominent psychological explanation for the existence of conditional cooperation is that individuals are reciprocal, that is, they have a desire to reward kind intentions with kindness and punish unkind intentions with unkindness (see Fehr and Schurtenberger (2018) for a review). Hence, as reciprocity is the behavioural response to perceived kindness or unkindness (Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006)), a crucial question is how participants evaluate actions of others in terms of (un)kindness, and whether these evaluations differ across games. In particular, if people perceive payoff-equivalent actions differently in terms of kindness across Maintenance and Provision, this could trigger game-specific reciprocal responses which, in turn, could explain the observed differences in conditional cooperation across the two setups.

To test this conjecture, we conducted two online studies in which we elicited kindness perceptions about other people’s contribution behaviour for both types of social dilemmas (see Falk and Fischbacher (2006) for a related exercise and Wilson (2012) for a cautionary note). In the questionnaire, we explained participants either a Maintenance or a Provision dilemma, and then asked them to evaluate the kindness of average effective contributions of three other group members on a scale from -100 to +100 (where -100 corresponds to ‘very unkind’ and +100 corresponds to ‘very kind’). We asked participants to evaluate the kindness of a low, an intermediate, and a high effective contribution of 0, 10, and 20, respectively (for the exact wording, see Online Appendix A3). We recruited $n = 185$ students from the University of Nottingham and $n = 401$ participants via MTurk. No participant was involved in any of our experimental sessions before. Results are in Figure 2.

Figure 2 reports the average kindness evaluation of others’ average effective contributions. The results from the two samples are remarkably similar. While low effective contributions of 0 are considered as significantly less kind in Provision than in Maintenance (two-sided t-tests, $p < 0.001$ and $p < 0.001$ for students and MTurkers, respectively), we observe the reverse pattern for medium and high effective contributions of 10 and 20, respectively. In these cases, payoff equivalent actions are considered as more unkind in Maintenance compared to Provision (two-sided t-tests, average others’ contribution = 10,

---

10 Since we asked participants for their personal judgments, answers were not incentivized. However, we did incentivize participation. Student participants were offered three randomly drawn prizes of £50 each. MTurkers received a flat payment of $2. According to Cubitt, et al. (2011b) who studied moral judgments in social dilemmas, incentivizing participation does not affect moral judgments, making it unlikely that it affects kindness evaluations.
p = 0.045 and p = 0.001 for students and MTurkers, respectively; average others’ contribution = 20, p = 0.007 and p = 0.062 for students and MTurkers, respectively).

Figure 2: Kindness perceptions in Provision and Maintenance of others’ effective contributions (±1 s.e.m).

Further support for this result comes from OLS regressions in which we use kindness evaluations as the dependent variable, and others’ average contributions, a dummy for the framing manipulation, and an interaction term of the last two as independent variables. We run this regression separately for each subject pool. The results are shown in Table B4. In line with the graphical impressions from Figure 2, we find a positive and significant coefficient for the interaction term, indicating greater responsiveness of kindness evaluations to others’ contributions in Provision than in Maintenance.

In sum, the higher share of conditional cooperators in Provision than in Maintenance seems to be rooted in a stronger kindness-contribution relationship in Provision compared to Maintenance: Complete free riding is perceived to be more unkind in Provision than in Maintenance. Positive contributions (of 10 and 20) are perceived as more kind in Provision than in Maintenance. This result suggests that both with respect to kindness and unkindness, individuals have stronger reactions in their perception of others’ contributions when they are framed as contributing to, rather than withdrawing from, a public good. These stronger reactions likely trigger a stronger need to reciprocate and hence can explain the higher frequency of conditional cooperators in Provision compared to Maintenance. In the next

---

11 This result is reminiscent of Cubitt, et al. (2011b) who found that people’s moral judgment of contributing 0 is more negative in Provision than in Maintenance. A possible explanation is that withdrawing 20 out of 80 is perceived as less bad than giving nothing out of 20.
section, we test to what extent these conclusions are confounded by misperceptions of the game form, which could affect perceptions of kindness as well as revealed preferences.

7. Game form misperceptions

One alternative explanation for our results, which has been proposed in previous literature (Fosgaard et al. (2017)), is that participants may have systematic misperceptions of the game form and that these misperceptions may be treatment specific. If this were the case, differences in conditional cooperation may not be due to different social preferences but due to differences in the understanding of the incentives of the game. In this section, we assess to what extent game-form misperceptions can explain our results.

7.1. Conceptualization of game form misperceptions

Our conceptualization of game-form misperceptions draws on Cason and Plott (2014). They analyse the tension between standard theory, which assumes that preferences are only influenced by elements of the game form (i.e., the set of actions, the set of material consequences, and the links between actions and consequences), and non-standard theories, which postulate that preferences may depend on elements outside the game form such as how the game form is described. In their example, Cason and Plott investigate anomalous bidding behaviour in the Becker et al., (1964) mechanism. While observed bids are consistent with frame-dependent preferences, Cason and Plott show that this effect is driven by a subset of participants who mistakenly perceive the situation as a first-price auction rather than a second-price auction. They conclude that in their case, the description of the decision situation affected participants’ perception of the game form, which, in turn, led them to implement ‘wrong’ behavioural responses given their underlying preferences. Cason and Plott’s general conclusion is that researchers should be careful of interpreting choices as revealed preferences, an issue that Koszegi and Rabin (2008) also point out.

The implication of Cason and Plott’s argument for our context is that interpreting differences in behavioural responses across treatments as evidence for frame-dependent preferences might be erroneous because such differences can be due to frame-dependent misperceptions of the game form. In particular, if failure of correct game-form recognition is also at work in our setup, which is possible given previous evidence on confusion (see introduction), then the observed distribution of cooperation types might not reflect participants’ true cooperation preferences as some of the participants might have mistakenly
implemented behaviour different from their preferred one. For example, if some participants erroneously believe that in order to maximize their individual income, they should increase their contribution if the contributions of other group members increase, this might lead to an inflated rate of conditional cooperation. Moreover, if this type of game-form misperception is more frequent in Provision than in Maintenance, this could explain the observed framing effect of a higher frequency of conditional cooperation and a lower frequency of free riding in the former than the latter. Some evidence for this possibility comes from Fosgaard et al. (2017) who, in an online study with a representative Danish sample, find that many participants fail to recognize the dominant strategy of full free riding, and that this type of mistake occurs more frequently in Provision than in Maintenance. In the next subsection, we describe the details of a new experiment that was specifically designed to examine the role of misperceptions in our context.

7.2. Measurement of game form misperceptions

We measure game-form misperceptions in a new experiment with \( n = 696 \) Nottingham students who had not participated in any of our experiments before. The experiment followed the structure presented in Section 3, except that there was no direct-response experiment after the strategy-method experiment. Instead, after participants answered the standard set of ten control questions, we asked them two additional sets of four incentivized questions (see Table 1), paying £0.1 per correct answer. After that, the experiment proceeded with the elicitation of cooperation preferences using the strategy method.

The first set of questions, which we label payoff questions, are akin to standard control questions in which participants have to calculate earnings for various contribution scenarios. We asked participants to determine their own and others’ monetary earnings in case (i) they would contribute their whole endowment but the other group members would contribute nothing and (ii) they would contribute nothing but each of the other group members would contribute their whole endowment (20 tokens) (see Q1 – Q4 in Table 1 for the exact wording of questions). Our first measure of game-form misperception classifies a subject as misperceiving if they make at least one mistake in the payoff questions (see Bartling et al., (2015) for a similar approach in a value elicitation task).

---

12 To avoid any income effects when eliciting cooperation preferences, incentives were modest, and participants were informed about the number of questions they answered correctly only at the very end of the experiment.
Table 1: Incentivized misperception questions and percentage of correct answers across Maintenance (M, \( n = 320 \)) and Provision (P, \( n = 376 \)).

Q1-Q4: payoff questions; Q5-Q8: goal questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>% correct answers</th>
<th>M</th>
<th>P</th>
<th>( \chi^2 ) – test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: Assume that you contribute ([\text{withdraw}] 20 {0} \text{ tokens to } [\text{from}]) the project and the other three group members contribute nothing to ([\text{withdraw} 20 \text{ tokens each from}]) the project. What will your total income (in points) be?</td>
<td>95.3</td>
<td>95.5</td>
<td>( p = 0.917 )</td>
<td></td>
</tr>
<tr>
<td>Q2: Assume that you contribute ([\text{withdraw}] 20 {0} \text{ tokens to } [\text{from}]) the project and the other three group members contribute nothing to ([\text{withdraw} 20 \text{ tokens each from}]) the project. What will the total income (in points) of each of the other group members be?</td>
<td>90.9</td>
<td>84.3</td>
<td>( p = 0.009 )</td>
<td></td>
</tr>
<tr>
<td>Q3: Assume that you contribute ([\text{withdraw}] 0 {20} \text{ tokens to } [\text{from}]) the project and each of the other three group members contribute 20 tokens ([\text{withdraw nothing from}]) the project. What will your total income (in points) be?</td>
<td>94.4</td>
<td>92.8</td>
<td>( p = 0.405 )</td>
<td></td>
</tr>
<tr>
<td>Q4: Assume that you contribute ([\text{withdraw}] 0 {20} \text{ tokens to } [\text{from}]) the project and each of the other three group members contribute 20 tokens ([\text{withdraw nothing from}]) the project. What will the total income (in points) of each of the other group members be?</td>
<td>95.9</td>
<td>93.1</td>
<td>( p = 0.103 )</td>
<td></td>
</tr>
<tr>
<td>Q5: Suppose the other group members contribute ([\text{withdraw}] ) on average 0 {20} tokens to ([\text{from}]) the project. How much should a person who wants to make as much money as possible for him/herself contribute to ([\text{withdraw from}]) the project?</td>
<td>92.8</td>
<td>95.0</td>
<td>( p = 0.239 )</td>
<td></td>
</tr>
<tr>
<td>Q6: Suppose the other group members contribute ([\text{withdraw}] ) on average 20 {0} tokens to ([\text{from}]) the project. How much should a person who wants to make as much money as possible for him/herself contribute to ([\text{withdraw from}]) the project?</td>
<td>93.8</td>
<td>85.6</td>
<td>( p = 0.001 )</td>
<td></td>
</tr>
<tr>
<td>Q7: Suppose the other group members contribute ([\text{withdraw}] ) on average 0 {20} tokens to ([\text{from}]) the project. How much should a person who wants that the group as a whole makes as much money as possible contribute to ([\text{withdraw from}]) the project?</td>
<td>79.1</td>
<td>77.1</td>
<td>( p = 0.539 )</td>
<td></td>
</tr>
<tr>
<td>Q8: Suppose the other group members contribute ([\text{withdraw}] ) on average 20 {0} tokens to ([\text{from}]) the project. How much should a person who wants that the group as a whole makes as much money as possible contribute to ([\text{withdraw from}]) the project?</td>
<td>93.1</td>
<td>92.0</td>
<td>( p = 0.581 )</td>
<td></td>
</tr>
<tr>
<td>Total*</td>
<td></td>
<td>91.9</td>
<td>89.4</td>
<td>( p = 0.030 )</td>
</tr>
</tbody>
</table>

Notes: * For testing the total effect (last row) we use logistic regressions with standard errors clustered at the individual level.

