Monetary and Non-Monetary Punishment in Public Goods Games: Comparing Teams with Individuals*

Mir Adnan Mahmood
Department of Economics
The Ohio State University

Christina Gore
Department of Agricultural, Environmental and Development Economics
The Ohio State University

John H. Kagel
Department of Economics
The Ohio State University

9/30/2022

Abstract

Results from an experiment comparing monetary and non-monetary punishment (sanctions) in a voluntary contribution mechanism (VCM) public good game are reported. Monetary punishment (MP) increased contribution rates more for teams than for individuals, with teams targeting three times as much punishment to below average contributors compared to individuals. Non-monetary punishment (NMP) increased contribution rates for individuals but had no significant effect for teams. Absent punishment, teams had significantly greater reductions in end game contributions than individuals.

Key words: Voluntary contribution mechanism, teams and individuals, monetary and non-monetary punishment.
JEL codes: C72, C73, C92, H41

*We thank Maddie Miller and Tiffany Tang for research assistance. PJ Healy, OSub Kwon, John Rehbeck, Yaron Azrieli, Jason Tayawa, Matt Weinberg and Hal Arkes provided helpful comments and suggestions, as did participants at the 2021 Midwest Economic Association Conference. We thank Caleb Cox and Marcel Thum whose zTree code which was modified for this research and Charles Noussair for the instructions from Masclet et al. (2003) for nonmonetary punishment. Research support was provided by the Decision Sciences Collaborative at The Ohio State University, the JMCB small research grant and NSF grants SES 1630288 and SES 2020704. Opinions, findings, conclusions, or recommendations offered here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

In voluntary contribution mechanism (VCM) public good games, monetary punishment has been shown to substantially increase contributions to the public good.[1] Masclet et al. (2003) extended this result, demonstrating that non-monetary punishment (sanctions) also increased contribution rates. These results have typically employed individual decision makers (DMs). The present experiment extends this line of inquiry to two person teams, comparing results between teams and individuals with respect to both monetary punishment (MP) and non-monetary punishment (NMP). The motivation for this extension is two-fold: (i) many economic decisions are made by groups of individuals, so that it is important to document both quantitative and qualitative differences between the two and (ii) within team discussions can provide insights into the motivation underlying the observed behavior.

Our results show that under a straight VCM game, with no opportunity for punishment, there were no significant differences in initial contribution rates to the public good between teams and individuals. However, reductions in end game contributions started earlier for teams, and were significantly larger than for individuals.[2] With MP, both teams and individuals increased contributions to the public good substantially, with significantly larger increases in contributions for teams. Teams punished less with MP than individuals, but punishment was more targeted at low contributors than for individuals. Results under NMP differed significantly between teams and individuals. Contribution rates increased significantly for individuals, but had no significant effect on contribution rates for teams, even though teams sanctioned free riding more than individuals. That sanctions had little to no effect on contribution rates for teams, but increased contribution rates for individuals, is consistent with previous experiments in economics and psychology that teams are more self-serving than individuals, with teammates providing support for self-serving behavior.[3]

With monetary punishment, end game contributions for teams were just over 95% of the maximum, with teams continuing to punish free riding even in the last play of the game. With MP teams punished free riding at higher rates than individuals, with little if any anti-social punishment (i.e., punishing above average contributors), compared to small positive rates of anti-social

[1] See Ledyard (1995) and Chaudhuri (2011) for reviews of the literature on VCM games including the role of monetary punishment in increasing contribution rates.
[2] See Cox and Stoddard (2018) for similar results.
[3] See, for example, Wildschut and Insko (2007) and Nielsen et al. (2019) and a number of other cases discussed in the concluding section of the paper.

punishment for individuals. Team discussions showed that under NMP teams did not take sanctions seriously, but nevertheless continued to punish free riders.

Prior Research:

There is a large experimental literature on VCM public good games. Ledyard (1995) provides a survey of early experimental work.[4] The summary of past research reported on here is restricted to games with a "partners" design, where group composition remains fixed throughout, like those conducted here. [5] The literature on public good games shows that contributions are sensitive to the marginal return for contributing to the public good (MPCR): The higher the MPCR, the higher the contributions to the public good. There are consistent end game effects, where contributions drop off as the end point of the game approaches, with the strength of these end game effects sensitive to the MPCR rate.[6] There is a much more limited literature investigating the effects of teams on contributions in public good games, as well as with the effect of non-monetary "sanctions" on free riding, two of the main topics investigated here.

Cox and Stoddard (2018) is one of the first experiments to compare individuals with teams. They employed 2 person teams within fixed groups of 3 decision makers (DMs), an MPCR of 0.75, and a chat box for teams to decide on team choices. They studied straight VCM games with no punishment. In comparing teams to individuals, they found that teams initially contributed more than individuals, but exhibited a significantly stronger downward trend. Analysis of team chats showed that many teams anticipated lower end game contributions by others and reacted accordingly.

Auerswald et al. (2018) compared 3 player teams with individuals employing an MPCR of 0.4, investigating the effect of monetary punishment on contribution rates. Experimental sessions started with a single one-shot game, followed by 10 periods of a standard VCM game, followed by 10 periods with monetary punishment along the lines of Fehr and Gächter (2002). There was no direct communication between teammates, rather each member of the team proposed allocation, which was reported back to team members, with this process repeating until two teammates made the same proposal (majority rule treatment) or everyone made the same proposal (unanimity rule treatment).[7] Teammates were fixed within each set of 10 period games,

[4] See Chaudhuri (2011) for a survey updating results since the Ledyard survey.
[5] Stranger sessions, where group composition changes between periods, consistently show lower contribution rates than comparable partners sessions both with and without punishment (Fehr and Gächter, 2002; Masclet et al., 2003)
[6] The lower the MPCR, the stronger the end game effect.
[7] The number of proposal rounds was capped at 10 with default options specified in case of no agreement. Agreement was always satisfied with majority rule case, but not with unanimity.

but team composition changed with the introduction of punishment. Without punishment contributions decreased continuously over time across treatments at essentially the same rate for teams and individuals. With MP there was an immediate increase in contributions, with modest increases after that until the end periods when contributions decreased a bit. Teams contributed more than individuals under unanimity, with no difference between the two under majority rule.

Masclet et al. (2003) introduced the idea of comparing non-monetary punishment (sanctions) compared to monetary punishment. They used individuals, and a partners' design with 4 individuals in each group and an MPCR of 0.4.[8] Each session had an ABA design with no punishment in periods 1-10, followed by either monetary or nonmonetary punishment in periods 11-20, with a return to no punishment in periods 21-30. NMP used the same punishment structure (points) as MP, designed to allow members to show their approval or disapproval of "each other group member's decision". They report that MP increased contributions more than NMP, but that after subtracting out the costs of punishment, total income was consistently higher under NMP than MP.

