



Cooperation in Groups of Different Sizes: The Effects of Punishment and Reputation-Based Partner Choice

Junhui Wu^{1*}, Daniel Balliet², Leonard S. Peperkoorn², Angelo Romano³ and Paul A. M. Van Lange²

¹ Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Institute of Developmental Psychology, Beijing Normal University, Beijing, China,

² Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, ³ Max Planck Institute for Research on Collective Goods, Bonn, Germany

OPEN ACCESS

Edited by:

Lin Zhang,
Ningbo University, China

Reviewed by:

Shaojie Lv,
Shandong Jiaotong University, China
Matjaž Perc,
University of Maribor, Slovenia

*Correspondence:

Junhui Wu
junhui.wu@bnu.edu.cn

Specialty section:

This article was submitted to
Evolutionary Psychology,
a section of the journal
Frontiers in Psychology

Received: 09 November 2019

Accepted: 12 December 2019

Published: 21 January 2020

Citation:

Wu J, Balliet D, Peperkoorn LS, Romano A and Van Lange PAM (2020) Cooperation in Groups of Different Sizes: The Effects of Punishment and Reputation-Based Partner Choice. *Front. Psychol.* 10:2956. doi: 10.3389/fpsyg.2019.02956

Reputation and punishment are two distinct mechanisms that facilitate cooperation among strangers. However, empirical research on their effectiveness is mainly limited to relatively small groups and does not address how they enhance cooperation in relatively larger groups. We address this gap in the literature by testing hypotheses from competing perspectives about the extent to which reputation-based partner choice and punishment enhance cooperation in both small and large groups. Prior work recognizes that an increase in group size is accompanied by a change in the incentive structure, which determines whether the temptation (extra benefit for each person from non-cooperation over cooperation, regardless of others' choices) or gain (extra benefit for each person from full cooperation over full non-cooperation) remains constant or varies with group size. Thus, we first test how group size affects cooperation when temptation or gain increases with group size (Study 1), and then move on to testing predictions on the effectiveness of reputation and punishment across different group sizes (Study 2). In Study 1 ($N = 820$), we randomly assigned participants to play an online one-shot public goods game in groups of 4, 20, or 40, while keeping the marginal group return or marginal per capita return fixed across groups, in which case the temptation or gain increased with group size. In Study 2 ($N = 1,132$), we further compared a public goods situation involving a punishment or reputation mechanism with an anonymous situation across group sizes, while the marginal group return was fixed across groups. Overall, we found that when temptation increased with group size, 20-person groups cooperated significantly less than 4-person groups in one-shot interactions, and that this effect was explained by lower expectation of others' cooperation, less perceived collective efficacy, and greater perceived conflict. However, 40-person and 4-person groups did not vary in one-shot cooperation. Importantly, reputation-based partner choice and punishment invariably promoted one-shot cooperation in groups of different sizes. These findings suggest no simple effect of group size on cooperation and underscore the utility of reputation and punishment in fostering cooperation (at least in one-shot interactions) regardless of the size of groups.

Keywords: group size, public goods game, reputation-based partner choice, punishment, cooperation

INTRODUCTION

The exponential population growth in human history has facilitated a transition from kin-based small-scale societies to large-scale societies with frequent interactions between strangers (Durand, 1977). This transition may facilitate group living in some aspects (e.g., easier labor division in large ventures), but may also pose challenges to cooperation (e.g., overharvesting). Indeed, cooperation is often costly because it requires individuals to sacrifice their personal interests for the collective good (Rand and Nowak, 2013). However, people often cooperate to benefit their group, and even cooperate with non-kin others on large scales—a puzzle that has attracted extensive attention across disciplines (e.g., Fehr and Gächter, 2002; Tomasello and Vaish, 2013; Perc et al., 2017). To date, a number of mechanisms have been proposed to explain cooperation between strangers, such as reciprocal altruism (Trivers, 1971), indirect reciprocity (Nowak and Sigmund, 2005), costly signaling (Gintis et al., 2001), and social norm enforcement (e.g., costly punishment; Yamagishi, 1986; Clutton-Brock and Parker, 1995; Fehr et al., 2002). In particular, reputation and punishment are assumed to be especially relevant in promoting large-scale cooperation (Panchanathan and Boyd, 2004; Chudek and Henrich, 2011). However, empirical research on these two mechanisms is mainly limited to relatively small groups (e.g., Fehr and Gächter, 2002; Milinski et al., 2002) and does not address how reputation and punishment enhance cooperation in larger groups. In fact, there is scant empirical evidence from behavioral experiments about their effectiveness in groups of different sizes (for one exception in the context of punishment, see Xu et al., 2013).

In this paper, we address this gap in the literature by testing predictions on the effectiveness of reputation and punishment in promoting cooperation in larger (versus smaller) groups. However, studying cooperation in groups of different sizes is complex because an increase in group size is accompanied by changes in the incentive structure (e.g., the individual and group payoff resulting from each member's behavior). Therefore, we first analyze the incentives to cooperate across groups of different sizes. Next, we forward hypotheses about how the two fundamental pillars of human cooperation—reputation and punishment—may foster cooperation across different group sizes. Finally, we discuss the proximate psychological processes that may account for variations in cooperation across groups.

Theoretically, the larger the group, the more difficult it is to track others' behavior and to identify free riders, and the more likely that coordination and cooperation would fail due to a lack of group efficacy (Kerr, 1989). Despite a widespread belief that cooperation declines with greater group size (e.g., Dawes, 1980; Van Lange et al., 2013), there is also research showing that larger group size increases (Carpenter, 2007; Szolnoki and Perc, 2011; Barcelo and Capraro, 2015), decreases (Suzuki and Akiyama, 2005), or does not influence cooperation (Kerr, 1989; Zelmer, 2003). This mixed evidence on group size and cooperation may be driven by the payoff structure in a specific interaction (Bonacich et al., 1976; Nosenzo et al., 2015). To illustrate this idea, consider an n -person public goods game ($n \geq 2$). In this game, each person contributes x ($0 \leq x \leq E$) out of the initial E monetary

units (MUs) to the group account and keeps $(E - x)$ MUs for themselves. The total contribution is multiplied by k ($1 < k < n$) and divided equally among n persons. The two parameters in this game—marginal group return ($MGR = k$) and marginal per capita return ($MPCR = k/n$)—represent the group payoff and individual payoff from each MU contributed. Variations in these parameters determine whether the *temptation* (i.e., the extra benefit for each person from non-cooperation over cooperation, regardless of others' choices) or *gain* (i.e., the extra benefit for each person from full cooperation over full non-cooperation) remains constant or varies with group size (n), and thereby may elicit differences in cooperation (Bonacich et al., 1976). Specifically, when the group payoff (MGR) is fixed, gain remains constant but temptation increases with group size, such that the personal benefit from cooperation decreases in larger groups. In contrast, when the individual payoff ($MPCR$) is fixed, larger groups produce greater added benefits when all cooperate (vs. no one cooperates), such that temptation remains constant but gain increases with group size, making cooperation more likely to occur in larger groups (see Isaac et al., 1994; Barcelo and Capraro, 2015; Shank et al., 2015). Thus, people may adjust their cooperation level depending on how temptation and gain shift as a function of the size of the interacting group. We predict that people will be *less* cooperative in larger (vs. smaller) groups when temptation increases with group size (*Hypothesis 1a*), but will be *more* cooperative in larger (vs. smaller) groups when gain increases with group size (*Hypothesis 1b*).

As noted earlier, reputation and punishment are two distinct but non-mutually exclusive pathways that facilitate large-scale cooperation among unrelated individuals. Regarding the role of reputation, the theory of indirect reciprocity states that cooperators who gain a good reputation are more likely to be reciprocated by other third parties, and that this process allows cooperation to evolve in large groups of genetically unrelated strangers (Yamagishi and Kiyonari, 2000; Nowak and Sigmund, 2005). Moreover, the theory of competitive altruism posits that people prefer to partner with the best cooperators, who will receive more benefits than less cooperative ones (Barclay and Willer, 2007; Van Vugt et al., 2007). Indeed, prior work reveals that people tend to select and cooperate with partners whose reputation is positive (Capraro et al., 2016), and that reputational cues (e.g., gossip) can effectively promote cooperation (Feinberg et al., 2014; for a review, see Wu et al., 2016b). Thus, we expect more cooperation in a situation with reputation-based partner choice than in an anonymous situation (*Hypothesis 2*).

Is reputation-based partner choice more (or less) effective in promoting cooperation when groups become larger, particularly when temptation increases with group size? Different perspectives have different answers to this question. The theory of competitive altruism suggests that people in larger (vs. smaller) groups would face more competition in selecting the best cooperators as partners and advertising themselves as more cooperative than others (Van Vugt et al., 2007). Similarly, the biological market theory argues that larger groups involve more competition for a good reputation and a lower chance to be chosen as potential partners (Barclay, 2013). Thus, signaling one's cooperativeness is more important in larger groups than in

smaller ones. Notably, costly cooperative behaviors reflect one's genuine concern for others and thus attract more long-term cooperative partners (Smith and Bliege Bird, 2000). When there is more temptation in larger (vs. smaller) groups (i.e., fixed MGR), cooperation becomes costlier, so reputation-based partner choice should promote cooperation more effectively when implemented in larger (vs. smaller) groups (*Hypothesis 2a*). However, another perspective argues that larger groups are more vulnerable to free riding and involve more difficulty in identifying others' accurate reputation (e.g., Olson, 1965). Importantly, indirect reciprocity enables cooperation to evolve when the probability of knowing others' reputation is sufficiently high (Rand and Nowak, 2013), which may be less likely to occur in larger groups. Thus, reputation-based cooperation may be more difficult to evolve as groups become larger (e.g., Suzuki and Akiyama, 2005; dos Santos and Wedekind, 2015). This leads to an opposite prediction that reputation-based partner choice is less likely to promote cooperation in larger (vs. smaller) groups when temptation increases with group size (*Hypothesis 2b*).