In the other four questions (compare Q5 – Q8 in Table 1), which we label goal questions, we follow a similar strategy as Fosgaard et al. (2017) and ask participants what a person who wants to implement a specific goal should do. The first goal was individual payoff.
maximization; participants were asked how much a person who “wants to make as much money as possible for him/herself” should contribute given the other group members contribute either 0 or 20. The second goal was group payoff maximization: we asked participants how much a person who “wants that the group as a whole makes as much money as possible” should contribute given the other group members contribute either 0 or 20. We classify a participant as misperceiving if they make at least one mistake in the goal questions. This constitutes our second measure of game-form misperception. Compared to the payoff questions, which require an understanding of the incentive structure as well as sufficient calculation skills, the goal questions require the ability to put oneself into the shoes of another person that might have different objectives than oneself, a task that is arguably more difficult than just calculating payoffs.

Finally, our third measure checks whether a subject is a mistaken conditional cooperator, that is, whether they think that maximizing their own income requires increasing own contribution if others’ contributions increase (from 0 to 20), that is, if their response to Q6 is strictly higher than their response to Q5. Such a mistake could lead participants to believe that they face incentives akin to a coordination game rather than a social dilemma game. As a result, similar to the case of Cason and Plott (2014), participants may then implement “wrong” behavioural responses given their underlying preferences. That is, while the behavioural response of such a misperceiving subject in the strategy method might look like evidence for prosocial, reciprocal preferences, such behaviour is also consistent with a model in which a purely selfish subject maximizes their misperceived payoffs.

7.3. Misperceptions are frame specific

Table 1 summarises the degree of misperception across Maintenance and Provision separately for each question. It reveals that in both treatments there is an overall very low level of misperception. With a few exceptions, the percentage of correct answers is above 90% for each single question and treatment. On average, participants answer 91% of the questions correctly, 92% in Maintenance (7.35 out of 8) and 89% in Provision (7.15 out of 8). Despite the overall treatment differences being small, they are statistically significant at the 5% level ($p = 0.030$). When comparing the fraction of correct answers between Maintenance and Provision for each question separately, we find significant differences for two out of the eight questions, Q2 ($p = 0.009$) and Q6 ($p = 0.001$).
To investigate where these differences come from, we now turn to an individual-level analysis of mistakes by applying our three measures of game-form misperceptions as described above. As shown in Table 2, the level of misperception varies across our different measures – in both treatments it is highest for the goal questions and lowest for mistaken conditional cooperation. The high fraction of misperception in the goal questions is mainly due to Q7 (Table 1), which asks subjects to state what is best for the group when others contribute 0. Moreover, for all measures, we find that misperceptions are significantly more frequent in Provision than in Maintenance. The number of people misperceiving is between 8 and 9 percentage points higher in Provision than in Maintenance ($\chi^2$-test, all $p < 0.023$, compare Table 2). For Measure 1 this is mainly due to differences in Q2, while for Measures 2 and 3 it largely due to Q6 (see Table 1). We hence find evidence that different descriptions of the same game form cause different degrees of misperceptions.

Table 2: Percent of participants classified as misperceiving in Maintenance and Provision

<table>
<thead>
<tr>
<th>Measure</th>
<th>Maintenance [n = 320]</th>
<th>Provision [n = 376]</th>
<th>$\chi^2$-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure 1 – At least one mistake in the payoff questions Q1-Q4</td>
<td>13% [n = 40]</td>
<td>22% [n = 84]</td>
<td>$p = 0.001$</td>
</tr>
<tr>
<td>Measure 2 – At least one mistake in the goal questions Q5-Q8</td>
<td>29% [n = 93]</td>
<td>37% [n = 140]</td>
<td>$p = 0.023$</td>
</tr>
<tr>
<td>Measure 3 – Mistaken conditional cooperation</td>
<td>5% [n = 17]</td>
<td>13% [n = 47]</td>
<td>$p = 0.001$</td>
</tr>
</tbody>
</table>

7.4. Misperceptions and cooperation preferences

In the following, we analyse the connection between misperception and elicited cooperation preferences. If there was none, i.e., if mistakes were randomly distributed across types, then the different degrees of misperception across Maintenance and Provision should have no effect on the distribution of types. If, instead, the likelihood of game-form misperception is correlated with displaying a certain cooperation type, this could explain the differences in conditional cooperation we observed across our two treatments.

Table 3 reports the fraction of misperceiving subjects conditional on being classified as a certain type. We report these numbers separately for Maintenance and Provision and our three measures of misperception. Table 3 reveals that the null hypothesis of no relationship
between types and misperceptions can be rejected for *Maintenance* and *Provision* according to Measure 2 (goal questions) and Measure 3 (mistaken conditional cooperation), but not for Measure 1 (payoff questions). It further reveals that while in *Maintenance* mainly subjects classified as others have some form of misperceptions, in *Provision* it is particular conditional cooperators who are more likely than the other types to display misperceptions.

**Table 3: Fraction of misperceiving subjects in Maintenance and Provision by type**

<table>
<thead>
<tr>
<th></th>
<th>Maintenance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measure 1</td>
<td>Measure 2</td>
<td>Measure 3</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.103</td>
<td>0.252</td>
<td>0.028</td>
</tr>
<tr>
<td>Cooperators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Riders</td>
<td>0.095</td>
<td>0.230</td>
<td>0.024</td>
</tr>
<tr>
<td>Others</td>
<td>0.195</td>
<td>0.425</td>
<td>0.126</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measure 1</td>
<td>Measure 2</td>
<td>Measure 3</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.247</td>
<td>0.447</td>
<td>0.200</td>
</tr>
<tr>
<td>Cooperators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Riders</td>
<td>0.165</td>
<td>0.262</td>
<td>0.029</td>
</tr>
<tr>
<td>Others</td>
<td>0.241</td>
<td>0.337</td>
<td>0.072</td>
</tr>
</tbody>
</table>

*χ²* - tests  

<table>
<thead>
<tr>
<th></th>
<th>p = 0.066</th>
<th>p = 0.005</th>
<th>p = 0.002</th>
</tr>
</thead>
</table>

|                  | p = 0.247 | p = 0.006 | p < 0.001 |

Note: For definitions of Measures 1 to 3 see Table 2.

These results show that the framing of the decision situation not only affects the overall level of misperception, but also how perceptions interfere with preferences. This provides a strong case for the need of controlling for misperceptions before interpreting behavioural differences as frame-dependent preferences. In the following, we therefore test whether accounting for the different types of misperceptions can explain the observed framing effect in the distribution of cooperation preferences. We report this analysis in Table 4.

Panel A of Table 4 shows the distribution of types in the full sample for *Maintenance* and *Provision*. In line with our results from Section 5, we find again a highly significant difference in the distribution of preferences across treatments (*χ²*(2) = 21.23, *p* < 0.001), with significantly fewer conditional cooperators (*χ²*(1) = 20.65, *p* < 0.001) and significantly more free riders (*χ²*(1) = 11.24, *p* = 0.001) in *Maintenance* than in *Provision*.¹³

¹³ The relative frequency of types is somewhat different compared to Figure 1. We find a significant change in the distribution of preferences in both *Provision* (*χ²*(2) = 15.02, *p* = 0.001) and *Maintenance* (*χ²*(2) = 13.59, *p* = 0.001). In the new experiment, we find more free riders (*Provision*: 27% vs. 16%, *χ²*(1) = 8.00, *p* = 0.005, *Maintenance*: 39% vs. 23%, *χ²*(1) = 13.57, *p* < 0.001) and fewer conditional cooperators (*Provision*: 68% vs. 51%, *χ²*(1) = 14.75, *p* = 0.000, *Maintenance*: 34% vs. 43%, *χ²*(1) = 4.42, *p* = 0.036). We believe the reason for this result is that we administered the incentivized misperception questions before the elicitation of preferences. The set of incentivized questions likely made the incentive structure of the social dilemma situation even clearer. If the difference between *Maintenance* and *Provision* is caused by misperceptions of incentives, then this intervention alone could weaken the treatment difference in cooperation preferences.
In panels B, C, and D, we compare the distribution of types across *Maintenance* and *Provision* after dropping participants who are classified as misperceiving according to Measures 1, 2, and 3, respectively. The results reveal that our main result of different distributions of cooperation preferences across *Maintenance* and *Provision* is robust to the exclusion of participants who do not fully understand the game form. That is, the distribution of types is significantly different across *Maintenance* and *Provision* across all subsamples of non-confused participants ($\chi^2$-tests, all $p < 0.05$).

**Table 4: Distribution of cooperation preferences in Maintenance and Provision after controlling for different types of misperceptions**

<table>
<thead>
<tr>
<th>Type</th>
<th>Maintenance ($n = 320$)</th>
<th>Provision ($n = 376$)</th>
<th>$\chi^2$-test</th>
<th>Maintenance ($n = 280$)</th>
<th>Provision ($n = 292$)</th>
<th>$\chi^2$-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional Cooperators</td>
<td>34%</td>
<td>51%</td>
<td>$p &lt; 0.001$</td>
<td>34%</td>
<td>49%</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Free Riders</td>
<td>39%</td>
<td>27%</td>
<td>$p = 0.001$</td>
<td>41%</td>
<td>29%</td>
<td>$p = 0.005$</td>
</tr>
<tr>
<td>Others</td>
<td>27%</td>
<td>22%</td>
<td>$p = 0.118$</td>
<td>25%</td>
<td>22%</td>
<td>$p = 0.332$</td>
</tr>
</tbody>
</table>

In particular, we find significantly more conditional cooperators ($\chi^2$-tests, all $p < 0.05$) and significantly fewer free riders ($\chi^2$-tests, all $p < 0.02$) in *Provision* than in *Maintenance*.
(the difference in others is never significant, \( \chi^2 \) – tests, all \( p > 0.117 \); see also Table B5).

Notably, however, we find that once we control for misperceptions, the differences in the distribution of types become smaller compared to the full sample. The percentage difference in conditional cooperators decreases from 17 percentage points in the full sample, to 10 to 15 percentage points depending on the misperception measure. The difference in the fraction of free riders, in contrast, remains stable, varying between 10 and 12 percentage points.

Given that our approach is related to Fosgaard et al. (2017), it is instructive to compare our results to theirs. Fosgaard, et al. use a measure of misperception that is similar to our Measure 2; in six questions they ask participants to indicate what a person “who only cares about their own earnings” (or “who only cares about other’s earnings”) will choose given a low, medium, or high average contribution of the other group members. As in our case, participants were rewarded for each correct answer. Different from our setup, however, Fosgaard, et al. asked these questions only at the end of the experiment rather than before the elicitation of cooperation preferences as we do. Furthermore, while their experiment was conducted online using a representative sample of the Danish population, our experiment was conducted in the laboratory using university students.

