Games and Economic Behavior 135 (2022) 338–355

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Preferences and perceptions in Provision and Maintenance public goods

Simon Gächter [a,b,c,*], Felix Kölle [d,*], Simone Quercia [e,*]

[a] University of Nottingham, Nottingham NG7 1BW, United Kingdom
[b] CESifo, 81679 Munich, Germany
[c] IZA, 53113 Bonn. Germany
[d] University of Cologne, 50923 Cologne, Germany
[e] University of Verona, 37129 Verona, Italy

ARTICLE INFO

Article history:
Received 28 October 2021
Available online 15 July 2022

JEL classification:
C92
H41

Keywords:
Maintenance and provision social dilemmas
Conditional cooperation
Kindness
Misperceptions
Experiments
Framing

ABSTRACT

We study two generic versions of public goods problems: in Provision problems, the public good does not exist initially and needs to be provided; in Maintenance problems, the public good already exists and needs to be maintained. In four lab and online experiments ($n = 2{,}105$), we document a robust asymmetry in preferences and perceptions in two incentive-equivalent versions of these public good problems. We find fewer conditional cooperators and more free riders in Maintenance than Provision, a difference that is replicable, stable, and reflected in perceptions of kindness. Incentivized control questions administered before gameplay 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 behavioral analyses.

1. Introduction

In this paper, we study 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 rule-bending, rent-seeking, and corruption.

As the examples illustrate, "provision" and "maintenance" public goods differ along many dimensions. Crucially, however, for selfish players, 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). Although the comparison between these two problems has been studied both in economics and psychology (see, e.g., Sell and Son, 1997 and Dufwenberg et al., 2011), until recently, most studies only investigated

* Corresponding authors.
E-mail addresses: simon.gaechter@nottingham.ac.uk (S. Gächter), felix.koelle@uni-koeln.de (F. Kölle), simone.quercia@univr.it (S. Quercia).

https://doi.org/10.1016/j.geb.2022.06.009

cooperative *behavior*, and less the psychological *mechanisms* that produce cooperation.[1] Here, we ask whether, from the perspective of social preferences and perceptions, maintenance and provision dilemmas are psychologically different social dilemmas.

Studying *preferences* and *perceptions* as drivers of cooperation and understanding whether their impact on cooperative behavior differs across maintenance and provision dilemmas is important from both a theoretical and practical point of view. From a theoretical viewpoint, studying *preferences* and *perceptions* is interesting because it can help explain why 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 (see, e.g., Fischbacher and Gächter, 2010; and Gächter et al., 2017 for evidence from the lab; and, e.g., Frey and Meier, 2004; Alpizar et al., 2008; Rustagi et al., 2010; Fehr and Leibbrandt, 2011 for evidence from the field). Do people cooperate because they misperceive their incentives in a social dilemma, or do they have a 'genuine' preference for cooperation? If people misunderstand their incentives, they may implement choices they otherwise would not. Thus, observing cooperation without controlling for perception of incentives may not be a conclusive revelation of a preference for cooperation (for related arguments see, e.g., Koszegi and Rabin, 2008 and Cason and Plott, 2014). Apart from potential differences in the understanding of the incentive structure, maintenance and provision might also differ in the way people perceive others' actions, e.g., in terms of kindness. Perceptions of the kindness of other players' actions are important because, in social dilemmas, they can explain why some people cooperate in the first place (e.g., Falk and Fischbacher, 2006).

The practical relevance of our research question comes from the fact that any policy intervention aimed at fostering cooperation must rest on accurate behavioral mechanisms. If these mechanisms are dilemma-specific, this would imply that maintenance and provision might require different approaches to overcome the tragedy of the commons.

To answer these questions, and to provide a comprehensive understanding of the fundamental nature of cooperation in maintenance and provision dilemmas, we present the results from four experiments, in which we study preferences and perceptions in conjunction. 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.[2]

To ascertain the role of preferences and perceptions in influencing cooperation in these two dilemmas, we proceed in four steps that we summarize in Table 1.

Our *first step* is to ensure that what we study is a replicable phenomenon (Drazen et al., 2021). This is the purpose of Experiment 1. In this experiment, we replicate the one-shot results of Gächter et al. (2017) using a diverse online subject pool (MTurk). We find that in a simultaneous one-shot game, people contribute 52% of their endowment in *Provision* compared to 39% in *Maintenance*, a difference that is highly statistically significant. Furthermore, using the Fischbacher et al. (2001) strategy-method experiment to separate beliefs from preferences, we replicate that there are systematically fewer conditional cooperators and more free riders in *Maintenance* than *Provision*. In two additional experiments (Experiments 2a and 2b in Table 1), we collect new evidence to test whether, 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.

In Steps 2 and 3 we turn to our central question of how people perceive others' behavior (Step 2) and the incentives in the dilemmas (Step 3). Our *second step* measures people's perceptions about the kindness of others' behavior. While differences in kindness perceptions could support 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 of previous evidence (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).

Testing for misperception of incentives is our *third step*. To this end, 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 prefer-

[1] The behavioral 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).

[2] We focus sharply on the social dilemma dimension of provision and maintenance public goods, and abstract from technological features (e.g., resource rivalry in common-pool resources vs. non-rivalrous public goods) and institutional details (rules and regulations) that define real-world social dilemmas (e.g., Ostrom, 1990; Cornes and Sandler, 1996; Poppe, 2005; Apesteguía and Maier-Rigaud, 2006; Levin, 2014).

Table 1
Four steps to test whether Maintenance and Provision are different dilemmas.

Exp.	Experiment	Number of participants	Subject pool	Purpose
Step 1: Establishing the replicability and stability of cooperative preferences (Section 3)				
0	Gächter et al. (2017)*	($n = 703$)	Students (UoN)	These data provide a previous benchmark result, which we replicate in Experiment 1
1	Replication study	$n = 704$	US citizens (MTurk)	Assess whether Experiment 0 with students can be replicated in a non-student subject pool
2a	Temporal stability (5 months delay)	$n = 119$	Students (UoN, sampled from experiment 0)	Assess the role of stability of conditionally cooperative preferences over time
2b	Predictive power of cooperation attitudes	$n = 116$	Students (UoN, sampled from experiment 4)	Assess the predictive power of conditionally cooperative preferences plus beliefs to explain actual cooperation levels
Step 2: Measuring perceptions of kindness (Section 4)				
3	Kindness survey	$n = 185$	Students (UoN, new participants)	Measure how kind or unkind people perceive a certain cooperation level to be. Measured on a scale of -100 (=very unkind) to +100 (= very kind)
		$n = 401$	US citizens (MTurk, new participants)	
Step 3: Measuring game form misperceptions and controlling for them (Section 5)				
4	Dilemma-specific game-form misperceptions	$n = 696$	Students (UoN, new participants)	Measure with 8 incentivized questions about payoffs and goals how people perceive the incentives in the public good game. Control for misperceptions to test for potential differences in cooperative preferences
Step 4: Assessing theoretical explanations (Section 6)				

* Data taken from https://doi.org/10.5061/dryad.8d9t2.

ences using the strategy method by Fischbacher et al. (2001).[3] Controlling for people's misunderstanding will then allow us to test whether cooperation preferences continue to differ statistically significantly between *Maintenance* and *Provision*. Our *fourth step* is to discuss how various theories of social preferences can explain our results.

Our paper offers several contributions to the literature. Our methodology of holding the nature of the social dilemma constant relates us to literatures on framing and context effects in other-regarding behavior.[4] More specifically, our focus on provision/maintenance of public goods leads us naturally to a design that is a version of what psychologists (Dawes, 1980) have called "take-some" vs. "give-some" dilemmas (e.g., Sell and Son, 1997; van Dijk and Wilke, 1997; Sonnemans et al., 1998; Messer et al., 2007; Cubitt et al., 2011a; Dufwenberg et al., 2011; Cox, 2015; Cox and Stoddard, 2015; Fosgaard et al., 2014; Fosgaard et al., 2017; Khadjavi and Lange, 2015; Isler et al., 2021). Our maintenance/provision design differs from designs that manipulate whether the positive externality of contributing to the public good or the negative externality of not contributing is emphasized. Papers in this line of research are Andreoni (1995b); Park (2000); and Fujimoto and Park (2010). For a comparative discussion of give/take or positive/negative externality framing effects and an overview of studies see Cartwright (2016).[5]

Different from most early literature on give-some and take-some games, our analysis consists of separating preferences and perceptions as determinants of cooperation rather than focusing only on cooperation decisions. With regard to preferences, our paper joins the small literature that elicits preferences for conditional cooperation in maintenance or provision problems (e.g., Frackenpohl et al., 2016; Gächter et al., 2017; Fosgaard et al., 2014; Fosgaard et al., 2017; Isler et al., 2021). Our experiments also measure the perception of incentives *before* we elicit preferences for cooperation. This heeds arguments by Koszegi and Rabin (2008) and Cason and Plott (2014) that measuring preferences requires controlling for the perception of incentives.[6]

Our goals relate us to Fosgaard et al. (2017). They also measure preferences for conditional cooperation and misperceptions albeit after the elicitation of preferences and with fewer questions. Theirs is a representative subject pool from Denmark ($n = 2,042$), whereas we present evidence from a student subject pool in the UK and online workers (MTurk) from the general population in the US (total $n = 2,105$). More importantly, unlike Fosgaard et al. (2017), we report evidence on the temporal stability of preferences, perceptions of kindness and guilt, and link our results to theories of social preferences. Our four-step analysis that combines lab and online experiments, replications, between and within-subject sta-

[3] 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 incentivized post-experimental questionnaire.