We look at the effect of MP and NMP for teams and individuals. The motivation for extending these results is that teams are known to be more self-serving than individuals, with teammates providing support for selfish choices (see, for example, Wildschut and Insko, 2007, Nielsen et al. 2019). This raises the question of whether groups will respond differently to sanctions, which if true, is important to know as sanctions, or targets, are often employed in group settings. More generally, there is a growing literature in experimental economics comparing individuals and teams (e.g., Cooper and Kagel, 2005; Kocher and Sutter, 2005; Feri, Irlenbusch, and Sutter, 2010; Charness and Sutter, 2012; Kugler et al., 2012; Maciejovsky, Sutter, Budescu, and Bernau, 2013; Casari, Zhang, and Jackson, 2016). The typical finding is that teams are more "rational" than individuals as teams are more likely to use a theoretically optimal strategy and are faster to learn to maximize own payoffs. This suggests that in addition to being less responsive to sanctions than individuals, they will be less likely to employ MP for free riding, particularly in end periods, where there is no room for the punishment to increase future public good contributions.

I Experimental Design and Procedures

Two person teams and individuals (DMs) participated in a standard linear public goods game with an initial endowment of 60 experimental currency units (ECUs) and a marginal rate of

[8] There were several sessions that mixed in strangers as well. These are not discussed here.

return for the public good of 0.75. DMs were in groups of 4 throughout so that the per period payoffs were

$$\pi_i = 60 - c_i + 0.75\left(\sum_{j=1}^{4} c_j\right)$$

where π_i is own payoff in each period, c_i is DM i's own contribution to the public good and c_j is DM j's contribution to the public good. Both teammates and groupings remained fixed throughout.

There were both crossover and control sessions. The crossover sessions started with 10 periods of the VCM game without punishment, followed by 10 periods with punishment within the same fixed groups. With punishment, in stage 1 DMs decided on how much to allocate to the public good, just as in the VCM game without punishment. In stage 2, after all DMs in the group determined their contributions, these were reported back to group members, after which each DM had the opportunity to assign between 0 and 15 punishment points to each of the other DMs in their group. With MP each punishment point cost the sender 1 ECU with a cost of 3 ECUs per point to the punished.[9] At the end of stage 2, DMs were told the total punishment points they received, the total punishment points assigned to other group members, and their earnings net of punishment costs.[10] Non-monetary punishment followed the same procedures except punishment points were referred to as "… a measure of your disapproval of what each team did", with no impact on earnings. "They just represent your disapproval of other teams' (individuals') choice in stage 1."[11]

Control sessions had essentially the same structure as the crossover sessions, starting with 10 periods of the VCM game without punishment. These were followed by a second set of 10 periods with no punishment. For both punishment and control sessions subjects were told there would be two parts to the experiment, but with no information about part 2 until after the completion of part 1.[12] Subjects were also told the number of periods (10) in each part.

Teammates had a chat box to communicate with each other to reach agreement on the team choice lasting for 2 minutes in each period. Individuals had 1 minute to decide on their choices in

[9] The total costs from receiving punishment could not exceed stage 1 earnings.
[10] DMs were only shown the total punishment they received, so there was no chance for retaliation.
[11] The wording here is taken from Masclet et al. (2003).
[12] Instructions are available upon request.

the baseline treatment. For stage 2 in the punishment periods, individuals (teams) had an additional 45 (90) seconds to decide on what, if any, punishment points to allocate.[13]

Subjects were randomly assigned to sets of 4 DMs whose composition stayed the same throughout the experimental session. There were 6 groups of 4 DMs for both the team and individual MP sessions, with another set of 6 groups each for the NMP sessions.[14] The control sessions had 6 sets of teams and 5 sets of individuals.[15]

Subjects were recruited from the undergraduate student population at Ohio State University using ORSEE (Greiner, 2015). Subjects participated in a single session, with sessions lasting 45 minutes for individuals and 75 minutes for teams. The exchange rate for ECUs was 100 ECUs = \$1. Subjects were paid based on total earnings from the two parts which averaged \$29.50 a session including a \$3 show up fee. Each member of the two-person teams received the team's earnings. The experiment was programmed in zTree (Fischbacher, 2007).

II. Experimental Results

II.1 *Contribution Rates without Punishment*

Figure 1 reports average contribution rates, by groups, for both teams and individuals with data pooled from periods 1-10 of the crossover sessions and periods 1-10 of the control sessions. Average contribution rates were essentially the same for teams and individuals over the first 8 periods, with a strong reduction in contributions for teams in periods 9 and 10. And a smaller reduction in contributions just for period 10 for individuals.

[13] Failure to meet the deadline resulted in a random allocation of the public good, along with notifying all group members that the contribution was randomly determined. Most of these cases occurred in the first or second period (27 out of 43 occurrences). We control for this in the regressions by using a dummy variable for random contribution. Similarly, the computer allocated punishment points randomly if DMs failed to meet that deadline.

[14] Due to a low number of show ups in one of the MP sessions, one of the experimenters had to be randomly matched with a subject to form a team. The experimenter was instructed to simply follow through with what their partner suggested and was not allowed to offer suggestions. We include the data for this team in the analysis.

[15] This was not intentional, resulting from fewer subjects showing up for the individual control sessions.

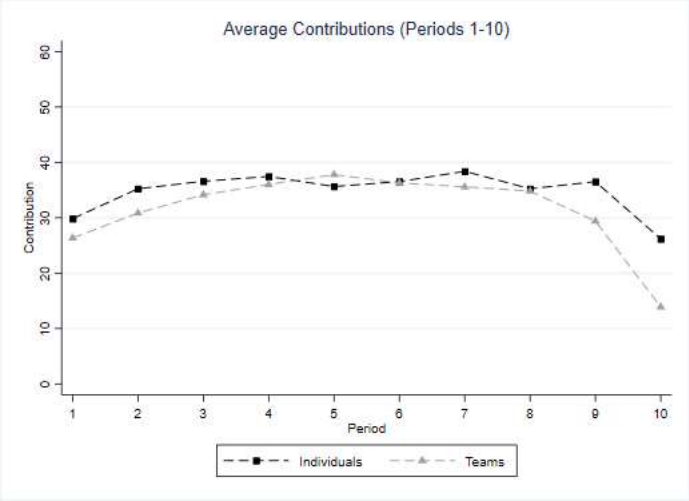


Figure 1: Differences in Baseline Contribution Rates (periods 1-10)

The following regression supports these observations, where the dependent variable is individual DM contributions, with bootstrapped standard errors, clustered at the group level.

$$Contribution = \underset{(4.65)}{35.6^{***}} - \underset{(5.52)}{0.8}\, Team - \underset{(2.18)}{4.3^{*}}\, EndBase - \underset{(3.82)}{8.9^{**}}\, Team \times EndBase$$