Despite these differences, Fosgaard et al. find a similar difference in the likelihood of misperception between Maintenance and Provision, although at a much higher overall level. In their sample, 41% and 51% of participants exhibit some form of misperception in Maintenance and Provision, respectively, compared to 29% and 37% in our case. Like us, using their full sample they also find a significantly higher fraction of conditional cooperators in Provision than in Maintenance. Different from us, their framing effect in cooperation preferences becomes insignificant once they control for misperceptions.

7.5. Misperceptions do not affect perceptions of kindness

In our final step, we present additional evidence that misperceptions cannot fully explain the differences across Maintenance and Provision. We show that the different perceptions of kindness that we presented in Section 6 are not a consequence of game-form misperceptions. This is important because if differences in kindness perceptions are indeed the main trigger of differences in conditional cooperation across the two social dilemmas, we should observe that perceptions of kindness still differ when controlling for misperceptions. If, instead, the different kindness perceptions across Maintenance and
Provision disappear when controlling for misperceptions, then the elicited kindness perceptions may not be considered a relevant explanation for the framing effect.

To test this, in some sessions of the misperception experiment reported in the previous two subsections, we included the kindness questionnaire at the end of the experiment before participants received feedback about the outcome of the game. Hence, while the kindness results reported in Section 6 (see Figure 2) were elicited using non-involved participants, we can now test whether the results hold when participants have experienced the decision situation. Furthermore, we can test whether the differences in kindness perceptions across Maintenance and Provision are robust to the exclusions of participants who exhibit some misperception. Our sample comprises n = 200 participants, n = 80 in Maintenance and n = 120 in Provision. The results are in Figure 3.

![Figure 3: Kindness perceptions in Provision and Maintenance of participants who experienced the decision situation (±1 s.e.m).](image)

The first (upper left) panel of Figure 3 depicts the comparison between the kindness schedules between Maintenance and Provision for the full sample. The second, third, and fourth panel show the same data for the subset of participants without misperceptions according to our three measures. Strikingly, the figure shows that the differences in kindness schedules across Maintenance and Provision are not only very similar across the four panels, but also very similar to the ones reported in Figure 2. This indicates that the differences in kindness perceptions across Maintenance and Provision are robust to having experienced
the decision situation beforehand and, more importantly, to the exclusion of participants who exhibit some form of game-form misperception.14

In sum, we find that while misperceptions of the game form explain some part of the observed framing effect of more conditional cooperators in Provision than in Maintenance, a sizeable difference that is both economically and statistically significant remains even after controlling for misperceptions. We have provided further empirical evidence that the framing of the decision situation affects perceptions of kindness of others’ actions and that misperceptions cannot explain the differences in kindness perceptions. In the next section we return to the social preferences models introduced in Section 4 and assess which of the models can reconcile our results.

8. Which theory of social preferences can explain our results?

Recall from Section 4 that only two models, guilt aversion (Battigalli and Dufwenberg (2007)) and revealed altruism (Cox et al. (2008)), are consistent with dilemma-specific differences in conditional cooperation. We discuss these two models in turn.

8.1. Guilt aversion

As discussed in Chang et al., (2011) and Bellemare et al., (2019), there are different approaches to test guilt aversion. Here, we rely on eliciting ex post feelings of guilt.15 In particular, using an online questionnaire, we test whether being a free rider generates stronger feelings of guilt in Provision compared to Maintenance. Similar to the elicitation of kindness perceptions, we conducted two studies using two different subject pools. In the first study, we recruited n = 347 students from the University of Nottingham, while in the second

---

14 Parametric estimates further corroborate these results. Using regression analyses in which we regress kindness evaluations on others’ average contributions, a treatment dummy, and an interaction term between the last two, we find that the Provision dummy is significantly negative and the interaction term between the Provision dummy and others’ average contributions is positive and significant. The size of these effects, which we report in Table B6, is similar to the ones reported in Section 6, where we analyse the kindness evaluations of uninvolved participants (compare also Table B4).

15 A potential alternative would have been to elicit second-order beliefs. However, given that we would need second-order beliefs for each cell of the strategy method, an incentive-compatible elicitation mechanism would require the elicitation of 21 contribution decisions, 21 first-order beliefs, and 21 second order beliefs. We believe that this procedure would have been too lengthy and too hard to understand for our participants. For this reason, we rely on the elicitation of ex post feelings of guilt. Research on guilt aversion in games also supports our approach. Chang, et al. (2011) show that subjects who match actions with second-order beliefs are in fact more likely to experience ex post feelings of guilt and conclude that the two measures point to the same psychological construct. Beranek, et al. (2015) find that survey-based measures of guilt are positively correlated with advantageous inequality aversion, which can explain conditional cooperation. Bellemare, et al. (2019) also find that a survey measure of guilt is correlated with game behaviour.
study we recruited \( n = 402 \) participants via MTurk (none had participated in any of our experimental sessions before). In the questionnaire, we first explained participants either a *Maintenance* or a *Provision* dilemma. We then asked them on a scale from 0 to 100 (where 0 corresponds to “not guilty at all” and 100 corresponds to “very guilty”) to assess how guilty they would feel if as a response to others’ contributions (withdrawals) of 0, 10, or 20, they would contribute 0 (withdraw 20) tokens (for the exact wording, see Online Appendix A4).

| Table 5: OLS regressions on feelings of guilt as a consequence of free riding |
|-----------------------------|-----------------------------|
|                            | (1)                      | (2)                      |
|                            | Lab                      | MTurk                    |
| Maintenance                | -4.961**                 | -2.064                   |
|                            | (2.483)                  | (2.698)                  |
| Others’ average contribution | 2.609***                | 2.403***                 |
|                            | (0.135)                  | (0.149)                  |
| Maintenance × Others’ average contribution | 0.593***               | -0.017                   |
|                            | (0.185)                  | (0.215)                  |
| Constant                   | 17.138***                | 20.390***                |
|                            | (1.889)                  | (1.999)                  |
| \( N \)                    | 1041                     | 1206                     |
| \( R^2 \)                  | 0.422                    | 0.268                    |

*Note:* Dependent variable: Guilt score. Others’ average contribution = 0 is omitted category. Robust standard errors clustered on the individual level are in parentheses. * \( p < 0.1 \), ** \( p < 0.05 \), *** \( p < 0.01 \). Model (1) includes survey responses from 347 students and model (2) includes survey responses from 402 MTurkers.

In Table 5 we report estimates from two regression models, one for each sample, in which we regress the guilt score on a treatment dummy for *Maintenance*, others’ average contributions, as well as an interaction term of the last two variables. As indicated by the significant coefficient for others’ contributions, the results show that people feel more guilty about free riding the more others contribute. Furthermore, even when others contribute nothing, people feel somewhat guilty about free riding as indicated by the constant, which is significantly larger than zero. Finally, and most importantly, we find no systematic differences in guilt between *Maintenance* and *Provision*.

\[ \text{16 Like with the elicitation of kindness perceptions, answers were not incentivized because we elicited personal judgments. We did, however, incentivize participation. As before, student participants were offered three randomly drawn prizes of £50 each, and MTurk participants received a flat payment of $2.} \]
For the MTurk sample, both the Maintenance dummy as well as the interaction term are not significantly different from zero, indicating no difference in feelings of guilt between the two treatments. For the student sample, in contrast, we find a significantly lower intercept and a significantly steeper slope in Maintenance compared to Provision. This suggests that free-riding on others’ effective contribution of zero triggers lower feelings of guilt in Maintenance than Provision, but that free-riding on others’ contributions of 10 or 20 tokens generates more guilt in Maintenance than Provision. As a consequence, if guilt aversion alone would be the driver behind the differences in conditional cooperation across the two dilemmas, we should observe more conditional cooperation in Maintenance than in Provision, which is the opposite of what we find. As such, our results suggest that guilt aversion is not the main explanation for the observed differences in cooperation preferences.

8.2. Revealed altruism

As discussed in Section 4, the model of revealed altruism by Cox et al. (2008) is consistent with a higher frequency of conditional cooperators in Provision than Maintenance. The reason for this is Axiom S, which assumes that the status quo (i.e., where resources are allocated initially) can affect an individual’s preference for reciprocity. Applied to our context, Axiom S implies that second movers in the strategy method will be less altruistic towards first movers in Maintenance than in Provision, because in the former any positive withdrawal decision of the first movers decreases the payoff opportunities of the second mover, while in the latter any positive contribution decision increases the second mover’s payoff opportunities. This asymmetry triggers different degrees of reciprocity. In the language of our parametrization of the Cox et al. (2008) model (see Section 4 and Online Appendix C), this is manifested in a stronger preference for reciprocity in Provision compared to Maintenance (i.e., \( \gamma^p \geq \gamma^M \)). This can explain our finding of more conditional cooperators and fewer free riders in Provision than in Maintenance.

Note that the results from our kindness surveys (Figures 2 and 3; Tables B4 and B6 in Online Appendix B) indicate that the mechanism through which the default position of the public good can affect preferences for cooperation is not only limited to differences in the degree of reciprocity (the size of the reciprocity parameter) but is also related to perceptual differences of others’ actions. Hence, if one interprets the Cox et al. (2008) model literally, our results uncover a different mechanism for the differences in reciprocity across Maintenance and Provision. However, given that perceptions of kindness are a trigger for
reciprocity, the Cox et al. (2008) model can be interpreted as a reduced form model. And if one is willing to assume that the difference in the reciprocity parameter stems from different perceptions of kindness, our findings would be in line with the predictions of the model.\textsuperscript{17}

9. Discussion and Conclusion

In this paper, we provided a behavioural anatomy of voluntary cooperation in two generic and incentive-equivalent social dilemmas of voluntary cooperation: providing and maintaining public goods. We focused on two fundamental dimensions: social preferences and (mis)perceptions of others’ intentions and the game form. We reported two important asymmetries. First, with regard to perceptions, we found that perceptions of the kindness of others’ actions differ between Maintenance and Provision because withdrawing everything from the public good is seen as less unkind than failing to contribute to the public good; and contributing everything is considered kinder in Provision than in Maintenance. With regard to perceptions of the game form, we found that misperceptions are game-specific: misunderstandings are more likely in Provision than Maintenance. Second, even after controlling for misperceptions, we observe substantial and significantly different degrees of conditional cooperation in Maintenance and Provision. Hence, conditional cooperation is not just mistaken cooperation (as argued, e.g., by Burton-Chellew et al. (2016)), but a true preference that also differs between Maintenance and Provision. In our setting, Maintenance and Provision are not just different frames, but also different games because, despite identical game forms, the two dilemmas trigger different social preferences.

Two models of social preferences, guilt aversion and the theory of revealed altruism, are potentially consistent with Maintenance and Provision being different games. Our test of guilt aversion, however, suggests that differences in guilt aversion cannot explain the treatment differences because elicited guilt does not differ across treatments. Hence, the only model of social preferences that is consistent with our results is the revealed altruism model by Cox et al. (2008), in which initial resource allocation before decisions are made affects the degree of reciprocity. Here, we provide evidence for the underlying mechanism of this the status quo affect. In particular, we show that the differences in conditional cooperation

\textsuperscript{17} Formally, this can be obtained in the model by assuming that Axiom S does not alter the degree of reciprocity (the MAT partial ordering in the language of the model) but has an effect on the perception of generosity of actions (i.e., the MGT partial ordering, see Online Appendix C for details).
are rooted in game-specific kindness perceptions, suggesting that framing effects occur at a fundamental (perceptual) level.

