TECHNISCHE UNIVERSITÄT DRESDEN

CEPIE Working Paper No. 02/16

Center of Public and International Economics

TEAMS CONTRIBUTE MORE AND PUNISH LESS

May 2016

Heike Auerswald
Carsten Schmidt
Marcel Thum
Gaute Torsvik

Editors: Faculty of Business and Economics, Technische Universität Dresden.

This paper is published on the Open Access Repository Qucosa.
The complete Working Paper Series can be found at the CEPIE Homepage.

ISSN 2510-1196

Teams Contribute More and Punish Less*

Heike Auerswald
Technische Universität Dresden

Carsten Schmidt
University of Mannheim

Marcel Thum†
Technische Universität Dresden

Gaute Torsvik
University of Oslo

May 2016

Abstract

Challenges in global politics like climate change, maritime piracy and fighting highly contagious diseases concern global public goods. The related policy decisions are mostly made by teams. In contrast, economic models of global public goods typically assume a single rational decision-maker. We use a laboratory experiment to compare team decisions to decisions of individuals in a finitely repeated public good game with and without a costly punishment option. Teams of three participants coordinate on decisions either by majority or unanimity rule. We find that in absence of a punishment option teams contribute more to the public good than individuals. With a punishment option subsequently to the contribution decision team treatments exhibit a less frequent use of anti-social punishment and lower levels of social as well as anti-social punishment. Extreme preferences for punishment are eliminated by the majority decision rule. Overall, team decisions are closer to the social optimum and teams yield higher net payoffs when compared to individuals.

JEL classification: C72; C92; H41

Keywords: Public Good; Group Decision-Making; Punishment; Experiment

*We thank Tim Cason, Martin Kocher, Robert Lehmann, Wolfgang Luhan, Florian Morath, Vai-Lam Mui, Bernd Suessmuth, Matthias Sutter, Michael Weber and Bengt-Arne Wickström for helpful comments. Financial support from the German Federal Ministry of Education and Research is gratefully acknowledged.

†Corresponding author. Faculty of Business and Economics, Technische Universität Dresden, 01062 Dresden, Germany. E-mail: marcel.thum@tu-dresden.de

1. Introduction

International policy challenges often concern global public goods. Climate change, as one example, is partly attributed to the emission of greenhouse gases. The mitigation of these emissions is a contribution to a global public good. Each country bears the cost of its own mitigation effort but the benefits accrue to all countries. Another example of an international public good is the fight against maritime piracy. Each country that sends convoying ships to the Horn of Africa makes the passage for all cargo vessels in this area safer. Again, the benefits accrue to traders from all over the world while the costs are born by the countries providing the convoying ships. For such global public goods standard economic theory predicts an under-provision due to free-rider behavior in absence of a global regulation. A huge experimental literature has shown that private contributions to public goods are indeed below the socially optimal level; however, contributions are much higher than predicted by orthodox game-theoretic considerations. Explanations for this phenomenon range from confusion (Andreoni 1995, Houser and Kurzban 2002) and warm-glow (Palfrey and Prisbrey 1997), to altruism (Goeree et al. 2002) and conditional cooperation (Fischbacher, Gächter and Fehr 2001; Fischbacher and Gächter 2010).

Public goods experiments usually consider single decision makers, while decisions on global public goods are typically made by teams, as the examples above indicate. The ministry of the environment decides on the national mitigation policy, the defense ministry and its experts allocate the resources for fighting maritime piracy. This raises the question whether the results from individual decision-making on public good contributions carry over to team decisions. Do teams show the same behavior regarding the private provision of public goods as individuals do? One might expect that the deliberation in teams fosters strategic considerations and that other-regarding aspects are weakened. With altruism, for instance, individuals may care about other individuals but not about more or less anonymous teams. There is a rapidly growing (experimental) literature on team decisions analyzing how teams make decisions and whether they make different decisions than individuals do (Charness and Sutter 2012, Kugler, Kausel and Kocher 2012). This literature, however, has not covered the politically important case of public good provision.

Our focus is on teams that jointly provide a single public good. We analyze whether teams contribute significantly more or less than individuals in a standard public goods game with and without a punishment option. We extend the standard public goods game with four single decision makers to a setting with four teams who jointly provide the public good. Each team consists of three players who have to coordinate on a team decision. The set of four teams that interact in the provision of the public good is referred to as a *group*. We analyze the contributions to the public good, the use of punishment and the final payoff. For the coordination among team members, we use a structured method of team decision-making in the spirit of Gillet, Schram, and Sonnemans (2009). This method allows team members to make proposals but it does not allow for communication. The proposals are aggregated to a team decision following either a majority rule or a unanimity rule. The lack of direct communication certainly eliminates an important aspect of real-world team decisions. However, it allows us to isolate the effects of the decision-making process itself. Therefore, the analysis is a first stepping stone in understanding complex team decisions. Further steps will have to consider direct communication among team members which will then allow isolating the effects of aggregation in teams from communication effects.

Teams in our experiment contributed significantly more to the public good than individuals if there was no punishment option. In treatments with a punishment option, teams punished significantly less than individuals. In particular, we find less anti-social punishment in team treatments. In terms of net profits, teams performed better than individuals in all treatments.
In section 2, we develop our working hypotheses. Section 3 describes the experimental setting in detail. Section 4 discusses the treatment effects. In Section 5.1 we scrutinize our finding that teams punish less by analyzing disaggregated data on the team level. Section 5.2 provides possible explanations for the somewhat surprising result that teams contribute more if there was no punishment option. Section 6 concludes.

2. Literature and Hypotheses

Charness and Sutter (2012) and Kugler, Kausel and Kocher (2012) provide recent surveys on team decision experiments. According to Charness and Sutter, the majority of research concludes that teams tend to behave more in line with game-theoretic predictions than individuals. They identify three reasons for the difference between decisions made by teams and individuals (p. 171). First, individual knowledge is aggregated in teams, and thus teams make qualitatively better decisions (e.g., investment decisions). Second, teams exhibit more detailed reasoning when making strategic decisions (e.g., in the beauty contest game and the trust game). Teams are better able to anticipate the reaction of the other player and his/her best strategy. Third, teams have a stronger focus on payoffs. Fairness and reciprocity seem to play a minor role in team decision-making.

Kugler, Kausel and Kocher (2012) refer also to results from social psychology dealing with team decision-making. Most of this literature analyses behavior in a prisoner's dilemma situation and stresses that individuals behave differently as sole decision-makers compared to deciding as members of a team. In social psychology, this difference in behavior is usually referred to as *discontinuity effect*. Two main motives for the discontinuity effect are 'greed' and 'fear' (Wildschut et al. 2003). 'Greed' refers to a player's stronger focus on payoffs in team decisions. It is explained by either the higher anonymity within a team that provides shelter from social punishment or the social support within a team for self-interest behavior (social support of shared self-interest hypothesis, see also Kugler et al. 2007). The second motive, 'fear', refers to the expectation of decision-makers that teams act more competitively and less cooperatively (schema-based distrust hypothesis). In a prisoner's dilemma situation, decision-makers tend to expect teams to act more selfishly and thus to defect more often; they fear being exploited and protect themselves by choosing defection as well. In the public good game, which is a variant of the prisoner's dilemma problem, the discontinuity effect would suggest that teams contribute less.

Economic research has also investigated team decisions and generally finds that teams behave more in line with game-theoretic predictions compared to individuals. The interactive tasks for which this effect was found include the ultimatum game (Robert and Carnevale 1997, Bornstein and Yaniv 1998), the dictator game (Luhan et al. 2009), the beauty contest game[1] (Kocher and Sutter 2005), the centipede game (Bornstein et al. 2004), the gift-exchange game (Kocher and Sutter 2007), the trust game (Kugler et al. 2007), the finitely repeated prisoner dilemma game (Kagel and McGee 2016) and the sequential market game (Stackelberg duopoly, Cardella and Chiu 2012).[2] There are also at least four studies in which teams behave less in line with game theory or are less able to process information efficiently. Cason and Mui (1997) were among the first to study experimentally the decisions of teams in an economic framework. In a dictator game, they found: *"[…] that when a team consists of members who have made different individual offers* [in a previous individual stage of the game, authors' note]*, the team offer tends to be dominated by the more other-regarding member"* (Cason and Mui 1997, p. 1477). They used an experimental setting with 2-person teams and face-to-face communication. Luhan, Kocher and Sutter (2009) repeated the game in a different environment (communication via electronic chat, 3-person teams). They found that teams act more selfishly and that the most

[1] Groups tend to perform better during the game due to improved reasoning abilities, but not in the first period.

[2] Cooper and Kagel (2005) find that teams are closer to the predicted equilibrium strategy in a signalling game with limit pricing and market entrance.

selfish player in a team has the strongest influence. The second study that deviates from the main stream of literature is that by Cox and Hayne (2006), who studied a common value auction with risky outcomes. They identified a *curse of information*. If additional information is provided on the value of the auctioned item, individuals and teams bid less rationally. In addition, this curse of information effect is even stronger for teams. Thirdly, Sutter et al. (2009) show that teams suffer more often from the winner's curse as teams stay longer in auctions and pay higher prices than individuals. The fourth and most recent study is by Müller and Tan (2013). They set up a sequential 2-player market game (Stackelberg duopoly) and found no significant difference between individuals and teams in a one-shot game. In the repeated game, team decisions were less in line with game-theoretic predictions compared to individual decisions.