When individuals live in stable groups where group members have conflicting interests, punishment strategies are likely to evolve to maintain cooperation (Yamagishi, 1986; Clutton-Brock and Parker, 1995). Punishment in this context often involves the intentional imposition of a sanction on a free rider that is costly to the punisher but even more costly to the free rider (typically in a fixed ratio, such as paying one MU to deduct three MUs from the free rider), so it discourages others' subsequent free riding behaviors. Indeed, people are motivated to punish free riders and norm violators (Ostrom et al., 1992; Fehr and Fischbacher, 2004), and punishment promotes cooperation in both one-shot and repeated interactions (Fehr and Gächter, 2002; for a meta-analysis, see Balliet et al., 2011). Thus, we also expect more cooperation in a situation with punishment than in an anonymous situation (*Hypothesis 3*).

If larger groups cooperate less when temptation increases with group size, as argued earlier, then how does punishment facilitate cooperation in larger groups? The gene-culture coevolutionary theory proposes that naturally occurring intergroup competition creates selection pressures among groups to promote ingroup cooperation and maintain group stability. Such selection pressures would favor altruistic punishment and strong reciprocity (i.e., the tendency to cooperate with others and punish free riders) that are individually costly but benefit the group, and thus allow altruistic punishment and cooperation to sustain in large groups (Bernhard et al., 2006; Henrich and Boyd, 2016; Richerson et al., 2016). In fact, the individual cost of punishing a free rider declines as the number of punishers increases in larger groups (Boyd et al., 2010). Thus, argued from this theory, punishment and social norm enforcement should be more conducive to promoting cooperation in larger (vs. smaller) groups. Indeed, norms and institutions (e.g., punishing free riders) have been shown to facilitate and maintain large-scale cooperation (Boyd et al., 2003; Henrich et al., 2010; Mathew and Boyd, 2011), and people in larger and more complex societies engage in more third-party punishment than those in small-scale societies (Marlowe et al., 2008). However, another perspective on trust and social capital would predict the

opposite—punishment more strongly promotes cooperation in smaller groups (Coleman, 1988; Portes, 1998; Ahn and Ostrom, 2008). This perspective suggests that people in high-trust groups believe that others will enforce norms (Coleman, 1988). Moreover, people consider smaller groups as more trustworthy (Wheelan, 2009; La Macchia et al., 2016), and punishment more strongly promotes cooperation in high-trust groups (Balliet and Van Lange, 2013). Thus, we will test whether punishment is more effective (*Hypothesis 3a*) or less effective (*Hypothesis 3b*) in promoting cooperation in larger groups than in smaller ones when temptation increases with group size.

Group size, as well as reputation and punishment, may affect cooperation through three proximate psychological processes: (a) expected others' cooperation (Pruitt and Kimmel, 1977), (b) perceived collective efficacy (i.e., group members' belief that they can solve their problem through collective effort; Kerr, 1989), and (c) perceived conflict of interest (Kelley et al., 2003). First, people in larger groups often feel more uncertain about others' decisions and thus show lower trust in others (Wheelan, 2009; La Macchia et al., 2016), which predicts less cooperation in social dilemmas (Pletzer et al., 2018). Second, the higher levels of anonymity and uncertainty in larger groups may weaken individuals' belief that the group can maximize the collective interest through joint effort (Kerr, 1989). Such collective efficacy belief facilitates team performance and group cooperation in highly interdependent tasks (Katz-Navon and Erez, 2005). Third, situations often vary in the degree of corresponding and conflicting interests (Kelley et al., 2003). When temptation increases with group size (i.e., fixed MGR), larger groups involve more conflict of interest, which inhibits cooperation in social interactions (Gerpott et al., 2018). Therefore, when temptation increases with group size, people may expect less cooperation from others, perceive less collective efficacy and more conflict, and thus cooperate less in relatively larger groups, and the opposite may occur when gain increases with group size. Our work will be the first to simultaneously test these psychological processes underlying the effect of group size on cooperation. In addition, we will examine whether the different effects of reputation, as well as punishment, on cooperation across group sizes can be explained by changes in any of these psychological processes.

Taken together, our major goal is to test whether reputation and punishment can promote cooperation more effectively in larger groups where there is a stronger temptation to free ride than in smaller groups. Using a one-shot public goods game, we first examine how cooperation varies with the number of group members (Study 1), and then move on to testing predictions on the roles of reputation and punishment across different group sizes (Study 2). Based on previous work (Isaac et al., 1994; Xu et al., 2013), we used groups of 4, 20, and 40 to represent relatively small, medium, and large groups. Study 1 initially tested the hypothesized group size effect when temptation (i.e., fixed MGR) or gain (i.e., fixed MPCR) increased with group size. Study 2 further compared a public goods situation involving *reputation-based partner choice* or *punishment* opportunities with a *control* condition to test how reputation and punishment promote cooperation across three group sizes. In both studies, we also tested whether differences in expected others' cooperation,

perceived collective efficacy, and perceived conflict could explain the decline in cooperation in larger groups with more temptation (i.e., fixed MGR) and/or the increase in cooperation in larger groups with more gain (i.e., fixed MPCR). Both studies were conducted without deception. All participants provided their informed consent and participated voluntarily.

STUDY 1

Materials and Methods

Participants and Design

Based on an *a priori* power analysis (Cohen, 1969, p. 348; Faul et al., 2007), a sample of 540 would result in 80% statistical power to detect a small-to-medium effect ($f = 0.15$) of group size on cooperation. We recruited 820 participants (455 women; $M_{age} = 37.87$ years, $SD = 12.15$) in the United States via Amazon Mechanical Turk (MTurk) and randomly assigned them to one of five conditions: 4-person group, 20-person group with either a fixed MGR or a fixed MPCR, and 40-person group with either a fixed MGR or a fixed MPCR¹. All participants were paid US\$1.00, and 34 of them received an extra 2-dollar bonus based on their decisions during the study.

Procedure

Participants were randomly assigned into an interacting group of 4, 20, or 40 persons, and were informed to interact with other members online in a decision-making task (i.e., a one-shot public goods game). Each person initially received 20 tokens and decided to contribute any tokens to the group account, while keeping the remaining tokens for themselves. In the 4-person groups, the total contribution was multiplied by 1.6 (i.e., MGR) and then divided equally among four persons. Thus, each person received 0.4 (i.e., MPCR) tokens from each token contributed. In the 20-person and 40-person groups, we kept either the MGR fixed at 1.6 or the MPCR fixed at 0.4. Each token that participants earned in this task represented a 0.05% chance to win a 2-dollar bonus.

To ensure that participants understood the task, they had to correctly answer five comprehension check questions with multiple choices prior to making their decisions. After they made their contribution decisions, they completed the measures of expected others' cooperation (i.e., "How many tokens on average do you think the other 3/19/39 group members will contribute to the group account?") and perceived collective efficacy (i.e., "To what extent do you believe that your group can maximize the collective earnings?"; 1 = *not at all*, 7 = *very much*). Then they completed a 30-item Situational Interdependence Scale, including six items that measured conflict ($\alpha = 0.84$; three items were reverse-coded; e.g., "Our preferred outcomes in this situation are conflicting"; Gerpott et al., 2018) on a 5-point scale (1 = *completely disagree*, 5 = *completely agree*). Their average score across the six items was the measure of perceived conflict. Finally, participants reported their age and gender. We calculated

participants' earnings of tokens based on the payoff parameters and their decisions after randomly composing them into groups of 4, 20, and 40, and then randomly selected 34 bonus winners based on their chance.

Results and Discussion

Group Size Effect on Cooperation

To test our hypotheses about the group size effect, we created four simple contrasts with the four-person group as the reference group: *medium-versus-small contrast (fixed MGR or MPCR)*, *large-versus-small contrast (fixed MGR or MPCR)*. A one-way analysis of variance (ANOVA) on cooperation revealed a significant group size effect, $F(4, 815) = 4.80$, $p = 0.001$, $\eta_p^2 = 0.023$. Further planned comparisons revealed a significant medium-versus-small contrast (fixed MGR), $F(1, 815) = 11.59$, $p = 0.001$, $\eta_p^2 = 0.014$, but the other contrasts were not statistically significant ($ps > 0.23$; see **Table 1**). This indicated that 20-person groups were significantly less cooperative than 4-person groups when temptation increased with group size (i.e., fixed MGR). However, contrary to our predictions, there was no significant difference in cooperation between 40-person and 4-person groups when temptation increased with group size (i.e., fixed MGR), and no significant group size effect on cooperation when gain increased with group size (i.e., fixed MPCR).