Our results also suggest an important general lesson: the revealed preference approach, that is, using choices to infer social preferences and/or framing effects, requires controlling for perceptions.\(^{18}\) This includes potential misperceptions of the game form to ensure measurement of preferences over clearly understood alternatives. Administering simple understanding questions at the beginning of experiments is nowadays quite common in experimental economics. However, it might not be enough. Our evidence on the existence of misperceived conditional cooperation is a point in case.

Our results also have implications for future literature. Hitherto, behavioural investigations of public goods provision and common pool resource problems have largely developed in independent literatures, in particular with regards to conditional cooperation, which was mostly studied in the context of linear public goods provision games (see, e.g., Chaudhuri (2011) and Fehr and Schurtenberger (2018)). Our comparative analysis of preferences and perceptions in maintenance and provision problems with identical social dilemma incentives is only a first step in bringing these two literatures closer together.

Finally, our results are not only of theoretical significance but have some potential policy implications. If many people are conditional cooperators, any factor that shifts beliefs about others’ cooperativeness will shift cooperation – a fact that can be used for policy interventions (e.g., Gächter (2007)). The observation that conditional cooperation, even after being corrected for misperceptions, is weaker in Maintenance than Provision suggests that policy proposals that reckon with conditional cooperation (e.g., MacKay et al., (2015)) need to take into account that the extent of it is dilemma-specific. Some of the most pressing challenges for mankind such as stopping global warming, sustaining natural resources and biodiversity concern mainly Maintenance dilemmas (e.g., Fehr-Duda and Fehr (2016)). Our results suggest that the power of conditional cooperation may be limited in maintenance problems, at least in comparison with provision dilemmas. Other solutions such as punishment (Gächter et al. (2017)) or incentives may instead be needed.

\(^{18}\) Alternatively, when it is not possible to measure misperceptions directly (e.g., in representative surveys), econometric techniques such as those proposed by Goldin and Reck (2020) can be applied to correct for potential measurement error \textit{ex post.}
References


34


APPENDIX – Overview of Experimental Data

Table A1: Overview of conducted experiments and surveys

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Experiment</th>
<th>Section where results are reported</th>
<th>Number of participants</th>
<th>Subject pool</th>
<th>Participants had participated before</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Gächter et al. (2017)¹⁹</td>
<td>Section 5.1</td>
<td>(n = 703)</td>
<td>Students (University of Nottingham)</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>Replication study</td>
<td>Section 5.1</td>
<td>n = 704</td>
<td>US citizens (MTurk)</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Temporal stability (5 months delay)</td>
<td>Section 5.2</td>
<td>n = 119</td>
<td>Students (University of Nottingham)</td>
<td>Yes. Subjects sampled from participants of experiment 0</td>
</tr>
<tr>
<td>3</td>
<td>Predictive power of cooperation attitudes</td>
<td>Section 5.3</td>
<td>n = 116</td>
<td>Students (University of Nottingham)</td>
<td>Yes. Subjects sampled from participants of experiment 0</td>
</tr>
<tr>
<td>4</td>
<td>Kindness survey</td>
<td>Section 6 Section 7</td>
<td>n = 185</td>
<td>Students (University of Nottingham)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n = 401</td>
<td>US citizens (MTurk)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Frame-dependent game form misperceptions</td>
<td>Section 7</td>
<td>n = 696</td>
<td>Students (University of Nottingham)</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Guilt survey</td>
<td>Section 8</td>
<td>n = 347</td>
<td>Students (University of Nottingham)</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n = 402</td>
<td>US citizens (MTurk)</td>
<td></td>
</tr>
</tbody>
</table>

¹⁹ Data taken from https://doi.org/10.5061/dryad.8d9t2.
ONLINE APPENDIX to
Preferences and Perceptions in Provision and Maintenance Public Goods

Simon Gächter\textsuperscript{a,b,c}, Felix Kölle\textsuperscript{d}, and Simone Quercia\textsuperscript{e}

This version: 15 January 2021

Online Appendix A – Experimental Instructions

A.1 – Laboratory Experiments 2
A.2 – Online Experiments 9
A.3 – Kindness Survey 12
A.4 – Guilt Survey 13

Online Appendix B – Supplementary Analyses 14

B.1 – Disaggregating the type category ‘others’ 14
B.2 – Procedural details and further supporting evidence for Section 5 15
B.3 – Additional Tables and Figures 20

Online Appendix C – Theoretical Considerations 24

Supplementary References 28

\textsuperscript{a} School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK, e-mail: simon.gaechter@nottingham.ac.uk; \textsuperscript{b} CESifo Munich; \textsuperscript{c} IZA Bonn. \textsuperscript{d} University of Cologne, Albertus Magnus Platz, 50923 Cologne, e-mail: felix.koelle@uni-koeln.de. \textsuperscript{e} Economics Department, University of Verona, Via Cantarane 24, 37129 Verona, e-mail: simone.quercia@univr.it.
ONLINE APPENDIX A – Experimental Instructions

A.1 – Laboratory experiments

In the following we present the original instructions subjects received in our PROVISION treatment. Differences in the MAINTENANCE instructions are reported in square brackets and highlighted in italics. We also report instructions for the one-shot game conducted in part 3 and used to evaluate predictive power in Section 5.2.

Part 1

Instructions

You are participating in a study in which you will earn some money. The amount will depend on the outcome of a game you will play. The amount of money which you earned with your decisions will be paid to you in cash at the end of the experiment. We will not speak of Pounds during the experiment, but rather of points. At the end, the total number of points you have earned will be converted to Pounds at the following rate:

1 point = £0.2

These instructions are solely for your private information. You are not allowed to communicate during the experiment. If you have any questions, please raise your hand. A member of the experimental team will come to you and answer them in private.

All participants will be divided into groups of four members. Only the experimenters will know who is in which group.

The decision situation

We first introduce you to the basic decision situation. Then, you will complete a pre-study questionnaire on the screen in front of you, which is intended to help you understand the decision situation.

In each group, every member has to decide the allocation of 20 tokens. You can put these 20 tokens into your private account or you can put some or all of them into a project. [In each group, there are 80 tokens in a project. You can withdraw up to 20 tokens from the project and put them into your private account or you can leave them fully or partially in the project.] The other three members of your group have to make the same decision.

Your income from the private account

You will earn 1 point for each token you put into your private account. For example, if you put all 20 tokens into your private account, your income from your private account would be 20 points. If you put 6 tokens into your private account, your income from this account would be 6 points. No one except you earns anything from tokens you put in your private account.
Your income from the project

Each group member will profit equally from the amount you or any other group member put into [leave in] the project. The income for each group member from the project will be determined as follows:

\[
\text{Income from the project} = 0.4 \times (\text{sum of contributions}) \ [0.4 \times (80 - \text{sum of all tokens withdrawn from the project})]
\]

If, for example, the sum of all contributions to the project [tokens withdrawn from the project] by you and your other group members is 60 [20] tokens, then you and each other member of your group would earn 60 [80-20] \times 0.4 = 24 points out of the project. If the four members of the group contribute [withdraw] a total of 10 [70] tokens to [from] the project, you and the other members of your group would each earn 10 [80-70] \times 0.4 = 4 points.

Total income

Your total income is the sum of your income from your private account and from the project:

\[
\text{Your Total Income} = \text{Income from your private account} + \text{Income from the project} \\
= 20 - \text{your contribution to the project} + 0.4 \times \text{sum of all contributions to the project} \\
[= \text{Tokens withdrawn from the project by you} + 0.4 \times (80\text{-sum of all tokens withdrawn from the project})]
\]

Please answer all the following questions, to help you understand the determination of your income.

1. Each group member has 20 tokens. Assume that none of the four group members (including you) contributes anything to the project. [There are 80 tokens in the project. Assume that everyone in your group withdraws 20 tokens from the project.]

What will your total income (in points) be?

What will the total income (in points) of each of the other group members be?

2. Each group member has 20 tokens. You contribute 20 tokens in the project. Each of the other three members of the group also contributes 20 tokens to the project. [There are 80 tokens in the project. You withdraw 0 tokens from the project. Each of the other three members of the group also withdraws 0 tokens from the project.]

What will your total income (in points) be?

What will the total income (in points) of each of the other group members be?
3. Each group member has 20 tokens. The other three members contribute a total of 30 tokens to the project. [There are 80 tokens in the project. The other three members withdraw 30 tokens from the project.]

a) What will your total income (in points) be, if - in addition to the 30 tokens contributed by others - you contribute 0 tokens to the project? [What will your total income (in points) be, if - in addition to the 30 tokens withdrawn by others - you withdraw 20 tokens from the project?]

b) What will your total income (in points) be, if - in addition to the 30 tokens contributed by others - you contribute 8 tokens to the project? [What will your total income (in points) be, if - in addition to the 30 tokens withdrawn by others - you withdraw 12 tokens from the project?]

c) What will your total income (in points) be, if - in addition to the 30 tokens contributed by others - you contribute 15 tokens to the project? [What will your total income (in points) be, if - in addition to the 30 tokens withdrawn by others - you withdraw 5 tokens from the project?]

4. Each group member has 20 tokens. Assume you invest 8 tokens to the project. [There are 80 tokens in the project. Assume you withdraw 12 tokens from the project.]

a) What will your total income (in points) be, if the other group members - in addition to your 8 tokens - contribute another 7 tokens to the project? [What will your total income (in points)? be, if the other group members - in addition to your 12 tokens - withdraw another 53 tokens from the project.]

b) What will your total income (in points) be, if the other group members - in addition to your 8 tokens - contribute another 12 tokens to the project? [What will your total income (in points) be, if the other group members - in addition to your 12 tokens - withdraw another 48 tokens from the project?]  

c) What will your total income (in points) be, if the other group members - in addition to your 8 tokens - contribute another 22 tokens to the project? [What will your total income (in points) be, if the other group members - in addition to your 12 tokens - withdraw another 38 tokens from the project?]

**Part 2**

**The Experiment**

The experiment is based on the decision situation just described to you, conducted **once**. You will enter your decisions in the screen in front of you.

As you know, you will have 20 tokens at your disposal. You can put them into a private account or into a project. [As you know, there are 80 tokens in a project. You can withdraw tokens from the project which will be automatically placed into your private account or you can leave them in the project.] Each subject has to make **two types** of decisions in this
experiment, which we will refer to below as the “unconditional contribution [withdrawal]” and the “contribution [withdrawal] table”.

- In the unconditional contribution [withdrawal] you simply decide how many of the 20 [80] tokens you want to put in [withdraw from] the project. Please indicate your contribution [withdrawal] in the following screen (screenshot taken from the PROVISION treatment):

After you have determined your unconditional contribution [withdrawal], please click “OK”.

- Your second task is to fill in a “contribution [withdrawal] table” where you indicate how many tokens you want to contribute [withdraw] to [from] the project for each possible average contribution [withdrawal] of the other group members (rounded to the next integer). Here, you can condition your contribution [withdrawal] on that of the other group members. This will be immediately clear to you if you Maintenance a look at the following table.