The list of interaction tasks analyzed in this context still lacks the standard public goods game, which is the game-theoretical framing of numerous real-world problems; environmental protection, combating maritime piracy, military protection, vaccination programs and financial stability are just a few relevant examples. We implement team decision-making in a public good framework and compare it with the standard individual game. We also consider punishment as it is an important instrument to foster cooperation and because it is frequently used in real world politics; for example, trade restrictions are linked to non-cooperation in international environmental policy.[3] In the contribution as well as in the punishment stage, we apply two different decision-making rules: majority and unanimity rule. Both voting procedures have been intensively studied for instance in the context of jury decisions. Theoretical and empirical findings in this context show that the two decision making rules may imply different incentives for strategic voting and consequently lead to different results (see f. ex. Feddersen and Pesendorfer, 1998, 1999, Guarnaschelli, McKelvey and Palfrey, 2000, Bond and Eraslan, 2010, Goeree and Yariv, 2011).[4] These results mostly refer to voting situations where voters have common preferences and are therefore not directly applicable to our scenario, but the strategic interaction with majority and unanimity rules may be similar in our setting.

[3] In general, the restriction of trade per se has no benefit for the home country but is employed as a punitive measure if the foreign country does not contribute sufficiently to the public good of "global environmental quality". Among others, Barrett (2003) discusses this linkage in the context of international environmental agreements.

[4] Feddersen and Pesendorfer (1998) show, for instance, that due to strategic voting a jury deciding unanimously convicts an innocent defendant with a higher probability in comparison to a jury deciding by majority. In addition, Bond and Eraslan (2010) point out that this result may crucially depend on the exogeneity of the issue that is put to vote. Goeree and Yariv (2011) find that collective decisions differ with different decision making rules when communication is restricted (no deliberation allowed).

Our aim is to address the following research question: Do teams outperform individuals in a public goods setting with and without punishment? To answer this question, we examine the level of cooperation (contributions). We also test whether the punishment behavior of teams is closer to standard game theory with rational and selfish preferences.

Our point of reference is the standard finite public goods game, which assumes purely selfish individuals and predicts extensive free-riding. In this framework, it is individually rational to contribute nothing (subgame perfect Nash-equilibrium). And as all players are assumed to be symmetric, this result does not change when decisions are jointly made in teams.

In public good experiments with individual decision makers, however, positive contributions and a substantial level of cooperation are observed, which indicates that players are partly driven by other factors than own material payoff, such as altruism, fairness or reciprocity (see Ledyard 1997). Based on the experimental research on team decisions, we would expect teams to behave in a more competitive and self-oriented manner than individuals (see Charness and Sutter 2012). We also know that teams behave more in line with game-theoretic predictions in many interactive tasks. Both conclusions from the literature point to lower contributions by teams compared to individuals in a public goods game. Thus, our working hypothesis addressing cooperation is:

H1: Teams contribute less.

To evaluate this conjecture, we compare average contributions in team treatments with those in individual treatments.

Based on the reasoning behind hypothesis H1, we also expect teams to punish less in the presence of a punishment option. Punishment is costly and destroys resources. It also has the characteristics of a public good. If the investment in punishment by one team motivates others to contribute more, the whole group will benefit from the higher contributions. Thus, not punishing at all is rational for each player. Our second working hypothesis therefore is:

H2: For a given level of contributions, teams punish less.

While there is an incentive to free-ride on the others' contributions to a public good, contributions increase *welfare* from a collective perspective. In line with recent approaches (Gächter et al. 2008, Ambrus and Greiner 2012), we also account for the welfare effects of different treatments. For our welfare analysis, we abstract from potential warm glow effects and altruism, which may enter the welfare function, and focus on payoffs as our welfare measure. The overall outcome depends on both, contributions and punishment. In the absence of a punishment option, a corollary of hypothesis H1 is

H3: Teams yield lower aggregated payoffs in the absence of punishment.

In presence of punishment the overall outcome is ambiguous. Payoffs depend positively on contributions but are diminished by costly punishment. A corollary of hypothesis H1 is that teams should end up with lower payoffs. If hypothesis H2 is correct, teams should yield higher payoffs. As these two effects are countervailing and, moreover, punishment might depend on contributions, it is far from obvious whether teams will generate higher or lower aggregated payoffs compared to individuals.

3. Experimental Design and Procedure

The general framework is a standard public goods game with and without punishment (Ledyard 1997, Fehr and Gächter 2000 and 2002). An experimental group of four teams or four individuals was formed. Each team consisted of three players. Each team/individual received the same endowment (20 tokens) at the beginning of each period and had to decide how many (integer) tokens to invest in the public good. The marginal per capita return for each token invested in the public good was 0.4, which accrued to each of the four teams.[5] All tokens not invested in the public good were retained by the team/individual; this was equivalent to a marginal return of 1 for the investment in the private good.

In the public goods game with punishment, the contribution stage was followed by a punishment stage. In line with the recent literature on punishment, we implemented the following punishment technology: a 3-token reduction in the payoff of another player costs the punishing player 1 token (see Gächter, Renner and Sefton 2008). We set up three between-subject treatments: an individual treatment (IND), a majority team treatment (MAJ) and a unanimity team treatment (UNA). In the majority team treatment, a minimum of two out of three team members had to agree. In unanimity teams, all members had to agree on a decision.

[5] The individual payoff of a team member is equal to the team result. Technically, the experimenter tripled the team's profit and awarded an equal share to each team member. Thus, the incentives in the individual and team treatments were identical.

The experiment consists of three parts. In Part I, all participants played an individual, one-shot public goods game. This one-shot game reveals information about the type of the player, e.g., regarding her preferences for contributions. No feedback was provided in Part I to avoid learning at this stage. At the beginning of Part II, the instructions for this part were distributed and participants were randomly assigned to the treatments (IND, MAJ and UNA) and to the teams in case of the UNA and MAJ treatment. Then in Part II, participants played a 10-period standard public goods game as a team member or an individual player with fixed matching. After each period, each participant received feedback concerning his/her results and anonymized information on the decisions of the other teams/individuals. Part III was implemented as a surprise restart.[6] Participants stayed in their treatment and teams were re-matched within their matching group of 12 such that every participant was assigned to new teammates. This procedure was common knowledge. The re-matching ensured that participants had no prior information concerning the behavior of their teammates.

Participants played a 10-period fixed matching public goods game with costly punishment. After each contribution and each punishment stage, participants received feedback on their results and anonymized information on the decisions of the other teams/individuals. After Part III had been completed, we disclosed the results of Part I. Figure 1 summarizes the setting and the sequence of the different parts of the experiment.

Figure 1: Structure and Sequence of the Experiment

		Part I	Part II	Part III		
		Contribution	Contribution	Contribution & Punishment (Stage 1)	—	(Stage 2)
Treatments	Majority	Individual contribution decision	MAJ	MAJ	—	MAJ
	Unaninimity		UNA	UNA	—	UNA
	Individual		IND	IND	—	IND
	Periods	1	10	10		
	Feedback*	no	yes	yes		

Note: * Feedback after each proposal or decision.

[6] Subjects were allowed to quit after Part II (nobody did so): Part III payoffs were in addition to the previous parts.

We implemented a structured method of team decision-making in the spirit of Gillet, Schram, and Sonnemans (2009). At the beginning of each period, each team member could propose a contribution to his/her teammates. Specifically, each team member had to type in a number (0…20) as his/her proposed contribution to the public good. Once all three team members had entered their numbers, the proposals were shown to the team. Under majority rule, two out of the three proposals had to be equal to form a decision. Under unanimity rule, all three proposals had to be equal. Except for making proposals to the other team members, we did not allow for any further communication between team members. Admittedly, face-to-face communication is an important part of real-world team decisions. However, this minimal, structured type of communication in the spirit of a pure coordination allows us to isolate the effects of the decision-making process (Gillet et al. 2009).

