



Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

[Isler, Ozan](#), Gachter, Simon, Maule, A. John, & Starmer, Chris (2021)

Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics.

Scientific Reports, 11(1), Article number: 13868.

This file was downloaded from: <https://eprints.qut.edu.au/225944/>

© 2021 The Author(s)

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution 4.0

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

<https://doi.org/10.1038/s41598-021-93412-4>



OPEN

Contextualised strong reciprocity explains selfless cooperation despite selfish intuitions and weak social heuristics

Ozan Isler^{1,2}, Simon Gächter^{3,4,5}, A. John Maule⁶ & Chris Starmer³

Humans frequently cooperate for collective benefit, even in one-shot social dilemmas. This provides a challenge for theories of cooperation. Two views focus on intuitions but offer conflicting explanations. The Social Heuristics Hypothesis argues that people with selfish preferences rely on cooperative intuitions and predicts that deliberation reduces cooperation. The Self-Control Account emphasizes control over selfish intuitions and is consistent with strong reciprocity—a preference for conditional cooperation in one-shot dilemmas. Here, we reconcile these explanations with each other as well as with strong reciprocity. We study one-shot cooperation across two main dilemma contexts, provision and maintenance, and show that cooperation is higher in provision than maintenance. Using time-limit manipulations, we experimentally study the cognitive processes underlying this robust result. Supporting the Self-Control Account, people are intuitively selfish in maintenance, with deliberation increasing cooperation. In contrast, consistent with the Social Heuristics Hypothesis, deliberation tends to increase the likelihood of free-riding in provision. Contextual differences between maintenance and provision are observed across additional measures: reaction time patterns of cooperation; social dilemma understanding; perceptions of social appropriateness; beliefs about others' cooperation; and cooperation preferences. Despite these dilemma-specific asymmetries, we show that preferences, coupled with beliefs, successfully predict the high levels of cooperation in both maintenance and provision dilemmas. While the effects of intuitions are context-dependent and small, the widespread preference for strong reciprocity is the primary driver of one-shot cooperation. We advance the Contextualised Strong Reciprocity account as a unifying framework and consider its implications for research and policy.

Humans are an exceptionally cooperative species, willing to cooperate at personal cost in one-shot encounters with unrelated and even anonymous individuals¹. Understanding cooperation in one-shot situations poses an important theoretical challenge from an evolutionary theory point of view^{1–5} because none of the mechanisms known to support cooperation among self-interested people, such as repeated interactions and reputation, are available in these circumstances^{4,5}. An early explanation for this capacity was strong reciprocity^{1,6–9}, a widespread cooperative preference reflecting inequality aversion^{10,11}, concern for social efficiency¹², and in particular, a desire to respond in kind to the perceived intentions of others^{13,14}. Strong reciprocity differs from weak reciprocity, where cooperation can be rationalized by strategic thinking and selfish incentives⁶. In contrast, strong reciprocity amounts to a willingness to pay a personal cost for cooperating on the expectation that others do the same^{7,15}.

The Social Heuristic Hypothesis (SHH) proposed a cognitive process explanation for cooperation in one-shot public good games that does not rely on strong reciprocity but on the distinction between intuition and deliberation^{16–18}. Accordingly, self-interested people who benefit from cooperation in repeated interactions develop prosocial intuitions, which save cognitive effort at the risk of misapplying them in unfamiliar one-shot situations¹⁸. In the one-shot social dilemmas studied here, SHH therefore predicts that deliberation will tend to lower cooperation as people realise that cooperation does not pay¹⁸. However, the theoretical explanation of

¹School of Economics and Finance, Queensland University of Technology, Brisbane 4000, Australia. ²Centre for Behavioural Economics, Society and Technology, Brisbane 4000, Australia. ³School of Economics, University of Nottingham, Nottingham NG7 2RD, UK. ⁴CESifo, 81679 Munich, Germany. ⁵IZA, 53113 Bonn, Germany. ⁶Leeds University Business School, University of Leeds, Leeds LS6 1AN, UK. ✉email: ozan.isler@qut.edu.au; simon.gaechter@nottingham.ac.uk

this effect is contested^{19,20}, in particular because SHH conceives one-shot cooperation as a spill-over from weak reciprocity²¹ and thereby neglects the importance of strong reciprocity preferences for one-shot cooperation.

SHH has primarily been studied in provision dilemmas, where cooperation requires generosity to create or supplement a public good (e.g., donating blood, charitable giving, team work, collective action), and where spontaneous reactions tend to be cooperative^{16,17,22,23}, though the extent to which this is the case has been challenged by more recent research^{24–26}. However, cooperation often requires maintaining—that is, not exploiting—an already existing public good (e.g., clean environment, antibiotic efficacy, as well as civil society). Evidence from experiments comparing the provision and maintenance dilemmas suggests that ‘giving’ and ‘taking’ involve different psychological mechanisms^{27–29}. Specifically, maintenance does not require generosity but rather restraint²⁷ of selfish impulses to exploit the existing public good. Moreover, differences in the cognition of giving and taking can affect cooperation behaviour in provision and maintenance dilemmas through their influence on reciprocal motivations²⁸. Cooperativeness has been shown to be weaker in maintenance than in provision dilemmas³⁰. This raises the possibility that different cognitive processes, consistent with the Self-Control Account (SCA)^{26,31–36}, underlie behaviour in the maintenance dilemma. According to SCA³¹ as applied to cooperation in social dilemmas^{33,34}, people experience a selfish impulse to exploit the public good but can overcome this through the exercise of willpower to achieve cooperation. SCA therefore predicts that deliberation will increase cooperation by promoting resistance to selfish intuitions and enabling behaviour according to strong reciprocity preferences.

In this paper, we provide a comprehensive account of one-shot cooperation by combining the viewpoints of SHH and SCA that rely on the distinction between intuition and deliberation, with the preference-based theory of strong reciprocity. We use a general working definition of intuition that includes not only socially acquired heuristics (as in SHH) but also spontaneous emotional reactions (as in SCA). This is because the two cognitive process accounts posit different types of intuitions: while SHH emphasizes past social interactions in generating cooperative heuristics, SCA focuses on the role of visceral impulses for self-protection. In contrast, the strong reciprocity perspective does not consider the role of intuitions but assumes social preferences: it posits that cooperation reflects a motivation to reciprocate the cooperation expected from others. In addition to experimental manipulations involving time-limits, we provide analyses based on reaction times, social dilemma understanding, perceptions of social appropriateness, and beliefs about and preferences for cooperation.

Our formalization of the strong reciprocity explanation of one-shot cooperation predicts that one-shot cooperation is explained not only directionally but also quantitatively by a combination of preferences and beliefs. Incorporating SHH and SCH, our framework allows for the possibility that context-dependent (i.e., dilemma-specific) intuitions and perceptions affect preferences and beliefs. Consequently, we propose the Contextualised Strong Reciprocity (CSR) framework, which we support by experimental evidence across varying decision-making contexts. To achieve comparability across these contexts, we abstract from a variety of naturally occurring institutional and technological features^{37–46}, and study intuitions and strong reciprocity in two fundamental types of social dilemmas: the provision and maintenance of public goods.

Results

Design. To experimentally compare maintenance and provision dilemmas, we used two formally equivalent and incentivized^{47,48} one-shot public goods games played in groups of four ($N = 3653$). In provision (P), each group member has a monetary endowment of 10 tokens which they can keep for themselves or invest in the public good. In maintenance (M), 40 tokens are already invested in the public good and each group member can withdraw up to 10 tokens for themselves. Selfish incentives are to contribute nothing in P or to withdraw 10 in M, but collective benefits are maximized by contributing 10 in P or by withdrawing nothing in M. We use incentivized²³ between-subjects time-pressure (TP) and time-delay (TD) manipulations that prompt participants to respond within or after a particular time-limit to induce, respectively, relatively more intuitive or deliberated decisions. In doing so, we follow the original¹⁶ and the majority of studies on SHH^{24,49–51}, which eases comparisons with the literature. We ran two pre-registered studies, which differed in the strength of time-pressure applied on cooperation decisions. As part of these studies, we also elicited strong reciprocity measures (expectations and preferences for conditional cooperation) as well as associated cognitive process measures (response times, social dilemma understanding, perceptions of social appropriateness and cognitive reflection) to investigate cooperation in the two social dilemmas.

Are intuitions selfish or cooperative? *Hypotheses on SHH and SCA.* Predictions of SHH and SCA diverge regarding the role of deliberation because SHH assumes that people tend to have cooperative intuitions but selfish preferences while SCA is consistent with a model of selfish intuitions and cooperative preferences. Our first objective is to test the predictions of SHH in P and M dilemmas and compare these with alternative predictions based on SCA. Drawing on previous literature^{30,52–55}, we conjectured that contributions to the public good (labelled C: contributions in P or what is not withdrawn in M) will be higher in P than in M ($H_1: C_P > C_M$). SHH predicts¹⁶ that intuitive decisions will be more cooperative than deliberated decisions ($H_2: C_{TP} > C_{TD}$). Assuming that people have selfish preferences, SHH further predicts¹⁷ that deliberation will lower cooperation when intuitions are cooperative (as we assume in P) more so than when intuitions are selfish (as we assume in M). We therefore conjectured that differences in contributions between P and M (ΔC) will be driven by intuitive decisions and will dissipate with deliberation ($H_3: \Delta C_{TP} > \Delta C_{TD}$).

As an alternative hypothesis (H_A) to H_2 , and in contrast to SHH, SCA predicts deliberation to increase cooperation ($H_A: C_{TP} < C_{TD}$). Self-control should be particularly important in M^{56,57}, where intuitions are expected to be more selfish; hence SCA also predicts H_3 . However, SCA suggests a different cognitive process, one where deliberation increases cooperation by promoting self-control.

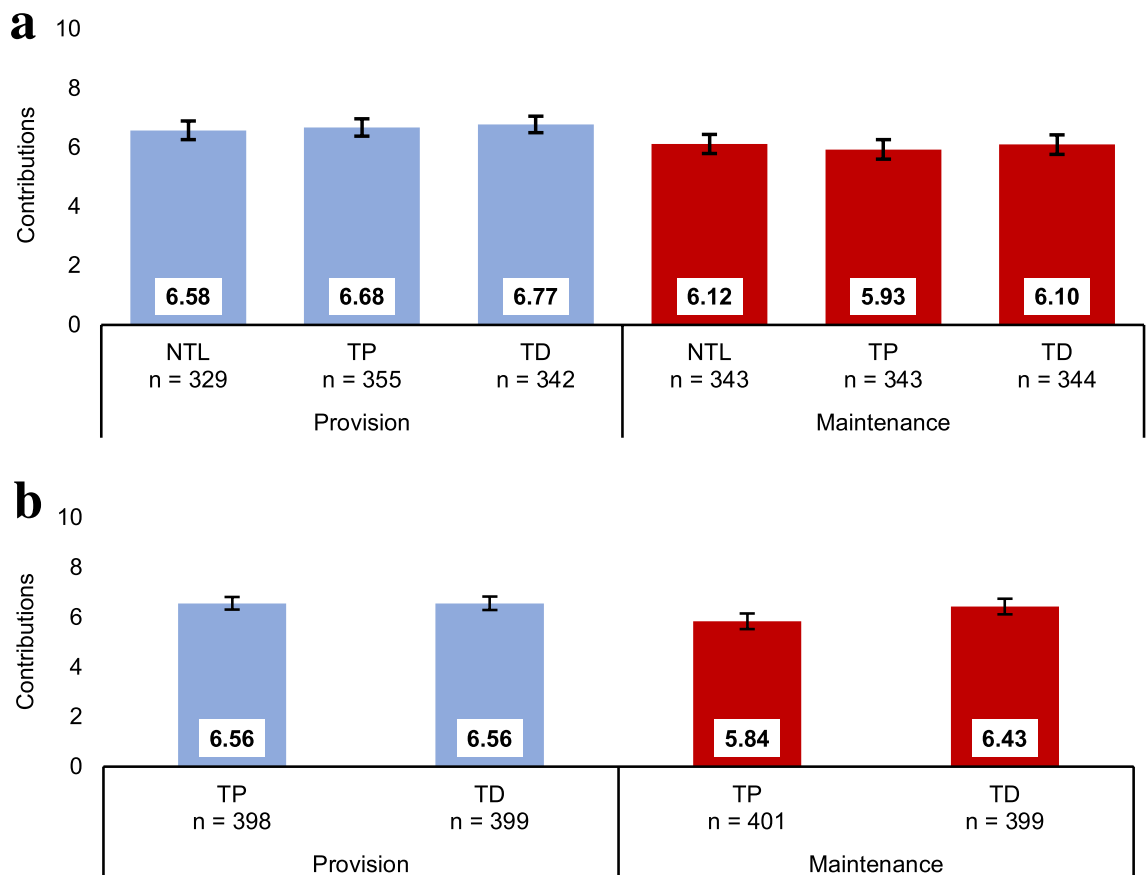


Figure 1. Cooperation in provision and maintenance dilemmas by time-limit conditions. Average cooperation (i.e., tokens contributed to or left in the public good, out of an endowment of 10) for provision (P) and maintenance (M) dilemmas. Numbers on bars are levels of cooperation. (a) Study 1: no time-limit (NTL), 10 s time-pressure (TP) and 10 s time-delay (TD) conditions. (b) Study 2: 5 s time-pressure (TP) and 10 s time-delay (TD) conditions. Error bars show 95% confidence intervals.