This table will be presented to you in the experiment (screenshot taken from the Provision treatment):
The numbers to the left of the blue cells are the possible (rounded) average contributions [withdrawals] of the other group members to the project. You have to insert how many tokens you want to contribute to [withdraw from] the project into each input box – conditional on the indicated average contribution [withdrawal] by the other members of your group. You must enter a number between 0 and 20 inclusive in each input box. For example, you have to indicate how much you contribute [withdraw from] the project if the others contribute [withdraw] 0 tokens on average to [from] the project; how much you contribute [withdraw] if the others contribute [withdraw] 1, 2, or 3 tokens on average; etc. Once you have made an entry in each input box, click “OK”.

After all participants of the experiment have made an unconditional contribution [withdrawal] and have filled in their contribution [withdrawal] table, a random mechanism will select one member from every group. For this group member, it is his contribution [withdrawal] table that will determine his actual contribution [withdrawal]; whereas, for the other three group members, it is their unconditional contributions [withdrawals] that will determine their actual contributions [withdrawals]. You will not know whom the random mechanism will select when you make your unconditional contribution [withdrawal] and fill in your contribution [withdrawal] table. You must therefore think carefully about both decisions because either could determine your actual contribution [withdrawal]. Two examples should make this clear.

**EXAMPLE 1**: Suppose that the random mechanism selects you; and that the other three group members made unconditional contributions [withdrawals] of 0, 2, and 4 [20, 18, and 16] tokens, respectively. The average contribution [withdrawal] of these three group members is, therefore, 2 [18] tokens. If you indicated in your contribution [withdrawal] table that you will contribute [withdraw] 1 [19] token[s] if the others contribute [withdraw] 2 [18] tokens on average, then the total contribution to the project is given by 0+2+4+1=7 [the total number of tokens left in the project is given by 80-(20+18+16+19)=7] tokens. Each group member would, therefore, earn 0.4×7=2.8 points from the project plus their respective income from their own private account. If, instead, you indicated in your contribution [withdrawal] table that you would contribute [withdraw] 19 tokens [1 token] if the others contribute [withdraw] 2 [18] tokens on average, then the total contribution of the group to the project would be given by 0+2+4+19=25 [the total number of tokens left in the project would be given by 80-(20+18+16+1)=25] tokens. Each group member would earn 0.4×25=10 points from the project plus their respective income from their own private account.

**EXAMPLE 2**: Suppose that the random mechanism does not select you; and that your unconditional [withdrawal] contribution is 16 [4] tokens, while those of the other two group members not selected by the random mechanism are 18 [2] and 20 [0] tokens, respectively. Your average unconditional contribution [withdrawal] and that of these two other group members is, therefore, 18 [2] tokens. If the group member whom the random mechanism did select indicates in her contribution [withdrawal] table that she will contribute [withdraw] 1 [19] token[s] if the other three group members contribute [withdraw] on average 18 [2] tokens, then the total contribution of the group to the project is given by 16+18+20+1=55 [the total number of tokens left in the project is given by 80-(4+2+0+19)=55] tokens. Each group member will therefore earn 0.4×55=22 points from the project plus their respective income from their own private account. If, instead, the randomly selected group member indicates in her contribution [withdrawal] table that she contributes [withdraws] 19 [7] if the others contribute [withdraw] on average 18 [2] tokens, then the total contribution of the
group to the project is 16+18+20+19=73 [the total number of tokens left in the project is 80-(4+2+0+1)=73] tokens. Each group member would therefore earn 0.4×73=29.2 points from the project plus their respective income from their own private account.

The random selection of the group member whose contribution [withdrawal] table will determine his actual contribution [withdrawal] will be made as follows. Each group member is assigned a Group Member ID between 1 and 4, which denote his/her number inside his group. Moreover, participant number 2 was randomly selected at the very beginning of the experiment. This participant will draw a ball from an urn after all participants have made their unconditional contribution [withdrawal] and have filled out their contribution [withdrawal] table. Each ball in the urn has a different colour and each colour corresponds to a Group Member ID: orange=1, blue=2, yellow=3, green=4. The resulting number will be entered into the computer. If participant 2 draws the Group Member ID that was assigned to you, then your contribution [withdrawal] table will determine your contribution [withdrawal] and their unconditional contributions [withdrawals] will determine the contribution [withdrawals] of the other group members. Otherwise, your unconditional [withdrawal] contribution determines your contribution [withdrawal].

**Part 3**

**Instructions**

You are now taking part in a second experiment. The money you earn in this experiment will be added to what you earned in the first one. As before, we will not speak of Pounds during the experiment, but rather of points. At the end, the number of points you have earned will be converted to Pounds at the following rate:

1 point=£0.2

As in the previous experiment you are in a group composed by 4 people. However, the composition of the group is entirely new. None of the participants who were in your group in the second experiment will be in your group in this experiment.

The decision situation is the same as the one described on the first instruction sheet of the previous experiment. Each member of the group has to decide about the usage of the 20 tokens. [In each group there are 80 tokens in a project.] You can put these 20 tokens into your private account or you can put them fully or partially into a project. [You can withdraw up to 20 tokens from the project or you can leave them fully or partially in the project.] Each token you do not put into the project [withdraw from the project] is automatically placed into your private account. Your income will be determined in the same way as before.

Reminder:

\[
\text{Your Total Income} = \text{Income from your private account} + \text{Income from the project} \\
= 20 - \text{your contribution to the project} + 0.4 \times \text{sum of all contributions to the project} \\
[=\text{Tokens withdrawn from the project by you} + 0.4 \times (80-\text{sum of all tokens withdrawn from the project})]
\]
The decision screen looks like this (screenshot taken from the PROVISION treatment):

1. First you have to decide on your contribution to [withdrawal from] the project. That is, you have to decide how many of the 20 tokens you want to contribute to the project, and how many tokens you want to put into your private account. [you have to decide how many of the 80 tokens you want to withdraw from the project and put into your private account.] Each other member of your group has to make the corresponding decision. This is the only contribution [withdrawal] decision that you or they make in this experiment. There is no contribution [withdrawal] table.

2. Afterwards you have to estimate the average contribution to [withdrawal from] the project (rounded to an integer) of the other three group members. You will be paid for the accuracy of your estimate:
   - If your estimate is exactly right (that is, if your estimate is exactly the same as the actual average contribution [withdrawal] of the other group members), you will get 3 points in addition to your other income from the experiment.
   - If your estimate deviates by one point from the correct result, you will get 2 additional points.
   - A deviation by 2 points still earns you 1 additional point.
   - If your estimate deviates by 3 or more points from the correct result, you will not get any additional points.
A.2 – Online experiments (MTurk)

In this decision problem you will form a group with three other people from MTurk. To determine your bonus payment, we will first record your earnings in points and then exchange the sum of points you earned into a dollar amount for your bonus payment.

Your bonus in Dollars will be determined as follows: \( \text{Earnings in Dollars} = \frac{\text{Earnings in Points}}{20} \).

In each group, every group member has an endowment of 20 tokens. You can put these 20 tokens into your private account or you can contribute them fully or partially to a project. [In each group, there are 80 tokens in a project. You can withdraw up to 20 tokens from the project and put them into your private account or you can leave them fully or partially in the project.] The other three members of your group have to make the same decision.

You will earn an income from your private account and from the project.

**Your income from the private account**

You will earn 1 point for each token you put into your private account. For example, if you put 20 tokens into your private account, your income from your private account is 20 points. If you put 6 tokens into your private account, your income from this account is 6 points. **No one except you earns anything from tokens you put into your private account.**

**Your income from the project**

Each group member will profit equally from the amount you or any other group member contributes to [leaves into] the project. All tokens contributed to [left in] the project will be **increased by 60 percent (a factor of 1.6) and split equally** among the four group members. That is, for every token contributed [left] by any group member, you and all three other group members will receive: \( 1 \times 1.6 / 4 = 0.4 \) points each.

If, for example, the sum of all tokens contributed to [left into] the project by you and your other three group members is 60 tokens, then you and each other member of your group would earn \( 60 \times 1.6 / 4 = 60 \times 0.4 = 24 \) points.

If the four members of the group contribute [leave] a total of 10 tokens in the project, you and the other three members of your group would each earn \( 10 \times 1.6 / 4 = 10 \times 0.4 = 4 \) points.

**Total income**

Your total income is the sum of your income from your private account and from the project.

The graphic below shows a summary of the interaction (figure from the Provision treatment only):
Please answer the following questions to check your understanding of the situation.

Assume that all four group members (including you) contribute 0 [withdraw 20] tokens each to [from] the project. What will your total point earnings be (= point earnings from private account + point earnings from project)?

Assume you contribute 20 [withdraw 0] tokens to [from] the project. Each of the other three members of the group also contributes 20 [withdraws 0] tokens to [from] the project. What will your total earnings be (= point earnings from private account + point earnings from project)?

Assume you contribute 0 [withdraw 20] tokens to [from] the project and the other group members contribute [leave] in total 60 tokens to [in] the project. What will your total earnings be (= point earnings from private account + point earnings from project)?

Assume you contribute 20 [withdraw 0] tokens to [from] the project and the other group members contribute in total 0 tokens to the project [leave 0 tokens in the project]. What will your total earnings be (= point earnings from private account + point earnings from project)?

You are now ready to make your decisions. Your task is based on the decision problem described above.

As you know, you will have 20 tokens at your disposal. You can put them into a private account or into a project. [As you know, there will be 80 tokens in a project. You can withdraw up to 20 tokens and put them into a private account or leave them in the project.]

All group members have two tasks, which we will refer to below as the “unconditional contribution [withdrawal]” and the “contribution [withdrawal] table”.

In the unconditional contribution [withdrawal] task you simply decide how many tokens (up to 20) you want to contribute to [withdraw from] the project.
Your second task is to fill in a “contribution \[ withdrawal \] table” where you indicate how many tokens you want to contribute to \[ withdraw from \] the project for each possible average contribution \[ withdrawal \] of the other group members (rounded to the next integer). Here, you can condition your contribution \[ withdrawal \] on that of the other group members. You can see the table that you will have to fill in if you scroll down.

This is a one-off decision problem that is finished once you have made both decisions.

**How your bonus will be determined**

When all participants in your group have made their decisions, we will randomly select three group members for whom the unconditional contribution \[ withdrawal \] will be relevant for their earnings. For the non-selected group member, the contribution \[ withdrawal \] table will be relevant for his/her earnings. This means that you should *Maintenance* both the unconditional contribution \[ withdrawal \] and the contribution \[ withdrawal \] table equally seriously because you don't know yet which one will be relevant for calculating your bonus.