Expectations, Collective Efficacy, and Conflict

One-way ANOVAs revealed significant group size effects on expected others' cooperation, $F(4, 815) = 2.76$, $p = 0.03$, $\eta_p^2 = 0.01$, perceived collective efficacy, $F(4, 815) = 3.75$, $p = 0.005$, $\eta_p^2 = 0.02$, and perceived conflict, $F(4, 815) = 2.48$, $p = 0.04$, $\eta_p^2 = 0.01$. Further planned comparisons revealed that only the medium-versus-small contrast (fixed MGR) significantly predicted expected others' cooperation, $F(1, 815) = 9.52$, $p = 0.002$, $\eta_p^2 = 0.01$, perceived collective efficacy, $F(1, 815) = 9.94$, $p = 0.002$, $\eta_p^2 = 0.01$, and perceived conflict, $F(1, 815) = 6.18$, $p = 0.013$, $\eta_p^2 = 0.01$, but the other contrasts were not statistically significant ($ps > 0.07$). Specifically, when temptation increased with group size (i.e., fixed MGR), participants expected others to be less cooperative, perceived less collective efficacy, and more conflict in 20-person groups ($Ms = 9.64, 4.52, \text{ and } 2.49$) than in 4-person groups ($Ms = 11.48, 5.11, \text{ and } 2.25$), but did not vary in these measures in 40-person and 4-person groups. In addition, when gain increased with group size (i.e., fixed MPCR), there was no significant group size effect on these measures (see **Table 1**).

Mediation Analyses

We further tested whether the three measures mediated the observed group size effect on cooperation using the bootstrapping method based on 5,000 bootstrap samples (Preacher and Hayes, 2008; Hayes and Preacher, 2014). Prior to the analysis, we created four dummy variables ($D_{mediumMGR}$, $D_{mediumMPCR}$, $D_{largeMGR}$, $D_{largeMPCR}$) with the four-person group as the reference group². The relative indirect effect of

² $D_{mediumMGR}$ and $D_{mediumMPCR}$ compared the 20-person groups with the 4-person groups when MGR was fixed at 1.6 and when MPCR was fixed at 0.4, respectively. $D_{largeMGR}$ and $D_{largeMPCR}$ compared the 40-person groups with the 4-person groups when MGR was fixed at 1.6 and when MPCR was fixed at 0.4, respectively.

¹One participant who put 1989 as age was treated as age 29 at the time of data collection.

TABLE 1 | Means (and standard deviations) of the key measures in each group size condition (Study 1).

Measures	4-person group	20-person group (fixed MGR)	20-person group (fixed MPCR)	40-person group (fixed MGR)	40-person group (fixed MPCR)
Cooperation	13.63 (6.81) ^a	11.02 (7.14) ^b	14.08 (6.31) ^a	12.71 (7.40) ^{ab}	13.16 (6.90) ^a
Expected others' cooperation	11.48 (5.70) ^a	9.64 (4.96) ^b	11.00 (5.44) ^{ab}	10.87 (5.32) ^{ab}	11.08 (5.47) ^{ab}
Perceived collective efficacy	5.11 (1.64) ^a	4.52 (1.69) ^b	5.10 (1.61) ^a	4.77 (1.75) ^{ab}	5.02 (1.71) ^{ab}
Perceived conflict	2.25 (0.82) ^a	2.49 (0.88) ^a	2.27 (0.83) ^a	2.38 (0.89) ^a	2.25 (0.89) ^a
<i>n</i>	159	169	162	164	166

MGR = marginal group return, MPCR = marginal per capita return. In the 4-person groups, MGR = 1.6 and MPCR = 0.4. In the other conditions, either MGR was fixed at 1.6 or MPCR was fixed at 0.4. Means with different superscripts per row for the key measures are statistically different ($p < 0.05$) based on post hoc comparisons using Bonferroni corrections.

$D_{mediumMGR}$ on cooperation was significant through expected others' cooperation, $b = -1.14$, 95% confidence interval (CI) $[-1.87, -0.43]$, perceived collective efficacy, $b = -0.51$, 95% CI $[-0.90, -0.18]$, and perceived conflict, $b = -0.32$, 95% CI $[-0.64, -0.07]$. However, the relative indirect effects of the other three dummy variables on cooperation were not statistically significant through these measures. These results suggested that when temptation increased with group size (i.e., fixed MGR), people in 20-person groups expected less cooperation from others, perceived less collective efficacy and more conflict, and thus cooperated less than those in 4-person groups.

Overall, we only found less cooperation in 20-person (but not 40-person) groups compared with 4-person groups in one-shot public goods game when MGR was fixed. This mixed evidence only partly supported Hypothesis 1a that people cooperate less in larger groups when temptation increases with group size. Hypothesis 1b was not supported, given the null effect of group size when MPCR was fixed. Despite this, our results suggest that having a fixed MPCR across group sizes (i.e., temptation is constant but gain increases with group size) may buffer against a potential negative effect of group size on cooperative behavior. Notably, the observed lower level of one-shot cooperation in 20-person groups than in 4-person groups was explained by lower expectations of others' cooperation, less perceived collective efficacy, and more perceived conflict.

STUDY 2

Study 2 was designed to test how reputation-based partner choice and punishment promote cooperation in groups of different sizes. We also sought to replicate our findings in Study 1 on the group size effect when temptation increased with group size (i.e., fixed MGR) and its underlying psychological processes.

Materials and Methods

Participants and Design

An *a priori* power analysis revealed that a sample of 536 would result in 80% statistical power to detect a small-to-medium interaction effect ($f = 0.15$) between group size and mechanism (three conditions for each; Faul et al., 2007). We used TurkPrime to recruit 1,199 participants in the United States with no experience in our prior study (Litman et al., 2017), and randomly

assigned them to one of nine conditions of a three (group size: 4, 20, 40) \times 3 (mechanism: reputation, punishment, control) between-participants design. All participants were paid US\$1.00, and 16 of them received an extra 2-dollar bonus based on their decisions during the study. Sixty-seven participants attempted to complete the study multiple times before actually completing it and were thus exposed to instructions from different conditions. We excluded these participants from data analyses, leading to a final sample of 1,132 participants (610 women, $M_{age} = 35.05$ years, $SD = 10.91$)³.

Procedure

Participants were randomly assigned into an interacting group of 4, 20, or 40 persons, and were informed to interact with other members online in a one-shot public goods game with two options. Each person decided whether to contribute 20 tokens to a group account or keep these tokens as their own. The total contribution was multiplied by 1.6 and then divided equally among all members. Participants learned that others may make their decisions before or after them, and that decisions within the same group would be matched at the end of the study. The tokens they earned determined their chance to win a 2-dollar bonus.

Afterward, participants were randomly assigned to reputation, punishment, or control condition with different instructions. Participants in the reputation condition learned that their decision would be made public to "some or all members", and then each member would choose their preferred partner for a new task. They could not continue with this new task if no one selected them as partners (adapted from Barclay, 2004; Van Vugt and Hardy, 2010). Participants in the punishment condition were instructed that after learning about "some or all members' decisions", each member could assign up to 10 deduction points to other members. Each deduction point they assigned to others cost them one token but decreased three tokens from others (see also Fehr and Gächter, 2002). Participants in the control condition learned that their decisions were anonymous to others.

³Participants were instructed and aware that they would be screened out of the study if they fail to answer the comprehension check questions correctly after two trials. However, some participants still completed the study through the survey link after they were initially screened out. We paid these participants but excluded them from the data analyses, because they had been exposed to instructions for other conditions in the initial trials. One participant who reported her age as 6 was excluded in the descriptive analysis of age.

Prior to the decision stage, participants had to correctly answer several comprehension check questions with multiple choices within two trials. After they made their contribution decisions, they reported the expected number of cooperators (i.e., “Out of the other 3/19/39 group members, how many of them do you think will contribute 20 tokens to the group account?”), from which we calculated the expected proportion of cooperators to represent expected others’ cooperation. They also completed the measures of perceived collective efficacy (one item) and perceived conflict (six items, $\alpha = 0.87$) used in Study 1.

Different from the control condition, the reputation and punishment conditions included a second stage during which participants made four decisions assuming that three (out of 3/19/39) other members’ decisions were CCC, CCD, CDD, or DDD (C = contribute, D = do not contribute; *strategy method*)⁴. In this stage, participants in the reputation condition chose their preferred partner to interact with in a new task. In this new task, they were assigned six additional tokens and decided whether to give these tokens to the selected partner. If they did, the partner would receive the tripled amount of 18 tokens (adapted from the indirect reciprocity game in Rockenbach and Milinski, 2006); participants in the punishment condition received an additional 10 deduction points, which they could assign to other members to reduce these others’ earnings (see **Figure 1**). Finally, participants reported their age and gender. We selected 16 bonus winners based on all participants’ decisions after we randomly composed them into groups of 4, 20, and 40.