The base group captures individual DM contributions. Team is a dummy variable equal to 1 for teams in periods 1-10, with EndBase a dummy variable equal to 1 for periods 9 and 10. TeamxEndBase accounts for the greater reduction in end game contributions for teams. The team dummy is small and not statistically significant, indicating no significant difference in contribution rates between teams and individuals, in periods 1-8. The EndBase dummy is negative ($p < 0.10$) accounting for the reduced contributions in periods 9-10, with the TeamxEndBase dummy accounting for the greater reduction in end game contributions for teams ($p < 0.05$) - about twice the size of the end game effect for individuals.[16]

[16] Running the regression with separate end game dummies for periods 9 and 10 showed that teams significantly reduced contributions in period 9 by -5.4 ECUs ($p < 0.10$), with no significant reduction for individuals. In period 10 individuals had an average reduction in contributions of -9.4 ECUs ($p < 0.01$) compared to periods 1-8, compared to teams with an average reduction in contributions of -20.9 ($p < 0.01$). Cox and Stoddard (2018) also reported a significantly larger end game effect for teams compared to individuals.

Conclusion 1: Contributions rates for teams and individuals were essentially the same in periods 1-8 for the straight VCM game without punishment. However, end game reductions in contributions started earlier, and were substantially larger, for teams compared to individuals.

II.2 Effects of Monetary Punishment

Figure 2 (top panel) compares public good contributions for both teams and individuals prior to and after the introduction of monetary punishment. The bottom panel compares these to the public good contributions in the control sessions for individuals and teams, respectively. It is clear from the top panel that MP increased contributions for both teams and individuals, more so for teams. The control sessions suggest what would have happened absent MP in periods 11-20. For teams, there was a positive restart effect (Andreoni, 1988; Croson, 1996) followed by a sharp end game effect that began earlier, and was more severe, than in periods 1-10. For individuals, the control sessions exhibit a modest *negative* restart effect, with end game contributions decreasing modestly after that.

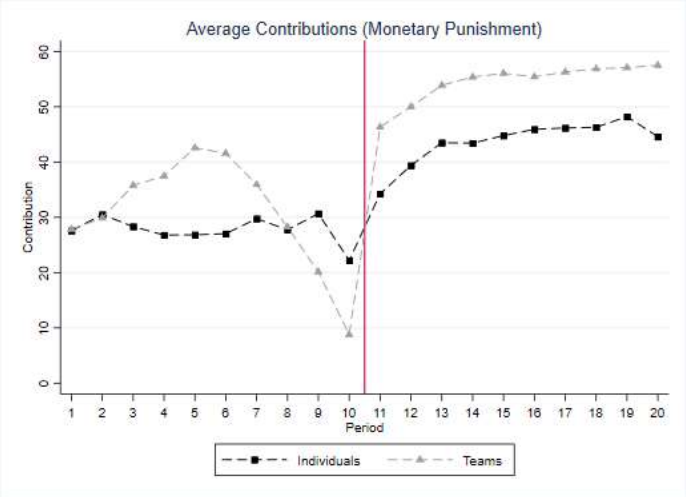
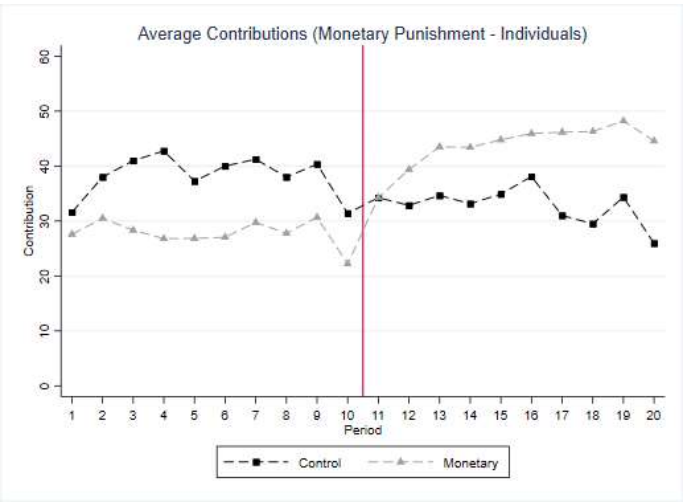


Figure 2 (top): Differences in contributions in MP



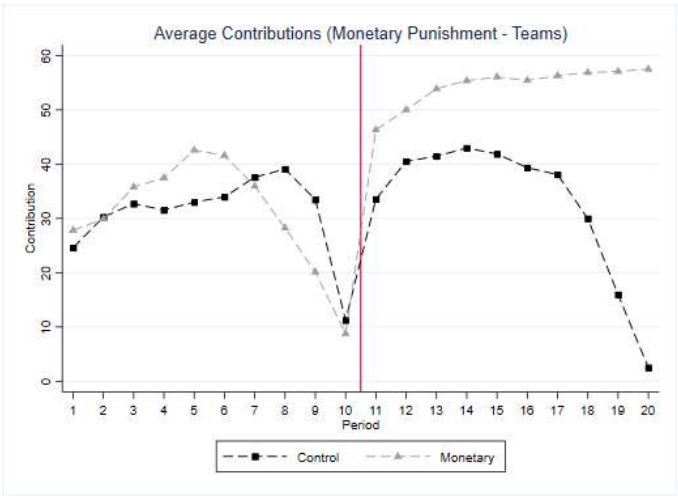


Figure 2 (bottom): Differences in contributions between control and MP

The regressions reported in the first two columns of Table 1 provide a detailed analysis of the effects of MP, with the same specification for both teams and individuals.

$$Contribution_{it} = \beta_0 + \beta_1 Control + \beta_2 EndBase + \beta_3 Control \times EndBase + \beta_4 Second + \beta_5 Control \times Second + \beta_6 EndFinal + \beta_7 Control \times EndFinal$$

The dependent variable is DM contributions to the public good in period *t*, with bootstrapped standard errors clustered at the group level. The constant captures DMs contributions to the public good in periods 1-8 of the crossover sessions. *Control* is a dummy variable for Periods 1-10 in the control sessions. *EndBase* is a dummy variable for periods 9-10. *Second* is a dummy variable for Periods 11-20, with *EndFinal* a dummy variable indicating Periods 19 and 20.

For individuals, contributions were somewhat higher in periods 1-8 for the controls, compared to the same period in the crossover sessions, but the difference is not significant ($p > 0.10$). Further, there were no significant differences between the controls and the crossover sessions in periods 9-10. MP increased contributions in periods 11-18 in the crossover sessions compared to periods 1-8 by 14.9 ECUs ($p < 0.01$). While for the control sessions contributions were 5.2 ECUs lower in periods 11-18 compared to periods 1-8. There was a further small increase in contributions of 3.4 ECUs in periods 19-20 in the crossover sessions ($p > 0.10$).