[4] Our focus is on public goods games, but also relates to the importance of context effects. For instance, 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; Bicchieri et al., 2022). Also, the approval of egoistic behavior seems to be context-dependent too, e.g., in markets vs. non-market settings (Bartling et al., 2021).

[5] Another dimension of framing effects is due to attaching labels to games (e.g., "Wallstreet vs Community game"; e.g., Ellingsen et al., 2012 and Dufwenberg et al., 2011). In this paper we use neutral labels.

[6] Our focus on social preferences and (mis-)perception does not deny the possibility that cognitive ability, risk preferences and loss aversion might matter too (e.g., De Dreu and McCusker, 1997; Iturbe-Ormaetxe et al., 2011), but we leave this for future research, not least to keep this paper manageable.

bility tests, measurement of perceptions of kindness as well as of the incentives in the games, establishes that *Maintenance* and *Provision* are different social dilemmas even when sharing the same game form and when controlling for possible misperceptions of incentives.

2. 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 members. Equation (1) describes the material incentives of individual i:

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

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 (partially or fully) withdraw (w_i) from the project. Material incentives are described by equation (2):

$$\pi_i = w_i + \frac{1.6}{4}(80 - \sum_{j=1}^{4} w_j). \tag{2}$$

If people are only motivated by material incentives, (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 several parts. In the first part of each experiment, 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-subjects 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.

The exact design of the remaining parts differed across our experiments.[7] In most of our experiments, in the second part we implemented a strategy-method public goods game (described below) through which we measure *cooperation attitudes*, our main proxy for cooperation preferences. Some of the sessions in these experiments included a third part in which participants played a direct-response game in which they simultaneously had to state their contribution decision and belief about others' contributions. In the experiments in which we elicited game-form misperceptions, the strategy-method game in part 2 was preceded by a set of incentivized control questions. In the following, we explain how we elicited cooperation attitudes, which is our main variable of interest. 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 (or 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 (or withdrawal) decision for *each* possible (rounded) average contribution (or 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 (or 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 (or withdraw); and (iii) as *other* if none of the criteria in (i) & (ii) apply.[8]

[7] 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 later parts could therefore not affect behavior in previous parts.

[8] 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.

Our data come from four experiments and three main sources: the CeDEx lab at the University of Nottingham; the online labor market platform Amazon Mechanical Turk (MTurk); and online experiments conducted with students at the University of Nottingham (see Table 1 for an overview of 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 (average age 20.2 years; 58% female) 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, participants (all US residents) were 31.9 years old and 41% were female.[9] Average payments were £20.60 for lab sessions, and $2.60 for MTurk sessions (corresponding to an hourly wage of $13.00).

3. Step 1: replicability and stability of preferences in Maintenance and Provision

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.[10] 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 gameplay and compare it across *Maintenance* and *Provision*. All procedural details and further supporting evidence are in Online Appendix B1.

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

The left panel of Fig. 1 summarizes the main relevant finding for our paper 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*.[11] In a one-shot direct response game played after the type elicitation, Gächter et al. (2017) further found that cooperation rates were significantly higher in *Provision* than in *Maintenance* (41% vs. 30%; two-sided t-test: $p = 0.007$).

Our replication for the purposes of this paper was conducted on MTurk with $n = 703$ US participants (instructions are in Online Appendix A).[12] The results match our previous findings. While the levels in the frequency of types are different compared to our UK student sample – we observe more conditional cooperators (73% vs. 53%, $\chi^2(1) = 62.21$; $p < 0.001$) and less free riders (13% vs. 22%, $\chi^2(1) = 23.26$; $p < 0.001$) and others (14% vs. 24%, $\chi^2(1) = 24.97$; $p < 0.001$) on MTurk - treatment differences are highly significant.[13] Specifically, as shown in the right panel of Fig. 1, in line with Gächter et al. (2017), we find a significantly different distribution of types across treatments ($\chi^2(2) = 15.96$, $p < 0.001$) with 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*.[14]

Like in our previous study, we also find that effective cooperation rates (after contributions/withdrawals), 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

[9] 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 (2021) demonstrate that behavior in a variety of games is similar on MTurk and the lab.

[10] See Maniadis et al. (2014), Camerer et al. (2016), Camerer et al. (2019), and Drazen et al. (2021) on the importance of replicability in experimental economics.

[11] The category 'others' contains "unconditional cooperators" who contribute a constant positive amount irrespective of what other group members contribute, "anti-conditional cooperators" whose cooperation depends negatively on the cooperation of other group members, "triangle cooperators" who are conditionally cooperative up to a certain level when they turn into anti-conditional and the rest. See Online Appendix B1 for further details and the relative frequencies of these subtypes.

[12] 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 $n = 700$.

[13] The different levels of the frequency of types across our two studies is not surprising given the different cultural and sociodemographic background of the participants. We note, however, that the results from our MTurk study are very similar to Kocher et al. (2008) who elicited cooperation types among US students using a Provision public goods game: When comparing their results to ours, 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' in our MTurk data shows similar results than in Gächter et al. (2017). See Table B1 (Panel B) in Online Appendix B.

[14] We note that the differences across *Maintenance* and *Provision* are somewhat less pronounced in the MTurk sample compared to the student sample: the difference in the share of conditional cooperators amounts to 13 and 20 percentage points, respectively; the difference in the share of free riders is 8 and 11 percentage points, resp.; and the difference in the share of others is 4 and 11 percentage points, resp.. To test whether these differences across the two samples are significant, we run logistic regressions in which we use the different types as dependent variable, a treatment dummy, a MTurk dummy, and an interaction between the latter two as independent variables. The results, reported in Table B2 in Online Appendix B, confirm that there are significantly more conditional cooperators and significantly less free riders and others in *Provision* than *Maintenance* and on MTurk, but that there are no significant interaction effects (Provision × MTurk).

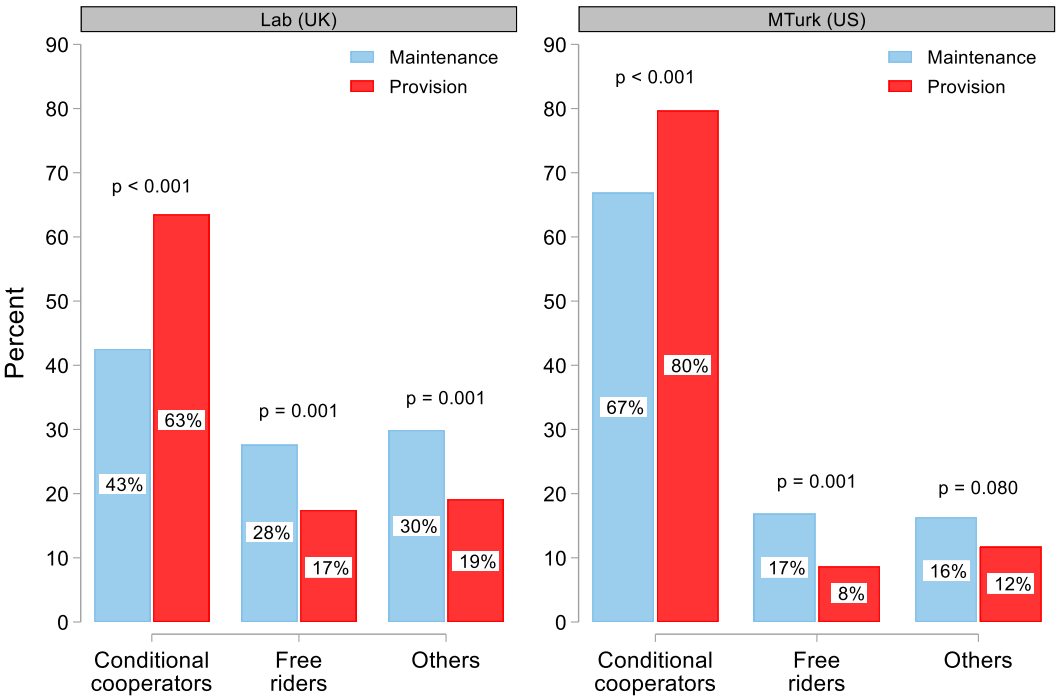

Fig. 1. Distribution of cooperation types. Left panel: $n = 704$ students from the UK (source: Gächter et al. (2017)). Right panel: $n = 703$ US participants from MTurk (source: new experiments). p-values from χ^2-tests.

to differ significantly across both dilemmas (Lab: *Provision*: 42%, *Maintenance:* 34%, two-sided t-test: $p = 0.003$; MTurk: *Provision*: 53%, *Maintenance:* 38%, two-sided t-test: $p < 0.001$).