If an agreement was reached according to the team decision rule, it was automatically implemented as the team decision. If no agreement was reached, a new round of proposals started. After a maximum of 10 rounds of proposals without reaching a decision, a default rule applied. Then, the decisions of another team that had reached a compromise was randomly selected as the undecided team's choice and the team's payoff for this period was set to zero. Participants had full information about the default rule. The restriction to 10 rounds was non-binding for all majority team decisions and in the vast majority of cases of unanimous decision-making. With the unanimity rule, we observed disagreement in less than 2 % of the decisions. At the contribution stage, coordination failures occurred in only 15 cases out of a total of 800 unanimity team decisions. At the punishment stage, teams failed to agree on the assignment of punishment points in 8 out of 400 cases of unanimous decisions.[7]

At the punishment stage, players (individuals/teams) could penalize other players at a cost. When all decisions regarding contributions had been made, each participant received anonymous feedback concerning the contributions of the others. Then, players (individual/teams) could decide about the allocation of punishment points to the other three players. Reducing the payoff of another player by three tokens costs the punishing player 1 token. In team decisions, each member of a team could propose punishment expenditures, for

[7] The default rule is a compromise between manageability and avoiding feedback effects from the default rule on the team decisions. In the majority treatment, a single player can never cause a default. In the unanimity treatment, however, a default implements non-extreme contributions for the defaulting team on average. If there are two players with extreme preferences in a team and one with preferences near the average, the third player could use the default option to achieve a contribution closer to his own preferences. We observe this behavior in 6 out of 15 cases. Most defaults are simply due to coordination failures.

example, team A: 0 token, team B: 1 token and team C: 3 tokens. If the punishment schedules of two (three) team members were identical, they formed the team decision. This process was automatically implemented by the program according to majority (unanimity) rule. As in the contribution stage, the default rule was designed in a way to minimize strategic behavior of team members. . If a team had not found an agreement after 10 rounds of proposals, the software randomly selected one of the decisions of the other teams in the group. The team had to bear the costs of this random decision.[8]

The experiment was programmed and conducted using z-tree software (Fischbacher 2007). Sessions were held in the experimental laboratory at the University of Mannheim from September to November 2012. Subjects were recruited using a database of volunteers (Greiner 2003, 2015). In all treatments, subjects were invited to participate for 90 minutes. In general, individual treatment sessions lasted approximately 60 minutes, sessions with the majority treatment lasted 75 minutes and those with the unanimity treatment up to 90 minutes. In total, 280 students, approximately half of which were bachelor-level business administration students, participated in the experiment. After being randomly seated in separated cubicles in front of a computer, which was subsequently used to run the experiment, participants received the instructions for Part I. The experiment began after all participants had passed a quiz ensuring they all had understood the instructions. The instructor answered questions individually and in private. No communication among participants was allowed during the experiment. The exchange rate for tokens was 2 Eurocent per token. Payments were made anonymously in cash immediately following the experiment with an average payoff of 11 Euro.

4. Results

Table 1 contains the summary statistics on the group level for average contributions, average punishment (when applicable) and average net profits in each treatment. Each matching group of four individuals/teams provides an independent observation. Thus, we have 10 observations for each part of the experiment (no punishment/punishment) and each decision rule (unanimity/majority/individual).

Altogether, average contributions are higher and average punishment is lower in the team treatments compared to the individual treatment. This results in higher net profits in the team

[8] Fixing a certain level of punishment, e.g. ‘no punishment’, as default would distort the team decision. It would give the player who prefers the default outcome an incentive to make sure that the team finds no agreement.

treatments as well. As stated earlier, we observed coordination failures only in the unanimity team treatment and here in less than 2 % of the decisions. Taking a closer look at the decision-making process within teams, we find that coordination was achieved fairly rapidly in all treatments.

Under majority rule, the teams required on average 1.6 and 1.3 rounds to reach a decision without and with punishment, respectively. Moreover, even in the much more complicated treatment with unanimity, 80 percent of teams reached agreement in 5 rounds or fewer. Here, the average number of rounds was 3.9 and 2.7 without or with punishment, respectively. One reason for the high coordination rate – despite the broad range of possible contributions from 0 to 20 – is certainly that there are clear focal points. Most participants suggested a contribution of 0, 10 or 20.

We now test the 3 hypotheses from Section 2.

Table 1: Summary Statistics: Average Contributions, Punishment and Net Profits

		Part I	Part II		Part III		
		One-shot PG w/o Feedback	10-period PG with Feedback		10-period PG with Punishment and Feedback		
	N	Avg. Contri-butions	Avg. Contri-butions	Avg. Net Profits	Avg. Contri-butions	Avg. Punish-ment Costs[a]	Avg. Net Profits
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
UNA	10[c]	9.08 (4.12)	9.78 (5.14)	25.24 (3.98)	16.24 (5.75)	0.53 (0.76)	27.27 (5.63)
MAJ	10[c]	9.62 (3.31)	8.98 (5.89)	25.39 (3.54)	14.95 (5.28)	0.59 (0.75)	26.62 (4.72)
IND	10	8.95 (3.55)	7.56 (5.31)	24.54 (3.19)	14.89 (6.17)	1.11 (1.39)	24.51 (7.39)
Teams[b]	20	9.35 (3.71)	9.38 (5.53)	25.31 (3.75)	15.60 (5.54)	0.56 (0.75)	26.95 (5.19)

Note: standard deviations in parentheses, N: number of matching groups (a matching group consists of 4 participants in the individual treatment and 12 participants in a team treatment), UNA: unanimity treatment, MAJ: majority treatment, IND: individual treatment

a Average number of assigned punishment points, equaling the average cost of punishment

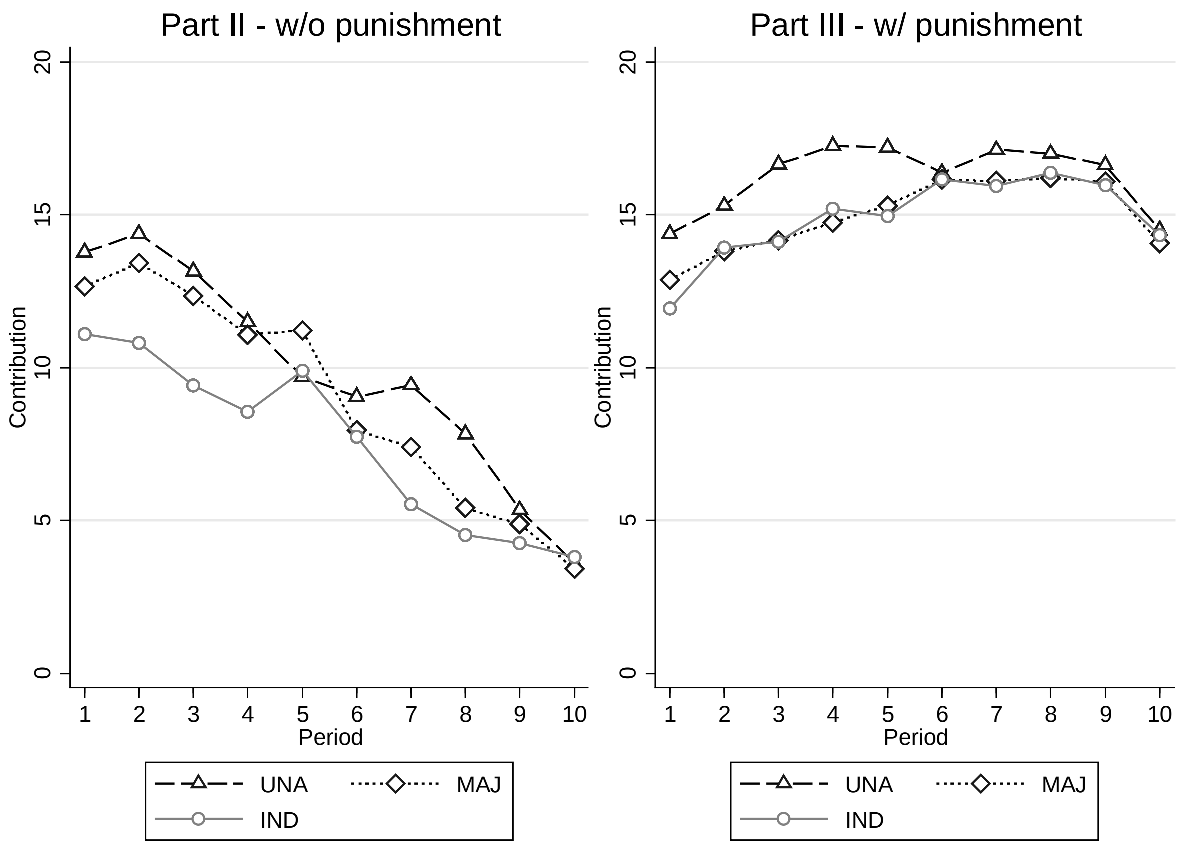
b "Teams" consists of the majority and unanimity treatments.

c the 240 subjects later assigned to the teamtreatments played in 60 groups in part I.

H1: Teams contribute less.

Figure 2 depicts the average contribution for each round for the two team treatments (MAJ, UNA) and the individual treatment (IND). The left panel refers to Part II without a punishment option and the right panel to contributions in Part III with punishment. The two panels exhibit well-known properties in finite horizon, repeated public good games. Without punishment, cooperation erodes over time and contributions decline. With punishment, cooperation can be sustained (Ledyard 1997, Fehr and Gächter 2000 and 2002). Our focus is on the effects of team decisions. An initial inspection of the curves for the two team treatments and the individual treatment suggests that teams contribute more in almost all periods. Moreover, for Part II there is no visible difference between the UNA and the MAJ treatment, whereas in Part III the UNA teams seem to have contributed on average more in comparison to the MAJ teams and the individuals. As we will see later, these differences are not statistically significant. In the following statistical analysis we therefore aggregate the two team treatments to one single treatment.

