Potential Pareto Public Goods*

Sagi Dekel Sven Fischer Ro'i Zultan†

Abstract

Potential Pareto Public Goods create an aggregate benefit to society while harming some members of the community. As the overall benefit outweighs the harm incurred, provision may lead to Pareto improvement if the gains from cooperation are used to compensate the harmed parties. Such situations are ubiquitous, e.g., in not-in-my-backyard (NIMBY) problems. We study experimentally voluntary contributions to Potential Pareto Public Goods, in which provision is efficient but harms a minority in the group. We test the effects of punishment and reward institutions, with and without communication. We find that contributions to Potential Pareto Public Goods are not viewed as unequivocally socially desirable and do not increase with communication or punishment. With the reward institution, communication facilitates compensation, undoing the harm imposed on the minority player by majority contributions. Consequently, contributions are no longer viewed as socially undesirable, and majority contributions increase. Taken together, our results establish that perceptions and behavior in voluntary contributions to Potential Pareto Public Goods are dramatically different than with universal public goods that benefit all members of the community. We suggest that the underlying mechanism is team reasoning: individuals consider what is good for the group, and play their part in achieving that goal.

Keywords: public goods, public bads, punishment, reward, externalities, team reasoning.

JEL classification: C72, C92, H41

---

*We thank Tomer Blumkin, Aniol Llorente-Saguer, Bradley Ruffle, and participants at the IMEBESS meeting in Toulouse, the ESA meetings in Heidelberg and Dallas, and CREED seminar for comments and discussion. Feedback from the editor and two anonymous referees has greatly improved the paper. Financial support by the Max Planck Society is gratefully acknowledged. DOI: 10.1016/j.jpubeco.2016.12.007.

†Dekel: Department of Economics, Ben-Gurion Univesity of the Negev; Fischer: Newcastle University Business School; Zultan: Corresponding author, Department of Economics, Ben-Gurion Univesity of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel. E-mail: zultan@bgu.ac.il. Tel: +972(0)86472306.

1 Introduction

Some public goods are *universal*, in the sense that they benefit all members of the community. Still, many others are only *Potential Pareto* public goods:[1] they create an aggregate benefit to society but harm some members of the community. Nonetheless, the overall benefit outweighs the harm incurred, so that provision may lead to Pareto improvement if the gains from cooperation are used to compensate the harmed parties.

Potential Pareto Public Goods are ubiquitous. One widespread category is Not-in-my-back-yard (NIMBY) situations (Schively, 2007). For example, people who live in proximity to highways, trains, and airports—a classic public good—incur negative externalities despite the overall social-welfare enhancing effect of transportation infrastructure. Theoretically, such cases can be resolved using compensation (Kunreuther et al., 1987; O'Hare, 1977). However, the conditions under which compensation schemes are effective in overcoming the NIMBY problem are not yet fully understood (Frey and Oberholzer-Gee, 1996; Frey et al., 1996). More generally, any land development and (re)zoning requires communities to cooperate. For example, in petitioning zoning committees to advance a socially beneficial development (Babcock and Siemon, 1985; Fischel, 2015, 1985; Nelson, 1977). Such cooperation is hindered by potential harms imposed on some members of the community in two ways. First, by making members of the community reluctant to advance development and rezoning that harms their peers. Second, when the harmed parties directly use existing institutions to obstruct development. In the case of the New Hampshire pulp mill studied by Fischel (1979), for example, stringent opponents of the mill forced open meetings with the mill representative, thereby effectively precluding compromises that may be beneficial for the majority of the community (Fischel, 1985).

Other examples of Potential Pareto Public Goods include large scale irrigation systems. Duflo and Pande (2007) have shown that construction of large irrigation dams in India benefits those districts that are located downstream from the dam, but harm the district where the dam is located. Duflo and Pande (2007) interpret this result as evidence for a failure to generate Pareto improvement by sharing the benefits resulting from the dam construction. Similar situations arise with smaller irrigation systems that are created and maintained by voluntary communal effort. (e.g., Balasubramanian and Selvaraj, 2003; Lam, 1998). Irrigation systems designed for a specific use may interfere with water consumption for other uses. A system designed for agricultural use reduces the availability and water quality for

[1] We thank an anonymous referee for suggesting this formulation.

household use, including cooking, cleaning and horticulture.[2] Fujiie et al. (2005) indeed found that a high share of nonfarm households in the community impedes collective action by Philippines farmers.

The spatial structure of farming communities also gives rise to conflicting interests. Headenders—those whose lands are located near the source of the water system—have more access to water, and are likely to resist cooperating with tailenders in irrigation systems that lead to more efficient and egalitarian water allocation (Tang, 1992). Successful cooperation in constructing and maintaining such systems, therefore, depends on finding "ways in which headenders can be better off and see themselves as better off with less water." (Chambers, 1988).