Table 1: Regressions comparing Control to Crossover

	(1) Indiv - MP	(2) Team - MP	(3) Indiv - NMP	(4) Team - NMP
Control	10.7	-2.6	-1.8	-2.1
	(11.94)	(7.03)	(10.70)	(6.44)
EndBase	-1.6	-21.6***	-8.1*	-7.3
	(2.88)	(3.93)	(4.57)	(6.57)
Control x EndBase	-1.2	10.5*	5.2	-3.7
	(4.25)	(5.62)	(5.92)	(7.81)
Second	14.9***	17.8***	10.0*	4.2
	(3.59)	(2.60)	(5.47)	(3.25)
Control x Second	-20.1***	-12.5***	-15.2**	1.1
	(5.13)	(4.29)	(7.11)	(4.84)
EndFinal	3.4	3.4*	-4.7	-17.4***
	(2.75)	(2.02)	(3.58)	(6.28)
Control x EndFinal	-6.8	-32.9***	1.3	-12.1
	(7.23)	(6.42)	(7.69)	(9.04)
Constant	28.1***	36.0***	40.6***	35.4***
	(8.73)	(4.65)	(7.17)	(3.94)
Observations	880	960	880	960

Bootstrapped standard errors in parentheses, clustered at the Group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

For teams there were no significant differences in public good contributions between own baseline periods 1-8 and the control sessions for the same period. However, own baseline end game contributions (periods 9-10) decreased more in the crossover sessions (-21.6 ECUs, $p < 0.01$), compared to a decrease of -11.1 ($p < 0.05$) in the control sessions. The introduction of MP in periods 11-18 increased contributions by 17.8 ECUs per period compared to periods 1-8 prior to the crossover ($p < 0.01$), with a further increase of 3.4 ECUs in periods 19-20 ($p < 0.10$). In contrast, for the controls there was a modest increase of 5.3 ECUs ($p > 0.1$) in Periods 11-18, relative to periods 1-8, followed by a decrease in contributions of 13.1 ECUs in periods 19-20 compared to periods 9-10 ($p < 0.01$).[17] However, contributions dropped sharply in the last two periods for the controls, so that in periods 19-20 public good contributions averaged 57.3 ECUs

[17] Computed by taking the difference of the relevant dummies for the control sessions.

for the crossover sessions compared to 9.2 ECUs for the control sessions with the difference highly significant ($p < 0.01$).

Figure 3 top panel compares teams with individuals for NMP in the crossover sessions. The two bottom panels compare outcomes under NMP with data from the control sessions. NMP clearly increased contributions for individuals in the crossover sessions relative to own contributions prior to the crossover, as well as relative to the control sessions in periods 11-20. In contrast for teams, contributions under NMP were higher than prior to the crossover in periods 12-14, after which there were sharp decreases throughout, along with little distinction under NMP compared to the control sessions for periods 11-20.

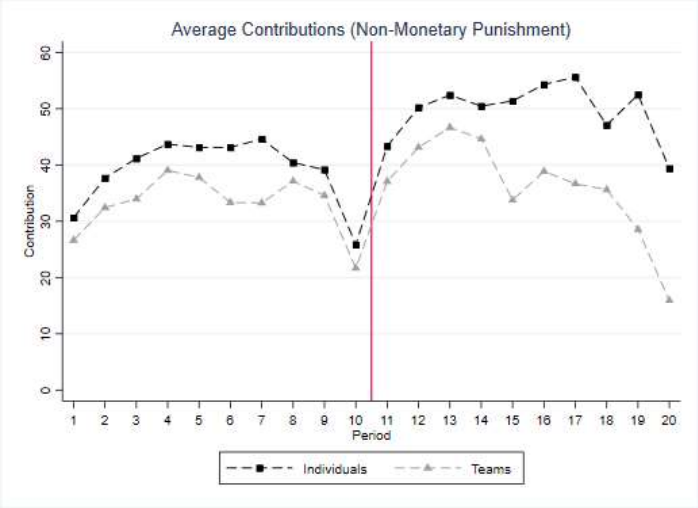
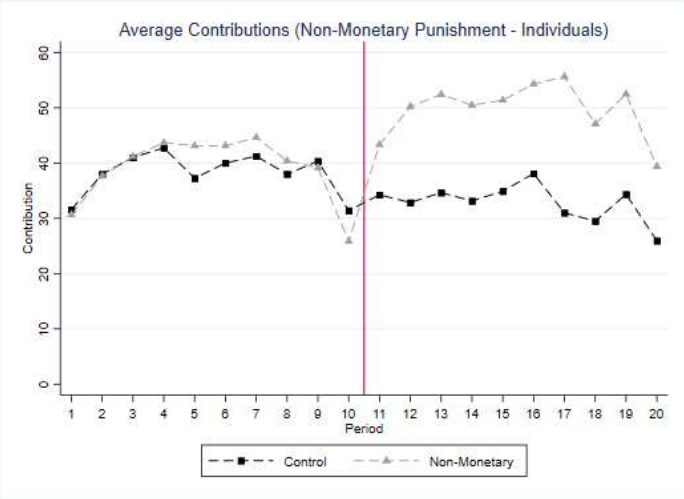


Figure 3 (top): Differences in contributions in NMP



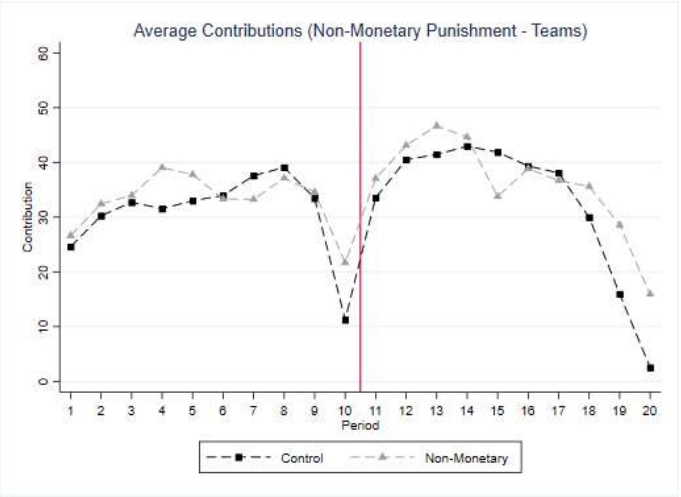


Figure 3 (bottom): Differences in contributions between control and NMP

The regressions reported in columns 3 and 4 of Table 1 provide a formal analysis of the effects of NMP, using the same specification as used for MP. The constant captures public good contributions in periods 1-8 of the crossover sessions. For individuals in the crossover sessions there were no significant differences between contributions in periods 1-8 and contributions in the control sessions for the same period. There was a relatively large marginal decrease in contributions of -8.1 ECUs ($p < 0.10$) in the crossover sessions in periods 9-10, with a smaller decrease of -2.9 ECUs ($p > 0.10$) for the controls. Periods 11-18 saw an increase in contributions relative to periods 1-8 of 10.0 ECUs ($p < 0.10$) in the crossover sessions, with greater average contributions compared to the controls in periods 11-18 by 17 ECUs ($p < 0.10$). Periods 19-20 showed a marginal decrease of -4.7 ECUs ($p > 0.10$) in contributions under NMP compared to periods 11-18, with contributions compared to the control sessions higher in periods 19-20. However, this difference was no longer statistically significant ($p > 0.10$).