Figure 2: Average Contributions in the Treatments with and without Punishment



The general pattern in the graphs are corroborated by a cross-section as well as a panel regression analysis. The cross-section analysis gives us the average difference between the team and the individual treatments over all periods. We implement a standard OLS regression on the group level with period dummies to take into account learning effects (see Appendix A for details). The results for the cross-section are reported in Table 2 (Columns 1 and 3). We explain the average contributions on the group level with a team dummy, the contributions in the one-shot game and nine period dummies (coefficients of the period dummies are not reported in the table).[9,10] Without punishment (Part II), teams contribute significantly more than individuals. The effect is not only statistically but also economically significant. On average, teams contribute 20 percent more than individuals.[11] This stands in contrast to other experiments and requires a more detailed analysis (see section 5.2). With punishment, the team effect is much smaller (3 percent) and statistically not significant. The regression analysis also reveals that the type of a player matters. We take the contribution from the one-shot game in Part I as an indicator of a participant's selfishness or other-regarding preferences. A low contribution indicates an orientation towards selfish behavior. In the case of teams, we calculate the average one-shot contribution across the group. The economic effect is quite large. A one-token increase in the contribution in the one-shot game translates into a 0.8-token greater contribution in the repeated game (Part II).

[9] We have also analysed the two team treatments, separately. The effects are mostly the same as in our aggregate team treatment but become mostly insignificant at standard levels due to the low number of observations. As stated above, there is no statistically significant difference between the MAJ and UNA treatment. Therefore, we decided to pool the majority and the unanimity treatments at this point of the analysis.

[10] For each part, treatment and period, we have 10 groups playing the public good game. These 10 groups provide us with 10 independent observations for each period. Therefore, no clustering on the group level in addition to the period dummies is needed.

[11] The coefficient of the team dummy is 1.499. The average contribution in the individual treatment without punishment amounts to 7.56 (see Column 3 in Table 1). This yields a team effect of roughly 20 percent (≈ 1.499/7.56 x100).

Table 2: Regression Results (group level)

	Part II 10-period PG with Feedback		Part III 10-period PG with Punishment and Feedback			
	Contributions[a]	Net Profits[a]	Contributions[a]	Punish-ment Costs[b]	Punish-ment Costs[b]	Net Profits[a, c]
	(1)	(2)	(3)	(4)	(5)	(6)
Team	1.499*** (0.443)	0.586* (0.292)	0.452 (0.631)	-0.676** (0.235)	-0.746 (0.847)	2.209** (0.764)
Contributions				-0.092*** (0.022)	-0.066 (0.045)	
Contribution in one-shot game[d]	0.795*** (0.071)	0.473*** (0.047)	0.639*** (0.129)	-0.036 (0.042)	-0.055 (0.043)	0.569*** (0.135)
Team x Contributions					-0.041 (0.051)	
Intercept	4.181*** (0.909)	22.76*** (0.587)	6.867*** (1.541)	2.287*** (0.557)	2.066** (0.712)	17.417*** (1.552)
Period dummies	yes	yes	yes	no	no	yes
Observations	300	300	300	300	300	300
R^2	0.52	0.43	0.12	0.05[e]	0.05[e]	0.14

+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001, robust standard errors in parentheses,
a OLS regression with period dummies; coefficients for the period dummies not reported here; robust standard errors,
b Tobit (marginal effects), robust standard errors,
c Profits net of punishment costs and punishment received,
d Group average.
e pseudo R^2

Table 3 reports the results of the corresponding panel regression on the team level. We use a random effects model as the particpants have been randomly sorted into teams and we thus expect no fixed effects on this level. We also include group and period dummies to control for group specific effects and learning (see Appendix A for details). Columns 1 and 3 in Table 3 show that teams contribute more than individuals in Part II and III, but the effect is only significant for Part III.

Table 3: Regression Results- Random Effects (team level)

	Part II 10-period PG with Feedback		Part III 10-period PG with Punishment and Feedback	
	Contributions	Net Profits	Contributions	Net Profits
	(1)	(2)	(3)	(4)
Teams	3.200 (3.283)	1.320 (3.744)	10.450*** (2.794)	17.700*** (1.709)
Intercept	10.087*** (1.478)	26.262*** (1.471)	5.873* (2.666)	9.176*** (1.694)
Period dummies	yes	yes	yes	yes
Group dummies	yes	yes	yes	yes
Observations	1,200	1,200	1,200	1,200

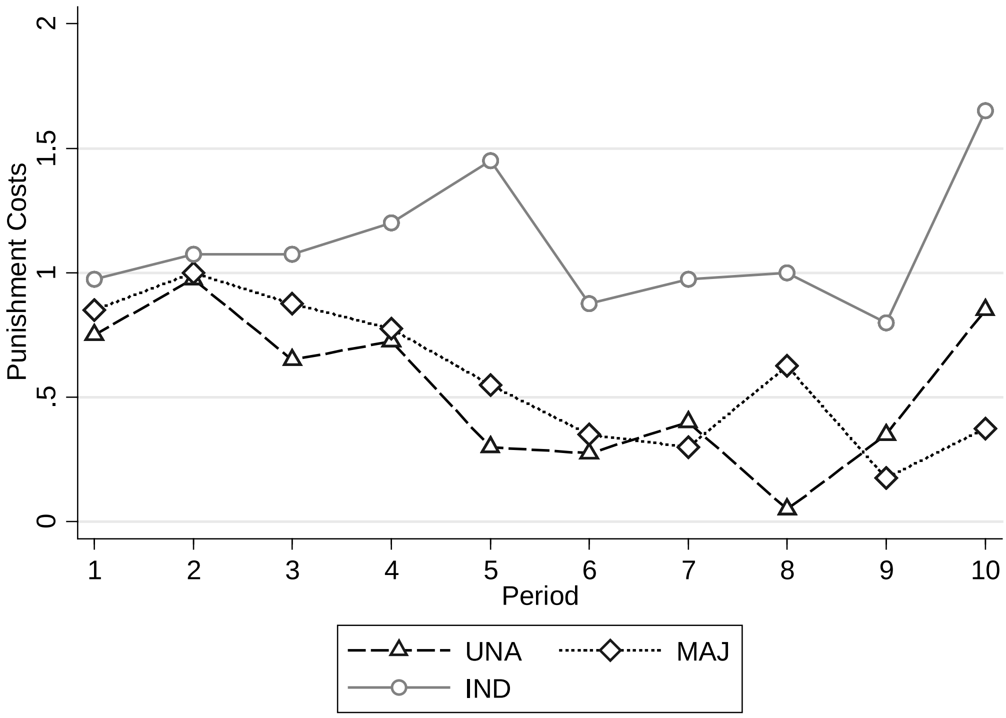
+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001, robust standard errors in parentheses.

H2: For a given level of contributions, teams punish less.

If teams behave closer to standard game-theoretic predictions, we should observe less punishment for a given level of contributions in a team setting. Punishment is costly for the punisher and the punished. Independent of whether players are selfish or have other-regarding preferences, destroying resources should not be in the interest of a rational player. Figure 3 depicts the average punishment costs[12] in the team and individual treatments for each period. The level of punishment remains nearly constant over all periods in the individual treatment and declines slightly in the two team treatments.

[12] The term “punishment costs” in our setting always refers to the expenditures of the punisher (the assigned punishment points) not to the amount of token that are taken away from the punishee. The latter can be easily calculated by tripling the punishment costs.

Figure 3: Average Punishment Costs



We explore this difference using a Tobit analysis as the data is left-censored.[13] The dependent variable is the average punishment cost incurred by the teams. Column 4 in Table 2 indicates that punishment was less common in team treatments than in individual decisions. As expected, a higher contribution level goes along with lower punishments. Column 5 includes the interaction effect between the team dummy and contributions. This suggests that teams punish less for a contribution of zero and they reduce their punishment more quickly when contributions are increased. However, the effects are no longer significant.

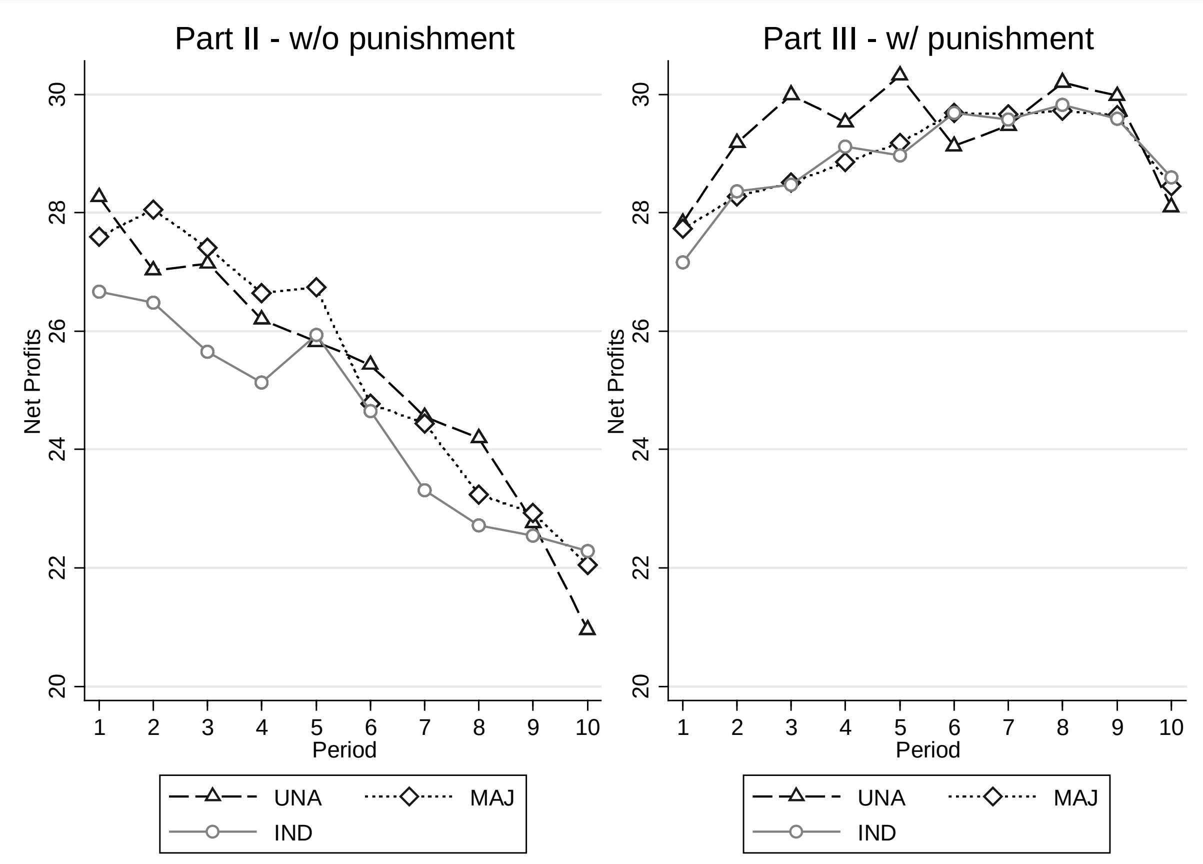
H3: Teams yield lower aggregated payoffs in the absence of punishment.