Study 1 found higher cooperation in P than M but supported neither SHH nor SCA. Study 1 included a no time-limit benchmark (NTL) alongside two time-limit conditions (TL), 10 s TP and 10 s TD. Compliance with time-limits was high in both TD (88.9%) and TP (93.6%). Indicating a successful experimental manipulation of behaviour, response times (RT) were faster in TP (mean, $M=6.15$ s) than in TD ($M=28.17$ s), $t(1382)=15.59$, $P<0.001$, $d=0.84$.

We found robust evidence for the effect of dilemma type on cooperation in the NTL and the TL samples (H_1 : $C_P > C_M$) but found neither a time-limit nor an interaction effect (Fig. 1a). In NTL, the average tokens contributed was 4.6 percentage points (pp) higher in P ($M=6.58$) than in M ($M=6.12$), $t(670)=1.99$, $P=0.048$, $d=0.15$. In TL, a two-way ANOVA showed 7.1 pp difference between P ($M=6.73$) and M ($M=6.01$), $F(1, 1380)=20.34$, $P<0.001$, $\eta_p^2=0.015$, but no difference in cooperation between TP ($M=6.31$) and TD ($M=6.43$) (H_2), $F(1, 1380)=0.68$, $P=0.408$, $\eta_p^2<0.001$; we also found no interaction effect (H_3), $F(1, 1380)=0.05$, $P=0.825$, $\eta_p^2<0.001$.

Survey questions on the extent to which decisions were based on deliberated vs. intuitive processes suggest that the 10 s time-limit was not short enough to disable deliberation relative to NTL (see Manipulation Checks in "Methods"). We therefore ran a second study using a stronger time-limit manipulation, with 5 s TP and 10 s TD.

Study 2 supported SCA—intuitions are selfish and more so in M than in P. In Study 2, manipulation checks supported the interpretation that decisions were made more intuitively in TP than in TD. The 5 s TP lowered RTs ($M=5.17$ s) as compared to the 10 s TP condition in Study 1 ($M=6.15$ s), $t(1495)=3.00$, $P=0.003$, $d=0.16$. RTs were faster in TP ($M=5.17$ s) than in TD ($M=28.45$ s), $t(1595)=23.91$, $P<0.001$, $d=1.20$ (see "Methods" for additional checks).

A two-way ANOVA indicated significant main effects of dilemma type and time-limits on cooperation (Fig. 1b). Supporting H_1 ($C_P > C_M$) and consistent with Study 1, contributions were higher in P ($M=6.56$) than in M ($M=6.14$), $F(1, 1593)=8.41$, $P=0.004$, $\eta_p^2=0.005$. However, the direction of the effect of time-limits on contributions ($M_{TP}=6.20$ vs. $M_{TD}=6.50$) was the opposite to that predicted by SHH (H_2 : $C_{TP} > C_{TD}$) and in line with SCA's prediction of selfish intuitions and increased cooperation with deliberation (H_3 : $C_{TP} < C_{TD}$), $F(1, 1593)=4.09$, $P=0.043$, $\eta_p^2=0.003$.

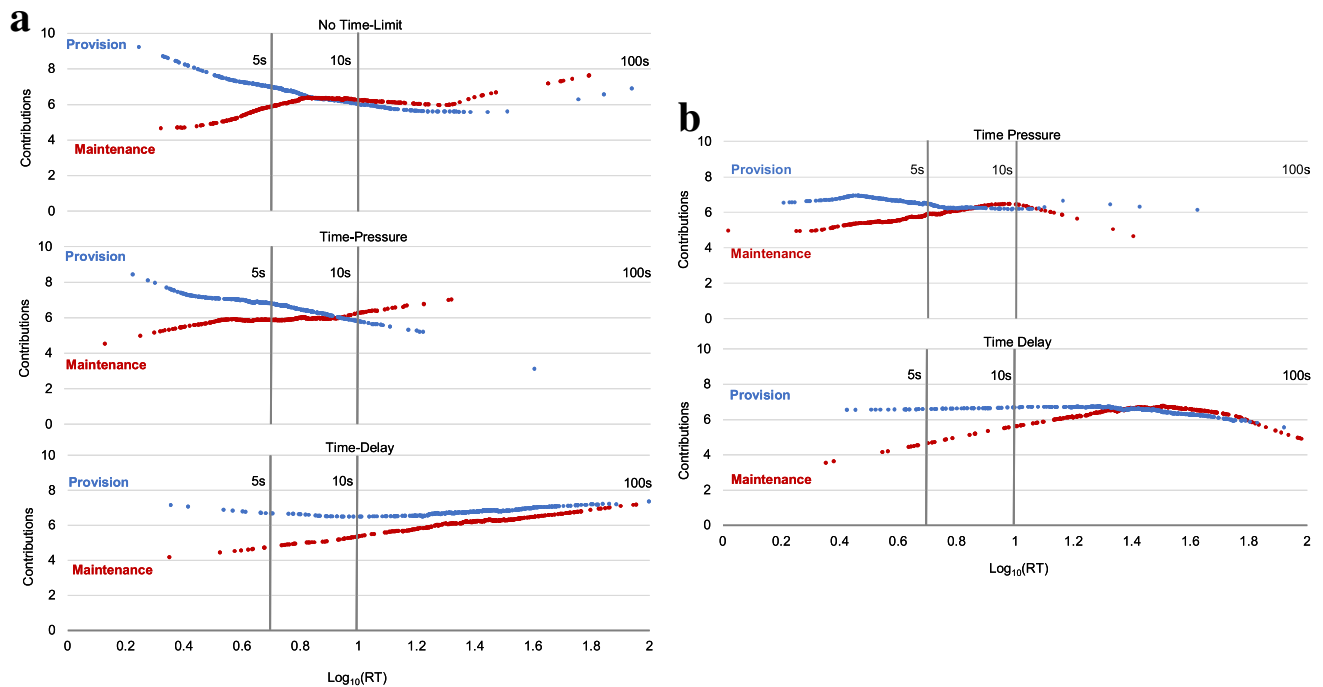


Figure 2. Cooperation by response times. **(a)** Study 1: Plotted are cooperation and response times (RT) in NTL, TP and TD. **(b)** Study 2: Plotted are cooperation and RT in TP and TD. The graphs show LOWESS estimates, representing the relationship between tokens contributed to the public good and RTs, shown for each social dilemma. Each series is composed of OLS estimates in the locality of each RT observation (bandwidth 0.8). Estimates for three responses (0.2%) that took longer than 100 s are not shown. Corresponding linear model estimates are in Tables S1 and S2.

The interaction between time-limit and dilemma type was also significant, indicating that intuitions are context-dependent. As predicted (H_3 : $\Delta C_{TP} > \Delta C_{TD}$), the difference in contributions between the two dilemmas was higher under TP ($\Delta C_{TP} = 7.2$ pp) than under TD ($\Delta C_{TD} = 1.3$ pp), $F(1, 1593) = 3.99$, $P = 0.046$, $\eta_p^2 = 0.002$. This was because average contributions in M were significantly lower under TP than under TD, $t(798) = 2.63$, $P = 0.009$, $d = 0.19$, whereas no effect could be identified in P, $t(795) = 0.02$, $P = 0.985$, $d < 0.01$.

Study 2 showed limited evidence for SHH in P. Finally, we compared the effect of time-limits on the prevalence of zero contributions in P and complete withdrawal in M because some interpretations of SHH make predictions about the likelihood of free-riding behavior rather than the extent of cooperation¹⁸. While this exploratory analysis provided no further insights to the null results in Study 1 (the prevalence of zero contributions in P being 3.1% in TP vs. 2.0% in TD and in M being 10.5% in TP vs. 10.5% in TD), we found support in Study 2 for SHH in P (prevalence of zero contributions being 0.8% in TP vs. 4.8% in TD) but not in M (11.2% in TP vs. 8.3% in TD). Specifically, a logit model of free-riding behaviour in Study 2, $\chi^2(3, n = 1597) = 50.37$, $P < 0.001$, indicated a significant interaction between dilemma type and time-limits ($P < 0.001$) such that time delay increased the likelihood of zero contributions in P ($OR = 6.58$, $P = 0.003$) but not in M ($OR = 0.71$, $P = 0.161$).

In short, confirmatory tests supported SCA's prediction that time-delay increases cooperation, an effect that was stronger in M than in P. In contrast, we only found limited, exploratory evidence based on measures of free-riding for SHH's predictions in P. Overall, the effect of intuitions on contributions depended on the social dilemma type but tended to be small ($d < 0.20$). These findings suggest that cognitive processes driving cooperation differ across the P and M dilemmas. Next, we explore this difference further by investigating the relationship between cooperation decisions and response times, which are considered indicative of the extent of deliberation underpinning a decision^{58,59}.

Fast decisions are more selfish in M than in P. We log-transform RTs (to base 10) to account for data skewness¹⁶ and depict the relationship between RT and cooperation across the time-limit conditions using local nonparametric estimates of contributions over time. Visual comparison of the two dilemmas in both Study 1 (Fig. 2a) and Study 2 (Fig. 2b) suggests that fast decisions, when compared to slow decisions, tend to be more prosocial in P and more selfish in M. Comparing the two dilemmas, contributions seem to be more prosocial in P and more selfish in M especially among fast decisions. The corresponding linear model estimates verify these visual trends (Tables S1 & S2). Considering the two studies together, contributions in P were higher than in M by 13.5 pp for decisions made within the first 5 s, $t(1017) = 7.54$, $P < 0.001$, $d = 0.47$; by 3.8 pp for decisions made in 5 s to 10 s, $t(1048) = 2.01$, $P = 0.044$, $d = 0.12$; and by 1.4 pp for decisions longer than 10 s, $t(1582) = 0.92$, $P = 0.357$, $d = 0.05$.

Two alternative interpretations have been offered for the RT analysis of cooperation decisions: the earlier argument that fast RTs indicate intuition^{16,58,59} and the more recent view that they reflect lack of conflict during decision-making due to strong preference for one option over others⁶⁰. While it may be difficult to disentangle these two influences in a correlational analysis^{60,61}, our RT analysis is consistent with the experimental findings in Study 2 of intuitive selfishness in M and modest deliberated free-riding in P. Moreover, we do not find clear evidence for decision conflict accounts predicting that faster decisions are more extreme (contributing nothing or everything)⁶² or predicting that TP increases random decision error^{61,63,64} or arguing for congruency of choices with preferences^{65,66} (see Supplementary Materials for details, including a revised decision conflict account model consistent with our data).