Results and Discussion

Cooperation

Overall, 73.14% of all participants cooperated by contributing 20 tokens. After dummy-coding group size into D_{medium} (20 vs. 4) and D_{large} (40 vs. 4) and mechanism into $D_{reputation}$ (reputation vs. control) and D_{punish} (punishment vs. control), we conducted a hierarchical logistic regression on cooperation (see **Table 2**). Replicating the results of Study 1, we found a significantly lower cooperation rate in 20-person groups (69.21%) than in 4-person groups (77.46%), $b = -0.44$, Wald $\chi^2(1) = 6.90$, $p = 0.009$, but no significant difference in cooperation rate between 40-person groups (73.32%) and 4-person groups (77.46%), $b = -0.24$, Wald $\chi^2(1) = 1.80$, $p = 0.18$. Supporting Hypotheses 2 and 3, both reputation-based partner choice, $b = 0.82$, Wald $\chi^2(1) = 23.04$, $p < 0.001$, and punishment, $b = 0.53$, Wald $\chi^2(1) = 11.13$, $p = 0.001$, significantly promoted cooperation.

Unexpectedly, we found no significant interactions between group size (D_{medium} and D_{large}) and mechanism (D_{punish} and $D_{reputation}$) in predicting cooperation ($ps > 0.07$; see **Table 2**). These findings suggest that both reputation-based partner choice and punishment can invariably promote cooperation in 20-person and 40-person groups, in comparison to 4-person groups.

⁴We used the strategy method (see also Barclay, 2006) with incomplete information about others’ possible decisions to increase our ecological validity, as people in their daily life situations often cannot directly observe all others’ behavior in large groups.

Previous evidence suggests that reputation promotes cooperation more than punishment (Grimalda et al., 2016; Wu et al., 2016a). However, other researchers proposed that punishment may be relatively more crucial than reputation and gossip to maintain large-scale cooperation (Jordan et al., 2013). To test both possibilities, we further compared the cooperation rates in the reputation and punishment conditions across groups of different sizes. Prior to the analysis, we re-coded mechanism into $D_{compare}$ (reputation vs. punishment) and $D_{control}$ (control vs. punishment) with the punishment condition as the reference group. A similar hierarchical logistic regression on cooperation revealed no significant effect of $D_{compare}$, $b = 0.29$, Wald $\chi^2(1) = 2.56$, $p = 0.11$, $D_{medium} \times D_{compare}$ interaction, $b = 0.19$, Wald $\chi^2(1) = 0.16$, $p = 0.69$, or $D_{large} \times D_{compare}$ interaction, $b = 0.56$, Wald $\chi^2(1) = 1.29$, $p = 0.26$. If anything, the cooperation rate was slightly (but not statistically significantly) higher in response to reputation-based partner choice (82.61%) than punishment (72.50%) in 40-person groups, $b = 0.59$, Wald $\chi^2(1) = 0.39$, $p = 0.066$ (see **Figure 2**). This pattern of results was consistent when comparing 20-person or 40-person groups with 4-person groups.

Expectations, Collective Efficacy, and Conflict

Multivariate analyses of variance revealed that group size had a significant effect on expected others’ cooperation, $F(2, 1123) = 25.52$, $p < 0.001$, $\eta_p^2 = 0.04$, perceived collective efficacy, $F(2, 1123) = 6.00$, $p = 0.003$, $\eta_p^2 = 0.01$, and perceived conflict, $F(2, 1123) = 6.56$, $p = 0.001$, $\eta_p^2 = 0.01$. Mechanism also had a significant effect on expected others’ cooperation, $F(2, 1123) = 6.53$, $p = 0.002$, $\eta_p^2 = 0.01$, perceived collective efficacy, $F(2, 1123) = 6.56$, $p = 0.001$, $\eta_p^2 = 0.01$, and perceived conflict, $F(2, 1123) = 4.20$, $p = 0.02$, $\eta_p^2 = 0.01$. However, there were no significant Group Size \times Mechanism interactions on these three measures ($ps > 0.54$). Further planned comparisons revealed that participants in the 20-person and 40-person groups expected others to be less cooperative, perceived less collective efficacy, and perceived more conflict than those in 4-person groups ($ps < 0.05$). These results slightly differed from Study 1 that only revealed significant differences in these measures between 20-person groups and 4-person ones. Moreover, compared to an anonymous situation, both reputation-based partner choice and punishment made participants expect more cooperation from others ($ps < 0.05$), but only reputation-based partner choice could enhance participants’ perceived collective efficacy and reduce conflict ($ps < 0.01$).

Mediation Analyses

Similar to Study 1, we further tested whether the observed psychological differences could explain the group size effect on cooperation while coding group size into D_{medium} and D_{large} . The relative indirect effect of D_{medium} on cooperation was significant through expected others’ cooperation, $b = -0.42$, 95% CI $[-0.63, -0.26]$, perceived collective efficacy, $b = -0.19$, 95% CI $[-0.33, -0.07]$, and perceived conflict, $b = -0.20$, 95% CI $[-0.35, -0.09]$. The relative indirect effect of D_{large} on cooperation was also significant through expected others’ cooperation, $b = -0.48$, 95% CI $[-0.69, -0.30]$, perceived collective efficacy, $b = -0.13$, 95%

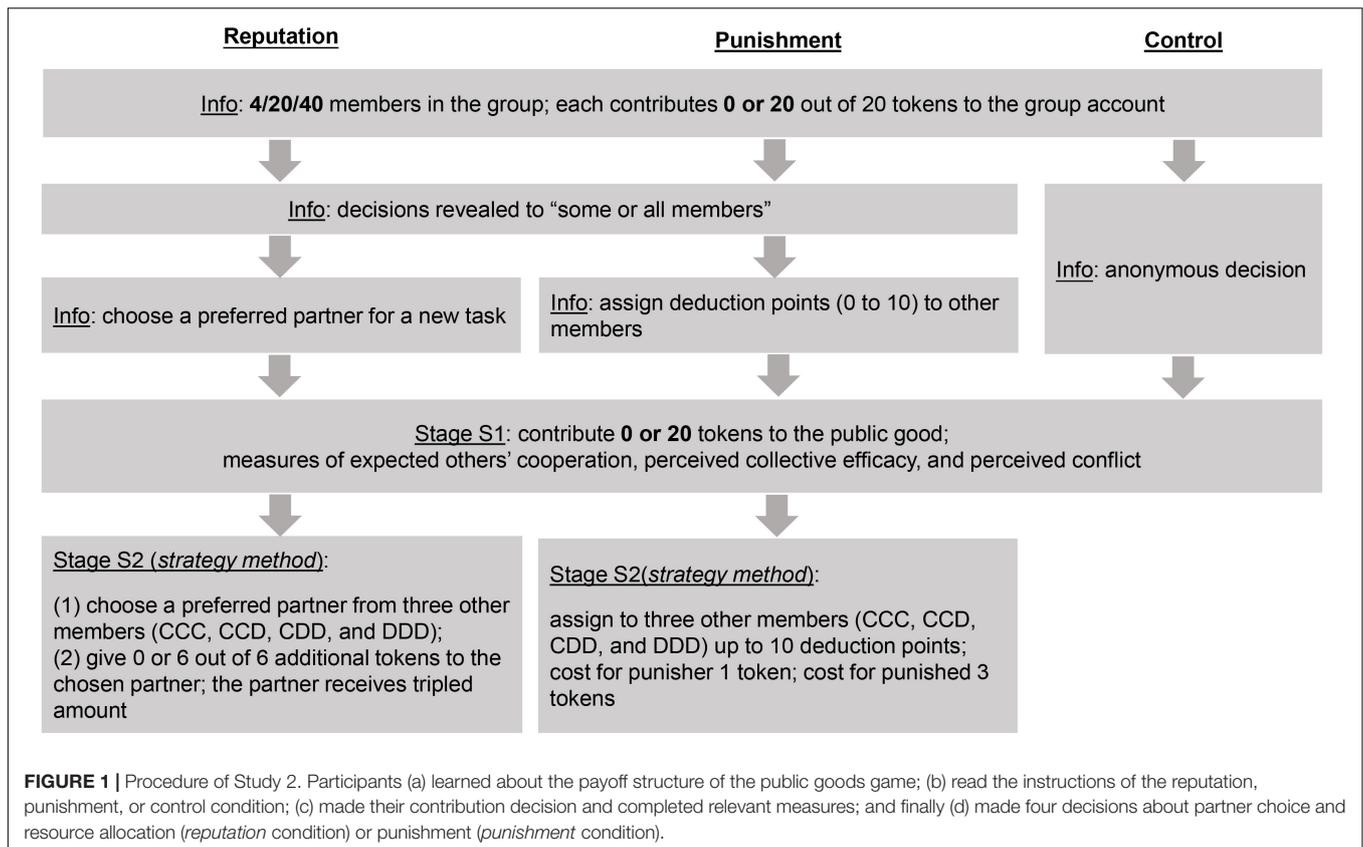


TABLE 2 | Hierarchical logistic regression on cooperation (Study 2).