Next, we compare the net profits realized in the different treatments. The net profits of each team/individual are calculated by adding the returns from the public good to the retained tokens and then subtracting the costs of punishing and being punished. Figure 4 displays the average net profits over the 10 periods by treatments with and without punishment (the right and left panels, respectively). Overall, teams performed better. In the treatment without punishment, the net profit curves of the two team treatments lie slightly above that of individuals except for the last period. The better performance results from the higher contributions of teams.

[13] As we pool all groups over all periods in the cross-section analysis, we have enough observations to implement a Tobit regression in the cross-section. A respective Tobit regression with panel data is not possible given the low number of observations per treatment and period.

In Part III with punishment, the individual and team contributions did not differ significantly. In this case, it is mostly the lower level of punishment imposed by teams that leads to increased net profits. Especially teams deciding by unanimity rule derive higher net profits due to slightly higher contributions and lower punishment. If we interpret the results in terms of economic welfare, teams generate higher welfare.

Figure 4: Net Profits



Again, we complement the investigation of net profits with a cross-section and a panel regression analysis. Columns 2 and 6 in Table 2 indicate that on average the team dummy is positive in both treatments and statistically significant in Part II. The regression also illustrates the already familiar pattern in the other variables. A higher contribution in the one-shot game in Part I led to higher net profits in Part II and III.[14]

The results from the random effects regression reported in Column 2 and 4 in Table 3 also confirm the first impression from Figure 4. In Parts II and III, teams yield higher net profits but

[14] Note that "contributions in the one-shot game" captures the average of the entire group. Thus, a group consisting of less selfish players generated higher profits, but this does not imply that a single individual/team gained from being less selfish.

this effect is statistically significant in Part III only. This goes back to the difference in contributions and punishment that have been discussed before.

5. Discussion

5.1 Team Decision-Making in the Punishment Stage

In the previous section, we have shown that there is less punishment in team treatments. Here we shed some more light on this finding by analyzing the punishment data in detail.

The lower punishment by teams can be traced back to two effects. Teams might use the punishment option less frequently, or they might choose lower levels of punishment. To distinguish the two effects we have to analyze disaggregated data on the team level (rather than on the group level). Table 4 shows the average punishment of individuals and teams in column 2. On average teams have punished significantly less than individuals. But as this does not allow us to tell whether teams punish at lower levels or are less likely to punish at all, we decompose punishment costs. First, we have calculated the share of punishment incidences over the 10 periods for each team and each individual (Table 4 column 3). The probability that an individual exercised the punishment option is 0.345. For teams, this probability is only 0.245. This difference, however, is statistically not significant at conventional levels. In addition, we have computed the level of punishments (Table 4 column 4). This is the average number of tokens spent on punishment conditional on having decided to punish at all. Here, the difference between individuals and teams is statistically significant. Whenever individuals used the punishment option, they spent some 3 tokens on average on punishment, while teams only spent a little bit more than 2 tokens. Hence, the difference between the average punishment by teams and individuals is more due to the chosen level of punishment rather than the probability of choosing punishment at all. Teams are more parsimonious with respect to their expenditures on punishment.

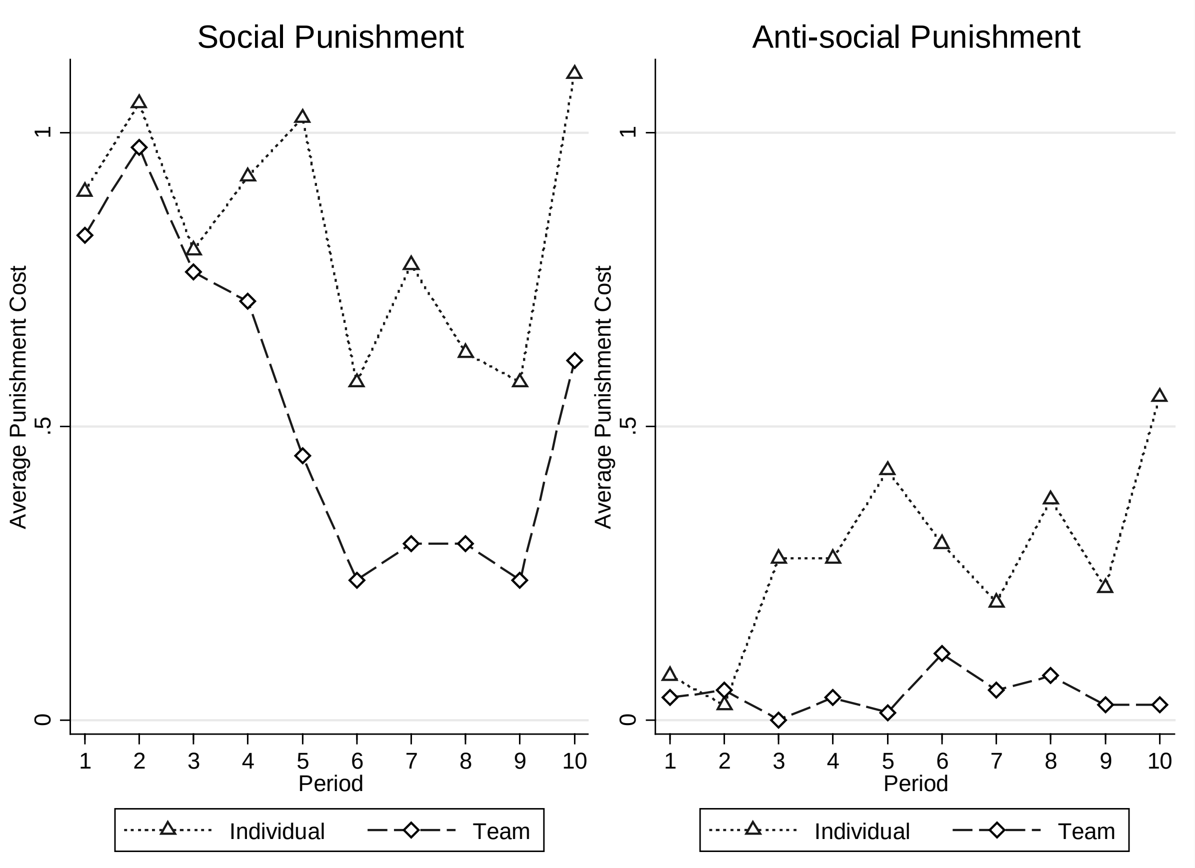
Table 4: Decomposition of Punishment Costs

	N	Avg. Punishment Costs	Probability of Punishment (on Individual/Team Level)	Level of Punishment (Absolute Punishment Costs)
	(1)	(2)	(3)	(4)
Individual	10	1.11 (1.105)	0.345 (0.300)	3.174 (1.154)
Team	20	0.56 (0.373)	0.245 (0.170)	2.191 (0.897)
Mann-Whitney-U-test[a]		p=0.355	p=0.551	p=0.026

Note: standard deviation in parentheses, due to a different level of aggregation standard deviations reported in column (2) slightly differ from those reported in Table 1 column (6).
a also known as Wilcoxon rank-sum test.