**Example:**

- Imagine that the unconditional contributions \[ withdrawals \] of group members 1, 2, 3, and 4 are 20, 15, 10 and 0, respectively.
- Assume that for group members 1, 3 and 4 the unconditional contributions \[ withdrawals \] are relevant for their earnings and for group member 2 the contribution \[ withdrawal \] table will be used to calculate earnings.
- Then we calculate the average of the three unconditional contributions \[ withdrawal \] -- in our example: \((20 + 10 + 0)/3 = 10\).
- To determine the contribution \[ withdrawal \] of group member 2 we will *Maintenance* the contribution \[ withdrawal \] this group member indicates in his/her contribution \[ withdrawal \] table if others contribute \[ withdraw \] on average 10.
- Imagine that this group member contributes \[ withdraws \] 12 if others contribute \[ withdraw \] 10 on average. Then the total sum of contributions to \[ withdrawals from \] the project is \(20 + 12 + 10 + 0 = 42\) and earnings are calculated as explained above.

We now ask you to make the unconditional contribution \[ withdrawal \] decision, followed by filling in the contribution \[ withdrawal \] table.
A.3 – Kindness Survey

After having read the general decision situation of either Maintenance or Provision (see above), participants were asked the following (results are in Sections 6 and 7):

In the following we ask you in various scenarios to evaluate the kindness of the other three group members on a scale from -100 to +100 where -100 corresponds to extremely unkind and +100 corresponds to extremely kind.

1. Assume that the other three group members contribute [withdraw] on average £0 [£20] to [from] the project.

How kind do you think they are?

2. Assume that the other three group members contribute [withdraw] on average £10 [£10] to [from] the project.

How kind do you think they are?

3. Assume that the other three group members contribute [withdraw] on average £20 [£0] to [from] the project.

How kind do you think they are?
A.4 – Guilt Survey

*After having read the general decision situation of either Maintenance or Provision (see above), participants were asked the following (results are in Section 8):*

In the following we ask you in various scenarios to evaluate how guilty you would feel by contributing £0 [withdrawing £20] to [from] the project given various average contributions [withdrawals] of the other three group members. Please indicate your answer on a scale from 0 to 100, where 0 means "not guilty at all" and 100 means "extremely guilty". Please click on the slider to submit your answer.

1. Assume that the other three group members move first, and you observe that they have contributed [withdrawn] on average £0 [£20] to [from] the project.

How guilty would you feel if you contribute £0 [withdraw £20] in response?

2. Assume that the other three group members move first, and you observe that they have contributed [withdrawn] on average £10 [£10] to [from] the project.

How guilty would you feel if you contribute £0 [withdraw £20] in response?

3. Assume that the other three group members move first, and you observe that they have contributed [withdrawn] on average £20 [£0] to [from] the project.

How guilty would you feel if you contribute £0 [withdraw £20] in response?
ONLINE APPENDIX B – Supplementary Analyses

B.1 – Disaggregating the type category ‘others’

In our main classification, described in Section 3 of the main text, we have only used three categories as our main focus is on conditional cooperators and free riders. The category “others”, however, can be broken down into different patterns of behavior. In particular, in addition to free-riders and conditional cooperators, we classify a subject as (i) unconditional cooperator if she contributes a constant positive amount irrespective of the others’ contributions, (ii) triangle cooperator if her contribution schedule is monotonically increasing up to a maximum of \( \kappa \) and thereafter monotonically decreasing, (iii) anti-conditional cooperator if either her contribution (withdrawal) schedule exhibits a (weakly) monotonically decreasing pattern, or if the Spearman correlation coefficient between her schedule and the others’ average contribution (withdrawal) is negative and significant at \( p < 0.01 \), (iv) other if none of the criteria above apply.

Table B1 reports the classification of attitudes types according to these criteria. Panel A reports the data from Gächter et al., (2017), and Panel B from our MTurk replication. In both samples the distribution of types is significantly different across treatments, in line with the results reported in the main text of the paper. Moreover, in both samples we find significantly more anti-conditional cooperators in Maintenance compared to Provision, and significantly fewer unconditional cooperators in Maintenance than in Provision in our MTurk sample.

### Table B1. Classification of types – disaggregating the category ‘others’

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>( \chi^2 ) test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>63%</td>
<td>43%</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Free riders</td>
<td>17%</td>
<td>28%</td>
<td>( p = 0.001 )</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>3%</td>
<td>3%</td>
<td>( p = 0.473 )</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>9%</td>
<td>11%</td>
<td>( p = 0.387 )</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>3%</td>
<td>7%</td>
<td>( p = 0.006 )</td>
</tr>
<tr>
<td>Others</td>
<td>5%</td>
<td>9%</td>
<td>( p = 0.065 )</td>
</tr>
<tr>
<td>Test overall distribution</td>
<td>( \chi^2(5) = 34.15, p &lt; 0.001 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Panel B - MTurk experiment (n = 704)

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>( \chi^2 ) test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>80%</td>
<td>67%</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Free riders</td>
<td>9%</td>
<td>17%</td>
<td>( p = 0.001 )</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>7%</td>
<td>3%</td>
<td>( p = 0.020 )</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>2%</td>
<td>4%</td>
<td>( p = 0.244 )</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>1%</td>
<td>7%</td>
<td>( p &lt; 0.001 )</td>
</tr>
<tr>
<td>Others</td>
<td>1%</td>
<td>1%</td>
<td>( p = 0.440 )</td>
</tr>
</tbody>
</table>

Test overall distribution \( \chi^2(5) = 38.61, p < 0.001 \)

B.2 – Procedural details and further supporting evidence for Section 5

Temporal Stability

To test the temporal stability of our revealed preference measure, we re-invited a subset of \( n = 288 \) participants who had participated in our first experiment reported in Gächter et al. (2017) and in the left panel of Figure 1.¹ This allows us to observe a participant’s cooperation preference at two different points in time and to assess its temporal stability. We report results from \( n = 119 \) participants; \( n = 65 \) in Provision and \( n = 54 \) in Maintenance who showed-up in both waves.² Table B2 reports the frequencies of types in the two waves for the two treatments.

¹ Our experiment was conducted in several waves as the third part differed across waves (see Gächter, et al. (2017)). The participants we re-invited were the ones who participated in the first wave. The data we report here as Wave 2 were not used in Gächter, et al. (2017).

² Due to attrition, we also invited additional participants in order to be able to form groups of four. For the ease of exposition, the data from these \( n = 48 \) additional participants are not reported here but are available upon request. The results from these participants, however, are very similar to the ones reported above. More importantly, we can rule out any selection effects with respect to our main variable of interest as the attrition we found between the two waves was not related to cooperation preferences. The distribution of types in Wave 1 does not significantly differ between participants who did or did not show up for the second experiment (Maintenance: \( \chi^2(2) = 0.35, p = 0.839 \); Provision \( \chi^2(2) = 2.77, p = 0.251 \)). Participants also did not differ with respect to important socio-demographic characteristics such as age, gender, nationality, or field of studies.
### Table B2: Distribution of types in Maintenance (M) and Provision (P) across waves.

<table>
<thead>
<tr>
<th></th>
<th>Wave 1</th>
<th></th>
<th>Wave 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>P</td>
<td>$\chi^2$ - test</td>
<td>M</td>
</tr>
<tr>
<td>Conditional Cooperators</td>
<td>43%</td>
<td>66%</td>
<td>$p = 0.010$</td>
<td>39%</td>
</tr>
<tr>
<td>Free Riders</td>
<td>18%</td>
<td>20%</td>
<td>$p = 0.839$</td>
<td>24%</td>
</tr>
<tr>
<td>Others</td>
<td>39%</td>
<td>14%</td>
<td>$p = 0.002$</td>
<td>37%</td>
</tr>
<tr>
<td>$\chi^2$-test M vs. P</td>
<td></td>
<td></td>
<td>$p = 0.006$</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the analyses reported in the main text, here we provide further information on the stability of types at the individual level. Our results are summarized in Table B3, displaying the relative frequency of all possible type combination across the two waves. As can be seen, the largest number of observations lie on the main diagonal, corresponding to subjects who display the same type in both waves. The numbers on the off-diagonal, in contrast, reveal changes in types across the two waves. As can be seen, we observe changes of types in all directions. For example, while some subjects classified as conditional cooperators in wave 1 become free riders in wave 2, some others display an opposite change.

To test whether some types are more likely to remain stable across the two waves, we can compare the fraction of ‘stable’ subjects across the different types. In Maintenance, we find that 57% of all subjects classified as conditional cooperators in wave 1, are classified as the same type in wave 2. Very similar numbers are observed for free riders and others, for which we observe consistency rates of 50% and 67%, respectively ($\chi^2(2) = 0.90; p = 0.636$). In Provision, consistency rates for conditional cooperators and free riders amount to 74% and 69%, respectively. These numbers are neither significantly different from each other ($\chi^2(1) = 0.14; p = 0.711$), nor are they different from the stability rates of the same type in Maintenance (conditional cooperators: $\chi^2(1) = 2.21; p = 0.137$; free riders: $\chi^2(1) = 0.88; p = 0.349$). The only difference we observe with regard to others, who in Provision display a consistency rate of only 22%, which is lower than the one observed in Maintenance ($\chi^2(1) = 5.00; p = 0.025$), and lower than the one observed for the other types in Provision ($\chi^2(1) = 9.00; p = 0.003$).
Table B3: Relative frequency of all 9 possible combination of types in the two waves.

<table>
<thead>
<tr>
<th></th>
<th>Maintenance</th>
<th></th>
<th>Provision</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>FR</td>
<td>OT</td>
<td>CC</td>
</tr>
<tr>
<td>Wave 2</td>
<td>24.1%</td>
<td>9.3%</td>
<td>9.3%</td>
<td>49.2%</td>
</tr>
<tr>
<td>Wave 1</td>
<td>FR</td>
<td>7.4%</td>
<td>9.3%</td>
<td>6.2%</td>
</tr>
<tr>
<td></td>
<td>OT</td>
<td>7.4%</td>
<td>5.6%</td>
<td>26.0%</td>
</tr>
</tbody>
</table>

To assess if individual stability occurs more frequently than would be expected if types would change randomly, we follow the approach of Volk et al., (2012) and simulate a distribution of types assuming that each player randomly picks a type in Wave 2 with a probability equal to the observed frequency in our data. In 100 runs of this simulation, we find that in Provision and Maintenance participants are of the same type in both waves on average in 46% and 35% of the cases, respectively. Testing the simulated distribution of stable types against the observed proportion in the experiment reveals that the hypothesis of random types can be rejected (two-sided t-test, \( p < 0.001 \)).

To further evaluate overall individual-level stability, we compare our results to the ones reported in Volk et al. (2012) who use a provision game in a similar design to ours. In line with our results, Volk et al. find that 64% of participants are classified as the same type between two waves that are 2.5 months apart. No such comparison is possible for Maintenance as (to the best of our knowledge) no previous study has investigated the stability of cooperation preferences using a maintenance game.

As a final step, we check whether the differences in the distribution of cooperation types is also robust when only considering the subset of participants who are classified as the same type in both waves. A significant difference in this subsample would indicate that the stability of the treatment effect across waves is systematic as it is due to differences in cooperation preferences that are stable at the individual level. Strikingly, the effect among this subsample is even stronger in terms of percentage point differences across types compared to the whole sample. We find 74% (41%) conditional cooperators, 21% (15%) free riders and 5% (44%) others in Provision (Maintenance) (\( \chi^2(2) = 16.92, p < 0.001 \)).