Whether driven by decision conflict or dual-process mechanisms⁶⁷, the differences in RTs suggest systematic differences in the cognitive processes underlying behaviour in the two dilemmas. Next, we present converging exploratory evidence that supports this view, showing dilemma-dependent differences in (1) understanding of social dilemmas, (2) perceptions of social appropriateness, and (3) reliance on deliberative rather than intuitive thinking.

Intuitions are consistent with other cognitive processes underlying cooperation. *M promotes self-gain understanding, P promotes group-gain understanding.* An important question is how people understand the P and M dilemmas^{64,68–70}. From a standard game-theoretic point of view, M and P are strategically identical: contributing nothing and withdrawing everything maximize self-gain, contributing everything and withdrawing nothing maximize group welfare. We measured participants' correct understanding of these strategies using two incentivized questions (Methods).

In both studies, the correct understanding of the self-gain maximizing strategy was higher in M (61.8% overall) than in P (44.2%), whereas understanding of the group-gain strategy was higher in P (74.7%) than in M (60.0%) (Figs. 3a, S1); χ^2 -tests, P s < 0.001. Across the two dilemmas, those who correctly understood the self-gain maximizing strategy tended to be less cooperative ($M = 6.13$) than those who misunderstood it ($M = 6.61$), $t(3651) = 4.96$, $P < 0.001$, $d = 0.16$, while those who correctly understood the group-gain maximizing strategy tended to be more cooperative ($M = 6.95$) than those who misunderstood it ($M = 5.14$), $t(3651) = 18.02$, $P < 0.001$, $d = 0.64$.

These contrasting effects of the two components of social dilemma understanding tended to cancel each other when we used—like related literature¹⁶—a composite understanding variable that equated one if both self- and group-gain questions were correctly answered and zero otherwise. Specifically, consistent with our preregistered two-way ANOVA model of Study 2, the effects of social dilemma type ($P = 0.003$), time-limits ($P = 0.049$) and their interaction ($P = 0.046$) were significant in an exploratory model that controlled for social dilemma understanding. In a less conventional second exploratory model that took correct answers to the two understanding questions separately, neither social dilemma type ($P = 0.326$) nor time-limits ($P = 0.171$) were significant. Nevertheless, consistent with the confirmatory evidence found for SCA in Study 2, the increase in contributions in TD over TP in M was similar for those with (5.1 pp) and without understanding (5.9 pp).

Freeriding is more socially appropriate in M than P. Perceptions of social appropriateness can guide cooperation decisions^{71,72}. Employing a standard approach that involves an incentivized coordination game⁷¹, participants in Study 1 estimated how socially appropriate other participants perceived various levels of contribution in P or withdrawal in M (Methods). Because our public good games were anonymous and one-shot, these perceptions are independent of the characteristics and actions of one's group members. Cooperation was negatively correlated with the perceived appropriateness of contributing nothing ($r_s = -0.25$, $P < 0.001$) or half of the endowment ($r_s = -0.35$, $P < 0.001$) and positively correlated with the perceived appropriateness of maximal contributions ($r_s = 0.28$, $P < 0.001$). Zero-contribution was perceived to be less socially appropriate in P than in M (Fig. 3b), $t(2054) = 11.09$, $P < 0.001$, $d = 0.49$. In contrast, half-contribution was perceived to be more socially appropriate in P than in M; $t(2054) = 3.77$, $P < 0.001$, $d = 0.17$. Likewise, social appropriateness of maximal-contribution was higher in P than in M, $t(2054) = 4.78$, $P < 0.001$, $d = 0.21$.

M increases reliance on intuitions more than P. The Cognitive Reflection Test⁷³ (CRT) was elicited at the end of Study 2 to test whether treatment effects depended on individual thinking styles (Methods). As expected, there was neither an effect of time limits on CRT scores—the total number of correct answers—nor an interaction between time-limits and dilemma types. However, CRT scores were significantly higher in P than in M (Fig. 3c), $F(1, 1593) = 7.32$, $P = 0.007$, $\eta_p^2 = 0.005$.

The dilemma-specific difference in CRT scores is surprising because the allocation of participants to P and M was random. The probability that this asymmetry was due to sampling bias is low (< 5%) as confirmed by simulations that randomly assigned each observed CRT score to one of two hypothetical experimental conditions and tested for difference in average scores (see Supplementary Materials). We also tested for sampling bias by re-inviting all Study 2 participants one-and-a-half years later to take part in a supplementary study that measured their CRT scores for a second time; as planned, we stopped data collection when we reached 50% of the original sample ($n = 800$). Among these re-invited participants, the initial CRT score was 9.7% higher for those who had experienced P than those who were in M, $\chi^2(3) = 9.95$, $P = 0.019$. In contrast, when we compared their second CRT scores, we found no difference, $\chi^2(3) = 2.71$, $P = 0.439$.

An alternative scoring of the CRT is to measure the total number of intuitive but incorrect answers (iCRT⁷⁴; Methods). Consistent with standard scoring, exposure to M resulted in 7.6% higher iCRT than P, $\chi^2(3) = 8.67$, $P = 0.034$, yet this difference was no longer evident the second time iCRT was measured, $\chi^2(3) = 1.42$, $P = 0.701$. These exploratory findings, which require replication in future research, suggest that dilemma type may influence

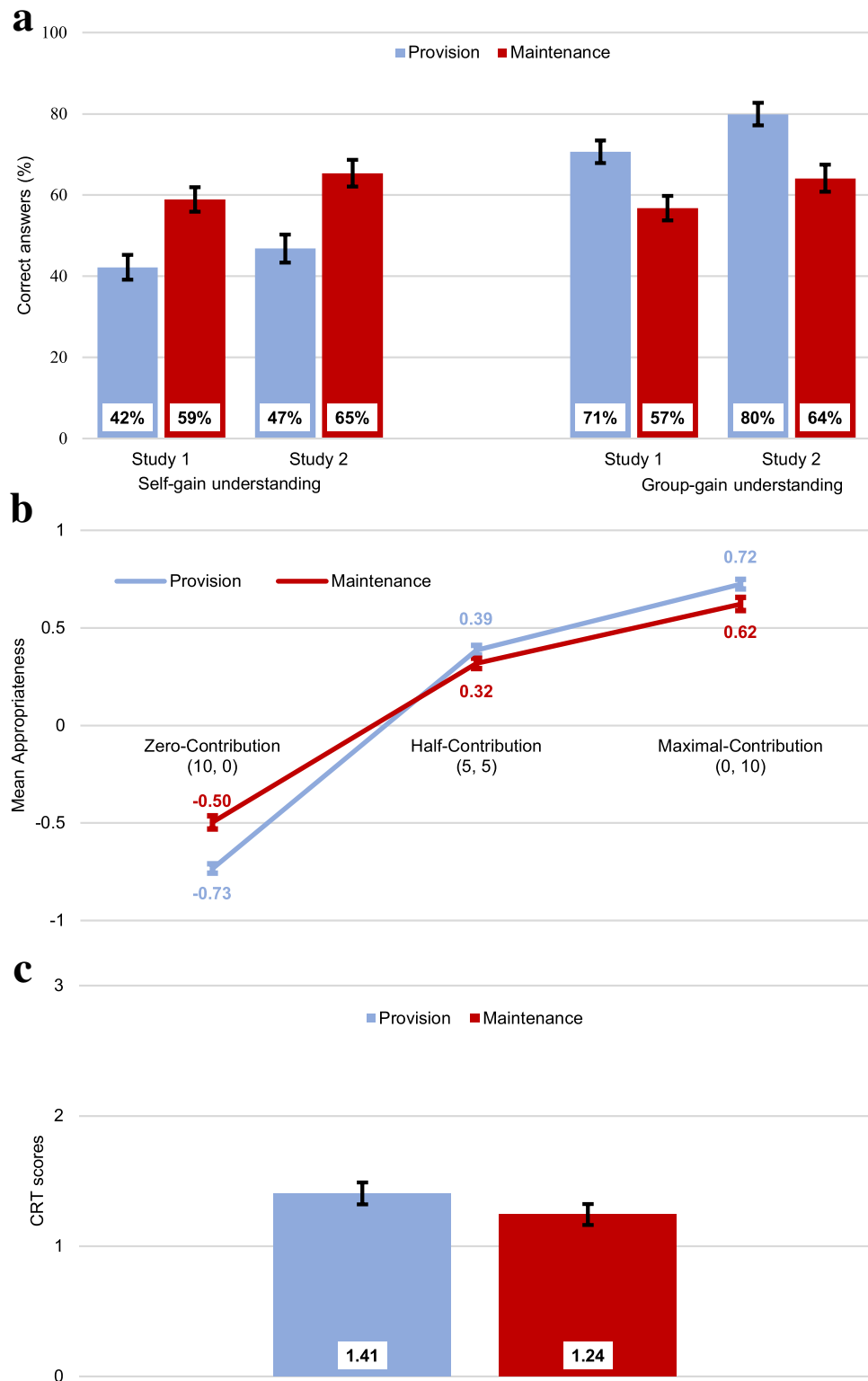


Figure 3. Dilemma-specific differences in cognitive processes. (a) Percentage of people who correctly answered the two questions on understanding about the self-gain maximization strategy and the group-gain maximization strategy. (b) Perceptions of social appropriateness of three decision scenarios: contributions of 0, 5 and 10 (or withdrawals or 10, 5 and 0), corresponding to “zero-contribution”, “half-contribution” and “maximal-contribution” (only elicited in Study 1). c, Number of correct answers on the Cognitive Reflection Test (only elicited in Study 2). Error bars show 95% confidence intervals. See “Methods” for detailed definition of measures and Supplementary Materials for additional analyses.

thinking styles. In particular, exposure to M may decrease reliance on reflection possibly by triggering stronger and more enduring intuitive reactions as compared to P.

These measures provide convergent evidence of systematic differences in the cognitive processes underpinning P and M dilemmas that are consistent with the differences in intuitions observed earlier. Overall, cooperation was higher in P than in M in all five time-manipulation conditions across the two studies (Fig. 1). Nevertheless, in contrast to SHH's assumption that preferences are selfish and its prediction that deliberation erodes cooperation, we found substantial levels of cooperation in the time-delay conditions across the two studies (67% of the endowment was contributed in P and 63% in M). We next test whether these high levels of cooperation can be explained by strong reciprocity^{1,6}.

Strong reciprocity explains one-shot cooperation. *Measures of strong reciprocity.* The key evidence for strong reciprocity is the preference for conditional cooperation despite net personal costs in anonymous one-shot social dilemmas^{9,15}. We elicited, in both studies, two incentivized measures of conditional cooperation—expectations and preferences⁷⁵. First, expectations are point estimates of the average expected contribution level of other group members (Methods). Expectations (here also referred to as expected cooperation) can be indicative of strong reciprocity motives in anonymous one-shot games because they are independent of how other participants actually behave^{75–80}. Second, individual preferences were measured by eliciting contribution schedules (i.e., a public good contribution decision for each possible average contribution level, from 0 to 10 tokens, of others in their group)⁷⁵. Based on pre-registered criteria about the relationship between own and other contributions, we classified each participant as conditional cooperator (strong positive correlation), free rider (zero contributions regardless of others' contributions) or other⁷⁵. Our protocol also provides a predicted contribution for each individual, found by combining their contribution schedule with their expectations (Methods).

Hypotheses on strong reciprocity. We test three preregistered hypotheses on dilemma-specific differences in strong reciprocity based on previous evidence³⁰: more people will be conditional cooperators (CC) in P than M (H_4 : $CC_P > CC_M$); expectations (E) will be higher in P than M (H_5 : $E_P > E_M$); and predicted contributions (PC) will be higher in P than in M (H_6 : $PC_P > PC_M$). Next, we present results of tests that combine data from the two studies (see "Methods" for additional hypotheses and Supplementary Materials for detailed results).