For teams there were no significant differences between own baseline contributions in periods 1-8 compared to the same period for the control sessions (-2.1 ECUs, $p > 0.10$). There were additional decreases in contributions in periods 9-10 for the crossover and the control sessions, but still no significant differences between the two. There was a small, but statistically insignificant, increase in contributions under NMP for the crossover sessions of 4.2 ECUs in periods 11-18 compared to contributions in periods 1-8 ($p > 0.10$) along with a marginal increase in contributions of 1.1 ECUs for the control sessions ($p > 0.10$), so that there were no significant

differences between the crossover and the control sessions in periods 11-18. Contributions in periods 19-20 averaged around 22.2 ECUs under NMP and around 9.2 ECUs under the control sessions with this difference being significant ($p < 0.05$).

Table 2 compares the effects of MP and NMP on public good contributions for teams and individuals, using own baseline contributions as the reference point.

$$Contribution = \beta_0 + \beta_1 Team + \beta_2 EndBase + \beta_3 Team \times EndBase + \beta_4 Second + \beta_5 Team \times Second + \beta_6 EndFinal + \beta_7 Team \times EndFinal$$

The dependent variable is individual DM contributions to the public good, with bootstrapped standard errors clustered at the group level. The constant captures individual contributions in periods 1-8. *Team* is a dummy variable for team sessions. *EndBase* and *EndFinal* are dummy variables that take on the value 1 for Periods 9-10 and 19-20, respectively. Finally, *Second* is a dummy variable that takes the value 1 for periods 11-20.

Table 2: Regressions comparing Teams to Individuals

	(1) Monetary	(2) Non-Monetary
Team	7.6	-5.6
	(9.03)	(7.49)
EndBase	-1.6	-8.1**
	(3.30)	(3.93)
Team x EndBase	-19.6***	1.3
	(5.52)	(8.17)
Second	14.9***	10.0**
	(4.13)	(4.70)
Team x Second	3.3	-5.4
	(5.10)	(5.25)
EndFinal	3.4	-4.7
	(2.83)	(3.30)
Team x EndFinal	0.0	-12.7*
	(3.25)	(7.53)
Constant	28.1***	40.6***
	(7.91)	(6.41)
Observations	960	960

Bootstrapped standard errors in parentheses, clustered at the group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

With MP there were no significant differences between teams and individuals in periods 1-8, but a significantly stronger marginal end game effect for teams in periods 9-10 of -19.6 ECUs ($p < 0.01$). MP increased contributions for individuals relative to periods 1-8 by an average of 14.9 ECUs per period ($p < 0.01$), with a marginally larger increase for teams of 3.3 ECUs ($p > 0.10$). Periods 19-20 saw a further small marginal increase in contributions for individuals ($p > 0.10$). with a negligible marginal increase for teams. The net effect was substantially higher average contributions to the public good in periods 19-20 for teams: 57.3 ECUs for teams versus 46.4 ECUs for individuals with the difference being significant ($p < 0.10$), based on the regression results in Table 2.

Differences between teams and individuals under NMP were quite different. For individuals there was a marginal increase in public good contributions in periods 11-18 of 10.0 ECUs per period compared to periods 1-8 ($p < 0.05$). In contrast, for teams the increase was 4.6 ECUs ($p > 0.10$) per period compared to 1-8. The gap in contributions between the two was large in periods 19-20: 45.9 ECUs for individuals versus 22.2 ECUs for teams ($p < 0.01$), based on the regression results in Table 2.

In comparing MP and NMP for individuals, after subtracting out the costs of punishment, total earnings within the group, net of costs, were *higher* for NMP averaged over the first 5 punishment periods (636.5 under NMP compared to 504.6 with MP, $p < 0.05$).[18] Although this difference in total earnings net of punishment costs decreased in periods 16-20, the difference was still statistically significant (638.2 under NMP versus 550 under MP, $p < 0.10$). These results replicate those reported in Masclet et al. (2003) for individuals. In contrast, total earnings for teams, net of costs, were significantly *lower* under NMP compared to MP in the first 5 punishment periods (624.3 under MP versus 568.4 under NMP, $p < 0.05$).[19] This difference was even larger in the last 5 punishment periods (674.4 under MP compared to 489.0 under NMP, $p < 0.01$).

The question is, what underlies this difference in total earnings for teams compared to individuals under NMP? The social psychology literature suggests that differences of this sort can be attributed to groups providing their members with support for acting in a self-benefiting manner, whereas individuals have no such support. That is, individual team members help to overcome

[18]Based on a Mann-Whitney test for total earnings within each group between the two treatments using group-period as the unit of observation.

[19] When random punishment was imposed because of a team's failure to coordinate on punishment, punishment for the group was rescaled based on the average punishment given by other teams in that group.

the pressure from social norms of equity, equality, and reciprocity in one-to-one interactions (see, for example, Insko et al., 1993 and Wildschut and Insko, 2007 from the psychology literature and Nielsen et al., 2019 and Kagel, 2018 from the economics literature).[20] This is reflected in discussions between teammates, many of which completely ignored or blew off non-monetary punishment.[21] For example, consider the following dialogue between teammates where changes in quotation marks distinguish between teammates:

> "seems like we got a fair amount of disproval haha … because we got a lot of earnings … i say we try zero (contributions) again… what do you think" "I don't think we need to care about points" "right i agree"

Or take this example for a different team:

> "are we sticking with 0 (contributions)?" "0 again? ya" "ok" "i don't think this part does anything" (referring to the NMP sessions)

Conclusion 2: MP increased contribution rates substantially for both teams and individuals relative to the VCM game without punishment, with teams having significantly higher contribution rates in the end periods than individuals.

Conclusion 3: NMP significantly increased contribution rates for individuals relative to the VCM game without punishment, but had little positive effect on contribution rates for teams. For individuals' total earnings, net of punishment costs, were significantly *higher* under NMP compared to MP. In contrast, for teams' total earnings net of punishment costs were significantly *lower* under NMP compared to MP. This difference between teams and individuals parallels the social psychology literature showing that teams provide their members with support for acting in a self-benefiting manner, whereas individuals have no such support.