There is one more striking difference in the punishment behavior of teams and individuals that may tell us something about the motives behind punishment. While anti-social punishment is quite common in the individual treatments, it hardly ever shows up in the team treatment. Anti-social punishment is defined as *"the sanctioning of people who behave prosocially"* (Herrmann, Thöni and Gächter 2008, p. 1362). In our case, anti-social punishment occurs when a player (team or individual) punishes another player who contributed more. Figure 5 splits up the punishment costs from Figure 3 on the individual and team level. The left-hand panel shows the average punishment costs for each period in all cases where the punished team or individual had contributed less than the punisher. The familiar pattern shows up. Individuals as well as teams use punishment in the first period to enforce cooperative behavior. The use of the option declines over time and more strongly so in the team treatments. The right-hand panel displays the spending on anti-social punishment. This includes all cases where the punisher had contributed less than the punished. The panel exhibits three striking features. First, anti-social punishment is much lower than social punishment as it occurs more rarely. Second, anti-social punishment increases over time in the individual treatment. Hence, the increasing punishment in Figure 3 is only due to this anti-social punishment. Third and most importantly for our investigation, there is hardly any anti-social punishment in team treatments. Decision-making in teams seems to eliminate the rare and fairly destructive anti-social behavior.

Figure 5: Social and Anti-social Punishment



In Table 5, we repeat the exercise from Table 4 but focus on the cases of anti-social punishment. The table shows average anti-social punishment in column 2 and the decomposed anti-social punishment costs in columns 3 and 4. The results reported in columns 3 and 4 show that teams do not only choose significantly lower levels of (anti-social) punishment as they do in the case of social punishment. They also exhibit a lower probability of choosing anti-social punishment at all.

Table 5: Decomposition of Anti-social Punishment Costs

	N	Avg. Anti-social Punishment Costs	Probability of Punishment (on Individual/Team Level)	Level of Punishment (Absolute Punishment Costs)
	(1)	(2)	(3)	(4)
Individual	10	0.273 (0.395)	0.078 (0.110)	2.946 (4.499)
Team	20	0.043 (0.078)	0.023 (0.040)	0.625 (0.930)
Mann-Whitney-U-test[a]		p=0.035	p=0.065	p=0.033

Note: standard deviation in parentheses
a also known as Wilcoxon rank-sum test.

A possible channel through which punishment might be reduced in team treatments is that the team decision making process prevents extreme preferences from being translated into final decisions. To be more precise: In the majority treatment, we expect that the team's median voter prevails.[15] If the median's willingness to punish is below the average willingness to punish, and thus the distribution is skewed, majority teams should exhibit lower punishment levels than the average individual decision. Suppose that the single peaked distribution of true preferences over punishment within a team is (0, 1, 5). Then, given that all team members strictly reveal their preferences, a majority vote would lead to a punishment of 1 (implying a reduction of the punished player's payoff by 3 tokens). If these three team members had participated in an equivalent individual treatment, we would have observed an average punishment of 2. Thus, with a skewed distribution, decision-making in majority teams may result in lower punishment. To test whether such skewness could be responsible for our result, we consider the team players' first proposals for punishments in the first round, as these most likely reflect their true preferences. Clearly, the first proposals within a team are not truly independent but as we focus on the first period, this is as closely as we can get to a true statement of individual preferences. Table 6 shows the difference between the highest/lowest proposal and the median proposal in majority treatments. As expected, the distribution is skewed. The median proposal is much

[15] As we have no clear prediction on the outcome in unanimity teams, we restrict the analysis to majority teams where we can expect the median voter to be decisive. The results of a simple regression analysis showed that the median voter's first proposal significantly explains the final decisions of teams.

closer to the minimum than to the maximum proposal. This suggests that teams show a higher degree of economic rationality by cushioning the extreme preferences for punishment.

Table 6: Skewness of Preferences: Initial Punishment Proposals to Team Members in the First Period

	Number of Teams	Initial Punishment Proposals
	(1)	(2)
Maximum – Median	40	0.517
Median – Minimum	40	0.225
Difference		0.292
Wilcoxon matched pairs signed-rank test		p=0.000

Note: Majority treatment only

5.2 Team Decision-Making in the Contribution Stage

We have tried to answer the question why teams punish less. We now explore the reasons behind the higher contribution levels that we observed in team treatments. Unfortunately the answer is not straightforward. We suggest three possible explanations, and we check to what extent they are consistent with our experimental evidence.

Skewness in the Preference Distribution

A skewness of preferences as it was at work in the punishment decisions could also be decisive for the higher contribution levels of teams. If the median team member's willingness to contribute is above the average willingness to contribute, team decisions with majority voting should lead to higher contributions compared to the individual treatment. Suppose that the three team members' preferred contributions are (0, 15, 15). Then, a majority vote would lead to a contribution of 15. In an individual treatment, the average contribution would have been 10. Thus, with a skewed distribution, team decision-making may lead to higher contributions. Again, we consider the team players' first proposals in the first round (contribution stage) as a proxy for the players' true preferences. Columns 2 and 3 in Table 7 show the difference between the highest/lowest proposal and the median proposal in majority teams (with and without punishment). The distribution turns out to be skewed; the median is above the team average. However, the difference is statistically not significant.

Table 7: Skewness of Preferences: Initial Contribution Proposals to Team Members in the First Period

		Part II 10-period PG with Feedback	Part III 10-period PG with Punishment and Feedback
	Number of Teams	Initial Contribution Proposals	Initial Contribution Proposals
	(1)	(2)	(3)
Maximum – Median	40	5.250	4.475
Median – Minimum	40	5.800	5.825
Difference		−0.550	−1.350
Wilcoxon matched pairs signed-rank test		p=0.937	p=0.526

Note: Majority treatment only

Elimination of Extreme Preferences Drives up Contributions

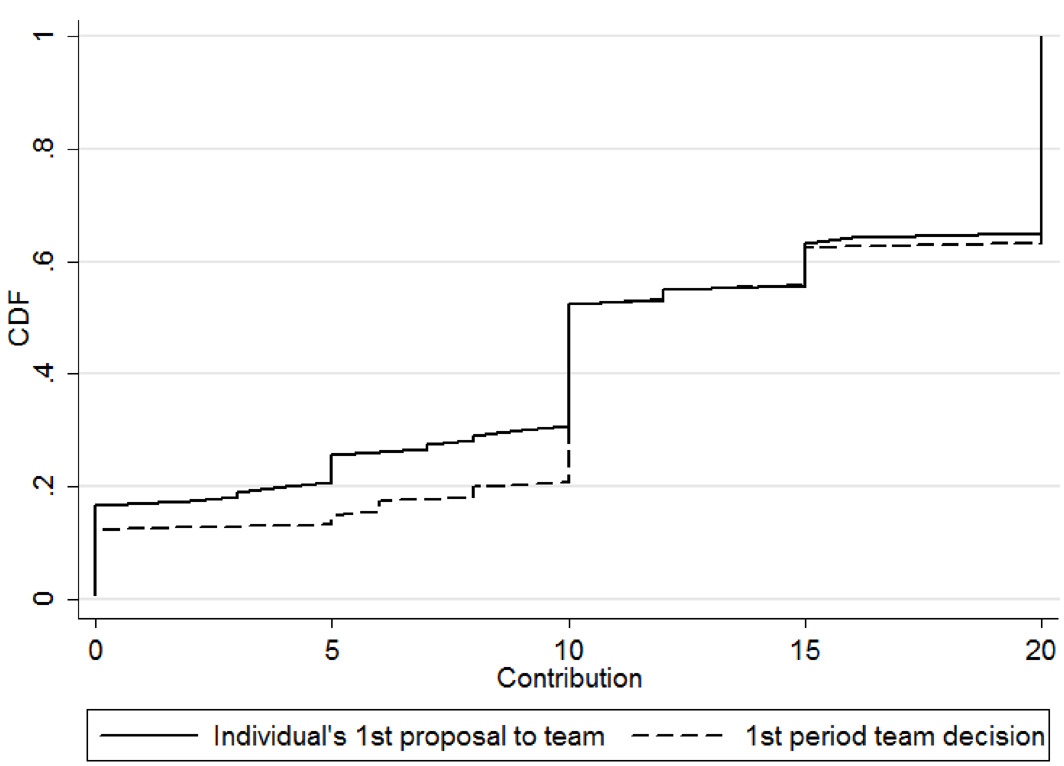
The second explanation focuses on the interaction of (majority) teams. Majority decisions may reduce the variance in the contributions and, in equilibrium, this may drive up the contributions of all teams in a group. In Appendix B, we sketch a simple model of public good provision where individuals gain utility from payoffs and from own contributions to a public good. We assume that individuals differ in terms of their willingness-to-contribute. This model shows that a reduction in the variance of the willingness-to-contribute parameter leads to higher contributions in equilibrium.[16] Team decision-making eliminates extreme preferences and thus may reduce the variance in final contributions. If this hypothesis has some empirical relevance in our experiment, we should observe (1) the variance of first proposals in teams to be larger than the variance of final contributions and (2) the variance of contributions to be lower in the majority team treatment than in the individual treatment.