Strong reciprocity is weaker in M than in P. As hypothesized (H_4), the frequency of conditional cooperators was significantly higher in P (63%) than in M (52%), $\chi^2(1, n = 3653) = 38.34, P < 0.001$, whereas preference for free-riding, though surprisingly rare overall⁸¹, was more prevalent in M (4%) than in P (2%), $\chi^2(1, n = 3653) = 16.08, P < 0.001$ (Fig. 4a). Across both studies, and consistent with H_5 , expectations about others' contributions were significantly higher in P ($M = 5.96$) than in M ($M = 5.05$), $t(3651) = 11.03, P < 0.001, d = 0.36$ (see Figs. 4b and S2). Although weaker, predicted contributions were also on average higher in P ($M = 5.76$) than M ($M = 5.56$), $t(3651) = 1.95, P = 0.051, d = 0.06$ (H_6). Expectations were elicited using the same time-limits that participants faced during contribution decisions. However, we found no effect of time-limits on expectations or predicted contributions. The lack of effect of time-limits on strong reciprocity measures is consistent with the view that motivation for strong reciprocity is equally high for intuitive and deliberated decisions. Next, we explore whether the dilemma-specific differences in strong reciprocity are related to how people cognitively process these two social dilemmas.

Understanding, social appropriateness, and reflection influence strong reciprocity. Linear regressions across the two studies (see Table S4) indicated that expectations regarding group members' contributions depended on social dilemma understanding (increasing with group-gain and decreasing with self-gain understanding) as well as on perceptions of social appropriateness (increasing with the perceived appropriateness of maximal contributions and inappropriateness of zero- and half-contributions; all P s < 0.001). In addition, logit models of cooperation preferences showed that the likelihood of being categorized as a conditional cooperator increased with social dilemma understanding, with the tendency for reflective thinking and with perceived appropriateness of cooperation. Hence, these measures of cognitive processes were consistently associated with measures that indicate motivation for strong reciprocity.

Strong reciprocity explains cooperation in both dilemmas. Irrespective of the dilemma-specific differences (Figs. 1, 2, 3), cooperation was high ($> 60\%$) in both dilemmas ($M_P = 6.63$ and $M_M = 6.09$). These high levels of cooperation need an explanation that SHH cannot provide. SHH predicts lower cooperation in TD than TP because deliberation should increase the saliency of free-riding. In fact, cooperation tended to increase slightly with time-delay across the two studies ($M_{TP} = 6.25$ vs $M_{TD} = 6.47$), $t(2979) = 2.01, P = 0.045, d = 0.07$. SHH also implies that cooperation should decrease with social dilemma understanding¹⁸, experience with one-shot social dilemma games⁸², and the Cognitive Reflection Test score¹⁷. However, all three variables showed either significantly positive or insignificant correlations with cooperation (Table S3).

The high cooperation levels in our data can instead be explained by strong reciprocity. Specifically, actual contributions were highly correlated with expectations ($r_s = 0.65, P < 0.001$) and predicted contributions ($r_s = 0.54, P < 0.001$) (Fig. 5). Expectations and predicted contributions remained strongly positively correlated with contributions when we controlled for the covariates of social dilemma understanding, experience and CRT score (Table S3). The correlations between expected and actual contributions were strongly positive for those with ($r_s = 0.63$) and without ($r_s = 0.66$) correct social dilemma understanding (Figs. S4, S5) and not only for conditional cooperators ($r_s = 0.65$) but also for free riders ($r_s = 0.68$) and those classified as "others" ($r_s = 0.62$; all P s < 0.001) (Fig. S6).

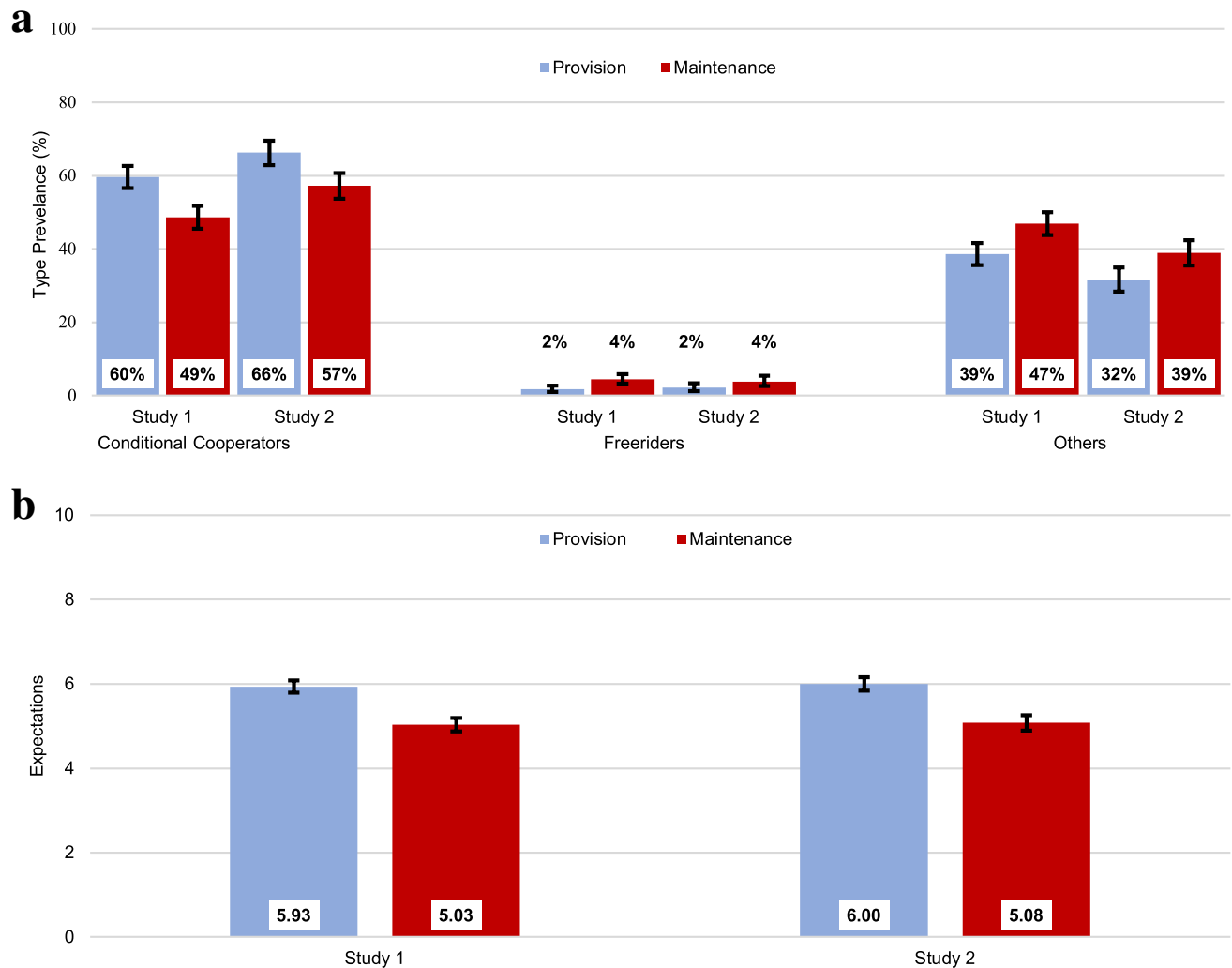


Figure 4. Strong reciprocity measures. (a) Distribution of social preference types: conditional cooperators (left panel), free riders (middle panel) and other social preference types (right panel). (b) Expectations about the average contribution made by others in the group. Error bars show 95% confidence intervals. See "Methods" for definition of expectation measures and preference type categorization.

Strong reciprocity explains both aggregate and individual cooperation. At the aggregate level, predicted contributions accounted for 87% of actual contributions in P and 91% of actual contributions in M. At the individual level, as preregistered, we categorized those with predicted contributions within 10% (i.e., ± 1 token) of their actual contributions as exhibiting predictive accuracy⁸³ (Methods). In both studies, 57% of participants showed predictive accuracy at the 10% level, and only 6% of all participants showed predictive inaccuracy by more than 50% (i.e., ± 6 or more tokens) (Fig. 6). The consistency of predicted contributions did not depend on dilemma type, time-limit manipulations or their interaction, since the corresponding logit models of predictive accuracy were overall insignificant in both Study 1 ($P=0.479$) and Study 2 ($P=0.452$).

Contextualised Strong Reciprocity (CSR) account of one-shot cooperation. Our results indicate contextualised (i.e., dilemma-dependent) strong reciprocity preferences explain anonymous one-shot cooperation. Our account has three key components. First, the context-specific features of the social dilemma (e.g., dilemma type) systematically shape intuitions as reflected in associated cognitive process measures (Figs. 2, 3). Second, expectations of others' cooperation and preferences for strong reciprocity are also influenced by the decision context (Fig. 4). Third, the prevalence of motivations for strong reciprocity explains the high levels of cooperation observed in one-shot anonymous social dilemmas (Figs. 5, 6).

We combine these components in the CSR account, a unified framework for understanding why and how people cooperate in one-shot social dilemmas (Fig. 7). Formally put, CSR explains cooperation as $a_i(f, b_i(f)) \rightarrow c_i$, where an individual i 's contribution level (c_i) is a function of i 's preference for conditional cooperation (a_i) and i 's expectation about other's cooperation (b_i). Both a_i and b_i are functions of the contextual features (f) of a given social dilemma (see Fig. 4). CSR posits that $a_i(f, b_i(f))|f$ predicts c_i equally well for all f —that is the prediction error, $c_i - c_i^*$, is equally distributed across f —a proposition that is supported by our data (Fig. 6). As discussed