3.2. Temporal stability

To test whether the observed difference in the distribution of types is also stable *within* participants, we ran an additional experiment in which we re-invited a subset of participants from the Gächter et al. (2017) sample (left panel of Fig. 1) four months after their first participation. Without knowing in advance, participants took part in sessions that were identical to the ones in which they participated before. We report results from $n = 119$ participants ($n = 65$ in *Provision* and $n = 54$ in *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 *within treatments* is very similar and does not significantly change between waves, neither in *Maintenance* ($\chi^2(2) = 0.51$, $p = 0.776$) nor *Provision* ($\chi^2(2) = 1.57$, $p = 0.456$). Consequently, when comparing the distribution of types *across treatments*, we find a significantly different distribution across *Maintenance* and *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 B3 in Online Appendix B).

Regarding *individual-level stability* of cooperation preferences, we find that in *Provision* 66% of participants are classified as the same type in both waves, compared to 59% in *Maintenance*, a difference that is not statistically significant ($\chi^2(1) = 0.60$, $p = 0.438$). While these numbers indicate that the stability of types across waves is clearly not perfect, for both treatments we find it to be significantly higher than chance (which amounts to 46% and 35%, in *Provision* and *Maintenance*, resp.; t-tests, both $p < 0.001$; see Table B4 in Online Appendix B). The stability rate in *Provision* is similar to Volk et al. (2012) who, using a similar setup and a gap of 2.5 months between waves, find a stability rate of 64%. No such comparison is possible for *Maintenance* because, as far as we are aware of, no previous study has investigated the stability of cooperation preferences using a maintenance dilemma.

3.3. 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 behavior in another comparable environment. To test this, we rely on the third part of our experiment in which a subset of participants took part 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 et al. (2012), by combining elicited cooperation attitudes with stated beliefs we can

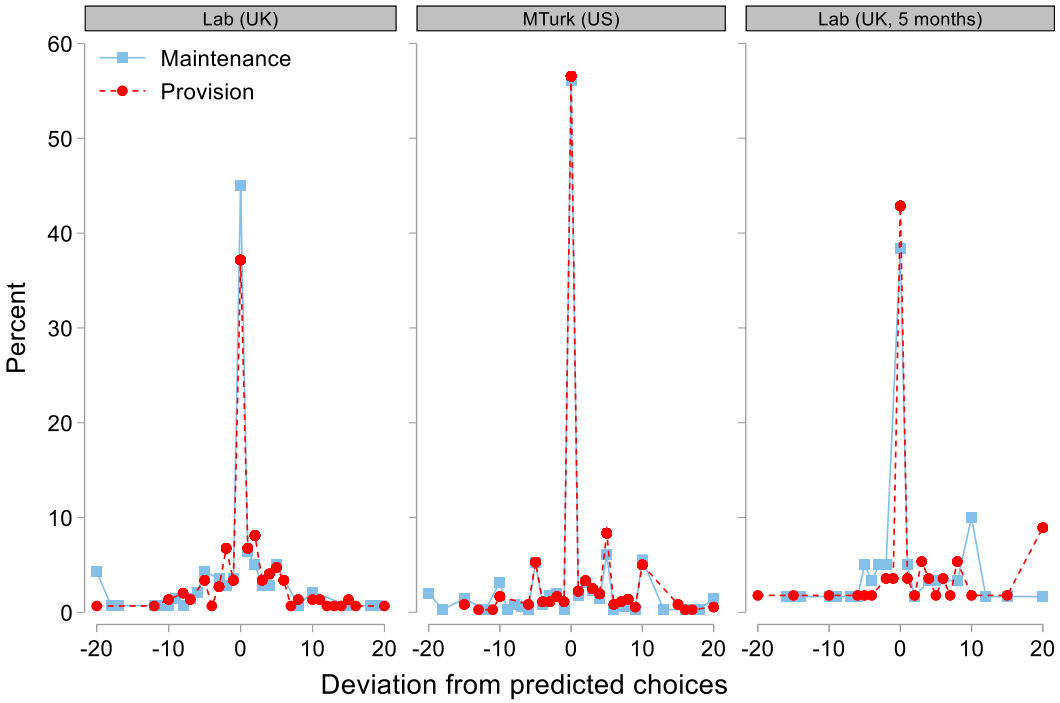

Fig. 2. Deviations from predicted choices in *Maintenance* and *Provision*. Left panel: Students ($n = 288$). Middle panel: MTurk ($n = 703$). Right panel: Students who participated in the direct-response experiment five months after the preference elicitation experiment ($n = 116$).

make a point prediction about the contribution decision, $\hat{c}_i$. We then compare $\hat{c}_i$ with c_i (*i's* actual contribution in the direct-response game), delivering an individual-level measure of consistency.

In total, we have (1) $n = 288$ observations from our Gächter et al. (2017) sample, and (2) $n = 703$ observations from our MTurk experiment.[15] 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 of temporal stability. Our results are shown in Fig. 2, depicting 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.

Fig. 2 reveals that in all cases the modal and the median deviation is zero, that is, participants' contribution decision 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 (see Online Appendix B for further details), 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): $p = 0.472$). Overall, this demonstrates that the elicited attitudes are, together with elicited beliefs, an equally good predictor of actual cooperation behavior in both *Maintenance* and *Provision*.

3.4. Discussion

Consistent with the evidence from Frackenpohl et al. (2016) and Fosgaard et al. (2017), in Gächter et al. (2017) we have shown that *Maintenance* and *Provision* dilemmas elicit systematically different cooperation attitudes with significantly fewer participants behaving conditionally cooperative in the former than in the latter. In Gächter et al. (2017) we have further shown that together with differences in the beliefs about others' cooperation, this translates into different levels of cooperation in both one-shot and repeated games. We extend this prior evidence by showing that (i) differences in cooperation attitudes across *Maintenance* and *Provision* are replicable across different subject pools, (ii) elicited attitudes in both dilemmas are equally stable within participants over a period of four months, and (iii) elicited attitudes are (jointly with beliefs) an equally good predictor of actual cooperation decision in both dilemmas, even after a delay of five months. We summarize these findings in our first result:

[15] The remaining participants from Gächter et al. (2017) played a repeated game. The results on the predicted power for the first-period contributions are similar to the one reported here (see Gächter et al., 2017).

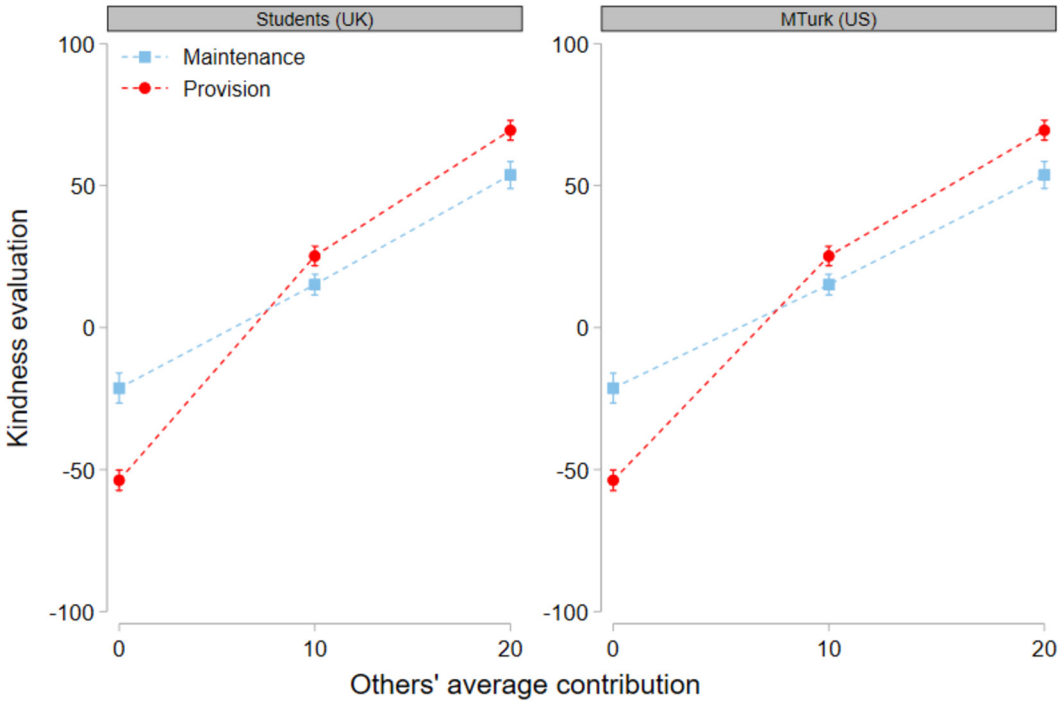

Fig. 3. Kindness perceptions in *Provision* and *Maintenance* of others' effective contributions (±1 s.e.m).