**Predictive Power**

To test the predictive power of our revealed preference measure we follow Fischbacher et al., (2012), combining the data from our strategy-method elicitation with the data from a one-shot direct-response game that followed immediately after. After participating in the strategy method experiment, subjects were re-matched in a perfect stranger protocol to play a one-shot simultaneous game where we also elicited beliefs about the average effective contribution of the other three members of their group. Using cooperation preferences and

---

3 Participants were paid for the accuracy of their beliefs. If their beliefs matched the other’s average contribution exactly, participants earned 3 points. When their belief deviated by 1 (2) point(s) from the correct
stated beliefs allows to make a point prediction about the contribution decision in the direct-response game, \( \hat{c}_t \). For each individual, we then compare the predicted contribution with their actual contribution in the direct-response game, \( c_t \), delivering an individual-level measure of consistency. For this exercise, we can use all the experiments where the strategy method experiment was followed by a one-shot direct-response game with belief elicitation. For this we have (1) \( n = 288 \) observations from participants from our first experiment reported in Gächter et al. (2017) and in the left panel of Figure 1, and (2) \( n = 703 \) observations from participants in our MTurk experiment. We further report data from (3) a set of \( n = 116 \) participants for which the elicitation of cooperation preferences and the direct-response game took place in two separate sessions that lay five months apart. \(^4\)

![Figure B1: Deviations from predicted choices in Maintenance and Provision. Left panel: Students (\( n = 288 \)). Middle panel: MTurk (\( n = 703 \)). Right panel: Students who participated in the direct-response experiment five months after the preference elicitation experiment (\( n = 116 \))](image)

estimate they earned 2 (1) points. If their estimation was off by more than two points, they received no additional money.

\(^4\) In this new set of experiments, we had a total of \( n = 696 \) participants. Out of these, five months after the first experiment we randomly re-invited \( n = 312 \) participants to participate in a one-shot direct-response experiment using the same frame, Maintenance or Provision. None of these participants had played a direct-response game in the first experiment. We report the elicited cooperation preferences of these \( n = 696 \) participants in Section 6. As before, we can rule out that attrition was related to cooperation preferences as the distribution of types in the sample of participants who did not show up five months later is not statistically different from the distribution for the ones who did show up (\( \chi^2(2) = 0.09, p = 0.953 \) and \( \chi^2(2) = 0.23, p = 0.894 \) in Maintenance and Provision, respectively). Participants also did not differ with respect to important socio-demographic characteristics such as age, gender, nationality, or field of studies.
Figure B1 depicts the distribution of individual deviations from predicted choices, $c_i - \hat{c}_i$, separately for Maintenance and Provision and for each of the three samples. Further to the analyses reported in the main text, we follow Fischbacher et al. (2012) and define a subject as consistent if their actual cooperation decision does not deviate by more than $\pm 2$ tokens (10 percent of their endowment) from their predicted contribution. We find that 63 and 62 percent of lab participants are consistent in Maintenance and Provision, respectively. The one percent difference is not statistically significant ($\chi^2(1) = 0.15; p = 0.903$). We observe similar numbers in the MTurk sample with no differences across treatments (Maintenance: 63%, Provision 65%, $\chi^2(1) = 0.18; p = 0.672$). In our experiment with a delay of five months, consistency is still remarkably high: 50% and 55% of participants are consistent in Maintenance and Provision, respectively. Compared to the case without delay, consistency is lower, but this difference is only marginally significant ($\chi^2(1) = 3.38; p = 0.066$). As before, we find no difference in the distribution of deviations across treatments ($\chi^2(1) = 0.33; p = 0.564$). We conclude that our proxy for cooperation preferences is an equally good predictor for actual game play in Maintenance and Provision.
### Table B4: OLS regressions on the evaluation of kindness of others’ contributions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lab</td>
<td>MTurk</td>
</tr>
<tr>
<td>Other's average contribution</td>
<td>3.752*** (0.400)</td>
<td>4.704*** (0.268)</td>
</tr>
<tr>
<td>Other's average contribution x Provision</td>
<td>2.410*** (0.483)</td>
<td>1.703*** (0.342)</td>
</tr>
<tr>
<td>Provision (1 if Provision, 0 otherwise)</td>
<td>-26.317*** (6.310)</td>
<td>-19.694*** (4.390)</td>
</tr>
<tr>
<td>Constant</td>
<td>-21.690*** (5.229)</td>
<td>-25.875*** (3.597)</td>
</tr>
<tr>
<td>N</td>
<td>555</td>
<td>1203</td>
</tr>
<tr>
<td>R²</td>
<td>0.541</td>
<td>0.562</td>
</tr>
</tbody>
</table>

*Note: Dependent variable: Kindness evaluations. Robust standard errors clustered on the individual level are in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01. Model (1) includes survey responses from n = 185 students and model (2) includes survey responses from n = 401 MTurkers.*
### Table B5. Classification of attitude types disaggregating the category ‘others’ and controlling for misperceptions

**Panel A – All (n = 696)**

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>$\chi^2$ test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>51%</td>
<td>33%</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Free riders</td>
<td>27%</td>
<td>39%</td>
<td>$p = 0.001$</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>2%</td>
<td>3%</td>
<td>$p = 0.409$</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>13%</td>
<td>14%</td>
<td>$p = 0.770$</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>3%</td>
<td>5%</td>
<td>$p = 0.158$</td>
</tr>
<tr>
<td>Others</td>
<td>4%</td>
<td>5%</td>
<td>$p = 0.409$</td>
</tr>
</tbody>
</table>

Test overall distribution $\chi^2(5) = 22.69, p < 0.001$

**Panel B – No mistake in payoff questions (n = 572)**

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>$\chi^2$ test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>49%</td>
<td>34%</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Free riders</td>
<td>29%</td>
<td>41%</td>
<td>$p = 0.005$</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>1%</td>
<td>2%</td>
<td>$p = 0.325$</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>14%</td>
<td>14%</td>
<td>$p = 0.933$</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>3%</td>
<td>4%</td>
<td>$p = 0.744$</td>
</tr>
<tr>
<td>Others</td>
<td>3%</td>
<td>5%</td>
<td>$p = 0.332$</td>
</tr>
</tbody>
</table>

Test overall distribution $\chi^2(2) = 14.53, p = 0.013$
Table B5 continued

**Panel C – No mistake in goal questions (n = 463)**

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>$\chi^2$ test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>45%</td>
<td>35%</td>
<td>$p = 0.042$</td>
</tr>
<tr>
<td>Free riders</td>
<td>32%</td>
<td>43%</td>
<td>$p = 0.019$</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>2%</td>
<td>2%</td>
<td>$p = 0.564$</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>14%</td>
<td>14%</td>
<td>$p = 0.919$</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>3%</td>
<td>2%</td>
<td>$p = 0.605$</td>
</tr>
<tr>
<td>Others</td>
<td>4%</td>
<td>4%</td>
<td>$p = 0.748$</td>
</tr>
<tr>
<td><strong>Test overall distribution</strong></td>
<td></td>
<td></td>
<td>$\chi^2(2) = 6.60, p = 0.252$</td>
</tr>
</tbody>
</table>

**Panel D – No mistaken conditional cooperation (n = 632)**

<table>
<thead>
<tr>
<th></th>
<th>Provision</th>
<th>Maintenance</th>
<th>$\chi^2$ test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional cooperators</td>
<td>46%</td>
<td>34%</td>
<td>$p = 0.002$</td>
</tr>
<tr>
<td>Free riders</td>
<td>31%</td>
<td>41%</td>
<td>$p = 0.007$</td>
</tr>
<tr>
<td>Unconditional cooperators</td>
<td>2%</td>
<td>3%</td>
<td>$p = 0.344$</td>
</tr>
<tr>
<td>Triangle cooperators</td>
<td>14%</td>
<td>14%</td>
<td>$p = 0.878$</td>
</tr>
<tr>
<td>Anti-conditional cooperators</td>
<td>3%</td>
<td>4%</td>
<td>$p = 0.679$</td>
</tr>
<tr>
<td>Others</td>
<td>4%</td>
<td>4%</td>
<td>$p = 0.830$</td>
</tr>
<tr>
<td><strong>Test overall distribution</strong></td>
<td></td>
<td></td>
<td>$\chi^2(2) = 11.25, p = 0.047$</td>
</tr>
</tbody>
</table>
Table B6: OLS regressions on the evaluation of kindness of others’ contributions controlling for misperceptions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All subjects</td>
<td>No mistake in payoff questions</td>
<td>No mistake in goal questions</td>
<td>No mistaken conditional cooperation</td>
</tr>
<tr>
<td>Other’s average contribution</td>
<td>4.684*** (0.444)</td>
<td>4.589*** (0.471)</td>
<td>4.791*** (0.592)</td>
<td>4.839*** (0.480)</td>
</tr>
<tr>
<td>Other’s average contribution × Provision</td>
<td>1.813*** (0.523)</td>
<td>1.824*** (0.563)</td>
<td>1.897*** (0.681)</td>
<td>1.631*** (0.565)</td>
</tr>
<tr>
<td>Provision</td>
<td>-23.623*** (7.342)</td>
<td>-23.484*** (8.150)</td>
<td>-22.325** (9.708)</td>
<td>-20.019** (7.981)</td>
</tr>
<tr>
<td>Constant</td>
<td>-16.244*** (5.928)</td>
<td>-12.779** (6.335)</td>
<td>-18.370** (7.760)</td>
<td>-18.509*** (6.418)</td>
</tr>
</tbody>
</table>

Observations 600 471 381 525

$R^2$ 0.583 0.568 0.564 0.574

Note: Dependent variable: Kindness evaluations. Robust standard errors clustered on the individual level are in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Model (1) includes responses from $n=200$ students for whom we elicited cooperation attitudes before the kindness questionnaire. Models (2) - (4) use only those subjects who are not classified as misperceiving according to our first, second and third measure, respectively.
ONLINE APPENDIX C – Theoretical Considerations

This appendix complements Section 4 in the main text. We first review models of social preferences that do not predict a difference between Maintenance and Provision. After that, we derive propositions applying the theory of revealed altruism (Cox et al., (2008)) to the two social dilemmas.

Models of altruism and distributional social preferences

Theories of inequity aversion (Fehr and Schmidt (1999); Bolton and Ockenfels (2000)) can explain conditional cooperation, e.g., if inequity aversion is strong enough, but do not predict any differences across the two social dilemmas in best replies and equilibrium behavior. The reason is that these theories are only based on payoff consequences, which are identical across the two incentive-equivalent social dilemmas of Maintenance and Provision.

Charness and Rabin (2002) model social preferences with a taste for efficiency and reciprocity. As the models above, their model can also explain conditional cooperation, but does not predict any difference between Maintenance and Provision because all the elements of the utility function are based on material consequences only.

A similar argument holds for models of altruism and fairness such as those by Levine (1998) and Cox et al., (2007), which assume that individuals attach some positive (or negative) weight on others’ payoffs. In the Levine model, this weight depends on an individual’s degree of altruism (or spite), as well as their current expectation about the other player’s type. In the Cox et al. (2007) model, the weight depends on the player’s emotional state, which is a function of reciprocity motivation (in a sequential game this is derived from the second mover’s maximum possible payoff given the first mover’s choice, relative to some neutral benchmark) and relative status (e.g., asymmetries in players’ claims or obligations). Importantly, while ad-hoc assumptions on potential differences in the weights attached to other players’ payoffs could potentially explain the observed differences between Maintenance and Provision, there is no explicit element in these models that predicts such differences.