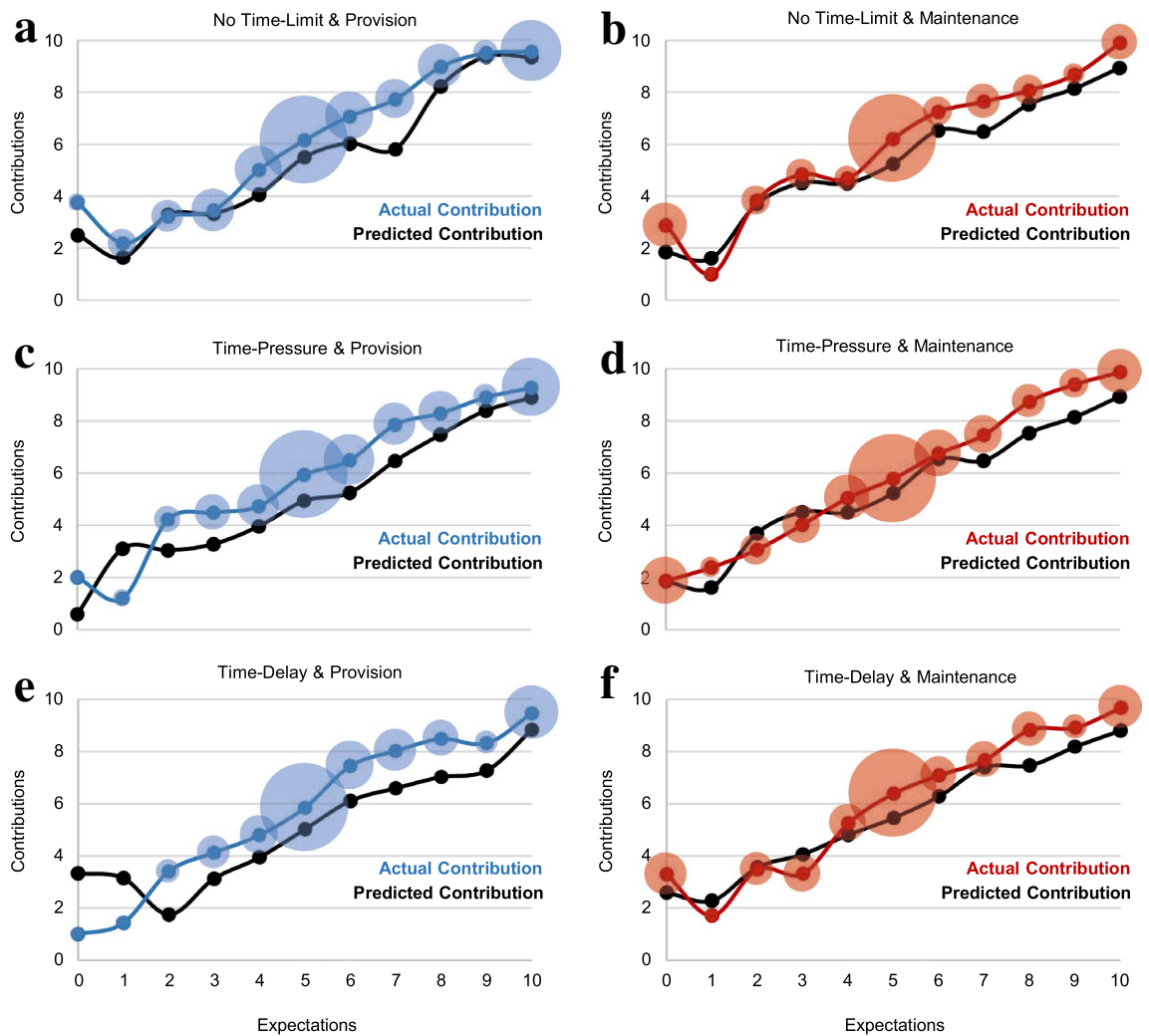


Figure 5. Actual and predicted contributions by expectations across experimental conditions. Coloured lines show average actual contribution at each expectation level reported by participants. The size of the bubble indicates the number of participants at that particular expectation level. The black lines denote average predicted contribution found by combining expectation with the deliberated contribution preference schedule (Methods). (a, b) Study 1 only; (c–f) Study 1 and 2 pooled. (a) Participants under no time-limit (NTL) in provision dilemma (P). (b) Participants under NTL in maintenance dilemma (M). (c) Participants under time-pressure (TP) in P. (d) Participants under TP in M. (e) Participants under time-delay (TD) in P. (f) Participants under TD in M. Figures S4 and S5 present this analysis separately for those with (Fig. S4) and without (Fig. S5) correct social dilemma understanding, with very similar results.

in the next section, CSR extends beyond the specifics of the games studied here, serving as a general framework for understanding cooperation in one-shot social dilemmas.

Discussion

In two experiments, we studied the psychological mechanisms underlying cooperation across maintenance and provision dilemmas. Our results provide converging evidence that, despite their formal equivalence, P and M dilemmas trigger systematic differences in cooperation behaviour and its underlying psychological processes. Cooperation is higher in P than in M (Fig. 1). We show that this gap is consistent with dilemma-specific intuitions, which are more selfishly oriented in M than in P (Figs. 2, 3). Measures of strong reciprocity (Fig. 4), being weaker in M than P, also depend to some extent on intuitions. Nevertheless, cooperation in both dilemmas remains high even with deliberation and is well-predicted by preferences for strong reciprocity (Figs. 5, 6).

Our time-limit manipulations allowed systematic comparison of the two main cognitive process accounts of cooperation: SHH which predicts intuitive cooperation and SCA which predicts intuitive selfishness. While deliberation increases free-riding in P, we found that it promotes cooperation in M. Hence, SHH's prediction of intuitive cooperation does not extend to M dilemmas, where SCA's prediction of intuitive selfishness applies. CSR integrates the contrasting mechanisms postulated in SHH and SCA by combining effects driven by different types of intuitions: SHH invokes social heuristics that tend to promote cooperation, whereas SCA invokes

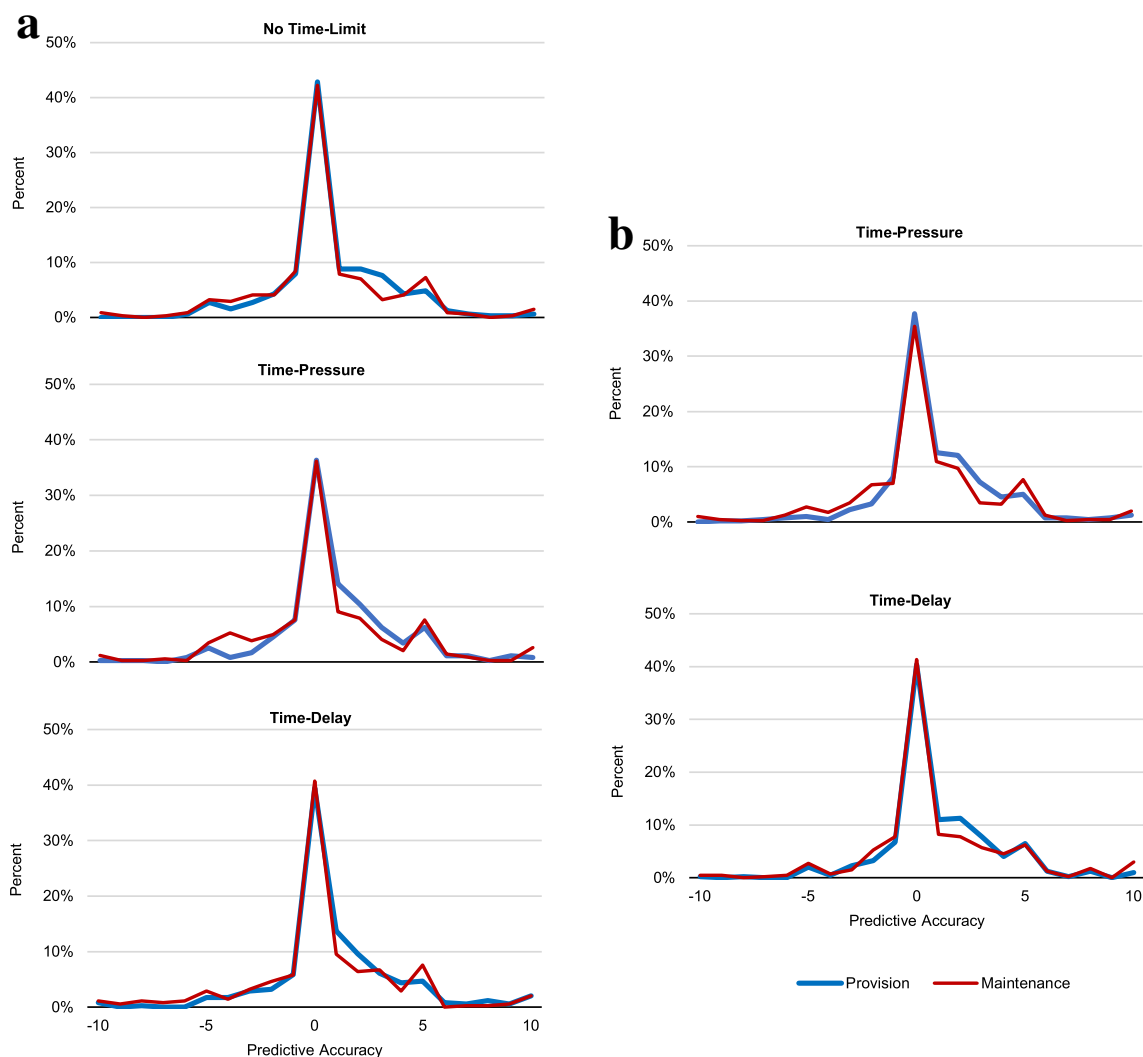


Figure 6. Accuracy of strong reciprocity measures in predicting cooperation. **(a)** Study 1: Distribution of predictive accuracy (the difference between actual and predicted contributions) for each participant) in NTL, TP and TD. **(b)** Study 2: Distribution of predictive accuracy in TP and TD. Note that zero indicates highest predictive accuracy possible. Figures S7 and S8 present this analysis separately for those with (Fig. S7) and without (Fig. S8) correct social dilemma understanding, with very similar results.

visceral, emotional reactions that tend to promote self-regard. Our study suggests that the relative importance of either type of intuition can depend systematically on contextual factors such as the social dilemma type, but further research on CSR is needed to understand the nature of these (and other) types of intuitions and the context-specific cues that trigger them. As these results are exclusively based on time-pressure manipulations (with methodological drawbacks such as high levels of non-compliance), their robustness should be tested using alternative, stronger cognitive process manipulations in future research^{84–87}.

CSR explains why finding evidence for SHH has proven difficult. Recent meta-analyses of the literature on SHH indicate a small effect size of promoting intuitive thinking on cooperation or no effect at all^{24,49–51,88}. One reason may be the coexistence of two independent and opposing intuitions in social dilemmas such that the negative effect of selfish intuitions (as argued by SCA) tends to cancel out the positive effect of social intuitions on cooperation (as argued by SHH)²⁶. Another reason may simply be that intuitions tend to have relatively small effects on cooperation, especially relative to the role that strong reciprocity plays in one-shot or ‘selfless’ cooperation.

Indeed, the preference for strong reciprocity is prevalent not only in our data (Fig. 4a) but across numerous studies^{81,89,90}. We observed high levels of deliberated cooperation, with time-delayed contributions being 63% in M and 67% in P (Fig. 1). Consistent with CSR’s formulation of $a_i(f, b_i(f))$, the prevalence of strong reciprocity preferences and high levels of expectations jointly predicted this behavioural outcome well, with predicted contributions of 56% in M and 57% in P for the TD conditions (Fig. 5). The high levels of deliberated cooperation motivated by strong reciprocity restricts the range for observing even higher cooperation due to social heuristics (e.g., under TP) to 37 pp in M and 33 pp in P. Isolating differences in cooperation levels supporting SHH in the P dilemma would therefore rely on intuitions being significantly more cooperative on average than

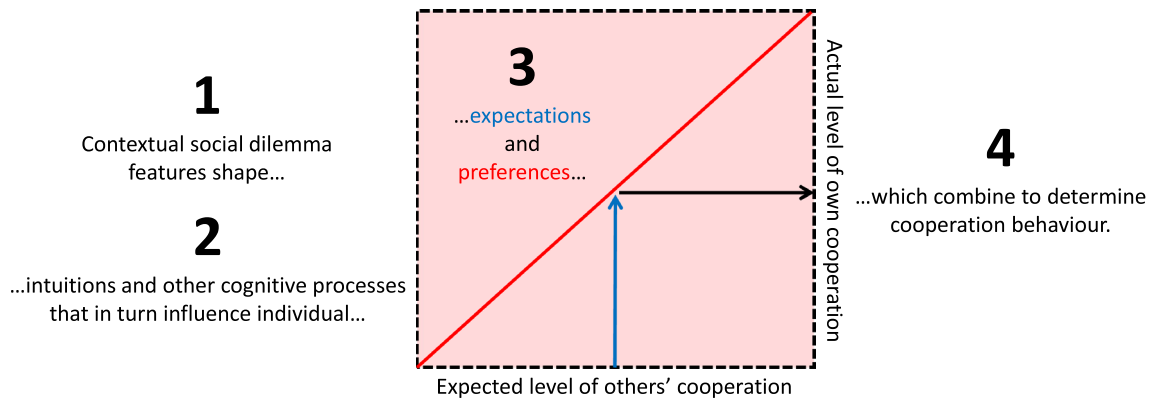


Figure 7. The Contextualised Strong Reciprocity (CSR) account of one-shot cooperation. CSR suggests four steps for explaining why people cooperate in anonymous one-shot social dilemmas: (1) The individual faces a specific decision context involving a social dilemma (e.g., Provision or Maintenance), which (2) shapes intuitions (e.g., social heuristics or selfish impulses) and other cognitive processes (e.g., influencing understanding of incentives, perceptions of social appropriateness of cooperation and the tendency for cognitive reflection). These in turn influence (3) expected cooperation by others (the horizontal axis represents the range of expectations and, for example, the blue arrow represents the expectation that others will contribute half of their endowments) as well as individual preferences for cooperation (the pink area represents the range of possible preferences and, for example, the red line represents the preferences of a perfect conditional cooperator). (4) Actual cooperation behaviour is a function of an individual's preferences, and in the case of strong reciprocity, also a function of the expected level of cooperation by others (as exemplified by the black arrow). In our studies, evidence for the relationship between steps 1 and 2 is in Figs. 2 and 3; evidence for the relationship between steps 2 and 3 is in Fig. 4; and evidence for the relationship between steps 3 and 4 is in Figs. 5 and 6.

this already high level. In this sense, a sample of individuals with relatively more selfish preferences (i.e., with lower deliberated cooperation) and with previous exposure to repeated cooperative interactions would be more likely to provide evidence for SHH.