Result 1: *Maintenance and Provision evoke systematically different cooperation attitudes. Most importantly, conditional cooperation is more frequent in Provision than Maintenance. The elicited attitudes are stable within individuals and, jointly with beliefs, predictive of actual cooperation decisions.*

While replicability, stability, and predictive power are necessary conditions to interpret the effects as differences 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 steps (and sections), we disentangle the relative importance of social preferences and misperceptions. 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.

4. Step 2: perceptions of kindness in 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 behavioral 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. 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 behavior 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 to 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 (see Online Appendix A3). We recruited $n = 185$ students from the University of Nottingham and $n = 401$ participants from MTurk. No participant was involved in any of our experimental sessions before.[16]

Fig. 3 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

[16] Since we asked participants for their personal perceptions, 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.

pattern for medium and high effective contributions of 10 and 20, respectively. In these cases, payoff equivalent actions are considered as unkinder in *Maintenance* compared to *Provision* (two-sided t-tests, average others' contribution = 10, $p = 0.045$ 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).

Further support comes from OLS regressions in which we use kindness evaluations as the dependent variable, others' average contributions, a dummy for *Provision*, and an interaction term of the last two as independent variables. We run this regression separately for each subject pool. The results are in Table B5 in Online Appendix B. In line with Fig. 3, 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*. We summarize these findings in our second result:

Result 2: *Maintenance and Provision evoke systematically different perceptions of kindness: Complete free-riding is perceived to be unkinder in Provision than in Maintenance, while positive contributions (of 10 and 20) are perceived as more kind in Provision than in Maintenance*.

Overall, Result 2 suggests that both with respect to kindness and unkindness, individuals have stronger reactions in their perception of others' contributions when contributing to, rather than withdrawing from, a public good. Result 2 is also consistent with Cubitt et al. (2011b) who found a similar pattern for moral judgments: Failing to contribute to the public good was perceived as morally worse than withdrawing everything. 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*. The result from our kindness survey thus favors the explanation that the differences in cooperation attitudes across treatments are rooted in differences in the underlying preferences.

5. Step 3: measuring and controlling for game form misperceptions

One alternative explanation for our results is that participants may have systematic misperceptions of the game form and that these misperceptions may be dilemma 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.

5.1. Conceptualization of game form misperceptions

Our conceptualization of game-form misperceptions draws on Cason and Plott (2014). They analyze 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 behavior 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' behavioral responses given their underlying preferences. Cason and Plott's general conclusion is that researchers should be careful when 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 behavioral responses across treatments as evidence for dilemma-dependent preferences might be erroneous because such differences can be due to dilemma-dependent misperceptions of the game form. 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 behavior different from their preferred one. For example, if some participants erroneously believe that 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 treatment 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 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.

5.2. 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 2, 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 2), paying £0.1 per

Table 2
Incentivized misperception questions and percentage of correct answers across Maintenance (M, $n = 320$) and Provision (P, $n = 376$).

	% Correct answers		
	M	*P*	χ^2 – test
Payoff questions:			
Q1: Assume that you contribute 20 tokens to the project and the other three group members contribute nothing to the project. What will your total income be?	95.3	95.5	$p = 0.917 (q = 0.917)$
Q2: Assume that you contribute 20 tokens to the project and the other three group members contribute nothing to the project. What will the total income of each of the other group members be?	90.9	84.3	$p = 0.009$ $(q = 0.035)$
Q3: Assume that you contribute 0 tokens to the project and each of the other three group members contributes 20 tokens to the project. What will your total income be?	94.4	92.8	$p = 0.405$ $(q = 0.649)$
Q4: Assume that you contribute 0 tokens to the project and each of the other three group members contributes 20 tokens to the project. What will the total income of each of the other group members be?	95.9	93.1	$p = 0.103$ $(q = 0.275)$
Goal questions:			
Q5: Suppose the other group members contribute on average 0 tokens to the project. How much should a person who wants to make as much money as possible for him/herself contribute to the project?	92.8	95.0	$p = 0.239$ $(q = 0.478)$
Q6: Suppose the other group members contribute on average 20 tokens to the project. How much should a person who wants to make as much money as possible for him/herself contribute to the project?	93.8	85.6	$p = 0.001$ $(q = 0.004)$
Q7: Suppose the other group members contribute on average 0 tokens to the project. How much should a person who wants that the group as a whole makes as much money as possible contribute to the project?	79.1	77.1	$p = 0.539$ $(q = 0.664)$
Q8: Suppose the other group members contribute on average 20 tokens to the project. How much should a person who wants that the group as a whole makes as much money as possible contribute to the project?	93.1	92.0	$p = 0.581$ $(q = 0.664)$
Total	91.9	89.4	$p = 0.030$

Notes: Shown are the questions in *Provision*. The questions for *Maintenance* were formulated equivalently. *q*-values correspond to *p*-values corrected for multiple comparisons using the Benjamini and Hochberg (1995) false discovery rate procedure. For testing the total effect (last row) we use logistic regressions with standard errors clustered at the individual level.

correct answer.[17] After that, the experiment proceeded with the elicitation of cooperation preferences using the strategy method. We asked the incentivized questions *before* we elicited participants' preferences to ensure maximal understanding of the situation and incentives.[18]

The first four 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 2 for the exact wording of questions). Our first measure of game-form misperception classifies a participant 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).

In the other four questions (compare Q5 – Q8 in Table 2), 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 participant 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, participants may then implement 'wrong' behavioral responses given their underlying preferences. That is, while the behavioral response of such a misperceiving participant in the strategy method might look like evidence of prosocial, reciprocal preferences, such behavior is also consistent with a model in which a purely selfish participant maximizes their misperceived payoffs.

5.3. Misperceptions are dilemma specific

Table 2 summarizes the percentages of correct answers separately for each question and for *Maintenance* and *Provision*. It reveals that, at the aggregate level, in both treatments there is an overall very low level of misperception. With a few

[17] 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.

[18] In this methodological aspect, our approach is akin to Plott and Zeiler (2005) who used a battery of experimental tools designed to maximize understanding before eliciting WTA-WTP valuations.

Table 3
Percent of participants classified as misperceiving in *Maintenance* and *Provision*.

	Maintenance [*n* = 320]	*Provision* [*n* = 376]	χ^2 - *test*
Measure 1 – At least one mistake in the payoff questions Q1-Q4	13% [*n* = 40]	22% [*n* = 84]	$p = 0.001$
Measure 2 – At least one mistake in the goal questions Q5-Q8	29% [*n* = 93]	37% [*n* = 140]	$p = 0.023$
Measure 3 – Mistaken conditional cooperation	5% [*n* = 17]	13% [*n* = 47]	$p = 0.001$

Table 4
Fraction of misperceiving participants in Maintenance and Provision by type.

	Maintenance			*Provision*		
	Measure 1	Measure 2	Measure 3	Measure 1	Measure 2	Measure 3
Conditional Cooperators	0.103	0.252	0.028	0.247	0.447	0.200
Free Riders	0.095	0.230	0.024	0.165	0.262	0.029
Others	0.195	0.425	0.126	0.241	0.337	0.072
χ^2 - *tests*	$p = 0.066$	$p = 0.005$	$p = 0.002$	$p = 0.247$	$p = 0.006$	$p < 0.001$

exceptions, the percentage of correct answers is above 90% for every 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 ($p = 0.001$).