II.2 Punishment Patterns and Responses to Punishment

Regressions reporting the distribution of punishment points along with responses to same are reported in Table 3.[22] Highlights are discussed in the text. The regressions include data for both teams and individuals, to be able to compare between the two.

[20] To be discussed in some detail in the concluding section of the paper.
[21] See Table 4 below.
[22] The empirical specification follows the one employed in Masclet et al. (2003).

For punishment points given (the left-hand side of Table 3), the regression specification is as follows:

$$P_{i,t}^{j} = \beta_0 + \beta_1 \max\{0, c_{i,t} - c_{j,t}\} + \beta_2 \max\{0, c_{j,t} - c_{i,t}\} + \beta_3 \max\{0, \bar{c_t} - c_{j.t}\} + \beta_4 \max\{0, c_{j,t} - \bar{c_t}\} + \beta_5 Team + \beta_6 (Team \times \max\{0, c_{i,t} - c_{j,t}\}) + \beta_7 (Team \times \max\{0, c_{j,t} - c_{i,t}\}) + \beta_8 (Team \times \max\{0, \bar{c_t} - c_{j.t}\}) + \beta_9 (Team \times \max\{0, c_{j,t} - \bar{c_t}\})$$

where $P_{i,t}^{j}$ are punishment points assigned to DM$_j$ in period t. Explanatory variables are positive and negative deviations in i's contribution relative to js [$\max\{0, c_{i,t} - c_{j,t}\}$ and $\max\{0, c_{j,t} - c_{i,t}\}$], the latter representing anti-social punishment. Positive and negative deviation in j's contribution relative to the group's average contribution, [$\max\{0, \bar{c_t} - c_{j,t}\}$ and $\max\{0, c_{j,t} - \bar{c_t}\}$], again the latter representing antisocial punishment. And a dummy variable for the team treatment (T).[23] Punishment points allocated by individual DMs serve as the omitted variable. A Tobit regression was used for MP, as punishment points were given in only 20% of all periods. Standard OLS was used for NMP as punishment was more common (45% of all observations having positive punishment points). Bootstrapped standard errors clustered at the group level in parentheses.

With MP, teams imposed an average of 4.85 *fewer* punishment points each period compared to individuals ($p < 0.05$). These were targeted at the same rate as individuals for deviations from own contributions. However, teams assigned three times as many punishment points to teams contributing less than the average contribution for their group ($p < 0.01$).[24]. In contrast under NMP teams imposed an average of 3.83 *more* punishment points per period than individuals ($p < 0.05$), with the marginal effect of punishing lower than own group average contributions not significantly different from individuals.

The right-hand side of Table 3 reports reactions to punishment. There are separate regressions for DMs who contributed more than their average group contributions in the previous period, from those that contributed less. The regression specification in both cases is as follows:

$$c_i^{t+1} - c_i^t = \beta_0 + \beta_1 P_i^t + \beta_2 (c_i^t - \bar{c}^t) + \beta_3 Team + \beta_4 (Team \times P_i^t) + \beta_5 (Team \times (c_i^t - \bar{c}^t))$$

Where the dependent variable $c_i^{t+1} - c_i^t$ measures the change in i's contribution between period t and $t+1$, P_i^t measures the number of punishment points i received in the previous period (t),

[23] We do not include observations in which DMs failed to decide in time which led to the computer randomly allocating punishment.
[24] The marginal coefficient value was twice as large for teams compared to individuals.

$(c_i^t - \bar{c}^t)$ measures the deviation in DM i's contribution in period t from the average contribution in their group, and $Team$ is a dummy variable for the team sessions. Standard errors are clustered at the DM level.

With MP, low contributing individuals receiving punishment increased their contributions an average of 1 ECU for every two punishment points ($p < 0.05$), with the $Team \times P_i^t$ marginal effect negative but not statistically significant.[25] The $c_i^t - \bar{c}^t$ coefficient was negative and statistically significant ($p < 0.01$), indicating that individuals with contributions closest to the group average increased contributions the least in response to punishment, with the marginal effect for teams not significant at conventional levels. In contrast, for DMs contributing more than the group average, punishment points per se had no significant effect on contributions.

For NMP, punishment points, per se, had no significant effect on contributions for both low and high contributors for both teams and individuals. However, for low contributors, individuals with below average contributions the $c_i^t - \bar{c}^t$ coefficient was negative and significant ($p < 0.01$), indicating that what effect there was from punishment was smallest for relatively high contributors. In contrast, the marginal effect for teams was positive, so that the net effect was not significant ($p = 0.38$). For above average contributors, none of the coefficients were significant, for both teams and individuals, indicating no response to NMP for above average contributors. This suggests that these above average contributors had a taste for contributing to the public good, as opposed to responding to non-monetary sanctions.

[25] The net effect for teams was positive (0.12) but, again, not significant.

Table 3: Punishment Points Given and Responses to Punishment

Punishment			Responses to Punishment				
	$P_{i,t}^{j}$			Below Average Contributors $(c_i^t - \bar{c}^t) < 0$		Above Average Contributors $(c_i^t - \bar{c}^t) > 0$	
	MP	NMP		MP	NMP	MP	NMP
$\max\{0, c_{i,t} - c_{j,t}\}$	0.24*** (0.06)	0.15*** (0.03)	P_i^t	0.52** (0.20)	-0.69 (0.52)	-0.16 (0.51)	-0.06 (0.27)
$\max\{0, c_{j,t} - c_{i,t}\}$	0.26*** (0.07)	0.08 (0.05)	$(c_i^t - \bar{c}^t)$	-0.41*** (0.12)	-1.43*** (0.51)	-0.27* (0.16)	-0.15 (0.13)
$\max\{0, \bar{c_t} - c_{j.t}\}$	0.20** (0.08)	0.17*** (0.05)	*Team* (T)	-1.72 (3.14)	2.28 (6.64)	5.19 (3.31)	-0.96 (5.05)
$\max\{0, c_{j,t} - \bar{c_t}$	-0.29 (0.28)	-0.08 (0.05)	Tx $P_{i,t}$	-0.40 (0.27)	0.62 (0.57)	-0.02 (0.57)	0.10 (0.33)
Team (T)	-4.85** (2.32)	3.83** (1.54)	$T \times (c_i^t - \bar{c}^t)$	-0.21 (0.30)	1.13* (0.61)	-0.14 (0.23)	-0.49 (0.34)
Tx $\max\{0, c_{i,t} - c_{j,t}\}$	-0.04 (0.15)	-0.09* (0.05)	Constant	-0.54 (2.49)	4.47 (3.80)	-0.33 (2.73)	-2.94 (2.78)
Tx $\max\{0, c_{j,t} - c_{i,t}\}$	0.02 (1.69)	0.01 (0.06)					
Tx $\max\{0, \bar{c_t} - c_{j.t}\}$	0.47** (0.20)	0.09 (0.07)					
Tx $\max\{0, c_{j,t} - \bar{c_t}\}$	0.06 (11.60)	-0.15** (0.07)					
Constant	-5.45** (2.44)	1.05 (0.78)					
N	1386	1407	*N*	99	133	112	183

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors in parentheses clustered at the group level for punishment regressions. Standard errors clustered at the DM level for reactions to punishment.