Figure 6 compares the first proposals in the first period (majority treatment without punishment) with the final team decision on contributions. The data is aggregated to obtain cumulative distribution functions (cdf). For each contribution level (horizontal axis), the solid line shows

[16] This effect depends on the slope of the reaction curves. Note that it is not our aim to show that this effect is always at work. We just want to demonstrate that team decision-making may alter the outcome when players differ in terms of their willingness-to-contribute.

the percentage share of first proposals below this level. 20 percent of the team members proposed a contribution below 5; 50 percent proposed a contribution of 10 or less. The dashed line shows the cumulative distribution function for the final contributions of teams. If the hypothesis above has some explanatory power, we should observe that the cdf of final contributions is steeper than the cdf of the first proposals. For contribution levels below 10, the dashed line is below the solid line. This implies that team decision-making reduces the variation in contributions below the median. Even without detailed statistical analysis it is obvious that this cushioning effect is fairly small.[17]

Figure 6: Cumulative Distribution Functions of Initial Proposals and First Period Contributions in the Majority Treatment



Social Approval

A third potential explanation is that team coordination triggers social conformity. It is a well-established result that a desire for social approval can increase cooperation among strangers

[17] In Appendix B, we compare standard deviations of contributions in the individual treatment and in the majority treatment for each period. Standard deviations in the majority treatment are not significantly smaller than in the individual treatment. Hence, there is little evidence, that the elimination of extreme preferences and the strategic interaction of teams drive the results.

(Rege and Telle 2004). In contrast to Rege and Telle's experiments, the identity of the player is not revealed to the others in our setting. However, the team members demonstrate their willingness to behave cooperatively to their fellow team members, and in contrast to individual treatments, this revelation has no immediate consequences for the payoffs. The weaker the link between one's own actions and the payoff, the more likely it is that individuals will behave according to social norms.[18] If this hypothesis has explanatory power, we should observe that players' first proposals are highest in the unanimity team treatment, lower in the majority team treatment and lowest in the individual treatment. While the proposal has immediate consequences for payoffs in the individual treatment, the probability that a high first proposal leads to high contributions is negligible when decisions are made unanimously as all three team members would have to make the same proposal. Hence, participants might be willing to seek social approval via high first proposals in team treatments.

A high first proposal alone cannot explain why teams contribute more in equilibrium. Here the literature on communication and announcements in public good games especially the work on numerical cheap talk (NCT) provides some interesting insights that might play a role in our setting. In general this research finds a positive effect of ex-ante communication on the contribution level. Ledyard (1997) gives a brief introduction to the effect of communication in public good experiments and concludes that communication increases contributions in standard public good games with small groups (N<15). The meta-studies by Sally (1995) and Balliet (2010) support this finding of a general positive effect of communication in social dilemma games. Balliet (2010) concludes that, in particular, face-to-face communication increases cooperation. For our analysis, the studies by Denant-Boemont, Masclet and Noussair (2011) and Berlemann, Dittrich and Markwardt (2009) are of special interest; they look into the effect of non-binding announcements on contributions in public good games. Both studies find a positive effect.

[18] See Goodin and Roberts (1975) for the example of the 'ethical voter'.

Table 8: Initial Proposals to Team Members versus Decisions of Individuals in the First Period of the Contribution Stage

		Part II 10-period PG with Feedback	Part III 10-period PG with Punishment and Feedback
	N	Initial Contribution Proposals	Initial Contribution Proposals
	(1)	(2)	(3)
IND	40	11.10 (1.12)	11.93 (1.11)
MAJ	40	11.82 (0.76)	12.33 (0.88)
UNA	40	12.74 (0.60)	12.38 (0.86)
Mann- Whitney- U-test[a]	IND-MAJ	p=0.621	p=0.880
	IND-UNA	p=0.230	p=0.915

Note: standard deviations are given in parentheses, N: number of participants, UNA: unanimity treatment, MAJ: majority treatment, IND: individual treatment
a also known as Wilcoxon rank-sum test.

Table 8 shows the average initial proposals in period 1 in the contribution stage of Parts II and III. The initial proposals increase when moving from individual to majority to unanimity treatments for both parts of the experiment. The Mann-Whitney-U-tests at the bottom of Table 8, however, reveal that the differences between proposals in the individual versus majority and versus the unanimity team treatment are not statistically significant.

6. Conclusions and Outlook

We analyze team decisions in a public goods setting. The outcomes of our experimental study confirm some but not all of the results of previous studies on team decisions in other games. First, as in prior work, we find that teams exhibit a higher degree of economic rationality. Teams punish significantly less than individuals; because punishment wastes the resources of the punisher and the punished, punishment is not a rational choice. Second, our results diverge from those of most other studies in that teams do not behave in a manner more similar to game-theoretic predictions. In contrast to our initial expectation and our working hypothesis, teams contribute more to the public good than individuals do. Overall, teams outperformed individual players.

This study suggests several promising avenues for future research. First, our analysis is only an initial attempt to explain the strikingly high contributions of teams. A more elaborated setting would allow the testing of further hypotheses regarding the high level of cooperation observed in the team setting. For instance, with punishment, cooperation evolves more quickly in team treatments. Second, mixed treatments with individuals playing against teams might be very fruitful in exploring the reasons for the divergent behavior reported here. For instance, we would like to know whether teams are less willing to punish or whether individuals find it more difficult to punish a team (rather than other individuals). Third, we would like to study team coordination in greater detail. This requires to allow for controlled communication among team members. Such a setting would help to gain a better understanding of the decision-making process under the requirements of unanimity.

Appendix A

Cross- section regression model
OLS on group level with period dummies and robust standard errors

$$x_{j,t}^{p} = \alpha_0 + \alpha_1 Team_j + \alpha_2 \bar{x}_{j,0} + \tau_t + \varepsilon_{j,t} \qquad \text{(A1-1)}$$

$x_{j,t}^{p}$	average contribution of group *j* in period *t*
$Team_j$	team dummy (0 for IND, 1 for MAJ and UNA)
$\bar{x}_{j,0}$	average contribution of group *j* in Part I (individual one-shot game)
τ_t	period dummies
$\varepsilon_{j,t}$	error term
t	periods 1 to 10
j	groups (10 for each treatment)

Panel regression model
Random effects on team level with period and group dummies and robust standard errors

$$x_{i,j,t} = \beta_0 + \beta_1 Team_i + a_j + \tau_t + u_i + \varepsilon_{i,j,t} \qquad \text{(A1-2)}$$

$x_{i,j,t}$	contribution of team *i* in group *j* in period *t*
$Team_i$	team dummy (0 for IND, 1 for MAJ and UNA)
a_j	group dummies
τ_t	period dummies
u_i	dummy for decision unit
$\varepsilon_{i,j,t}$	error term
t	periods 1 to 10
j	groups (10 for each treatment)
i	decision units (4 per group, either team or individual)

Appendix B

We set up a simple model of public good provision where individuals draw utility from monetary payoffs and own contributions (as a proxy for warm-glow). For simplicity, we restrict our model to two players. The Cobb-Douglas utility of player 1 amounts to

$$U_1 = [y - (1-g) \cdot x_1 + g \cdot x_2]^{1-b_1} \cdot [x_1]^{b_1},$$

where x_i denotes player i's contribution to the public good, y is the player's budget and g is the private return from investing in the public good. For a collectively profitable investment in the public good, we need > 0.5 . Finally, b_i reflects player i's willingness-to-contribute.

From the first-order condition, we obtain the individually optimal contribution for player 1 as

$$x_1 = b_1 \cdot \frac{y + g \cdot x_2}{1-g}.$$

The individual contribution increases in the return to the public good (g), the willingness-to-contribute (b_1) and the contribution of the other player (x_2). The same reaction curve can be obtained for player 2. Solving for the equilibrium contributions of the two players yields

$$x_1^* = b_1 \cdot y \cdot \frac{1-g \cdot (1-b_2)}{(1-g)^2 - b_1 \cdot b_2 \cdot g^2} \quad \text{and} \quad x_2^* = b_2 \cdot y \cdot \frac{1-g \cdot (1-b_1)}{(1-g)^2 - b_1 \cdot b_2 \cdot g^2}.$$

For simplicity, we discuss interior solutions only ($0 < x_1^*, x_2^* < y$), which always prevails for sufficiently small parameters b_i. In equilibrium, the total contribution of both players amounts to

$$x_{total}^* = y \cdot \frac{(b_1 + b_2) \cdot (1-g) + 2 \cdot b_1 \cdot b_2 \cdot g}{(1-g)^2 - b_1 \cdot b_2 \cdot g^2}.$$

Now, we change the heterogeneity of the players in terms of their willingness-to-contribute. Starting from identical players ($b = b_1 = b_2$), we increase one player's willingeness-to-contribute and reduce the other player's preference for contributions. Let $b_1 = b + d$ and $b_2 = b - d$. Then, we can write x_{total}^* as

$$x_{total}^* = 2 \cdot y \cdot \frac{b \cdot (1-g) + (b^2 - d^2) \cdot g}{(1-g)^2 - (b^2 - d^2) \cdot g^2}.$$

An increase in d implies an increase in players' heterogeneity in terms of willingness-to-contribute. Taking the derivative with respect to d yields

$$\frac{\partial x_{total}^*}{\partial d} < 0.$$

The total contributions to the public good decrease when society becomes more heterogeneous. Put differently, as the median voter prevails in majority decisions, one can expect that extreme preferences are more likely to be eliminated compared to individual decisions. Hence, when moving from individual to team decisions, preferences should become more homogeneous and total contributions to the public good should increase. We do not claim that this effect is always at work as it clearly depends on the precise type of preferences: For instance, with quasi-linear preferences, the individually optimal contribution is independent of the other player's contribution and the interaction effect vanishes.

Appendix C

Table A.1 compares standard deviations of contributions in the individual treatment and in the majority treatment without punishment for each period. Standard deviations in the majority treatment (column 5) are not smaller than in the individual treatment (column 3).