Next, we turn to an individual-level analysis of mistakes by applying our three measures of game-form misperceptions as described above. As shown in Table 3, 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* (χ^2 – tests, all $p < 0.023$).

We summarize these findings in our third result:

Result 3: *Provision dilemmas cause significantly higher levels of misperceptions of the game form than Maintenance dilemmas.*

5.4. Misperceptions and cooperation attitudes

In the following, we analyze the connection between misperceptions and the elicited cooperation attitudes. If there was none, i.e., if mistakes were randomly distributed across types, then the different degrees in the level of misperception across *Maintenance* and *Provision* should not affect 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 4 reports the fraction of misperceiving participants conditional on type classification. We report these numbers separately for *Maintenance* and *Provision* and our three measures of misperception. Table 4 reveals that the null hypothesis of no relationship between types and misperceptions can be rejected for both *Maintenance* and *Provision* according to Measure 2 (goal questions, χ^2 – tests, both $p < 0.007$) and Measure 3 (mistaken conditional cooperation, χ^2 – tests, both $p < 0.003$), but not for Measure 1 (payoff questions, χ^2 – tests, both $p > 0.065$). On top of that, our results reveal that the way misperceptions interact with the elicited attitudes is treatment-specific. In *Maintenance* we observe mainly the participants classified as others who display some form of misperceptions; compared to free riders their odds of displaying misperceptions (calculated as the ratio between misperceiving and non-misperceiving participants) are increased by a factor of 2.3 (Measure 1) up to 5.9 (Measure 3). In *Provision*, in contrast, we find that mainly the group of conditional cooperators exhibiting misperceptions; compared to free riders their odds of displaying misperceptions is increased by a factor of 1.6 (Measure 1) up to 8.3 (Measure 3).

These results show that the two types of dilemmas not only affect 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 behavioral differences as dilemma-dependent preferences. In the following, we therefore test whether accounting for the different types of misperceptions can explain the observed treatment differences in the distribution of cooperation preferences. We report this analysis in Table 5.

Panel A of Table 5 shows the distribution of types in the full sample (without controlling for misperceptions). In line with our results from Section 3, we find again a highly significant difference in the distribution of preferences across treatments ($\chi^2(2) = 21.23$, $p < 0.001$), with significantly fewer conditional cooperators ($\chi^2(1) = 20.65$, $p < 0.001$) and significantly more free riders ($\chi^2(1) = 11.24$, $p = 0.001$) in *Maintenance* than in *Provision*.[19]

[19] The relative frequency of types is somewhat different compared to the one found in our initial student sample as reported in Section 3.1. Specifically, we find a significant change in the distribution of types in both *Provision* ($\chi 2(2) = 14.26$, $p = 0.001$) and *Maintenance* ($\chi 2(2) = 11.01$, $p = 0.004$), with more free riders (*Provision*: 27% vs. 17%, $\chi 2(1) = 10.43$, $p = 0.001$, *Maintenance* 39% vs. 28%, $\chi 2(1) = 10.44$, $p = 0.001$) and fewer conditional cooperators

Table 5
Distribution of cooperation preferences in Maintenance and Provision after controlling for different types of misperceptions.

	Panel A: Full sample			Panel B: No mistake in payoff questions		
Type	Maintenance ($n = 320$)	Provision ($n = 376$)	χ^2-test	Maintenance ($n = 280$)	Provision ($n = 292$)	χ^2-test
Conditional Cooperators	34%	51%	$p < 0.001$	34%	49%	$p < 0.001$
Free Riders	39%	27%	$p = 0.001$	41%	29%	$p = 0.005$
Others	27%	22%	$p = 0.118$	25%	22%	$p = 0.332$
χ^2-test	$p < 0.001$			$p = 0.001$		
	Panel C: No mistake in goal questions			Panel D: No mistaken conditional cooperation		
Type	Maintenance ($n = 227$)	Provision ($n = 236$)	χ^2-test	Maintenance ($n = 303$)	Provision ($n = 329$)	χ^2-test
Conditional Cooperators	35%	45%	$p = 0.042$	34%	46%	$p = 0.002$
Free Riders	43%	32%	$p = 0.019$	41%	31%	$p = 0.007$
Others	22%	23%	$p = 0.743$	25%	23%	$p = 0.623$
χ^2-test	$p = 0.050$			$p = 0.006$		

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 (χ^2 – tests, all $p \leq 0.05$).

Regarding the distribution of types, we observe significantly more conditional cooperators (χ^2 – tests, all $p < 0.05$) and significantly fewer free riders (χ^2 – tests, all $p < 0.02$) in *Provision* than in *Maintenance* (the difference in others is never significant, χ^2 – tests, all $p > 0.117$; see also Table B7 in Online Appendix B). 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. This demonstrates that while misperceptions can account for some of the observed differences across treatments, even after controlling for misperceptions, we observe substantial and significantly different degrees of conditional cooperation in *Maintenance* and *Provision*. We summarize these findings in our fourth result:

Result 4: *Even after accounting for the different degrees of misperceptions across Maintenance and Provision, we find significantly more conditional cooperators and fewer free riders in Provision than in Maintenance.*

5.5. Misperceptions do not affect perceptions of kindness

The results above already strongly suggest that the observed differences in cooperation types across *Maintenance* and *Provision* are rooted in differences in the underlying social preferences. As we have argued above, these differences can be explained by differences in the perceived kindness of others' actions across the two treatments. As a final test of this argument, we provide evidence that the different perceptions of kindness that we presented in Section 4 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 differences in conditional cooperation across the two dilemmas.

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 4 (see Fig. 3) 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 shown in Fig. 4.

(*Provision*: 63% vs. 51%, $\chi 2(1) = 12.50$, $p < 0.001$, *Maintenance* 33% vs. 43%, $\chi 2(1) = 5.84$, $p = 0.016$) in the new experiment. We believe that the reason for this result is that we administered the incentivized control questions before the elicitation of cooperation attitudes. This might have made the incentive structure of the social dilemma situation even clearer, which, in turn, might have corrected some 'mistaken' conditional cooperation. Alternatively, it could be that the incentivized questions increased the salience of material incentives, thereby priming participants to be more self-interested. While we cannot rule out neither of these channels, we can ascertain whether the shifts in the distribution of types had any effect on the differences across treatments. To this end, similar to our analysis in which we compared our student with the MTurk sample, we run logistic regressions in which we use the different types as dependent variable, a treatment dummy, a sample dummy, and an interaction between the latter two as independent variables. The results, reported in Table B6 in Online Appendix B, show that there are no significant interaction effects between the treatment and the sample, indicating that adding the additional questions at the beginning of the experiment had no systematic effect on our treatment comparison.

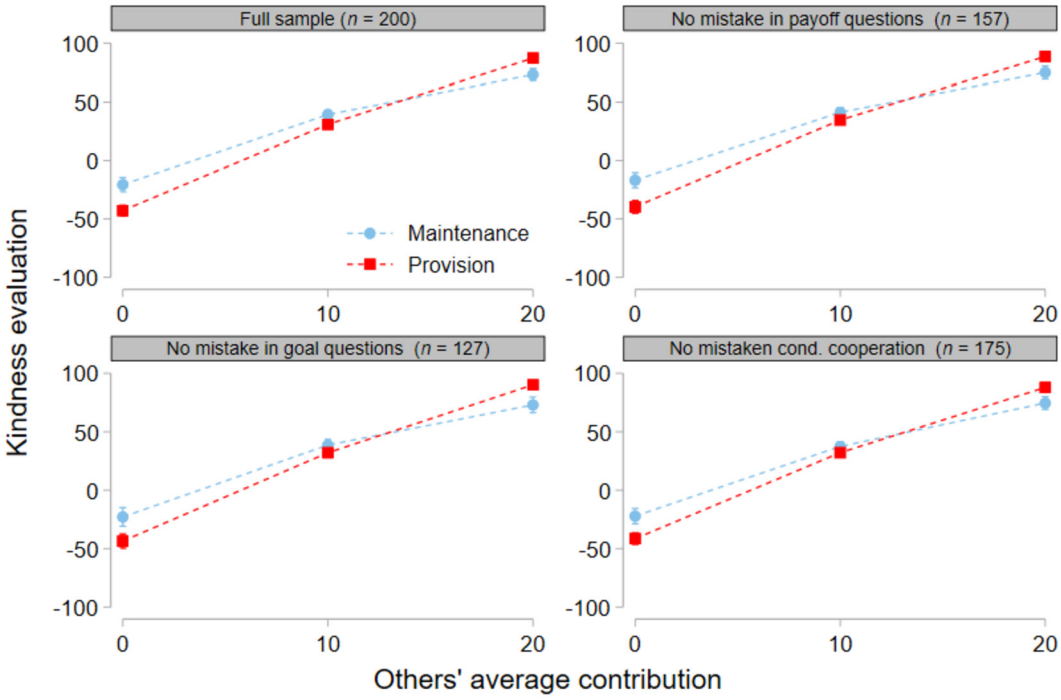

Fig. 4. Kindness perceptions in *Provision* and *Maintenance* of participants who experienced the decision situation (±1 s.e.m).