Theories of reciprocity

In theories of reciprocity (Rabin (1993); Dufwenberg and Kirchsteiger (2004); Falk and Fischbacher (2006)), agents’ motivations derive from their material payoff as well as a psychological payoff that depends on their first- or second-order beliefs about others’ actions. These theories are natural candidates to explain conditional cooperation because they postulate that agents want to reward kind actions (or intentions) with kindness and to punish hostile actions (or intentions) with unkindness. Kindness is evaluated with regard to a reference point that is a statistic derived from the set of material consequences (typically an equitable payoff calculated as an average between the maximum and the minimum payoff a player can get). Any reference point that is only based on material payoffs has to be the same across Maintenance and Provision because of equivalent game forms. Therefore, the
only channel through which simultaneous game play may differ across Maintenance and Provision is via differences in beliefs. As argued in Dufwenberg et al., (2011), in a linear public goods game the evaluation of others’ kindness only depends on first-order but not on second-order beliefs. Therefore, since in our strategy-method experiment first-order beliefs are fixed because participants condition their contributions on all possible average contributions of others, these models do not predict differences in cooperation preferences across Maintenance and Provision.

**Revealed Altruism**

Consider a generic player $i$ and define an opportunity set at a given node of the game as a subset of $\mathbb{R}^4_i$ that contains all feasible payoffs for player $i$. We start by stating a similar definition of Definition 2 from Cox et al. (2008), which allows an ordering of opportunity sets contained in $\mathcal{C}$ (the set containing all the opportunity sets for player $i$):

**Definition 2** (compare Cox et al. (2008), p. 36): Opportunity set $G$ is more generous than ($MGT$) opportunity set $F$ if (a) $\pi_{iG}^* - \pi_{iF}^* \geq 0$ and (b) $\pi_{iG}^* - \pi_{iF}^* \geq \bar{\pi}_{-iG}^* - \bar{\pi}_{-iF}^*$. In this case, we say $G \succ MGT F$.

where $\pi_{iF}^*$ stands for the maximum feasible payoff of player $i$ in opportunity set $F$ and $\bar{\pi}_{-iF}^*$ stands for the maximum feasible average payoff of the other three group members in opportunity set $F$. Notice that we assume that player $i$ will compare her earnings with the average earnings of the other three members, which is a slightly modified version of Definition 2 in Cox et al. (2008) that accounts for the fact that subjects in our experiment were confronted with possible averages contributions of the other group members and not with the entire vector of contributions.

Applying Definition 2 to Maintenance and Provision leads to our first proposition:

**PROPOSITION 1.** (a) In Provision, an opportunity set generated by an average contribution $\bar{c}_{-i}^A$ of the other three group members is more generous than ($MGT$) an opportunity set generated by another average contribution $\bar{c}_{-i}^B$ if and only if $\bar{c}_{-i}^A > \bar{c}_{-i}^B$. (b) In Maintenance, an opportunity set generated by an average withdrawal $\bar{w}_{-i}^A$ is more generous than an opportunity set generated by another average withdrawal $\bar{w}_{-i}^B$ if and only if $\bar{w}_{-i}^A < \bar{w}_{-i}^B$.

**Proof.** Applying Definition 2 to the Provision problem, we show that an opportunity set generated by an average contribution $\bar{c}_{-i}^A$ of the other three group members is more generous than ($MGT$) an opportunity set generated by another average contribution $\bar{c}_{-i}^B$ if $\bar{c}_{-i}^A > \bar{c}_{-i}^B$.

Consider the payoff function of the Provision game:

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

In the strategy method experiment, as the average contribution of the other group members goes from $\bar{c}_{-i}^B$ to $\bar{c}_{-i}^A$, a subject gains $0.4 \times 3 \times (\bar{c}_{-i}^A - \bar{c}_{-i}^B)$ on her maximum

25
feasible payoff which satisfies condition (a) of Definition 2. Regarding condition (b), her gain on the maximum feasible payoff \( \pi^*_i - \pi^*_{iF} \) is always greater than \( \pi^*_i - \pi^*_{iF} \), as the first term is positive and the second term is negative if \( \bar{c}^A_{-i} > \bar{c}^B_{-i} \).

Now consider the payoff function of the Maintenance problem:

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

Using \( \bar{c}^A_{-i} = 20 - \bar{w}^A_{-i} \) and \( \bar{c}^B_{-i} = 20 - \bar{w}^B_{-i} \), it follows that if \( \bar{c}^A_{-i} > \bar{c}^B_{-i} \) then \( \bar{w}^A_{-i} < \bar{w}^B_{-i} \). Given that the payoff functions of Provision and Maintenance are isomorphic, similar to above we can show that an opportunity set generated by an average withdrawal of \( \bar{w}^A_{-i} \) is more generous than an opportunity set generated by another average withdrawal \( \bar{w}^B_{-i} \) if \( \bar{w}^A_{-i} < \bar{w}^B_{-i} \). Q.E.D.

Next, we apply Axiom R of Cox et al. (2008), p. 40, which specifies how differences in generosity across opportunity sets translate into differences in preferences.

Formally, Axiom R states that if \( G, F \in C \) and \( G \text{ MGT } F \), then \( A_G \text{ MAT } A_F \), where \( A_G \) indicates the preferences induced by opportunity set \( G \) and \( \text{MAT} \) stands for “more altruistic than”.

Hence, for any two opportunity sets, \( G \) and \( F \) if \( G \text{ MGT } F \), then the preferences induced by \( G \) are more altruistic than (MAT) the preferences induced by \( F \). Applying Axiom R to our context leads to our second proposition:

**PROPOSITION 2.** (a) In Provision, the preferences induced by an average contribution \( \bar{c}^A_{-i} \) are more altruistic (MAT) than the preferences induced by \( \bar{c}^B_{-i} \) if and only if \( \bar{c}^A_{-i} > \bar{c}^B_{-i} \). (b) In Maintenance, the preferences induced by an average withdrawal \( \bar{w}^A_{-i} \) are more altruistic (MAT) than the preferences induced by \( \bar{w}^B_{-i} \) if and only if \( \bar{w}^A_{-i} < \bar{w}^B_{-i} \).

**Proof:** As shown before, Proposition 2 establishes that opportunity set \( G_A \text{ MGT } G_B \) if and only if \( \bar{c}^A_{-i} > \bar{c}^B_{-i} \) (\( \bar{w}^A_{-i} < \bar{w}^B_{-i} \)). Axiom R states that if \( G_A \text{ MGT } G_B \), then \( A_A \text{ MAT } A_B \), where \( A_A \) are the preferences induced by opportunity set \( G_A \). It follows that \( A_A \text{ MAT } A_B \) if and only if \( \bar{c}^A_{-i} > \bar{c}^B_{-i} \) (\( \bar{w}^A_{-i} < \bar{w}^B_{-i} \)). Q.E.D.

Finally, we derive the implications of Axiom S in Cox et al. (2008), p. 41. Assume \( C \) is composed by at least two opportunity sets, one of which is the status quo. Denote as \( A'_C \) the preferences induced by opportunity set \( C \) when \( C \) is the status quo and \( A_C \) when \( C \) is not the status quo. If \( C \) is singleton, preferences induced by the only feasible opportunity set \( C \) are indicated as \( A^*_C \). This is a case where the action is forced, i.e., there is no alternative than that particular opportunity set.
The first part of Axiom S (Cox et al. (2008), p. 41) states that if \( G \) \( \)\( MGT \) \( F \) and either \( G \) \or \( F \) is the status quo, then \( A_G \) \( MAT \) \( A'_G, A_G^* \) and \( A'_F, A_F^* \) \( MAT \) \( A_F \).^5

Hence, if there are two identical opportunity sets \( G \) and \( G' \) and the latter is the status quo, then the preferences induced by \( G \) are more altruistic than (MAT) the ones induced by \( G' \); and that if there are two identical opportunity sets \( F \) and \( F' \) and the latter is the status quo, then the preferences induced by \( F' \) are more altruistic than the ones induced by \( F \). Intuitively, Axiom S strengthens or weakens Axiom R, depending on whether the status quo opportunity set is more or less generous than the opportunity set under consideration. Applied to our context, we can derive the following proposition:

PROPOSITION 3. Consider an average level of cooperation by the other group members of \( \bar{c}' \) and the implied opportunity sets \( G_c \) for Provision and \( G_w \) for Maintenance. The preferences induced by \( G_c \) are (weakly) more altruistic than the ones induced by \( G_w \).

Proof. Consider an average contribution of \( \bar{c}'_i > 0 \) and an average withdrawal of \( \bar{w}'_i < 20 \), where \( \bar{c}'_i = 20 - \bar{w}'_i \). We prove that the best response to an average \( \bar{c}'_i \) is higher than the best response to an average of \( \bar{w}'_i \) (see Cox et al. (2013) and Frackenpohl et al. (2015) for similar proofs).

In the Provision problem, consider \( \bar{c}'_i > 0 \) and another average contribution \( \bar{c}'_i^* = \bar{c}'_i \) that is generated by nature (\( C \) is singleton). Assume that \( c_i \) is the best reply to the latter average contribution. Axiom S implies the following ranking of the best replies (br):

\[
br_{\bar{c}'_i} \geq c_i
\]

We turn now to the Maintenance problem. Consider the average withdrawal \( \bar{w}'_i \), where \( \bar{c}'_i = 20 - \bar{w}'_i \). Consider further another average withdrawal \( \bar{w}'_i^* = \bar{w}'_i \) that is generated by nature (\( C \) is singleton) and \( w_i \) is the best reply to that average withdrawal. From Axiom S it follows that:

\[
br_{\bar{w}'_i} \geq w_i
\]

From Axiom R, it follows that the two best replies to the contribution and withdrawal generated by nature are isomorphic. Hence, we can express \( w_i = 20 - c_i \) and combine the two inequalities above as follows:

\[
br_{\bar{c}'_i} \geq c_i \geq 20 - br_{\bar{w}'_i}
\]

The inequality proves that according to revealed altruism, best response for a given effective average contribution is (weakly) higher in Provision compared to Maintenance. Q.E.D.

The intuition for Proposition 3 is the following: in Provision, the status quo opportunity set (in which no tokens are yet contributed to the public good) is the least generous possible. Hence, any other opportunity set compared with the status quo will increase the effect of

---

^5 We only describe the first part of Axiom S for the ease of exposition. See Cox, et al. (2008) for the complete Axiom. All our predictions depend on the first part of the Axiom only.
Axiom R. Conversely, in Maintenance the status quo opportunity set (in which no token is yet withdrawn from the public good) is the most generous possible. Hence, any other opportunity set compared with the status quo will decrease the effect of Axiom R. This implies that for the same effective average contribution of the other three group members, preferences will be (weakly) more altruistic in Provision than in Maintenance.

Supplementary References