The potential applicability of CSR extends beyond the P and M dilemmas studied here. The contextual features (f) of any particular cooperation problem (e.g., social dilemma type) will trigger f -specific intuitions and strong reciprocity preferences. In combination, these provide a general framework to explain cooperation in any one-shot social dilemma.

At this early stage, CSR is theoretically mute on the potentially complex relationship between f and the specific preferences, intuitions, and other cognitive processes that it triggers. Instead, CSR is primarily a framework for explaining one-shot cooperation that, for a given f , emphasizes measurement of expectations and preferences. As part of this empirical analysis, auxiliary measures of cognitive processes provide insights into how the specific f of a social dilemma situation influences cooperation. For example, our measurements (Figs. 2, 3) suggest that intuitions and other cognitive processes are more selfishly oriented when f pertains to M than to P. This is consistent with CSR's theoretical prediction that preferences and expectations jointly explain cooperation and our empirical observation that behaviour is selfishly biased in M compared to P (Fig. 4). While expectations and preferences constitute the primary determinants of cooperation in CSR, intuitions and auxiliary measures are particularly relevant when comparing cooperation across contexts, when seeking insights into the cognitive processes underlying cooperation and when formulating frames for the effective delivery of public policies.

The stark cognitive and behavioural differences between P and M dilemmas show that researchers and policymakers should not presume that the formal equivalence of the two dilemmas on standard theoretical analysis implies behavioural equivalence and should instead distinguish between provision and maintenance problems both analytically and for policy purposes. The generalizability of our results to naturally occurring environments should be tested in field studies⁹¹, for example, in attempts to maintain environmental public goods. Serious global threats such as excessive energy consumption or antibiotic overuse involve maintenance problems, where selfish impulses threaten the collective interest. In these cases, our results indicate that policymakers should engage individuals in deliberation⁵⁷ to promote collective welfare. Crucially, as our CSR account shows, policymakers should primarily strive to harness the power of strong reciprocity to motivate cooperation.

Methods

Experimental design. We ran two experiments designed to investigate intuitions and strong reciprocity across the two main social dilemma types (i.e., the Provision and Maintenance dilemmas). Each study had four main parts. Part A elicited decisions in one-shot public good games (PGGs) with time-limits, and Parts B to D elicited various additional cognitive process measures. The methods of the two studies were common except in the few aspects detailed below. Both studies employed between-subject designs crossing dilemma type (P or M) by time-limit (Study 1: 10 s pressure, 10 s delay or no time-limit; Study 2: 5 s pressure or 10 s delay). We adapted the P and M dilemmas developed for lab-based research³⁰ for an online setting, and employed standard time-limit manipulations¹⁶ and incentivized compliance with time-limits²³. Simple randomization was used,

and participants were blind to the experimental conditions. Both studies were preregistered at the Open Science Framework (<https://osf.io/euw9d/>).

Ethics. Our research complies with all relevant ethical regulations. Ethics approval was obtained from the University of Nottingham School of Economics Research Ethics Committee, and informed consent was obtained from the participants.

Participants. We recruited participants via Prolific (www.prolific.ac)⁹², restricted to native English-speaking UK residents who were 18 years or older. Study 1 participants were not allowed to participate in Study 2. We use data from 2,056 participants in Study 1 (age: $M = 36.6$, $SD = 11.4$; female: 75.2%) and from 1,597 participants in Study 2 (age: $M = 34.2$, $SD = 11.6$; female: 65.0%), excluding 153 incomplete submissions across the two studies. Because experience with (or other knowledge of) public goods experiments may weaken the effect of time-limit manipulations⁸², we recruited participants from Prolific, where most participants reported no prior knowledge or experience with PGGs (89.9% in Study 1 and 89.2% in Study 2).

Stakes. Participants were paid according to their decisions^{47,48}. Including a participation fee of £0.50, average earning was £1.43 in Study 1 and £1.38 in Study 2 with a median completion time of 8 min in Study 1 and 10 min in Study 2. This corresponds to a fee of approximately £11 per hour, which is considerably higher than the minimum wage in the UK (about £8 at the time of the experiments).

Planned sample size. *Study 1.* In a one-shot PGG, previous research³⁰ observed higher cooperation in P than in M dilemma with an effect size (d) of 0.59. In a pilot study ($n = 92$), we observed a smaller effect in the same direction ($d = 0.28$). A related study on time-pressure²³ found time-pressured decisions to be more cooperative than time-delayed decisions ($d = 0.22$). We used the smallest of the three effects as our benchmark and calculated the sample size required to detect an effect in a two-tailed t -test to be at least $n = 338$ per treatment group ($\alpha = 0.05$, $1 - \beta = 0.80$). Considering possible incomplete submissions, we aimed to recruit 2100 participants.

Study 2. In a pilot study ($n = 91$), we established the feasibility of using a 5 s time-limit by comparing compliance in 4 s (27.3%), 5 s (64.3%) and 6 s (70.0%) TP limit conditions. Motivated by previous research⁹³, we used the effect size of the interaction term in the Study 1 sample from an OLS model of PGG contributions on dilemma type, binary response time (RT) variable (1 if $RT \geq 5$ s), and their interaction ($f = 0.0815$). With a more powerful test than in Study 1 ($\alpha = 0.05$, $1 - \beta = 0.90$), the sample size required to detect an interaction effect between time-manipulation and dilemma type in an ANOVA was found to be at least $n = 396$ per treatment group. Considering possible incomplete submissions, we aimed to recruit 1600 participants.

Randomization. Data on P and M dilemmas were simultaneously collected and Qualtrics survey software (www.qualtrics.com) was used to randomly allocate participants to treatments. Lack of significant differences in the distribution of the number of participants (χ^2 -tests, $P = 0.780$ for Study 1 and $P = 0.940$ for Study 2) across the experimental conditions in both studies as well as for demographic variables such as gender (χ^2 -tests, $P = 0.577$ for Study 1 and $P = 0.385$ for Study 2) and age (Kruskal–Wallis tests, $P = 0.106$ for Study 1 and $P = 0.243$ for Study 2) indicate that simple randomization worked as intended.

Procedures. We programmed the experiments using *Qualtrics*. Participants first received slider training (the mechanism used for recording participants' responses) and PGG instructions for either the P or the M dilemma, and then completed four decision-making parts. In Part A, participants played a one-shot linear PGG. In Study 1, Part A was completed under 10 s time-pressure (TP), 10 s time-delay (TD) or no time-limit (NTL), whereas in Study 2, Part A was completed under 5 s TP or 10 s TD. In Part B, we first elicited expectations regarding other group members' contributions (under TP, TD or NTL congruent with Part A) and then elicited understanding of the social dilemma. We elicited cooperation preferences in Part C. In Part D, we either elicited perceptions of social appropriateness⁷¹ for three scenarios (Study 1) or the three-item Cognitive Reflection Task (CRT)⁷³ (Study 2). Finally, participants completed a brief demographic survey. Participants were informed at the beginning that one of the four parts would be randomly selected at the end of the study to calculate their additional earnings. We detail the components of the procedure below.

Slider training. Participants were familiarized with the slider tool—later used to elicit contribution decisions—on a practice screen²³. To prevent anchoring, the tool did not have a default slider position. The training was intended to minimize differences in familiarity with the tool that would have otherwise occurred between time-pressured and other participants.

Public good game. We used a one-shot linear public good game with four-person groups. Each token kept (or withdrawn) earned this group member one token; each token contributed to (not withdrawn from) the public good returned half a token to each group member that is, two tokens for the group as a whole, constituting a social dilemma. Each group was randomly assigned to instructions for either the P or the M dilemma. The two tasks are equivalent in terms of the relationship between allocations to the public good and monetary earnings from the public good game³⁰. We opted for the brief instructional style introduced by Rand et al.¹⁶ without control questions prior to the game rather than the extended analytical style standard in lab experiments³⁰. This

was intended to minimize inducement of a calculative mindset, which may weaken the capacity of time-limits to induce intuitive thinking¹⁶.

Part A. *Time manipulations and PGG decision.* Participants were randomly assigned to a time-limit condition for their PGG decisions. In Study 1, we adopted the standard limits¹⁶—a 10 s threshold for both the TP and the TD conditions. In Study 2, we intensified the TP treatment by using a 5 s threshold while retaining a 10 s threshold for TD. The TP conditions prompted participants to “be quick” and decide “in less than 10 [5] seconds”, whereas the TD conditions prompted them to “carefully consider” their decision “for more than 10 s”. NTL did not use a time-limit prompt. To minimize noncompliance with time-limits, participants in the TP and TD conditions were initially informed of the upcoming time-limits and that noncompliant participants would be ineligible for earnings from the public good game²³. This transitory screen was displayed for a fixed period of 15 s, which was long enough to allow reading and short enough to prevent deliberation about the upcoming task.

Part B. *Expectations.* Participants next guessed the average number of tokens contributed to (P) or withdrawn from (M) the public good by the three other people in their group (i.e., expected cooperation). Congruent with the time-limit condition that was assigned during the public good decision, these expectations were elicited under either TP, TD or NTL, thus exactly mimicking the conditions of the public good game environment. If the task was selected for payment, participants were rewarded £0.50 for correct predictions that also complied with time conditions. Prior to the elicitation of beliefs, a transitory screen was displayed for a fixed period of 20 s describing the reward conditions. This screen was displayed for 5 s longer than the transitory screen prior to the public good decision to account for the slightly longer text. No additional time-limits were used in the study after this question.

Social dilemma understanding. Understanding of the social dilemma was measured by two questions displayed on the same screen in random order each asking participants to choose a contribution (withdrawal) level from 0 to 10 tokens to identify: the correct strategy for maximizing own monetary gain (no contribution in P or full withdrawal in M) and the correct strategy for maximizing group’s monetary gain (full contribution in P or no withdrawal in M). An initial screen informed the participants that the questions had correct answers each worth £0.50 and prompted those who previously were under time-limits to decide at their “own pace”. Following the literature¹⁶, we categorize a participant as having understood the social dilemma if both questions were answered correctly (41.0% in Study 1 and 47.4% in Study 2; these figures are comparable with previous studies such as 45.6% in Study 6 of Rand, Greene and Nowak, 2012, an online PGG experiment using time-limits).

Part C. *Measuring preferences for (conditional) cooperation.* The preference elicitation task^{15,30,94} involved a modified public good game, which asked participants to provide a contribution (P) or withdrawal (M) level for each possible rounded average level of contribution or withdrawal that could have been made by the other three people in the participant’s group. Using eleven sliders on the same screen—each corresponding to a possible scenario ranging from an average contribution of others of 0, 1, ..., 10 tokens (randomly presented either in ascending or descending order)—the participants indicated their preferred contribution or withdrawal for each possible level of others’ average contributions (i.e., conditional cooperation). If Part C was selected for payment, payments were calculated by randomly selecting one group member for whom the conditional contribution at the average of others’ one-shot PGG contributions was payoff-relevant; for the others their initial (unconditional) PGG contributions were used to calculate payoffs. This is an incentive compatible method to elicit preferences for conditional cooperation⁹⁴. Monetary costs and returns from contributing to the public good in this task were equivalent to those from the public good game described above.

The contribution profiles elicited in these tables are used to categorize each participant into one of three cooperation preference types³⁰. Those who consistently contributed nothing to (or withdrew everything from) the public good in all eleven scenarios are categorized as “free riders” (3.1% in Study 1 and 3.0% in Study 2). Participants are categorized as “conditional cooperators” if the eleven contribution (withdrawal) decisions show a weakly monotonically increasing pattern for own contribution (withdrawal) in relation to others’ average contributions (withdrawals) or if the Spearman rank correlation coefficient between own and others’ contributions (withdrawals) is positive and significant at the 1% level (54.1% in Study 1 and 61.7% in Study 2). Remaining participants are categorized as “other” (42.8% in Study 1 and 35.3% in Study 2).