The first (upper left) panel of Fig. 4 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. Fig. 4 shows that the differences in kindness schedules across *Maintenance* and *Provision* are not only similar across the four panels but also similar to the ones reported in Fig. 3. 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.[20]

We summarize these findings in our fifth result:

Result 5: *Even after accounting for the different degrees of misperceptions across Maintenance and Provision, we find that the two dilemmas evoke systematically different perceptions of kindness that can explain why there are more conditional cooperators and fewer free riders in Provision than in Maintenance*.

6. Step 4: which theory of social preferences can explain our results?

In our *fourth step*, we investigate which of the existing models of social preferences, using their original formulations, can reconcile the observation of a higher share of conditional cooperators and fewer free riders in *Provision* than in *Maintenance* as found in our experiments as well as in some previous research (Fosgaard et al., 2014; Frackenpohl et al., 2016; Gächter et al., 2017; Isler et al., 2021). We emphasize that our aim here is not to conduct a horse race between different models of social preferences (see, e.g., Miettinen et al., 2020, for such an analysis in a sequential PD), but to provide a discussion about whether, and under which assumptions, existing theories of social preferences can explain *dilemma-dependent* conditional cooperation (note that all theories we discuss here can explain conditional cooperation in a given social dilemma). In the following, we only describe the main arguments, more details and formal analyses are in Online Appendix C. In Section 6.1., we discuss how theories of social preferences might explain dilemma-specific conditional cooperation. In Section 6.2., we discuss the models also in relation to our results on kindness perceptions (Results 2 and 5).

[20] Parametric estimates further corroborate these results. Using regression analyses in which we regress kindness evaluations on others' average contributions, a *Provision* 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 B8 in Online Appendix B, is similar to the ones reported in Section 4 where we analyze the kindness evaluations of uninvolved participants (compare also Table B5).

6.1. Explaining dilemma-specific conditional cooperation

We start our discussion with theories of *distributional preferences* (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). While all these theories can explain conditional cooperation, e.g., if advantageous inequity aversion is strong enough, they do not predict any dilemma-specific conditional cooperation. 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*.[21] Note that this prediction hinges on the original assumption of these theories that individual preference parameters are game-independent. If one would be willing to relax this assumption by allowing preference parameters to differ across contexts, then these models could potentially rationalize our findings. For example, if one assumes, using Fehr and Schmidt (1999) preferences, that agents are more advantageous inequity-averse when facing a *Provision* rather than a *Maintenance* dilemma, then this could explain why we observe more conditional cooperators and fewer free riders in the former than in the latter. We note, however, that while such shifts in advantageous inequity aversion might be empirically relevant, there is no psychological mechanism in the formal assumptions of Fehr and Schmidt (1999) that postulate such shifts.

Next, we consider theories of *reciprocity* (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006), in which agents' motivations derive from their material payoff as well as a psychological payoff that depends on their first-order beliefs about others' actions. These theories can also explain conditional cooperation (Dufwenberg et al., 2011) because agents want to reward kind actions with kindness and punish hostile actions with unkindness. Kindness is thereby assumed to be evaluated relative to a reference point that is the midpoint between the maximum and minimum possible payoff (see Online Appendix C, and Dufwenberg and Kirchsteiger, 2019 for a recent discussion). Because the payoff sets are the same in *Maintenance* and *Provision*, the reference point must be the same. Therefore, beliefs are the only channel through which simultaneous gameplay may differ in *Maintenance* and *Provision*. Since in our strategy-method experiment (i) first-order beliefs are fixed because participants condition their contributions on all possible average contributions of others, and (ii) as argued by Dufwenberg et al. (2011), in a linear public goods game the evaluation of others' kindness only depends on first-order beliefs, these models do not predict dilemma-dependent conditional cooperation. This prediction also hinges on the assumption that the concern for reciprocity as measured by a reciprocity parameter does not vary across dilemmas. If one would be willing to relax this assumption to allow that the reciprocity parameter is stronger under *Provision* than *Maintenance*, then reciprocity theory could also rationalize our finding that conditional cooperation is more frequent in *Provision* than *Maintenance*. Note, however, that also in this theory there is no psychological mechanism in the formal assumptions that postulates such a shift.

The next theory we discuss is *guilt aversion*. We rely on a model of guilt by Battigalli and Dufwenberg (2007) that assumes that 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. Applied to the context of a public goods game (see also Dufwenberg et al., 2011), 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 (that is, her second-order beliefs). Given that in our strategy-method experiment first-order beliefs are fixed, differences in cooperation attitudes could be reconciled via dilemma-dependent second-order beliefs: If more participants in *Provision* than in *Maintenance* would have second-order beliefs that others expect them to reciprocate their contributions, the perceived guilt from not matching others' contributions would be stronger, which, in turn, could lead to a higher fraction of conditional cooperators in *Provision* than in *Maintenance*. However, guilt aversion theory does not explicitly model any mechanism for why there should be any difference in second-order beliefs between *Maintenance* and *Provision*. Similarly, the guilt-sensitivity parameter is also assumed to be the same across dilemmas and there is no psychological mechanism in the formal assumptions that would suggest dilemma-specific guilt sensitivity.[22]

In sum, all standard theories of social preferences discussed so far can explain conditional cooperation. However, to be able to explain dilemma-specific conditional cooperation they all require some adjustments not present in the original formulations of these theories, such as modeling mechanisms of dilemma-dependent beliefs or preference parameters.

One theory that has the potential to predict a difference between *Maintenance* and *Provision* is *revealed altruism* by Cox et al. (2008). This theory is governed by two axioms, the reciprocity Axiom R, and Axiom S, which allows for status quo effects. Applying Cox et al.'s Axiom R to our public goods games, the model predicts that participants will perceive higher contributions by the other group members as more generous towards them, and, therefore, 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).

In 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. If one assumes that players perceive the original allocation to the public good *before decision making* (0 in *Provision* and

[21] A similar argument holds for models of altruism and fairness such as Levine (1998) and Cox et al. (2007).

[22] Evidence from two online surveys in which we elicited *ex post* feelings of guilt further reveal that participants do not feel more guilty when free riding on others' contributions in *Provision* than in *Maintenance*, indicating that *ex post* feelings of guilt cannot explain the observed differences in conditional cooperation across the two dilemmas (see Online Appendix D for further details).

80 in *Maintenance*) as the status quo (as in Cox et al., 2013), then, applying Axiom S to our strategy-method experiment, the theory predicts that second movers will be less altruistic towards first movers in *Maintenance* than in *Provision*. The reason is that in *Provision* any positive contribution by the other three group members increases 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. This asymmetry triggers a stronger preference for reciprocity in *Provision* than *Maintenance*, which can explain our finding of more conditional cooperators and fewer free riders in *Provision* than in *Maintenance* (see Online Appendix C for further details).[23]

6.2. Explaining dilemma-specific kindness perceptions

As a next step, we discuss which of the above theories can reconcile the robust and replicable evidence on different perceptions of kindness between *Maintenance* and *Provision* (Results 2 and 5). First, notice that models of distributional preferences are only based on payoff consequences and therefore do not incorporate any evaluation of others' actions. Guilt aversion is also mute with respect to others' kindness as players evaluate their *own* action with respect to the distance from their second-order belief but make no evaluation about *others*' actions. Reciprocity models such as those by Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) are natural candidates to explain our kindness results: these theories incorporate the perception of other's (un)kindness as a central element into their models and the positive and negative kindness evaluations we observe are consistent with typical modeling assumptions. In these models, kindness is evaluated with respect to a reference point that is only based on material payoffs. Because material payoffs are identical in *Maintenance* and *Provision*, for a given effective contribution perceived kindness is predicted to be the same across *Maintenance* and *Provision*.