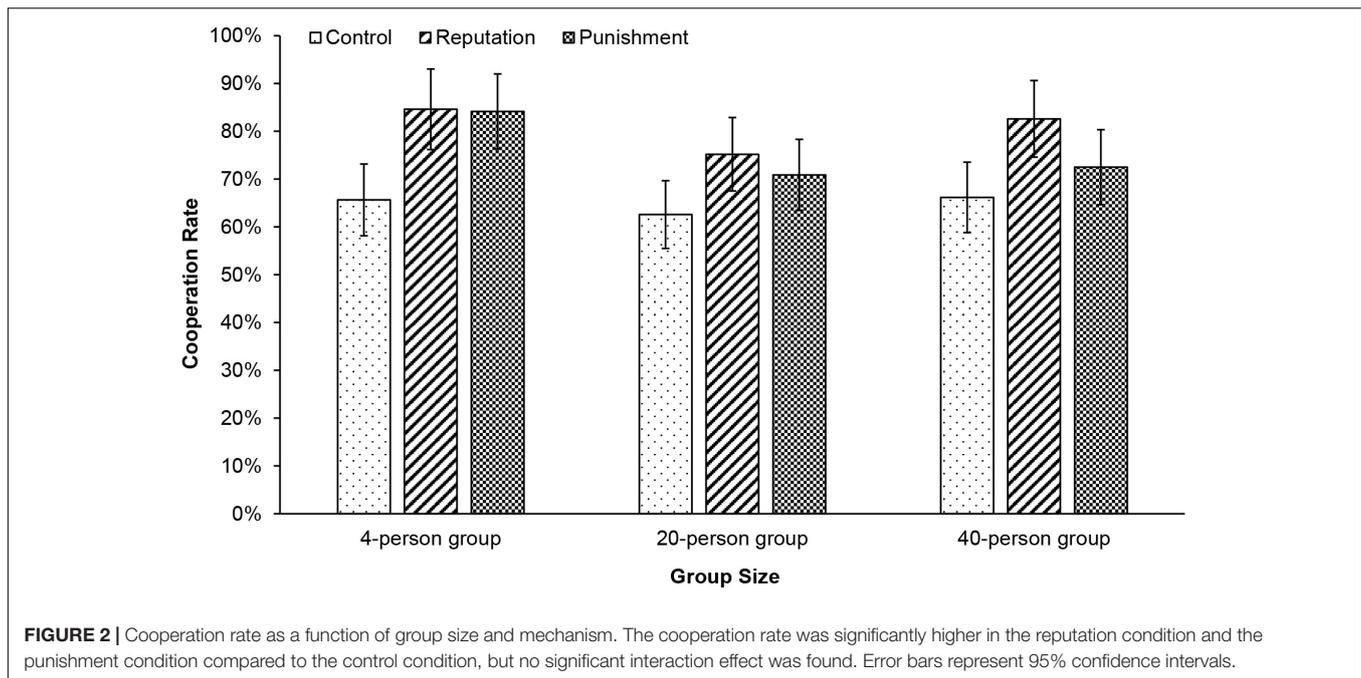
Predictor	Step 1				Step 2			
	<i>b</i>	Wald	<i>p</i>	OR [95% CI]	<i>b</i>	Wald	<i>p</i>	OR [95% CI]
Group size		6.93	0.03			0.47	0.79	
<i>D_{medium}</i>	−0.44	6.90	0.009	0.64 [0.46, 0.89]	−0.13	0.28	0.60	0.88 [0.54, 1.43]
<i>D_{large}</i>	−0.24	1.80	0.18	0.79 [0.56, 1.11]	0.02	0.008	0.93	1.02 [0.62, 1.70]
Mechanism		25.54	< 0.001			15.94	< 0.001	
<i>D_{reputation}</i>	0.82	23.04	< 0.001	2.28 [1.63, 3.18]	1.06	10.37	0.001	2.88 [1.51, 5.48]
<i>D_{punish}</i>	0.53	11.13	0.001	1.70 [1.25, 2.32]	1.02	10.86	0.001	2.78 [1.51, 5.11]
Group size × mechanism						4.42	0.35	
<i>D_{medium}</i> × <i>D_{reputation}</i>					−0.46	1.19	0.28	0.63 [0.27, 1.45]
<i>D_{medium}</i> × <i>D_{punish}</i>					−0.65	2.59	0.11	0.52 [0.24, 1.15]
<i>D_{large}</i> × <i>D_{reputation}</i>					−0.17	0.14	0.70	0.84 [0.35, 2.03]
<i>D_{large}</i> × <i>D_{punish}</i>					−0.73	3.07	0.08	0.48 [0.22, 1.09]

OR = odds ratio; CI = confidence interval. Cooperation was coded as 0 (contribute 0 tokens) and 1 (contribute 20 tokens). Group size was dummy coded into *D_{medium}* (20 vs. 4) and *D_{large}* (40 vs. 4). Mechanism was dummy coded into *D_{reputation}* (reputation vs. control) and *D_{punish}* (punishment vs. control).

CI [−0.26, −0.02], and perceived conflict, *b* = −0.12, 95% CI [−0.25, −0.005]. These findings suggest that when temptation increases with group size (i.e., fixed MGR), people expect others to be less cooperative, perceive less collective efficacy and more conflict, and thus cooperate less in relatively larger groups than in smaller ones. In addition, we found that reputation-based partner choice promoted cooperation through expected others’ cooperation, *b* = 0.24, 95% CI [0.10, 0.41], perceived collective efficacy, *b* = 0.20, 95% CI [0.09, 0.34], and perceived conflict,

b = 0.16, 95% CI [0.05, 0.30], whereas punishment promoted cooperation only through expected others’ cooperation, *b* = 0.15, 95% CI [0.009, 0.31].

Taken together, Study 2 replicated the findings of Study 1 that when temptation increased with group size (i.e., fixed MGR), people cooperated less in 20-person groups than in 4-person groups in one-shot public goods game, and that this was because they expected others to be less cooperative, perceived less collective efficacy and more conflict in 20-person groups than



in 4-person groups. Moreover, both reputation-based partner choice and punishment strongly promoted one-shot cooperation regardless of the size of groups.

GENERAL DISCUSSION

Over the past decades, several reviews of social dilemmas have suggested that cooperation tends to decrease with the size of the interacting group (e.g., Dawes, 1980; Van Lange et al., 2013). However, despite some studies that support this conclusion (Suzuki and Akiyama, 2005; Wheelan, 2009), other studies suggest that cooperation increases with group size (Isaac et al., 1994; Carpenter, 2007; Barcelo and Capraro, 2015) or does not relate to group size (Kerr, 1989; Zelmer, 2003). Our research contributed insights into the ongoing debate on the “how and why” of the association between group size and cooperation by taking into account the potential changes in the incentive structure across groups. More importantly, we examined the effective strategies that promote cooperation in groups of different sizes. We addressed these questions across two studies using one-shot public good games. In Study 1, we observed participants’ cooperation in an interacting group of 4, 20, or 40 while keeping the MGR or MPCR fixed across groups, in which case the temptation or gain increased with group size. In Study 2, participants interacting in groups of 4, 20, or 40 with a fixed MGR were further randomly assigned to reputation, punishment, or control condition. This setting allowed us to test hypotheses about the effectiveness of reputation-based partner choice and punishment in relatively small and large groups when the MGR was fixed. Overall, we provide novel evidence that when temptation increases with group size (i.e., fixed MGR), 20-person (but not 40-person) groups cooperate

less than 4-person groups in one-shot interactions. Moreover, both reputation-based partner choice and punishment can promote one-shot cooperation in groups of different sizes when temptation increases with group size.

Our first goal was to test whether larger groups would cooperate less when temptation increased with group size (H1a), but cooperate more when gain increased with group size (H1b). Partially supporting Hypothesis 1a, when temptation increased with group size, 20-person groups cooperated significantly less than 4-person groups, yet surprisingly, 40-person and 4-person groups did not vary in cooperation. Unexpectedly, inconsistent with previous research (Barcelo and Capraro, 2015; Shank et al., 2015), we found no statistical difference in cooperation across groups of 4, 20, and 40 when gain increased with group size (i.e., fixed MPCR). Thus, keeping the MPCR fixed may buffer against potential negative consequences of larger groups for cooperation and collective action.

It remains unclear why 40-person and 4-person groups did not vary in cooperation when temptation increased with group size. There are several potential explanations for this unexpected result. First, compared with the 20-person and 4-person groups, the payoff structure in the 40-person group situation may be more cognitively demanding, and thus requires longer decision time and may elicit more dropouts from the study, which may confound with the group size effect. Although we have no data confirming such dropouts, the non-significant difference in the survey completion time across group size conditions ($ps = 0.93$ and 0.08 in Studies 1 and 2) ruled out this alternative explanation. Second, we observed relatively higher levels of cooperation (64.5% of tokens contributed in Study 1 and 73.14% cooperators in Study 2) compared to previous research (e.g., 37.7% of total endowment; for a meta-analysis, see Zelmer, 2003), which may suggest a potential selection bias. Specifically, we speculate that

participants are more likely to drop out when interacting in the 40-person groups due to more conflict of interest and lower collective efficacy, and those dropouts are more likely to be low cooperators (e.g., prosely individuals who generally prioritize their own interest), while those who completed the survey (especially in the 40-person groups) are more likely to be high cooperators (e.g., prosocial individuals who generally care about the collective good and equality; Van Lange, 1999; Balliet et al., 2009). Since we had no behavioral measures for those dropouts or other relevant measures (e.g., social value orientation, Van Lange, 1999) for all participants, our studies left open this account that future research should address.