Table C.1: Average Contribution and Standard Deviation of Contributions in Part II

	Individual Treatment		Majority Team Treatment	
Period	Average	Standard Deviation	Average	Standard Deviation
(1)	(2)	(3)	(4)	(5)
1	11.10	7.09	12.65	6.87
2	10.80	7.54	13.43	7.76
3	9.43	7.49	12.35	8.15
4	8.55	7.77	11.08	8.40
5	9.90	7.58	11.23	8.36
6	7.75	7.44	7.95	8.52
7	5.53	6.71	7.40	8.26
8	4.53	6.46	5.40	8.04
9	4.25	6.33	4.88	7.50
10	3.80	6.85	3.43	6.22

Note: includes only data from Part II (no punishment).

References

Ambrus, Attila and Ben Greiner (2012): "Imperfect Public Monitoring with Costly Punishment – An Experimental Study", *American Economic Review* 102(7), 3317- 32.

Andreoni, James (1995): "Cooperation in Public-Goods Experiments: Kindness or Confusion?", *American Economic Review* 85(4), 891-904.

Balliet, Daniel (2010): "Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review", Journal of Conflict Resolution 54(1), 39-57.Baron, Robert S. and Norbert L. Kerr (2003): *Group Process, Group Decision, Group Action*, 2nd edition, Open University Press, Buckingham/Philadelphia, 2003.

Barrett, Scott (2003): *Environment and Statecraft: The Strategy of Environmental Treaty-Making*, Oxford University Press, New York, 2003.

Berlemann, Michael, Marcus Dittrich and Gunther Markwardt (2009): „The value of non-binding announcements in public good experiments: Some theory and experimental evidence", *The Journal of Socio-Economics* 38(2009), 421-428.

Blinder, Alan S. and John Morgan (2005): "Are Two Heads Better than One? Monetary Policy by Committee", *Journal of Money, Credit and Banking* 37(5), 789-811.

Bond, Philip and Hülya Eraslan (2010): "Strategic Voting over Strategic Proposals", Review of Economic Studies 77(2), 459-490.

Bornstein, Gary, Tamar Kugler and Anthony Ziegelmeyer (2004): "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?", *Journal of Experimental Social Psychology* 40(5), 599-605.

Bornstein, Gary and Ilan Yaniv (1998): "Individual and Group Behavior in the Ultimatum Game: Are Groups More 'Rational' Players?", *Experimental Economics* 1(1), 101-108.

Cardella, Eric and Ray Chiu (2012): "Stackelberg in the Lab: The Effect of Group Decision Making and 'Cooling-off' Periods", *Journal of Economic Psychology* 33(6), 1070-1083.

Cason, Timothy N. and Vai-Lam Mui (1997): "A Laboratory Study in Group Polarisation in the Team Dictator Game", *The Economic Journal* 107(444), 1465-1483.

Charness, Gary and Matthias Sutter (2012): "Groups Make Better Self-Interested Decisions", *Journal of Economic Perspectives* 26(3), 157–176.

Cooper, David J. and John H. Kagel (2005): "Are Two Heads Better than One? Team versus Individual Play in Signaling Games", *American Economic Review* 95(3), 477-509.

Cox, James C. and Stephen C. Hayne (2006): "Barking Up the Right Tree: Are Small Groups Rational Agents?", *Experimental Economics* 2006(9), 209-222.

Denant-Boemont, Laurent, David Masclet and Charles Noussair (2011): "Announcement, Observation and Honesty in the Voluntary Contributions Game", *Pacific Economic Review* 16(2), 207-229.

Federsen, Timothy and Wolfgang Pesendorfer (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting", *American Political Science Review* 92(1), 22-35.

Federsen, Timothy and Wolfgang Pesendorfer (1999): "Elections, Information Aggregation, and Strategic Voting", *Proceedings of the National Academy of Science* 96 (19), 10572-10574.

Fehr, Ernst and Simon Gächter (2000): "Cooperation and Punishment in Public Goods Experiments", *American Economic Review* 90(4), 980-994.

Fehr, Ernst and Simon Gächter (2002): "Altruistic Punishment in Humans", *Nature* 415 (10 January 2002), 137-140.

Fischbacher, Urs (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics* 10(2), 171-178.

Fischbacher, Urs and Simon Gächter (2010): "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments", *American Economic Review* 100(1), 541–556.

Fischbacher, Urs, Simon Gächter and Ernst Fehr (2001): "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters* 71, 397-404.

Gächter, Simon, Elke Renner and Martin Sefton (2008): "The Long-Run Benefits of Punishment", *Science* 322, 1510.

Gillet, Joris, Arthur Schram and Joep Sonnemans (2009): "The Tragedy of the Commons Revisited: The Importance of Group Decision-making", *Journal of Public Economics* 93, 785–797.

Goeree, Jacob K., Charles A. Holt and Susan K. Laury (2002): “Private Costs and Public Benefits: Unraveling the Effects of Altruism and Noisy Behavior”, *Journal of Public Economics* 83, 255–276.

Goeree, Jacob K. and Leeat Yariv (2011): “An Experimental Study of Collective Deliberation”, *Econometrica* 79(3), 893-921.

Goodin, Robert E. and Kevin W. S. Roberts (1975): “The Ethical Voter”, *American Political Science Review* 69(3), 926-928.

Greiner, Ben (2003): “An Online Recruitment System for Economic Experiments”. In: Kurt Kremer, Volker Macho (Eds.): *Forschung und wissenschaftliches Rechnen.* GWDG Bericht 63, Göttingen : Ges. für Wiss. Datenverarbeitung, 79-93.

Greiner, Ben (2015): “Subject Pool Recruitment Procedures: Organizing Experiments with ORSE”, *Journal of the Economic Science Association* 1(1), 114-125.

Guarnaschelli, Serena, Richard D. McKelvey and Thomas R. Palfrey (2000): “An Experimental Study of Jury Decision Rules”, *American Political Science Review* 94(2), 407-423.

Herrmann, Benedikt, Christian Thöni and Simon Gächter (2008): “Antisocial Punishment Across Societies”, *Science* 319, 1362-1367.

Houser, Daniel and Robert Kurzban (2002): ”Revisiting Kindness and Confusion in Public Goods Experiments*”, American Economic Review* 92, 1062-1069.

Kagel, John H. and Peter McGee (2016): “Team versus Individual Play in Finitely Repeated Prisoner Dilemma Games”, *Amercian Economic Journal: Microeconomics* 8(2), 253-276.

Kerr, Norbert L. and R. Scott Tindale (2004): “Group Performance and Decision Making”, *Annual Review of Psychology* 55, 623-655.

Kocher, Martin G. and Matthias Sutter (2005): “The Decision Maker Matters: Individual versus Group Behaviour in Experimental Beauty-contest Games”, *The Economic Journal* 115, 200-223.

Kocher, Martin G. and Matthias Sutter (2007): “Individual versus Group Behavior and the Role of the Decision Making Procedure in Gift-Exchange Experiments”, *Empirica* 34, 63-88.

Kugler, Tamar, Gary Bornstein, Martin G. Kocher and Matthias Sutter (2007): “Trust Between Individuals and Groups: Groups are less Trusting than Individuals but just as Trustworthy”, *Journal of Economic Psychology* 28(6), 646-657.

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* 3(4), 471-482.

Ledyard, John O. (1997): “Public Goods: A Survey of Experimental Research”, In: Kagel, John H. and Alvin E. Roth (Eds): *The Handbook of Experimental Research*, Princeton University Press, 1997.

Luhan, Wolfgang J., Martin G. Kocher and Matthias Sutter (2009): “Group Polarization in the Team Dictator Game Reconsidered”, *Experimental Economics* 12(1), 26-41.

Müller, Wieland and Fangfang Tan (2013): “Who Acts More Like a Game Theorist? Group and Individual Play in a Sequential Market Game and the Effect of the Time Horizon”, *Games and Economic Behavior* 82, 658-674.

Palfrey, Thomas R. and Jeffrey E. Prisbrey (1997): “Anomalous Behavior in Public Good Experiments: How Much and Why?” *American Economic Review* 87, 829-846.

Rege, Mari and Kjetil Telle (2004): “The Impact of Social Approval and Framing on Cooperation in Public Good Situations”, *Journal of Public Economics* 88, 1625-1644.

Robert, Christopher and Peter J. Carnevale (1997): “Group Choice in Ultimatum Bargaining”, *Organizational Behavior and Human Decision Processes* 72(2), 256-279.

Sally, David (1995): “Conversation and Cooperation in Social Dilemmas”, *Rationality and Society* 7(1), 58-92.

Sutter, Matthias (2009): “Individual Behavior and Group Membership: Comment”, *American Economic Review* 99(5), 2247-2257.

Sutter, Matthias, Martin G. Kocher, Sabine Strauss (2009): “Individuals and Teams in Auctions”, *Oxford Economic Papers* 61(2), 380-394.

Wildschut, Tim, Brad Pinter, Jack L. Vevea, Chester A. Insko and John Schopler (2003). “Beyond the Group Mind: A Quantitative Review of the Interindividual-Intergroup Discontinuity Effect”, *Psychological Bulletin* 129(5), 698-722.