Finally, we consider the revealed altruism model by Cox et al. (2008) that also incorporates kindness. However, like the reciprocity models, this model also does not predict any differences in kindness perceptions across the dilemmas. Instead, it postulates that the different status-quo allocation in *Maintenance* and *Provision* directly affect the reciprocity parameter (see Axiom S). If one would be willing to adjust the model by allowing Axiom S to also affect perceptions of kindness, then every effective contribution in *Provision* should be perceived as more generous than the corresponding effective contribution in *Maintenance*.[24] The results from our kindness survey reveal, however, that while this is indeed true for average effective contributions of 10 and 20, for effective contributions of 0 we find the opposite as contributing 0 in *Provision* is perceived as unkinder than withdrawing everything in *Maintenance*. Hence, although Cox et al. (2008) comes closest to explaining our results as it is the only theory that postulates an explicit mechanism for why one could expect stronger conditional cooperation in *Provision* than in *Maintenance*, some discrepancies between the theory and our data remain as we have seen in our results on kindness perceptions (Results 2 and 5).

7. Discussion and conclusion

In this paper, we provided a comprehensive behavioral analysis of two generic and incentive-equivalent social dilemmas of voluntary cooperation: providing and maintaining public goods. We first established a lower fraction of conditional cooperators (and higher fraction of free riders) in *Maintenance* than *Provision* (Result 1). We then focused on two fundamental dimensions: social preferences and (mis)perceptions of others' intentions and the game form. We reported two important asymmetries. First, regarding *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* (Result 2). Regarding perceptions of the game form, we found that misperceptions are game-specific: misunderstandings are more likely in *Provision* than *Maintenance* (Result 3). Second, even after controlling for misperceptions, we observe substantial and significantly different degrees of conditional cooperation in *Maintenance* and *Provision* (Result 4); perceptions of kindness remain unaffected by misperceptions (Result 5). Hence, conditional cooperation is not just mistaken cooperation (as argued, e.g., by Burton-Chellew et al., 2016), but a true preference that is less frequently found in *Maintenance* than in *Provision* dilemmas.

Our Result 4 somewhat differs from the results by Fosgaard et al. (2017). Like them, we find that the differences in conditional cooperation across dilemmas become smaller once accounting for misperceptions. Unlike us, in their case the differences become statistically insignificant while in ours they remain economically and statistically significant. One possible explanation for the different findings is that Fosgaard et al. (2017) asked their misperception questions, which are similar to our Measure 2, only at the end of the experiment rather than *before* the elicitation of cooperation preferences as we do. Another reason could be the different subject pools used across the two studies – our results are based on a UK student sample while theirs is based on a representative sample of the Danish population – which could also explain why

[23] Here we have assumed that, as in Cox et al. (2013), the initial resource allocation is taken as the status quo. We acknowledge, however, that different assumptions on the status quo could be plausible. For example, the status quo could be the wealth level of participants before entering the lab. In this case, the Cox et al. (2008) theory would predict no differences between *Maintenance* and *Provision*.

[24] Formally, Axiom S would need to be modified to allow the status quo allocation to affect not only the *MAT* (more altruistic than) partial ordering but also the *MGT* (more generous than) partial ordering (see Online Appendix C for further details).

the overall level of misperceptions is much higher in Fosgaard et al.: 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. Our finding that kindness perceptions remain different across dilemmas also after removing misperceiving subjects reinforces further a preference explanation for the difference between *Maintenance* and *Provision*.

Our Results 1 – 5 also suggest an important general lesson: the revealed preference approach, that is, using choices to infer social preferences and/or dilemma-specific effects, requires controlling for perceptions.[25] 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, behavioral investigations of public goods provision and common pool resource problems have largely been conducted in independent literatures, in particular regarding conditional cooperation, which was mostly studied in the context of linear public goods provision games (see, e.g., Chaudhuri, 2011; Fehr and Schurtenberger, 2018; and Thöni and Volk, 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 pressing challenges for mankind such as global warming and 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; Ramalingam et al., 2019) or incentives may instead be needed.

Data availability

The data and analysis code of this paper are available at https://osf.io/3jpgh/

Acknowledgments

This work was supported by the European Research Council [grant numbers ERC-AdG 295707 COOPERATION and ERC-AdG 101020453 PRINCIPLES] 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, Tim Cason, Gary Charness, Jim Cox, Robin Cubitt, Martin Dufwenberg, Urs Fischbacher, Maria Garcia-Vega, Werner Güth, Georg Kirchsteiger, Friederike Mengel, Charles Noussair, Elena Manzoni, Vjollca Sadiraj, Maroš Servátka, Chris Starmer, Robert Sugden, and referees and participants from various seminars and conferences for helpful comments.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2022.06.009.

References

Alpizar, F., Carlsson, F., Johansson-Stenman, O., 2008. Anonymity, reciprocity, and conformity: evidence from voluntary contributions to a national park in Costa Rica. J. Public Econ. 92, 1047–1060.
Andreoni, J., 1995a. Cooperation in public-goods experiments - kindness or confusion? Am. Econ. Rev. 85, 891–904.
Andreoni, J., 1995b. Warm glow versus cold prickle - the effects of positive and negative framing on cooperation in experiments. Q. J. Econ. 110, 1–21.
Apesteguía, J., Maier-Rigaud, F.P., 2006. The role of rivalry. Public goods versus common-pool resources. J. Confl. Resolut. 50, 646–663.
Arechar, A.A., Gächter, S., Molleman, L., 2018. Conducting interactive experiments online. Exp. Econ. 21, 99–131.
Bardsley, N., 2008. Altruism or artefact? A note on dictator game giving. Exp. Econ. 11, 122–133.
Bartling, B., Engl, F., Weber, R.A., 2015. Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. J. Econ. Sci. Assoc. 1, 72–85.
Bartling, B., Fehr, E., Özdemir, Y., 2021. Does market interaction erode moral values? Rev. Econ. Stat., 1–32.
Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Am. Econ. Rev. 97, 170–176.
Bayer, R.-C., Renner, E., Sausgruber, R., 2013. Confusion and learning in the voluntary contributions game. Exp. Econ. 16, 478–496.
Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9, 226–232.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300.

[25] 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 *ex post*.