Conclusion 4: With MP, teams distributed significantly fewer punishment points than individuals, other things equal, but targeted three times as many punishment points to below average contributors compared to individuals. In contrast, for NMP, other things equal, teams distributed significantly more punishment points than individuals.

Conclusion 5: For individuals, MP was quite effective at increasing contributions for DMs contributing less than the group average, with this effect decreasing the closer DMs were to the

group average. For teams, however, MP was not effective for low contributors. Punishment points per se had no significant effects on contributions under NMP.

II.3 Insights from team chats

Looking at the team chats provides an opportunity to better understand the process underlying the behavior reported. To do this, the authors read through a sampling of the team discussions to establish coding categories, with two undergraduate students recruited to do the coding. After going over the categories of interest with the students, they independently coded up one MP and one NMP session. After this, the authors went through the coding with the students to identify any differences in understanding between the two coders and to answer any questions the coders had. After which the coders worked on the whole data set for the crossover sessions (MP and NMP).

Table 4 reports the coding categories, the frequency with which either student coded that category, and Cohen's kappa (Cohen, 1960), a standard statistic used to measure inter-rater reliability when measuring qualitative categories. Below these measures are brief chat samples for each of the coding categories.[26] Coding categories 1-3 apply to baseline periods 1-10 only.

Teams rarely specified, in advance, the exact period they were planning to reduce contributions in the VCM game without punishment. However, in periods 1-8, one or both coders had 83.3% of teams discussing reduced contributions in "something like the last round". The average period in which these discussions took place was period 6.[27] In periods 1-10, 73% (79%) of the time one or both coders coded for contributing to the public good to develop a cooperative norm (to be pro social/wanting to be nice).[28]

[26] Grammar and punctuation have not been corrected for. Each category was coded for only the first period the category was identified.
[27] The same for both coders. Reduced end game contributions were also coded for in periods 11-18 for NMP. The mean period when these discussions took place was 14.5 for one coder and 15.1 for the other coder, a little earlier than in the baseline sessions.
[28] Teams could be coded in both categories.

Table 4: Coding Team Chats
(percentage of teams coded by one or both coders[29])

1. Discussed reduced end game contributions in periods 1-8:	83.3%, $k = .33$
2. Contributed to develop a cooperative norm:	72.9%, $k = .38$
3. Contributed to be pro-social/wanting to be nice:	79.2%, $k = .48$
4. MP only - periods 18- 20 Discussed not reducing contributions out of fear of Punishment:	20.8%, $k = .50$
5. NMP	
a. Punishing to increase contributions:	75.0%, $k = .32$
b. Blow off punishment points:	79.2%, $k = .22$

Examples of chats for coding categories:

1. "Now that I think about it, it makes the most sense for us to hold back all of our tokens on something like the last round …Like if everyone else contributes and we don't we keep more" "yah i was thinking that to hahaha"
2. "I think we should give it all. It's risky, but if we can convince all the teams to give 60. The money turnout is good"
3. "we put in way too few" "how much did they all put in" "like 30 each" "but yeah lets put more this time hahah" "should we do 30?" "im good with 30" "sounds good"
4. "I say last round we still donate 60" "me too" "too big risk of getting penalized "exactly im not tryna get screwed over"
5. a. "alright anyone that goes less than 60 getting 15 naughty points" "thatll show them"
   b. "LMAO" "we got 45 points LAMAO" "hit em with 0 again" (ECUs to contribute)

Under NMP 75% of teams were coded for discussing punishing free riders with the goal of increasing contributions. With 79.2% coded as blowing off punishment points, with many teams coded in *both* categories. While the same team being coded for both seems like a contradiction, as the quote offered indicates, punishing free riders under NMP was done with some irony: "anyone that goes less than 60 getting 15 naughty points" "thatll show them". Further, the same team was coded in the next period as blowing off punishment points: "i dont understand what the effect of the disapproval points is supposed to be" "yeah i agree other than to discourage people from making a decision.... but since it has no effect then why would people abide by the approval of other teams" "idk they must be testing some psychological thing or whatever".[30]

[29] Category 1 coded for periods 1-8. Categories 2 and 3 coded for periods 1-10.
[30] Further, this same team in period 20 contributed 0 to the public good while imposing 30 punishment points on other members of their group!

One thing that has been discussed in the literature is why, even with monetary punishment, do DMs continue to punish free riders, even though it is costly to the punisher, with no future benefits to be had by inducing the punished to contribute more? This argument holds in spades for teams since, as noted earlier, they tend to be smarter and earn more than individuals. But indeed, teams consistently punished below average contributors in periods 18-20, while never punishing those contributing the full 60 to the public good.

The psychology literature suggests two possible factors underlying this late period punishment. First, there is evidence from older electric shock experiments that punishment of a frustrating person tends to foster further aggression against the offending person (Geen et al, 1975; Verona and Sullivan, 2008).[31] Second, there is evidence that people tend to "overgeneralize", meaning that a rule that is usually adaptive may still be followed when it is not appropriate. This is very common in language but has been used to explain things like the sunk cost fallacy – the latter being an overgeneralization of the maxim "waste not want not" (Arkes, 1996).

Restart effect: Team discussions in the control sessions provide an opportunity for understanding the thinking behind the restart effect. Fifteen out of 24 teams contributed zero to the public good in period 10.[32] Focusing on these teams, the restart effect was reasonably strong in comparing choices in periods 1-5 to 11-15: contributions averaged 29.8 ECUS in periods 1-5 versus 42.0 for in periods 11-15. Team discussions suggest that, as in the baseline periods of the crossover sessions, contributions in periods 11-15 were driven by the hope that contributing to the public good would stimulate others to do so, resulting in higher earnings, along with some concern that other group members might contribute zero again. Below are some representative team chats (where change in quotation marks indicate different team members):

> Period 11: "I feel like we should start high ...45?" "sure" "set the bar high for the rest of the round". (Followed by a total group contribution of 140 ECUs.)
> Period 12: "I feel good staying high" "same …45 again?"

Or the following team:

> Period 11: "ok what do you think" "idrk…i wanna say 30 but at the same time i wanna do 0 …the other teams did 0 last time first and they ended up doing good" "we should do like 30 or 40 to show they were willing to work together and see what they do" … "okay true im just going to be so mad if they do 0" (Followed by a total group contribution of 140 ECUs)

[31] These experiments would likely not pass IRB approval these days.
[32] For individuals just half as many did so – 6 out of 20.