Predicted contributions. This measure³⁰ provides a prediction of public good contribution for each participant by combining the individual’s cooperation preference and expected cooperation measures: using the expectations measure, we first find the scenario in the preference elicitation task that is expected by the participant to be true, and we then identify the contribution level stated for that scenario as the participant’s predicted contribution.

Part D. *Perceptions of social appropriateness.* Social appropriateness was measured in Study 1 using an incentivized coordination game⁷¹ for three possible actions in the public good game: 0, 5 or 10 tokens contributed to (10, 5, 0 tokens withdrawn from) the public good. The three questions were randomly displayed on the same screen either in ascending or descending order. Using a six-point scale ranging from “very socially inappropriate” to “very socially appropriate”, participants were asked to estimate how socially appropriate most people in the study would find each action. Participants were informed that they would be randomly matched with someone in the study, that one of the three actions would be randomly chosen, and that if this task was chosen to

calculate extra earnings then they would each earn £1 if the two evaluations matched. The task thus constitutes a coordination game that incentivizes honest reporting of participants' perceptions of the social appropriateness of various actions in the public good game. Responses were converted to numerical scores⁷¹ so that 0 implies "neutral": "very socially inappropriate" = -1, "somewhat socially inappropriate" = -2/3, "socially inappropriate" = -1/3, "socially appropriate" = 1/3, "somewhat socially appropriate" = 2/3, "very socially inappropriate" = 1.

Cognitive reflection test. We elicited the three-item CRT⁷³ in standard order twice, first at the end of Study 2 and then one-and-a-half years later. If Part D in Study 2, which included the CRT, was selected for payment, then participants were rewarded £0.50 for completing the task. The second time the CRT was measured, we invited all Study 2 participants and continued data collection until 50% of participants were recruited. In this follow up experiment, CRT was measured in isolation, with no preceding play of P or M dilemma games. We calculated two types of CRT scores for each individual. The standard CRT scores (between 0 and 3) were calculated as the total number of correct answers. The alternative iCRT scores (also ranging from 0 to 3) were calculated as number of intuitive but incorrect answers (see Supplementary Materials for the CRT questions and further information)⁷⁴.

Hypotheses. Main hypotheses. We describe our main preregistered hypotheses in the main text. The first three hypotheses make predictions about contribution decisions. We predicted higher contributions (C) in P than in M ($H_1: C_P > C_M$). Consistent with SHH, we also predicted higher contributions in TP than in TD ($H_2: C_{TP} > C_{TD}$) and higher difference in contributions between P and M in TP than in TD ($H_3: \Delta C_{TP} > \Delta C_{TD}$). As an alternative to H_2 , SCA predicts higher contributions in TD than in TP ($H_A: C_{TP} < C_{TD}$). Our other three main hypotheses involve predictions about measures of strong reciprocity: conditional cooperators (CC) will be more prevalent in P than in M ($H_4: CC_P > CC_M$); beliefs about others' contributions, expectations (E), will be higher in P than in M ($H_5: E_P > E_M$); and predicted contributions (PC) will be higher in P than in M ($H_6: PC_P > PC_M$).

Additional tests. Extending H_2 ($C_{TP} > C_{TD}$) to strong reciprocity measures, we tested the hypotheses of higher conditional cooperation ($H_7: CC_{TP} > CC_{TD}$), expectations ($H_8: E_{TP} > E_{TD}$) and predicted contributions ($H_9: PC_{TP} > PC_{TD}$) in TP than in TD. Similarly, we extend H_3 ($\Delta C_{TP} > \Delta C_{TD}$) to reciprocity measures ($H_{10}: \Delta CC_{TP} > \Delta CC_{TD}$, $H_{11}: \Delta E_{TP} > \Delta E_{TD}$, $H_{12}: \Delta PC_{TP} > \Delta PC_{TD}$). However, cooperation schedules were elicited without time-limits. H_7 and H_{10} should therefore be viewed as testing spill-over of previous time-limits (applied during elicitation of C and E) on type elicitation. All twelve hypotheses were tested for both Study 1 and Study 2, and the results of additional hypotheses tests are reported in the Supplementary Materials.

Two additional hypotheses on the initial CRT scores were preregistered in Study 2. In particular, because contributions are expected to be higher in P than in M for intuitive decisions and because low CRT scores indicate tendency for intuitive thinking, we predicted that contributions in P will be higher than in M among those with a CRT score of zero (H_{13}) and that the effect of dilemma type on contributions will be weaker for higher CRT scores (H_{14}). Although we report these analyses for completeness in the Supplementary Materials, these cannot be interpreted as valid tests of H_{13} and H_{14} due to unexpected spill-over effects from dilemma type on subsequently elicited CRT scores.

Analysis plan. We report tests for our main hypotheses in the Results section of the main text and introduce exploratory analyses as appropriate. We provide detailed results in the Supplementary Materials for the main and the additional hypothesis tests (see Hypotheses). Our preregistered tests do not control for any of the covariates sometimes used in the literature as control measures (i.e., age, gender, experience with or knowledge of the public good game, compliance with time-limits and social-dilemma understanding). The results in both studies are robust when controlling for these covariates.

We recode all withdrawal-based decisions in the M condition as tokens kept in the public good game, and we refer to these values as contributions to the public good. First, we present our preregistered test results. As planned, we analyse the no time-limit (NTL) and the time-limit samples (TL: TP & TD) separately. To test the first twelve hypotheses as appropriate, we use χ^2 and t-tests in the NTL sample of Study 1, and we use two-way ANOVA and Logistic interaction models in the TL samples of Study 1 and Study 2. These models include the time-limit (TP vs. TD), the dilemma type (P vs. M) and their interaction term as covariates. The models differ by their dependent variable: contributions (C), expectations (E) or predicted contributions (PC) in ANOVAs and prevalence of conditional cooperators (CC) in Logistic regressions. To find main effects on CC in the TL samples, we use χ^2 tests. Apart from ANOVAs and χ^2 tests that are based on distributions with one tail, all tests are two-tailed.

Manipulation checks. As our main manipulation check, in both studies we compared the differences in mean response times between the time-limit conditions. In addition to this behavioral measure reported in the main text, we elicited and compared self-reported cognitive manipulation checks.

Study 1. A post-experimental survey question asked participants to compare how intuitively vs. deliberately they decided in the PGG on a 4-point scale ranging from very intuitively (1) to very deliberately (4). The exact wording of the question depended on the social dilemma condition: "Please indicate the degree to which you decided intuitively vs. deliberately while choosing how many tokens to contribute to [withdraw from] the group project." Although the difference in answers between TP ($M = 2.23$) and TD ($M = 2.32$) was significant,

$t(1382) = 2.08, P = 0.037, d = 0.11$, responses did not differ between NTL ($M = 2.29$) and either the TP or the TD conditions (NTL vs TP: $t(1368) = 1.27, P = 0.205, d = 0.07$; NTL vs TD: $t(1356) = 0.75, P = 0.453, d = 0.04$).

Study 2. This study introduced a cognitive manipulation check by eliciting two simple evaluations right after the measurement of expectations (i.e., the end of time-limits): (1) “I did not have time to think through my decisions” and (2) “I decided based on my ‘gut reaction’”. Ratings for the two statements ranged on a 5-point scale from “strongly disagree” (1) to “strongly agree” (5). The two questions were presented on the same screen in randomized order. Ratings of the first statement indicate that TP ($M = 3.37$) limited opportunities for deliberation compared to TD ($M = 1.55$), $t(1595) = 32.28, P < 0.001, d = 1.62$. Ratings of the second statement indicate that TP ($M = 4.06$) increased spontaneous decisions compared to TD ($M = 3.76$), $t(1595) = 5.60, P < 0.001, d = 0.28$.

Compliance with time-limits. We include all non-compliant participants in the analyses, since exclusion may bias estimates²⁵.

Study 1. Compliance with time-limits based on RTs was high in both TD (Overall: 88.9%; P: 89.8%; M: 88.1%), and TP (Overall: 93.6%; P: 94.9%; M: 92.1%).

Study 2. Due to increased time-pressure (5 s), the standard time-limit compliance rate based on the submission of the decision page (i.e., RT) was lower in TP (Overall: 58.9%; P: 63.6%; M: 54.4%) than in TD (Overall: 91.9%; P: 89.5%; M: 94.2%). Nevertheless, supporting measures indicate that participants in TP were strongly motivated to comply with the time limits. Specifically, the median RT among those who failed to comply with the TP limits was still relatively fast (6.4 s) such that more than a third of these participants (38.7%) submitted their responses within 5 to 6 s (i.e., within one additional second above the limit), nearly a quarter (22.5%) submitted their responses within 6 to 7 s, and overall, 87.5% of all noncompliant participants submitted their responses within 5 to 10 s.

Statistical analysis. As preregistered, we conducted most of our analyses using ANOVAs and t-tests. Our main results hold when we use nonparametric tests. The statistical tests used are described in the main text, and the analysis codes will be available upon publication at the Open Science Framework project site.

Data availability

The datasets, materials and analysis codes are available at the Open Science Framework project site (<https://osf.io/euw9d/>).

Received: 12 January 2021; Accepted: 15 June 2021

Published online: 06 July 2021

References

- Bowles, S. & Gintis, H. *A cooperative species: human reciprocity and its evolution* (Princeton University Press, 2011).
- Sober, E. & Wilson, D. S. *Unto others. The evolution and psychology of unselfish behavior* (Harvard University Press, 1998).
- Henrich, J. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter* (Princeton University Press, 2015).
- Rand, D. G. & Nowak, M. A. Human cooperation. *Trends Cogn. Sci.* **17**, 413–425 (2013).
- Henrich, J. & Muthukrishna, M. The origins and psychology of human cooperation. *Annu. Rev. Psychol.* **72**, 207–240 (2021).
- Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
- Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
- Gintis, H., Bowles, S., Boyd, R. & Fehr, E. *Moral sentiments and material interests. The foundations of cooperation in economic life* (MIT Press, 2005).
- Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25 (2002).
- Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
- Bolton, G. E. & Ockenfels, A. ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **90**, 166–193 (2000).
- Charness, G. & Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **117**, 817–869 (2002).
- Falk, A. & Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **54**, 293–315 (2006).
- Dufwenberg, M. & Kirchsteiger, G. A theory of sequential reciprocity. *Games Econ. Behav.* **47**, 268–298 (2004).
- Weber, T. O., Weisel, O. & Gächter, S. Dispositional free riders do not free ride on punishment. *Nat. Commun.* **9**, 2390 (2018).
- Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous giving and calculated greed. *Nature* **489**, 427–430 (2012).
- Rand, D. G. *et al.* Social heuristics shape intuitive cooperation. *Nat. Commun.* **5**, 3677 (2014).
- Bear, A. & Rand, D. G. Intuition, deliberation, and the evolution of cooperation. *Proc. Natl. Acad. Sci.* **113**, 936–941 (2016).
- Myrseth, K. O. R. & Wollbrant, C. E. Models inconsistent with altruism cannot explain the evolution of human cooperation. *Proc. Natl. Acad. Sci.* **113**, E2472–E2472 (2016).
- Bear, A. & Rand, D. G. Reply to Myrseth and Wollbrant: Our model is consistent with altruism, and helps to explain its evolution. *Proc. Natl. Acad. Sci.* **113**, E2473–E2473 (2016).
- Rand, D. G. & Peysakhovich, A. Habits of virtue: creating norms of cooperation and defection in the laboratory. *Manag. Sci.* **62**, 631–647 (2016).
- Everett, J. A. C., Ingbretsen, Z., Cushman, F. & Cikara, M. Deliberation erodes cooperative behavior—even towards competitive out-groups, even when using a control condition, and even when eliminating selection bias. *J. Exp. Soc. Psychol.* **73**, 76–81 (2017).
- Isler, O., Maule, J. & Starmer, C. Is intuition really cooperative? Improved tests support the social heuristics hypothesis. *PLoS ONE* **13**, e0190560 (2018).
- Kvarven, A. *et al.* The intuitive cooperation hypothesis revisited: a meta-analytic examination of effect size and between-study heterogeneity. *J. Econ. Sci. Assoc.* **6**, 26–42 (2020).
- Tinghög, G. *et al.* Intuition and cooperation reconsidered. *Nature* **498**, E1–E2 (2013).