Bicchieri, C., Dimant, E., Gächter, S., Nosenzo, D., 2022. Social proximity and the erosion of norm compliance. Games Econ. Behav. 132, 59–72.
Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. Am. Econ. Rev. 90, 166–193.
Burton-Chellew, M.N., El Mouden, C., West, S.A., 2016. Conditional cooperation and confusion in public-goods experiments. Proc. Natl. Acad. Sci. 113, 1291–1296.
Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. Science 351, 1433–1436.
Camerer, C.F., Dreber, A., Johannesson, M., 2019. Replication and other practices for improving scientific quality in experimental economics. In: Schram, A., Ule, A. (Eds.), Handbook of Research Methods and Applications in Experimental Economics. Edward Elgar Publishing, Cheltenham, pp. 83–102.
Cappelen, A.W., Nielsen, U.H., Sørensen, E. Ø, Tungodden, B., Tyran, J.-R., 2013. Give and take in dictator games. Econ. Lett. 118, 280–283.
Cartwright, E., 2016. A comment on framing effects in linear public good games. J. Econ. Sci. Assoc. 2, 73–84.
Cason, T.N., Plott, C.R., 2014. Misconceptions and game form recognition: challenges to theories of revealed preference and framing. J. Polit. Econ. 122, 1235–1270.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117, 817–869.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14, 47–83.
Cornes, R., Sandler, T., 1996. The Theory of Externalities, Public Goods and Club Goods. Cambridge University Press, Cambridge.
Cox, C.A., 2015. Decomposing the effects of negative framing in linear public goods games. Econ. Lett. 126, 63–65.
Cox, C.A., Stoddard, B., 2015. Framing and feedback in social dilemmas with partners and strangers. Games 6, 394–412.
Cox, J.C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games Econ. Behav. 59, 17–45.
Cox, J.C., Friedman, D., Sadiraj, V., 2008. Revealed altruism. Econometrica 76, 31–69.
Cox, J.C., Ostrom, E., Sadiraj, V., Walker, J.M., 2013. Provision versus appropriation in symmetric and asymmetric social dilemmas. South. Econ. J. 79, 496–512.
Cubitt, R., Drouvelis, M., Gächter, S., 2011a. Framing and free riding: emotional responses and punishment in social dilemma games. Exp. Econ. 14, 254–272.
Cubitt, R., Drouvelis, M., Gächter, S., Kabalin, R., 2011b. Moral judgments in social dilemmas: how bad is free riding? J. Public Econ. 95, 253–264.
Dal Bó, P., Fréchette, G.R., 2018. On the determinants of cooperation in infinitely repeated games: a survey. J. Econ. Lit. 56, 60–114.
Dawes, R.M., 1980. Social dilemmas. Annu. Rev. Psychol. 31, 169–193.
De Dreu, C.K.W., McCusker, C., 1997. Gain–loss frames and cooperation in two-person social dilemmas: a transformational analysis. J. Pers. Soc. Psychol. 72, 1093–1106.
Drazen, A., Dreber, A., Ozbay, E.Y., Snowberg, E., 2021. Journal-based replication of experiments: an application to “Being chosen to lead”. J. Public Econ. 202, 104482.
Dreber, A., Ellingsen, T., Johannesson, M., Rand, D., 2013. Do people care about social context? Framing effects in dictator games. Exp. Econ. 16, 349–371.
Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2011. The framing of games and the psychology of play. Games Econ. Behav. 73, 459–478.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47, 268–298.
Dufwenberg, M., Kirchsteiger, G., 2019. Modelling kindness. J. Econ. Behav. Organ. 167, 228–234.
Ellingsen, T., Johannesson, M., Mollerstrom, J., Munkhammar, S., 2012. Social framing effects: preferences or beliefs? Games Econ. Behav. 76, 117–130.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54, 293–315.
Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods, 1–17.
Fehr, E., Leibbrandt, A., 2011. A field study on cooperativeness and impatience in the Tragedy of the Commons. J. Public Econ. 95, 1144–1155.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114, 817–868.
Fehr, E., Schurtenberger, I., 2018. Normative foundations of human cooperation. Nat. Hum. Behav. 2, 458–468.
Fehr-Duda, H., Fehr, E., 2016. Game human nature: finding ways to adapt natural tendencies and nudge collective action is central to the well-being of future generations. Nature 530, 413–416.
Ferraro, P.J., Vossler, C.A., 2010. The source and significance of confusion in public goods experiments. B.E. J. Econ. Anal. Policy 10, 1–42.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for readymade economic experiments. Exp. Econ. 10, 171–178.
Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. Am. Econ. Rev. 100, 541–556.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett. 71, 397–404.
Fischbacher, U., Gächter, S., Quercia, S., 2012. The behavioral validity of the strategy method in public good experiments. J. Econ. Psychol. 33, 897–913.
Fosgaard, T.R., Hansen, L.G., Wengström, E., 2014. Understanding the nature of cooperation variability. J. Public Econ. 120, 134–143.
Fosgaard, T.R., Hansen, L.G., Wengström, E., 2017. Framing and misperception in public good experiments. Scand. J. Econ. 119, 435–456.
Frackenpohl, G., Hillenbrand, A., Kube, S., 2016. Leadership effectiveness and institutional frames. Exp. Econ. 19, 842–863.
Frey, B.S., Meier, S., 2004. Social comparisons and pro-social behavior. Testing ‘conditional cooperation’ in a field experiment. Am. Econ. Rev. 94, 1717–1722.
Fujimoto, H., Park, E.-S., 2010. Framing effects and gender differences in voluntary public goods provision experiments. J. Socio-Econ. 39, 455–457.
Gächter, S., 2007. Conditional cooperation: behavioral regularities from the lab and the field and their policy implications. In: Frey, B.S., Stutzer, A. (Eds.), Psychology and Economics: A Promising New Cross-Disciplinary Field (CESifo Seminar Series). The MIT Press, Cambridge, pp. 19–50.
Gächter, S., Herrmann, B., 2009. Reciprocity, culture, and human cooperation: previous insights and a new cross-cultural experiment. Philos. Trans. - R. Soc., Biol. Sci. 364, 791–806.
Gächter, S., Kölle, F., Quercia, S., 2017. Reciprocity and the tragedies of maintaining and providing the commons. Nat. Hum. Behav. 1, 650–656.
Goldin, J., Reck, D., 2020. Revealed-preference analysis with framing effects. J. Polit. Econ. 128, 2759–2795.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1, 114–125.
Hardin, G., 1968. The tragedy of the commons. Science 162, 1243–1248.
Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: conducting experiments in a real labor market. Exp. Econ. 14, 399–425.
Houser, D., Kurzban, R., 2002. Revisiting kindness and confusion in public goods experiments. Am. Econ. Rev. 92, 1062–1069.
Isler, O., Gächter, S., Maule, A.J., Starmer, C., 2021. Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics. Sci. Rep. 11, 13868.
Iturbe-Ormaetxe, I., Ponti, G., Tomás, J., Ubeda, L., 2011. Framing effects in public goods: Prospect Theory and experimental evidence. Games Econ. Behav. 72, 439–447.
Khadjavi, M., Lange, A., 2015. Doing good or doing harm: experimental evidence on giving and taking in public good games. Exp. Econ. 18, 432–441.
Kocher, M.G., Cherry, T., Kroll, S., Netzer, R.J., Sutter, M., 2008. Conditional cooperation on three continents. Econ. Lett. 101, 175–178.
Korenok, O., Millner, E.L., Razzolini, L., 2014. Taking, giving, and impure altruism in dictator games. Exp. Econ. 17, 488–500.
Koszegi, B., Rabin, M., 2008. Choices, situations, and happiness. J. Public Econ. 92, 1821–1832.
Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Roth, A.E., Kagel, J.H. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–181.
Levin, S.A., 2014. Public goods in relation to competition, cooperation, and spite. Proc. Natl. Acad. Sci. 111, 10838–10845.
Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dyn. 1, 593–622.
List, J.A., 2007. On the interpretation of giving in dictator games. J. Polit. Econ. 115, 482–493.

MacKay, D.J.C., Cramton, P., Ockenfels, A., Stoft, S., 2015. Price carbon — I will if you will. Nature 526, 315–316.
Maniadis, Z., Tufano, F., List, J.A., 2014. One swallow doesn't make a summer: new evidence on anchoring effects. Am. Econ. Rev. 104, 277–290.
Messer, K.D., Zarghamee, H., Kaiser, H.M., Schulze, W.D., 2007. New hope for the voluntary contributions mechanism: the effects of context. J. Public Econ. 91, 1783–1799.
Miettinen, T., Kosfeld, M., Fehr, E., Weibull, J., 2020. Revealed preferences in a sequential prisoners' dilemma: a horse-race between six utility functions. J. Econ. Behav. Organ. 173, 1–25.
Ostrom, E., 1990. Governing the Commons. The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge.
Ostrom, E., 2006. The value-added of laboratory experiments for the study of institutions and common-pool resources. J. Econ. Behav. Organ. 61, 149–163.
Park, E.-S., 2000. Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding. J. Econ. Behav. Organ. 43, 405–421.
Plott, C.R., Zeiler, K., 2005. The willingness to pay-willingness to accept gap, the "endowment effect", subject misconceptions, and experimental procedures for eliciting valuations. Am. Econ. Rev. 95, 530–545.
Poppe, M., 2005. The specificity of social dilemma situations. J. Econ. Psychol. 26, 431–441.
Rabin, M., 1993. Incorporating fairness into game-theory and economics. Am. Econ. Rev. 83, 1281–1302.
Ramalingam, A., Morales, A.J., Walker, J.M., 2019. Peer punishment of acts of omission versus acts of commission in give and take social dilemmas. J. Econ. Behav. Organ. 164, 133–147.
Rand, D.G., Nowak, M.A., 2013. Human cooperation. Trends Cogn. Sci. 17, 413–425.
Rustagi, D., Engel, S., Kosfeld, M., 2010. Conditional cooperation and costly monitoring explain success in forest commons management. Science 330, 961–965.
Sell, J., Son, Y., 1997. Comparing public goods and common pool resources: three experiments. Soc. Psychol. Q. 60, 118–137.
Selten, R., 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In: Sauermann, H. (Ed.), Beiträge zur Experimentellen Wirtschaftsforschung. J.C.B. Mohr (Paul Siebeck), Tübingen, pp. 136–168.
Snowberg, E., Yariv, L., 2021. Testing the waters: behavior across participant pools. Am. Econ. Rev. 111, 687–719.
Sonnemans, J., Schram, A., Offerman, T., 1998. Public good provision and public bad prevention: the effect of framing. J. Econ. Behav. Organ. 34, 143–161.
Spranca, M., Minsk, E., Baron, J., 1991. Omission and commission in judgment and choice. J. Exp. Soc. Psychol. 27, 76–105.
Thöni, C., Volk, S., 2018. Conditional cooperation: review and refinement. Econ. Lett. 171, 37–40.
van Dijk, E., Wilke, H., 1997. Is it mine or is it ours? Framing property rights and decision making in social dilemmas. Organ. Behav. Hum. Decis. Process. 71, 195–209.
Volk, S., Thöni, C., Ruigrok, W., 2012. Temporal stability and psychological foundations of cooperation preferences. J. Econ. Behav. Organ. 81, 664–676.
Wilson, B.J., 2012. Contra private fairness. Am. J. Econ. Sociol. 71, 407–435.