At the macroeconomic level, Potential Pareto Public Goods are closely related to issues of societal fractionalization and and its effect on public goods provision. Alesina and Ferrara (2005) show in a simple model that higher diversity in preferences for different public goods lowers the level of public goods provision chosen by a central social planner (see also Alesina and Spolaore, 1997). Indeed, fractionalization is negatively correlated with centralized public goods provision (Alesina et al., 1999, 2003), collective action (Alesina and La Ferrara, 2000; Vigdor, 2004) and growth (Easterly and Levine, 1997). This problem is exacerbated if public goods preferred by a large part of society are actually harmful to a minority.

While the role of inequality in economic benefits in public goods provision have been thoroughly studied (Baland and Platteau, 1997; Bardhan and Dayton-Johnson, 2002; Olson, 1965; Poteete and Ostrom, 2004; Varughese and Ostrom, 2001), the issue of harmed minorities remains understudied. How do communities perceive public goods that are not universal? How do the harmed minorities interact with the broad community? Can communication and compensation facilitate Pareto-improving public goods provision?

In this paper, we experimentally study voluntary public goods provision in Potential Pareto Public Goods. While Güth et al. (2011, 2014) found that a centralized revelation mechanism can lead to efficient outcomes in a similar situation, it is not at all clear that the same applies for a decentralized voluntary provision mechanism. Similarly, Engel and Rockenbach (2011, 2014) found that contributions did not decrease when *non-active* players are harmed by the public good (compared to other situations involving non-active players).[3] Delaney and Jacobson (2014) also found that the addition of a passive harmed minority did not affect contributions, this time when the net social benefit from contributions is negative.

[2]Such disparate preferences often exist within the same household, where the the man is responsible for agricultural activities and the woman for household uses (Meinzen-Dick and Bakker, 1999; Meinzen-Dick and Jackson, 1997).

[3]Cooperation in a standard Prisoner's Dilemma game, in contrast, does decrease when cooperation by either player harms a passive minority (Engel and Zhurakhovska, 2014).

Contributions did decrease, however, when the negative externalities were levied on active members of a separate group—such that each player is aware of potentially being in the role of the harmed outsiders for another group. In contrast to the existing studies, we argue that new considerations arise when there is a harmed minority that is an integral part of the active group, for several reasons that we discuss below.

A harmed minority pays a double cost when contributing. The direct cost of investment in the public good is augmented by the indirect harm imposed by the increased provision of the public good. Accordingly, the minority player has a justification for not contributing. Unlike the harmed third parties in Engel and Rockenbach (2011, 2014), who are not asked to make a contribution decision, the existence of an integral player who refrains from contributing may erode contribution norms and lead the majority players to reduce their contributions (de Oliveira et al., 2015). Whether majority players realize the predicament of the harmed minority and ignore the lack of contributions is an empirical question. This is tested in our first conjecture.

Conjecture 1. *Majority contributions respond positively to higher contributions by the minority player.*

We study the efficacy of punishment and reward institutions in Potential Pareto public goods. Such institutions have garnered substantial attention as ways to overcome the problem of collective action—as free riders are punished and contributors are rewarded (Fehr and Gächter, 2000, 2002; Fehr et al., 2010; Ostrom et al., 1992; Rand et al., 2009; Sefton et al., 2007; Sutter et al., 2010; Walker and Halloran, 2004). Punishment institutions are not effective, however, if contributors are punished as much as free riders (Herrmann et al., 2008). Such *anti-social punishment* was shown to exist as part of cultural norms (Gächter et al., 2008), or to emerge due to environmental uncertainty (Ambrus and Greiner, 2012; Bornstein and Weisel, 2010; Grechenig et al., 2010) or counter-punishment (Nikiforakis, 2008). In the context of Potential Pareto Public Goods, anti-social punishment can result if the punishment institution is (mis)used by the harmed minority member to deter others from contributing. Furthermore, pro-social punishment—that is, punishment of low contributors—may be hindered by the ambiguous contribution norms with respect to the harmed minority. As a minority player should be expected not to contribute, a majority player can free ride without being the lowest contributor. Typically, lower contributions are met with higher punishment. Thus, high contributors may be reluctant to punish a majority peer more than a minority player who contributed less. Previous studies have shown that heterogeneity may be enough to obscure whether and how contribution norms violations should be punished (Noussair and Tan, 2011; Tan, 2008). This problem may be exacerbated with harmed minorities. These considerations are reflected in our next conjecture.

Conjecture 2. *An environment with a harmed minority fosters anti-social punishment by the minority player and is unfavorable to pro-social punishment by the majority players. As a result, punishment institutions do not facilitate public goods provision.*

Finally, we turn our attention to reward institutions. Whereas punishment institutions are subject to misuse and mis-targeted punishment, reward and redistribution institutions can be used to compensate the harmed minorities as is often done centrally in cases of LULU's—Locally Unwanted Land Uses (Armour, 1991)—and NIMBY's. Redistribution effectively reinstates the possibility for mutual gain from cooperation, leading to Pareto improvement even in the presence of a harmed minority. This, however, requires a non-trivial agreement by which the minority contributes to generate surplus, part of which is transferred back by the majority players. Accordingly, we test whether communication between the players can foster such an agreement to boost contributions by the minority player.