Nevertheless, our findings suggest no simple relation between group size, incentives, and cooperation. Although we did not consider all possible group sizes and thus could not determine the optimal size of group that yields the highest level of cooperation, some previous studies can provide insights into this question. For example, Capraro and Barcelo (2015) assigned participants into 12 groups of different sizes (from 3 to 100) and found that 15-person groups were the most cooperative when gain (i.e., benefit for each person when all cooperate than when no one cooperates) increased linearly till groups of 20 and then remained constant. This finding slightly deviates from the argument that larger groups cooperate more when gain increases with group size, which would predict 20-person groups (instead of 15-person groups) to be the most cooperative in this setting. Other field research similarly suggests a non-linear effect of group size on cooperation (e.g., Agrawal and Goyal, 2001; Yang et al., 2013). Importantly, outside experimental settings, many social and ecological factors can drive the optimal size of a group that can manage resources successfully (Casari and Tagliapietra, 2018). Thus, people may not explicitly calculate the costs and benefits of cooperation, especially in complex situations (e.g., larger groups). However, the observed patterns of cooperative behavior across group sizes in our studies were largely consistent with those in the proposed psychological processes underlying cooperation (i.e., expected others' cooperation, perceived collective efficacy, and perceived conflict). In particular, when temptation increased with group size, people in 20-person groups expected less cooperation from others, had lower belief that their group could maximize the collective interest, and perceived more conflict compared with those in 4-person groups.

Our second goal was to replicate the effect of reputation on cooperation (Feinberg et al., 2014; Wu et al., 2016b) and test hypotheses about how reputation fosters cooperation in relatively small and large groups. One perspective predicts that reputation is more effective in larger groups because people prefer the best cooperators as partners, such that larger groups contain more competition for reputation and less chance to be selected as partners (Van Vugt et al., 2007; Barclay, 2013). Another perspective mainly based on computer simulations posits that larger groups are more vulnerable to free riding, so reputation is less capable of maintaining cooperation (Suzuki and Akiyama, 2007). Our findings suggested that reputation-based partner choice greatly enhanced cooperation and that this positive effect invariably occurred in both relatively small and large groups. Moreover, compared to an anonymous situation, having a

reputation mechanism promoted cooperation through eliciting higher expectations of others' cooperation, greater perceived collective efficacy, and less perceived conflict. Notably, these findings are based on an experimental manipulation of reputation that involves an extra partner choice stage with a resource allocation game. Thus, they do not necessarily contradict with findings from computer simulations that reputation-based cooperation becomes difficult to evolve as group size increases (Suzuki and Akiyama, 2005). Indeed, in simulations that allowed agents to condition their behavior on others' reputations across many generations, defecting with others whose reputation is bad may make reputation-based cooperation less likely to evolve in larger groups, as one's defection and bad reputation may elicit others' defection in turn (Suzuki and Akiyama, 2005). In contrast, participants in our research competed to be chosen by others in the reputation-based partner choice stage. To attract potential partners in this stage, it is important for participants to behave more cooperatively than others, especially in larger groups with more competition over a good reputation.

Our final goal was to replicate the effect of punishment on cooperation (Fehr and Gächter, 2002; Balliet et al., 2011) and test hypotheses on how punishment enhances cooperation in relatively small and large groups. The gene-culture coevolutionary theory suggests that punishment should more strongly promote cooperation in larger groups (Boyd et al., 2003; Henrich et al., 2010), yet it is also plausible that punishment loses its effectiveness in larger groups (Coleman, 1988; Ahn and Ostrom, 2008). We found that punishment invariably promotes cooperation in relatively small to larger groups, which did not support either prediction. Moreover, punishment promoted cooperation only through enhancing one's belief that others would cooperate. Similarly, previous research shows that public good contribution does not decrease with a larger group size as long as participants can sufficiently monitor and punish many other group members (Carpenter, 2007). Moreover, punishment, at least when it remains intact, induces more trust in others being externally motivated to cooperate (Mulder et al., 2006). Interestingly, we found that the punishment-cooperation relation (odds ratio = 1.70) had a smaller effect size than the reputation-cooperation relation (odds ratio = 2.28). Future research may examine the optimal size of group where reputation can more effectively enhance cooperation and group welfare than punishment.

Strengths, Limitations, and Directions for Future Research

The present work has some methodological strengths, limitations, and implications for future research. First, due to the inherent difficulties in recruiting people from organizations of different sizes and conducting such a large-scale interaction study in the lab, we chose to recruit relatively large and diverse samples through MTurk and manipulate group size in the experimental setting. Although this might lower our ecological validity, the anonymous online environment guarantees that only group size and payoff parameters (i.e., fixed MGR or MPCR) are the salient cues that participants rely on while making their

decisions. Moreover, participants in our studies had a chance to win an extra bonus that was determined by their own and others' decisions after we randomly composed them into groups at the end of the study. This setup provided them with real incentives to weigh their own interest against others' interests. One potential limitation of our studies was that participants could not feel others' physical presence or learn about others' actual decisions when interacting online. However, some evidence suggests that MTurk participants behave as if their partners are real even when doing so involves a financial cost, and are sensitive to subtle cues about their partners (Summerville and Chartier, 2013), which supports the plausibility of conducting social interaction experiments on MTurk. Moreover, many classic "group effects" (e.g., intergroup contact, bystander effect) have been observed without the physical presence of a group (Garcia et al., 2002; Crisp et al., 2009). Thus, others' physical presence is not necessary to study group effects on cognition, motivation, and behavior. Nevertheless, future research would enrich our findings by observing real-time interactions in the lab or organizations involving groups of different sizes.

Second, to simplify the procedure in Study 2, we only asked participants to interact once (instead of multiple times) in groups of different sizes, during which they could choose their preferred partner for a new task, punish others after their decisions, or have no other option. In such settings, participants may cooperate due to a motivation to establish a good reputation that increases their chance to be selected by others, or due to fear of punishment. However, when a group of people interact repeatedly, they may also respond differently to others' behaviors. For example, people may adjust their subsequent cooperation after being (un)selected or punished by others. Moreover, the potential occurrence of antisocial punishment (i.e., punishing people who behave prosocially; Herrmann et al., 2008) and retaliation (i.e., counter-punishing in response to a punishment; Niki forakis, 2008) in repeated interactions might make cooperation break down. In addition, the use of costly punishment may undermine the collective welfare in repeated interactions (Dreber et al., 2008). Future research needs to test how reputation and punishment may affect cooperation and group efficiency across group sizes in repeated interactions.

Third, different reputation-based strategies (e.g., defecting against or ostracizing free riders) and punishment strategies (e.g., centralized vs. decentralized) may affect cooperation differently. Previous evidence suggests that the possibility of being evaluated may not affect cooperation (Capraro et al., 2016), while reputation-based partner choice (i.e., competitive altruism) can promote cooperation (Sylwester and Roberts, 2013; Giardini et al., 2014), but the strategy of defecting with free riders undermines cooperation (Giardini et al., 2014). Thus, the opportunities to form coalitions with cooperators and to ostracize free riders from one's current group based on others' reputation may more robustly promote cooperation than strategies like defecting with free riders. Moreover, punishment can be implemented in a centralized way such that people who contribute less than others, or below a threshold, pay a fine (Kamijo et al., 2014), or in a decentralized way (i.e., group members punish each other). Although centralized punishment

is shown to be less effective than decentralized punishment in promoting cooperation (Balliet et al., 2011), the former may be easier to implement and can prevent anti-social punishment and retaliation. Moreover, punishment can be executed by people who cooperate or free ride (Helbing et al., 2010) and can vary in the punishment fine and probability of occurrence (Chen et al., 2015), which may affect the sustainability of cooperation. Notably, different strategies may complement each other in real-life situations. For example, people may coordinate their punishment through gossip with other punishers (Boyd et al., 2010), or switch between punishment and social exclusion in response to free riders (Liu et al., 2018). Thus, it is imperative for future research to test how different reputation-based strategies and punishment strategies enhance cooperation and how they complement each other in groups of different sizes over repeated interactions.

Finally, we only focused on three group sizes with a fixed parameter of MGR or MPCR, while allowing the other to vary. However, different combinations of group sizes and payoff parameters may elicit different perceptions and decisions. Indeed, previous research shows that larger groups cooperate more when MPCR is 0.3, but such effect disappears or is even reversed when MPCR is 0.75 (Isaac et al., 1994; Nosenzo et al., 2015). This pattern of results may be explained by the greater afforded opportunity to exploit others as the individual payoff from each unit of contribution (i.e., MPCR) increases. That is, when people perceive less conflict of interest and believe that others will cooperate, some of them may take this chance to exploit others by withholding their own resources and harvesting the benefit from the group. Future research can manipulate both group size and MPCR to test their potential interaction in predicting cooperation. Moreover, the ways that people are interrelated (i.e., network structure) in groups of different sizes may function differently. For example, people may have small circles of friends for private conversations but larger networks of acquaintances or colleagues for completing large projects (Dunbar, 2004). Variations in network structure and their functions in social groups may affect trust, collective efficacy, and the effective strategies that promote cooperation (Santos et al., 2008; Jiang et al., 2011; Apicella et al., 2012). We believe that future research that combines network structure with reputation and punishment in groups of different sizes would provide useful insights into explaining cooperation within social communities.