Period 12: “yayyyy' “lets do that again” “yeah”

Finally, note that in the last period of the control sessions 14 of the 15 teams contributing 0 to the public good in period 10 again contributed 0.[33]

III Summary and Conclusions

This experiment explored differences and similarities between two person teams and individuals in a VCM game without punishment, as well as the effects of both monetary and non-monetary punishment. Absent punishment, public good contributions were quite similar between the two until end game periods, when teams reduced contributions earlier, and substantially more than individuals. Monetary punishment increased contributions for both teams and individuals, with teams increasing contributions more than individuals did. Furthermore, teams punished lower than average contributions at three times the rate at which they were punished by individuals. With non-monetary punishment individuals increased contributions relative to the VCM game without punishment, whereas teams did not. Further, comparing monetary and non-monetary punishment, total earnings, net of punishment costs, were higher under non-monetary punishment. But for teams, this was reversed, with total earnings net of punishment costs higher under monetary compared to non-monetary punishment.

One of the important findings here is that teams were less cooperative than individuals under sanctions. The social psychology literature explains this difference because teammates provide support for acting selfishly, whereas individuals do not, typically acting based on social norms of kindness between individuals (Insko et al., 1993; Wildschut and Insko, 2007). There is support for this in several economic experiments beyond this one: Nielsen et al (2019) found that in a hidden action trust game, two person teams and individuals made promises to cooperate at the same rate, but that individuals lived up to these promises at a much higher rate, even though maximizing payoffs required forfeiting on the promises. Kagel (2018) reports a finitely repeated prisoners dilemma game with and without cheap talk, comparing individuals with two person teams. Absent cheap talk, with experience, cooperation rates converged to zero percent for both teams and individuals. With cheap talk, with experience, cooperation rates in the last play of the game were 47.1% compared to 0% for teams.

[33] The remaining team contributed 25 ECUs to the public good for a total group contribution of 25 ECUs.

References

Andreoni, James (1988) “Why free ride? Strategies and learning in public goods experiments” *Journal of Public Economics*, vol 37, pp 291-304.

Arkes, Hal R (1996). "The psychology of waste." *Journal of behavioral decision making*, Vol. 9, No. 3, pp. 213-224.

Auerswald, Heike, Carsten Schmidt, Marcel Thum, and Gaute Torsvik (2018) "Teams in a public goods experiment with punishment." *Journal of Behavioral and Experimental Economics* vol 72, pp 28-39.

Casari, Marco, Jingjing Zhang, and Christine Jackson (2016). "Same process, different outcomes: group performance in an acquiring a company experiment." *Experimental Economics*, Vol. 19, No. 4, pp. 764-791.

Charness, Gary and Matthias Sutter (2012) “Groups Make Better Self-Interested Decisions,” *Journal of Economic Perspectives*, Vol. 26, No. 3, pp. 157–76.

Chaudhuri, Ananish (2011) “Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature,” *Experimental Economics*, Vol. 14, No. 1, pp. 47–83.

Cohen, Jacob (1960) “A coefficient of agreement for nominal scales.”, *Educational and Psychological measurement*, vol 20 (1), pp. 37-46.

Cooper, David, J., and John H. Kagel (2005). "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games." *American Economic Review*, Vol. 95, No. 3, pp. 477-509.

Croson, Rachel (1996) “Partners and strangers revisited,” *Economic Letters*, vol. 53, 25-32.

Cox, Caleb A. and Brock Stoddard (2018) “Strategic thinking in public goods games with teams,” *Journal of Public Economics*, Vol. 161, pp. 31 – 43.

Fehr, Ernst and Simon Gächter (2002) “Altruistic punishment in humans,” *Nature,* Vol. 415, pp. 137–140.

Feri, Francesco, Bernd Irlenbusch, and Matthias Sutter (2010). "Efficiency Gains from Team Based Coordination—Large-Scale Experimental Evidence." *American Economic*

Review, Vol. 100, No. 4, pp. 1892-1912.

Fischbacher, Urs (2007) "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, Vol. 10, No. 2, pp. 171–178.

Green, Russell G., David Stonner, and Gary L. Shope (1975). "The facilitation of aggression by aggression: Evidence against the catharsis hypothesis." *Journal of personality and social psychology*, Vol. 31, No. 4: 721.

Greiner, Ben (2015) "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, Vol. 1, pp. 114-125

Insko, Chester, John Schopler, Stephen M. Drigotas, Kenneth A. Graetz, James Kennedy, Chante Cox, and Garry Bornstein (1993). "The Role of Communication in Interindividual-Intergroup Discontinuity." *The Journal of Conflict Resolution*, 37(1): 108-138.

Kagel, John (2018) "Cooperation Through Communication: Teams and Individuals in Finitely Repeated Prisoner Dilemma Games". *Journal of Economic Behavior and Organization*, 146, 55—64

Kocher, M.G. and Matthias Sutter. (2005), "The Decision Maker Matters: Individual Versus Group Behaviour in Experimental Beauty-Contest Games." *The Economic Journal*, Vol. 115, pp. 200-223.

Kugler, Tamar, Edgar E. Kausel, and Martin G. Kocher (2012) "Are groups more rational than individuals? A review of interactive decision making in groups," *WIREs Cognitive Science*, Vol. 3, No. 4, pp. 471–482.

Ledyard, John O. (1995) "Public Goods: A Survey of Experimental Research," in John H. Kagel & Alvin E. Roth ed. *The Handbook of Experimental Economics, Volume 1*, Princeton: Princeton University Press, Chap. 2, pp. 111–194.

Maciejovsky, Boris, Matthias Sutter, David V. Budescu, and Patrick Bernau (2013). "Teams make you smarter: How exposure to teams improves individual decisions in probability and reasoning tasks." *Management Science*, Vol. 59, No. 6, pp. 1255-1270.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval (2003) “Monetary and Non- monetary Punishment in the Voluntary Contributions Mechanism,” *The American Economic Review*, Vol. 93, No. 1, pp. 366–380.

Nielsen, Kirby, Puja Bhattacharya, John H. Kagel, and Arjun Sengupta (2019) “Teams promise but do not deliver,” *Games and Economic Behavior*, Vol. 117, pp. 420 – 432.

Verona, Edelyn, and Elizabeth A. Sullivan (2008). "Emotional catharsis and aggression revisited: heartrate reduction following aggressive responding." *Emotion* Vol. 8, No. 3, 331.

Wildschut, Tim and Chester A. Insko. (2007). “Explanations of the Interindividual-Intergroup Discontinuity: A Review of the Evidence.” *European Review of Social Psychology*, 18: 175-211