26. Isler, O., Yilmaz, O. & JohnMaule, A. Religion, parochialism and intuitive cooperation. *Nat. Hum. Behav.* **5**, 512–521 (2021).
27. Cain, D. M., Dana, J. & Newman, G. E. Giving versus giving in. *Acad. Manag. Ann.* **8**, 505–533 (2014).
28. Keysar, B., Converse, B. A., Wang, J. & Epley, N. Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychol. Sci.* **19**, 1280–1286 (2008).
29. Offerman, T. & Sonnemans, J. Learning by experience and learning by imitating successful others. *J. Econ. Behav. Organ.* **34**, 559–575 (1998).
30. Gächter, S., Kölle, F. & Quercia, S. Reciprocity and the tragedies of maintaining and providing the commons. *Nat. Hum. Behav.* **1**, 650–656 (2017).
31. Myrseth, K. O. R. & Fishbach, A. Self-control: a function of knowing when and how to exercise restraint. *Curr. Dir. Psychol. Sci.* **18**, 247–252 (2009).
32. Capraro, V. & Cococcioni, G. Rethinking spontaneous giving: extreme time pressure and ego-depletion favor self-regarding reactions. *Sci. Rep.* **6**, 27219 (2016).
33. Martinsson, P., Myrseth, K. O. R. & Wollbrant, C. Social dilemmas: when self-control benefits cooperation. *J. Econ. Psychol.* **45**, 213–236 (2014).
34. Kocher, M. G., Martinsson, P., Myrseth, K. O. R. & Wollbrant, C. E. Strong, bold, and kind: self-control and cooperation in social dilemmas. *Exp. Econ.* **20**, 44–69 (2017).
35. Lohse, J. Smart or selfish—when smart guys finish nice. *J. Behav. Exp. Econ.* **64**, 28–40 (2016).
36. Curry, O. S., Price, M. E. & Price, J. G. Patience is a virtue: cooperative people have lower discount rates. *Personal. Ind. Differ.* **44**, 780–785 (2008).
37. Van Lange, P. A. M., Balliet, D., Parks, C. D. & Van Vugt, M. *Social Dilemmas. The Psychology of Human Cooperation* (Oxford University Press, 2014).
38. Apesteguía, J. & Maier-Rigaud, F. P. The role of rivalry. Public goods versus common-pool resources. *J. Confl. Resolut.* **50**, 646–663 (2006).
39. Levin, S. A. Public goods in relation to competition, cooperation, and spite. *Proc. Natl. Acad. Sci.* **111**, 10838–10845 (2014).
40. Brewer, M. B. & Kramer, R. M. Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. *J. Pers. Soc. Psychol.* **50**, 543–549 (1986).
41. De Dreu, C. K. & McCusker, C. Gain-loss frames and cooperation in two-person social dilemmas: a transformational analysis. *J. Pers. Soc. Psychol.* **72**, 1093 (1997).
42. Van Dijk, E., Wilke, H., Wilke, M. & Metman, L. What information do we use in social dilemmas? Environmental uncertainty and the employment of coordination rules. *J. Exp. Soc. Psychol.* **35**, 109–135 (1999).
43. McCusker, C. & Carnevale, P. J. Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organ. Behav. Hum. Decis. Process.* **61**, 190–201 (1995).
44. Hauser, O. P., Rand, D. G., Peysakhovich, A. & Nowak, M. A. Cooperating with the future. *Nature* **511**, 220–223 (2014).
45. Cox, C. A. & Stoddard, B. Framing and feedback in social dilemmas with partners and strangers. *Games* **6**, 394–412 (2015).
46. Sonnemans, J., Schram, A. & Offerman, T. Public good provision and public bad prevention: the effect of framing. *J. Econ. Behav. Organ.* **34**, 143–161 (1998).
47. Kocher, M. G., Martinsson, P. & Visser, M. Does stake size matter for cooperation and punishment?. *Econ. Lett.* **99**, 508–511 (2008).
48. Amir, O., Rand, D. G. & Gal, Y. A. K. Economic games on the internet: the effect of \$1 stakes. *PLoS ONE* **7**, e31461 (2012).
49. Bouwmeester, S. *et al.* Registered replication report: Rand, Greene, and Nowak (2012). *Perspect. Psychol. Sci.* **12**, 527–542 (2017).
50. Rand, D. G. Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychol. Sci.* **27**, 1192–1206 (2016).
51. Rand, D. G. Intuition, deliberation, and cooperation: further meta-analytic evidence from 91 experiments on pure cooperation. SSRN <https://doi.org/10.2139/ssrn.3390018> (2019).
52. Fosgaard, T. R., Hansen, L. G. & Wengström, E. Understanding the nature of cooperation variability. *J. Public Econ.* **120**, 134–143 (2014).
53. Frackenhohl, G., Hillenbrand, A. & Kube, S. Leadership effectiveness and institutional frames. *Exp. Econ.* **19**, 842–863 (2016).
54. Andreoni, J. Warm glow versus cold prickle—the effects of positive and negative framing on cooperation in experiments. *Q. J. Econ.* **110**, 1–21 (1995).
55. Fiedler, S. & Hillenbrand, A. Gain-loss framing in interdependent choice. *Games Econom. Behav.* **121**, 232–251 (2020).
56. Fehr, E. & Leibbrandt, A. A field study on cooperativeness and impatience in the tragedy of the commons. *J. Public Econ.* **95**, 1144–1155 (2011).
57. Lohse, J., Goeschl, T. & Diederich, J. H. Giving is a question of time: response times and contributions to an environmental public good. *Environ. Resource Econ.* **67**, 455–477 (2017).
58. Rubinstein, A. Instinctive and cognitive reasoning: a study of response times. *Econ. J.* **117**, 1243–1259 (2007).
59. Spiliopoulos, L. & Ortmann, A. The BCD of response time analysis in experimental economics. *Exp. Econ.* **21**, 383–433 (2018).
60. Krajbich, I., Bartling, B., Hare, T. & Fehr, E. Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nat. Commun.* **6**, 7455 (2015).
61. Merkel, A. L. & Lohse, J. Is fairness intuitive? An experiment accounting for subjective utility differences under time pressure. *Exp. Econ.* **22**, 24–50 (2019).
62. Evans, A. M., Dillon, K. D. & Rand, D. G. Fast but not intuitive, slow but not reflective: decision conflict drives reaction times in social dilemmas. *J. Exp. Psychol. Gen.* **144**, 951–966 (2015).
63. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
64. Recalde, M. P., Riedl, A. & Vesterlund, L. Error-prone inference from response time: the case of intuitive generosity in public-good games. *J. Public Econ.* **160**, 132–147 (2018).
65. Mischkowski, D. & Glöckner, A. Spontaneous cooperation for prosocials, but not for proselves: social value orientation moderates spontaneous cooperation behavior. *Sci. Rep.* **6**, 21555 (2016).
66. Chen, F. & Krajbich, I. Biased sequential sampling underlies the effects of time pressure and delay in social decision making. *Nat. Commun.* **9**, 3557 (2018).
67. Caplin, A. & Martin, D. The dual-process drift diffusion model: evidence from response times. *Econ. Inq.* **54**, 1274–1282 (2016).
68. Fosgaard, T. R., Hansen, L. G. & Wengström, E. Framing and misperception in public good experiments. *Scand. J. Econ.* **119**, 435–456 (2017).
69. Goeschl, T. & Lohse, J. Cooperation in public good games. Calculated or confused?. *Eur. Econ. Rev.* **107**, 185–203 (2018).
70. Stromland, E., Tjøtta, S. & Torsvik, G. Cooperating, fast and slow: testing the social heuristics hypothesis. CESifo working paper series no. 5875, SSRN. <https://ssrn.com/abstract=2780877>.
71. Krupka, E. L. & Weber, R. A. Identifying social norms using coordination games: why does dictator game sharing vary?. *J. Econ. Assoc.* **11**, 495–524 (2013).
72. Fehr, E. & Schurtenberger, I. Normative foundations of human cooperation. *Nat. Hum. Behav.* **2**, 458–468 (2018).
73. Frederick, S. Cognitive reflection and decision making. *J. Econ. Perspect.* **19**, 24–42 (2005).
74. Cueva, C. *et al.* Cognitive (ir)reflection: new experimental evidence. *J. Behav. Exp. Econ.* **64**, 81–93 (2016).

75. Fischbacher, U. & Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public good experiments. *Am. Econ. Rev.* **100**, 541–556 (2010).
76. Croson, R. Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Econ. Inq.* **45**, 199–216 (2007).
77. Cubitt, R., Drouvelis, M. & Gächter, S. Framing and free riding: emotional responses and punishment in social dilemma games. *Exp. Econ.* **14**, 254–272 (2011).
78. Dufwenberg, M., Gächter, S. & Hennig-Schmidt, H. The framing of games and the psychology of play. *Games Econ. Behav.* **73**, 459–478 (2011).
79. Gächter, S. & Renner, E. Leaders as role models and ‘belief managers’ in social dilemmas. *J. Econ. Behav. Organ.* **154**, 321–334 (2018).
80. Ackermann, K. A. & Murphy, R. O. Explaining cooperative behavior in public goods games: how preferences and beliefs affect contribution levels. *Games* **10**, 15 (2019).
81. Thöni, C. & Volk, S. Conditional cooperation: review and refinement. *Econ. Lett.* **171**, 37–40 (2018).
82. Rand, D. G. Non-Naïvety may reduce the effect of intuition manipulations. *Nat. Hum. Behav.* **2**, 602–602 (2018).
83. Fischbacher, U., Gächter, S. & Quercia, S. The behavioral validity of the strategy method in public good experiments. *J. Econ. Psychol.* **33**, 897–913 (2012).
84. Persson, E., Heilig, M., Tinghög, G. & Capusan, A. J. Using quantitative trait in adults with adhd to test predictions of dual-process theory. *Sci. Rep.* **10**, 20076 (2020).
85. Koppel, L. *et al.* The effect of acute pain on risky and intertemporal choice. *Exp. Econ.* **20**, 878–893 (2017).
86. Capraro, V. The dual-process approach to human sociality: a review. SSRN working paper (2019). <https://doi.org/10.2139/ssrn.3409146>.
87. Isler, O., Yilmaz, O. & Doğruyol, B. Activating reflective thinking with decision justification and debiasing training. *Judgm. Decis. Mak.* **15**, 926–938 (2020).
88. Camerer, C. F. *et al.* Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nat. Hum. Behav.* **2**, 637–644 (2018).
89. Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J. & Sutter, M. Conditional cooperation on three continents. *Econ. Lett.* **101**, 175–178 (2008).
90. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **14**, 47–83 (2011).
91. Artavia-Mora, L., Bedi, A. S. & Rieger, M. Intuitive help and punishment in the field. *Eur. Econ. Rev.* **92**, 133–145 (2017).
92. Palan, S. & Schitter, C. Prolific.Ac—a subject pool for online experiments. *J. Behav. Exp. Finance* **17**, 22–27 (2018).
93. Myrseth, K. O. R. & Wollbrant, C. E. Cognitive foundations of cooperation revisited. *J. Behav. Exp. Econ.* **69**, 133–138 (2017).
94. Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).

Acknowledgements

We thank E. Ferguson, F. Kölle, L. Molleman and O. Yilmaz for helpful comments on the manuscript. This work was supported by the European Research Council [Grant Number ERC-AdG 295707 COOPERATION] and the Economic and Social Research Council [Grant Numbers ES/K002201/1 and ES/P008976/1]. O. Isler acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie fellowship number 658186 (<https://ec.europa.eu/>). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

All authors developed the study concept and contributed to the study design. O.I. programmed the experiment, conducted the study, and analysed the data. O.I. and S.G. wrote the manuscript. A.J.M. and C.S. provided critical comments. All the authors approved the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93412-4>.

Correspondence and requests for materials should be addressed to O.I. or S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021