CONCLUDING REMARKS

Our research is among the first attempts to test how reputation-based partner choice and punishment foster cooperation in relatively small and large groups. We find that when the temptation to free ride increases with group size, people cooperate less in 20-person (but not 40-person) groups than in 4-person groups in one-shot interactions, which is explained by lower expectations of others' cooperation, lower perceived collective efficacy, and higher perceived conflict in the interaction. Notably, both reputation-based

partner choice and punishment invariably promote one-shot cooperation in groups of 4, 20, and 40 persons, which supports their general effectiveness in promoting cooperation (at least in one-shot interactions). Thus, even in fairly large groups where direct reciprocity has its limits, punishment and reputation mechanisms are prominent solutions that enhance cooperation. We also provide some tentative insights into the psychological mechanisms underlying the effects of reputation and punishment. That is, punishment enhances the diminished trust (i.e., expectation about others' cooperation) that occurs in larger groups and thus may foster cooperation. It is worth mentioning that reputation-based partner choice promotes cooperation through enhancing both trust and collective efficacy, and reducing perceived conflict across groups of different sizes. These findings provide important insights into how people perceive social interactions involving groups of different sizes and the effective measures that can be taken to promote cooperation in these groups. Taken together, our findings suggest that there is no simple relation between group size and cooperation, but when the temptation to free ride increases with group size in one-shot interactions, reputation-based partner choice and punishment are both effective in promoting cooperation.

DATA AVAILABILITY STATEMENT

The datasets and syntax for the two studies can be found in the Open Science Framework: <https://osf.io/qvys6/>.

REFERENCES

- Agrawal, A., and Goyal, S. (2001). Group size and collective action: third-party monitoring in common-pool resources. *Comp. Polit. Stud.* 34, 63–93. doi: 10.1177/0010414001034001003
- Ahn, T. K., and Ostrom, E. (2008). "Social capital and collective action," in *The Handbook of Social Capital*, eds D. Castiglione, J. W. Van Deth, and G. Wolleb, (New York, NY: Oxford University Press), 70–100.
- Apicella, C. L., Marlowe, F. W., Fowler, J. H., and Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature* 481, 497–501. doi: 10.1038/nature10736
- Balliet, D., Mulder, L. B., and Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* 137, 594–615. doi: 10.1037/a0023489
- Balliet, D., Parks, C., and Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Process. Intergroup Relat.* 12, 533–547. doi: 10.1177/1368430209105040
- Balliet, D., and Van Lange, P. A. M. (2013). Trust, punishment, and cooperation across 18 societies: a meta-analysis. *Perspect. Psychol. Sci.* 8, 363–379. doi: 10.1177/1745691613488533
- Barcelo, H., and Capraro, V. (2015). Group size effect on cooperation in one-shot social dilemmas. *Sci. Rep.* 5:7937. doi: 10.1038/srep07937
- Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the "tragedy of the commons. *Evol. Hum. Behav.* 25, 209–220. doi: 10.1016/j.evolhumbehav.2004.04.002
- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evol. Hum. Behav.* 27, 325–344. doi: 10.1016/j.evolhumbehav.2006.01.003
- Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evol. Hum. Behav.* 34, 164–175. doi: 10.1016/j.evolhumbehav.2013.02.002

ETHICS STATEMENT

The studies were reviewed and approved by the Ethics Committee of Faculty of Psychology at Beijing Normal University. Participants provided their informed consent prior to taking part in the studies, and had the opportunity to withdraw at any time during the studies.

AUTHOR CONTRIBUTIONS

JW, DB, and PV had the initial idea for the studies and designed the studies. JW collected the data, conducted the data analyses, and wrote the first draft of the manuscript with the generous input by DB, LP, AR, and PV.

FUNDING

This research received support from the International Postdoctoral Exchange Fellowship Program by China Postdoctoral Council and the National Natural Science Foundation of China (71901028).

ACKNOWLEDGMENTS

We thank Yu Kou for helpful comments on an earlier draft of this manuscript.

- Barclay, P., and Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proc. Biol. Sci.* 274, 749–753. doi: 10.1098/rspb.2006.0209
- Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature* 442, 912–915. doi: 10.1038/nature04981
- Bonacich, P., Shure, G. H., Kahan, J. P., and Meeker, R. J. (1976). Cooperation and group size in the n-person prisoners' dilemma. *J. Conflict Resol.* 20, 687–706. doi: 10.1177/002200277602000406
- Boyd, R., Gintis, H., and Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328, 617–620. doi: 10.1126/science.1183665
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3531–3535.
- Capraro, V., and Barcelo, H. (2015). Group size effect on cooperation in one-shot social dilemmas II: curvilinear effect. *PLoS One* 10:e0131419. doi: 10.1371/journal.pone.0131419
- Capraro, V., Giardini, F., Vilone, D., and Paolucci, M. (2016). Partner selection supported by opaque reputation promotes cooperative behavior. *Judgm. Decis. Mak.* 11, 589–600.
- Carpenter, J. P. (2007). Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. *Games Econ. Behav.* 60, 31–51. doi: 10.1016/j.jgeb.2006.08.011
- Casari, M., and Tagliapietra, C. (2018). Group size in social-ecological systems. *Proc. Natl. Acad. Sci. U.S.A.* 115, 2728–2733. doi: 10.1073/pnas.1713496115
- Chen, X., Szolnoki, A., and Perc, M. (2015). Competition and cooperation among different punishing strategies in the spatial public goods game. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 92:012819. doi: 10.1103/PhysRevE.92.012819
- Chudek, M., and Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cogn. Sci.* 15, 218–226. doi: 10.1016/j.tics.2011.03.003
- Clutton-Brock, T. H., and Parker, G. A. (1995). Punishment in animal societies. *Nature* 373, 209–216. doi: 10.1038/373209a0

- Cohen, J. (1969). *Statistical Power Analysis for the Behavioural Sciences*. New York, NY: Academic Press.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *Am. J. Sociol.* 94, 95–120. doi: 10.1086/228943
- Crisp, R. J., Stathi, S., Turner, R. N., and Husnu, S. (2009). Imagined intergroup contact: theory, paradigm and practice. *Soc. Personal. Psychol. Compass* 3, 1–18. doi: 10.1111/j.1751-9004.2008.00155.x
- Dawes, R. M. (1980). Social dilemmas. *Annu. Rev. Psychol.* 31, 169–193. doi: 10.1146/annurev.ps.31.020180.001125
- dos Santos, M., and Wedekind, C. (2015). Reputation based on punishment rather than generosity allows for evolution of cooperation in sizable groups. *Evol. Hum. Behav.* 36, 59–64. doi: 10.1016/j.evolhumbehav.2014.09.001
- Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature* 452, 348–351. doi: 10.1038/nature06723
- Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Rev. Gen. Psychol.* 8, 100–110. doi: 10.1037/1089-2680.8.2.100
- Durand, J. D. (1977). Historical estimates of world population: an evaluation. *Popul. Dev. Rev.* 3, 253–296. doi: 10.2307/1971891
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87. doi: 10.1016/S1090-5138(04)00005-4
- Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* 13, 1–25. doi: 10.1007/s12110-002-1012-7
- Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a
- Feinberg, M., Willer, R., and Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* 25, 656–664. doi: 10.1177/0956797613510184
- Garcia, S. M., Weaver, K., Moskowitz, G. B., and Darley, J. M. (2002). Crowded minds: the implicit bystander effect. *J. Personal. Soc. Psychol.* 83, 843–853. doi: 10.1037/0022-3514.83.4.843
- Gerpott, F. H., Balliet, D., Columbus, S., Molho, C., and de Vries, R. E. (2018). How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *J. Personal. Soc. Psychol.* 115, 716–742. doi: 10.1037/pspp0000166
- Giardini, F., Paolucci, M., Villatoro, D., and Conte, R. (2014). “Punishment and gossip: sustaining cooperation in a public goods game,” in *Advances in Social Simulation*, eds B. Kamiński, and G. Koloch, (Berlin: Springer-Verlag), 107–118. doi: 10.1007/978-3-642-39829-2_10
- Gintis, H., Smith, E. A., and Bowles, S. (2001). Costly signaling and cooperation. *J. Theor. Biol.* 213, 103–119. doi: 10.1006/jtbi.2001.2406
- Grimalda, G., Ponderfer, A., and Tracer, D. P. (2016). Social image concerns promote cooperation more than altruistic punishment. *Nat. Commun.* 7:12288. doi: 10.1038/ncomms12288
- Hayes, A. F., and Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *Br. J. Math. Stat. Psychol.* 67, 451–470. doi: 10.1111/bmsp.12028
- Helbing, D., Szolnoki, A., Perc, M., and Szabó, G. (2010). Punish, but not too hard: how costly punishment spreads in the spatial public goods game. *New J. Phys.* 12:083005. doi: 10.1088/1367-2630/12/8/083005
- Henrich, J., and Boyd, R. (2016). How evolved psychological mechanisms empower cultural group selection. *Behav. Brain Sci.* 33, 28–29. doi: 10.1017/S0140525X15000138
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327, 1480–1484. doi: 10.1126/science.1182238
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science* 319, 1362–1367. doi: 10.1126/science.1153808
- Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group size and the voluntary provision of public goods: experimental evidence utilizing large groups. *J. Public Econ.* 54, 1–36. doi: 10.1016/0047-2727(94)90068-X
- Jiang, L. L., Perc, M., Wang, W. X., Lai, Y. C., and Wang, B. H. (2011). Impact of link deletions on public cooperation in scale-free networks. *Europhys. Lett.* 93:40001. doi: 10.1209/0295-5075/93/40001
- Jordan, F. M., van Schaik, C., François, P., Gintis, H., Haun, D. B. M., Hruschka, D. J., et al. (2013). “Cultural evolution of the structure of human groups,” in *Cultural Evolution: Society, Technology, Language, and Religion*, eds P. J. Richerson, and M. H. Christiansen, (Cambridge, MA: MIT Press), 87–116. doi: 10.7551/mitpress/9780262019750.003.0006
- Kamijo, Y., Nihonsugi, T., Takeuchi, A., and Funaki, Y. (2014). Sustaining cooperation in social dilemmas: comparison of centralized punishment institutions. *Games Econ. Behav.* 84, 180–195. doi: 10.1016/j.geb.2014.01.002
- Katz-Navon, T. Y., and Erez, M. (2005). When collective- and self-efficacy affect team performance: the role of task interdependence. *Small Group Res.* 36, 437–465. doi: 10.1177/1046496405275233
- Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., and Van Lange, P. A. M. (2003). *An Atlas of Interpersonal Situations*. Cambridge: Cambridge University Press.
- Kerr, N. L. (1989). Illusions of efficacy: the effects of group size on perceived efficacy in social dilemmas. *J. Exp. Soc. Psychol.* 25, 287–313. doi: 10.1016/0022-1031(89)90024-3
- La Macchia, S. T., Louis, W. R., Hornsey, M. J., and Leonardelli, G. J. (2016). In small we trust: lay theories about small and large groups. *Personal. Soc. Psychol. Bull.* 42, 1321–1334. doi: 10.1177/0146167216657360
- Litman, L., Robinson, J., and Abberbock, T. (2017). TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behav. Res. Methods* 49, 433–442. doi: 10.3758/s13428-016-0727-z
- Liu, L., Wang, S., Chen, X., and Perc, M. (2018). Evolutionary dynamics in the public goods games with switching between punishment and exclusion. *Chaos* 28:103105. doi: 10.1063/1.5051422
- Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., et al. (2008). More ‘altruistic’ punishment in larger societies. *Proc. Biol. Sci.* 275, 587–592. doi: 10.1098/rspb.2007.1517
- Mathew, S., and Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl. Acad. Sci.* 108, 11375–11380. doi: 10.1073/pnas.1105604108
- Milinski, M., Semmann, D., and Krambeck, H.-J. (2002). Reputation helps solve the “tragedy of the commons.” *Nature* 415, 424–426. doi: 10.1038/415424a
- Mulder, L. B., Van Dijk, E., De Cremer, D., and Wilke, H. A. (2006). Undermining trust and cooperation: the paradox of sanctioning systems in social dilemmas. *J. Exp. Soc. Psychol.* 42, 147–162. doi: 10.1016/j.jesp.2005.03.002
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Public Econ.* 92, 91–112. doi: 10.1016/j.jpubeco.2007.04.008
- Nosenzo, D., Quercia, S., and Sefton, M. (2015). Cooperation in small groups: the effect of group size. *Exp. Econ.* 18, 4–14. doi: 10.1007/s10683-013-9382-8
- Nowak, M. A., and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature* 437, 1291–1298. doi: 10.1038/nature04131
- Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.
- Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: self-governance is possible. *Am. Pol. Sci. Rev.* 86, 404–417. doi: 10.2307/1964229
- Panchanathan, K., and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432, 499–502. doi: 10.1038/nature02978
- Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., and Szolnoki, A. (2017). Statistical physics of human cooperation. *Phys. Rep.* 687, 1–51. doi: 10.1016/j.physrep.2017.05.004
- Pletzer, J. L., Balliet, D. P., Joireman, J., Kuhlman, D. M., Voelpel, S., and Van Lange, P. A. M. (2018). Social value orientation, expectations, and cooperation: a meta-analysis. *Eur. J. Pers.* 32, 62–83. doi: 10.1002/per.2139
- Portes, A. (1998). Social capital: its origins and applications in modern sociology. *Annu. Rev. Sociol.* 24, 1–24. doi: 10.1146/annurev.soc.24.1.1
- Preacher, K. J., and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891. doi: 10.3758/BRM.40.3.879
- Pruitt, D. G., and Kimmel, M. J. (1977). Twenty years of experimental gaming: critique, synthesis, and suggestions for the future. *Annu. Rev. Psychol.* 28, 363–392. doi: 10.1146/annurev.ps.28.020177.002051
- Rand, D. G., and Nowak, M. A. (2013). Human cooperation. *Trends Cogn. Sci.* 17, 413–425. doi: 10.1016/j.tics.2013.06.003

- Richerson, P., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., et al. (2016). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behav. Brain Sci.* 39:e30. doi: 10.1017/S0140525X1400106X
- Rockenbach, B., and Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444, 718–723. doi: 10.1038/nature05229
- Santos, F. C., Santos, M. D., and Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature* 454, 213–216. doi: 10.1038/nature06940
- Shank, D. B., Kashima, Y., Saber, S., Gale, T., and Kirley, M. (2015). Dilemma of dilemmas: how collective and individual perspectives can clarify the size dilemma in voluntary linear public goods dilemmas. *PLoS One* 10:e0120379. doi: 10.1371/journal.pone.0120379
- Smith, E. A., and Bliege Bird, R. L. (2000). Turtle hunting and tombstone opening: public generosity as costly signaling. *Evol. Hum. Behav.* 21, 245–261. doi: 10.1016/S1090-5138(00)00031-3
- Summerville, A., and Chartier, C. (2013). Pseudo-dyadic “interaction” on Amazon’s Mechanical Turk. *Behav. Res. Methods* 45, 116–124. doi: 10.3758/s13428-012-0250-9
- Suzuki, S., and Akiyama, E. (2005). Reputation and the evolution of cooperation in sizable groups. *Proc. Biol. Sci.* 272, 1373–1377. doi: 10.1098/rspb.2005.3072
- Suzuki, S., and Akiyama, E. (2007). Evolution of indirect reciprocity in groups of various sizes and comparison with direct reciprocity. *J. Theor. Biol.* 245, 539–552. doi: 10.1016/j.jtbi.2006.11.002
- Sylwester, K., and Roberts, G. (2013). Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evol. Hum. Behav.* 34, 201–206. doi: 10.1016/j.evolhumbehav.2012.11.009
- Szolnoki, A., and Perc, M. (2011). Group-size effects on the evolution of cooperation in the spatial public goods game. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 84:047102. doi: 10.1103/PhysRevE.84.047102
- Tomasello, M., and Vaish, A. (2013). Origins of human cooperation and morality. *Annu. Rev. Psychol.* 64, 231–255. doi: 10.1146/annurev-psych-113011-143812
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755
- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *J. Personal. Soc. Psychol.* 77, 337–349. doi: 10.1037/0022-3514.77.2.337
- Van Lange, P. A. M., Joireman, J., Parks, C. D., and Van Dijk, E. (2013). The psychology of social dilemmas: a review. *Organ. Behav. Hum. Dec. Process.* 120, 125–141. doi: 10.1016/j.obhdp.2012.11.003
- Van Vugt, M., and Hardy, C. L. (2010). Cooperation for reputation: wasteful contributions as costly signals in public goods. *Group Process. Intergroup Rel.* 13, 101–111. doi: 10.1177/1368430209342258
- Van Vugt, M., Roberts, G., and Hardy, C. (2007). “Competitive altruism: development of reputation-based cooperation in groups,” in *Handbook of Evolutionary Psychology*, eds R. Dunbar, and L. Barrett, (Oxford: Oxford University Press), 531–540.
- Wheelan, S. A. (2009). Group size, group development, and group productivity. *Small Group Res.* 40, 247–262. doi: 10.1177/1046496408328703
- Wu, J., Balliet, D., and Van Lange, P. A. M. (2016a). Gossip versus punishment: the efficiency of reputation to promote and maintain cooperation. *Sci. Rep.* 6:23919. doi: 10.1038/srep23919
- Wu, J., Balliet, D., and Van Lange, P. A. M. (2016b). Reputation, gossip, and human cooperation. *Soc. Personal. Psychol. Compass* 10, 350–364. doi: 10.1111/spc3.12255
- Xu, B., Cadsby, C. B., Fan, L., and Song, F. (2013). Group size, coordination, and the effectiveness of punishment in the voluntary contributions mechanism: an experimental investigation. *Games* 4, 89–105. doi: 10.3390/g4010089
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *J. Personal. Soc. Psychol.* 51, 110–116. doi: 10.1037/0022-3514.51.1.110
- Yamagishi, T., and Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Soc. Psychol. Q.* 63, 116–132. doi: 10.2307/2695887
- Yang, W., Liu, W., Viña, A., Tuanmu, M. N., He, G., Dietz, T., et al. (2013). Nonlinear effects of group size on collective action and resource outcomes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10916–10921. doi: 10.1073/pnas.1301733110
- Zelmer, J. (2003). Linear public goods experiments: a meta-analysis. *Exp. Econ.* 6, 299–310. doi: 10.1023/A:1026277420119

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wu, Balliet, Peperkoorn, Romano and Van Lange. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.