Conjecture 3. *When rewards are feasible, majority players compensate the harmed minority. Communication facilitates compensation, and enables Pareto improvement through minority contributions.*

Our results show that, although majority players are much more responsive to their majority's peer contribution, they are nonetheless somewhat sensitive to the contributions made by the minority player, providing evidence that harmed minority undermine contribution norms. Furthermore, many groups explicitly argue for refraining from contributions in the communication phase. Under the punishment institutions, punishment levels are uncharacteristically low, with minority players using punishment to deter contributions, and majority players administering very low punishment even to their free riding peers, except for the few cases in which the difference between the majority contributions is very extreme. Accordingly, contribution levels are not higher, and efficiency is lower, compared to the baseline without punishment.

Rewards are not used by the majority players for compensating the minority player when communication is not available. Rather, majority players reward each other without any efficiency gains, presumably as a trust-building mechanism. Communication fosters compensation, with majority players redistributing significant amounts to the harmed minority. However, it is the majority players, rather than the harmed minority, who increase their contributions with communication-mediated compensation. Compensation mostly eliminates discussions of refraining from contributing.

Overall, our results establish that the public good is no longer perceived as unequivocally desirable if it harms a minority of the relevant group. This perception undermines contributions as well as the efficacy of communication and sanctioning mechanisms in boosting contributions. The combination of communication and compensation, however, reinstates the public good as socially desirable. These observations are consistent with the notion of *team reasoning*, by which each member of the group does their part to promote the group's objective. The existence of a harmed minority undermine both the cohesiveness of the group—and consequently the chances of the individuals perceiving themselves as a team—and the existence of a joint group objective.

Our conclusions bear implications for real-life situations such as NIMBY's and LULU's, and any situation in which a communal effort may impose negative externalities on minorities, such as land development and construction and maintenance of irrigation systems. In such situations, attention should be given to open communication and compensation directed at potentially harmed minorities.

2 Experimental design and procedure

The experimental design was a 3×2 between-subjects design, varying the institution—standard voluntary contribution mechanism (VCM), punishment, or reward—and communication. The stage game was the standard linear public goods game. Participants interacted in groups of three. At the beginning of each round, each player decided how much out of an endowment of 20 tokens to contribute to a group account. The payoff at this stage is given by

$$\pi_i^1 = 20 - c_i + m_i \sum_{j \in N} c_j,$$

where c_i is player i's contribution to the group account, N denotes the group of three players, and m_i is player i's marginal return from the public good, which in the experiment can be negative. In the punishment and reward treatments, this stage was followed by a second stage, in which each player received additional eight tokens, which she could use for punishment or reward. In the punishment treatment, each punishment token reduced the payoff of the punished member by three tokens. The final payoff was thus

$$\pi_i^{pun} = \pi_i^1 + 8 - p_{ij} - 3 \sum_{j \neq i} p_{ji},$$

Table 1: Experimental design and number of independent observations.

	Institution		
Communication	Standard VCM	Punishment	Reward
No communication	$N = 11$	$N = 9$	$N = 9$
Communication	$N = 11$	$N = 9$	$N = 9$

where p_{ij} is the number of punishment points allocated by i to j. In the reward treatment, each player similarly received an additional 8 tokens in the second stage, which she could transfer to the other players such that the final payoff was

$$\pi_i^{red} = \pi_i^1 + 8 - r_{ij} + \sum_{j \neq i} r_{ji},$$

where r_{ij} is the number of reward points allocated by i to j.

Each group of three players consisted of two majority players with $m_1 = m_2 = 0.8$ and one minority player with $m_3 = -0.4$. These parameters were set with the following properties in mind. First, they imply an efficiency gain of 1.2 and an average marginal per capita return of 0.4, similar to many previous public goods experiments. In the punishment treatments, the minority player has enough leverage to destroy all gains from cooperation obtainable by the majority players.[4] In the reward stage, the majority players can compensate the minority player for contributing, hence minority contributions can lead to Pareto improvement.[5] We opted for a high punishment and a low reward leverage instead of equalizing the leverage in the two institutions.[6] This design has two advantages. First, while destroying resources is feasible, it is less plausible that agents can generate surplus when rewarding—presumably if such welfare improvement is possible it would take place as part of trade regardless of public goods provision. Monetary transfer, in comparison, are more plausible and reflect actual mechanisms used to solve, e.g., NIMBY and LULU problems. Second, our conjectures state that reward is more successful than punishment in increasing contributions and efficiency in the presence of a harmed minority. Accordingly, we design a situation in which

[4]By contributing their full endowment, the majority players can increase their payoffs from 20 to 32 each (not including the eight punishment tokens). The minority player can reduce their payoff back to 20 by allocating 4 punishment points to each of the other players.

[5]Full contributions cannot be Pareto improving, even if the majority players allocate all of their reward points to the minority player. Nonetheless, the parameters allow for the minority player to be fully compensated for the loss incurred by contributing half of her endowment.

[6]Walker and Halloran (2004) and Sefton et al. (2007) compared punishment and reward with a leverage of 1. Sutter et al. (2010) studied all four combinations of punishment/reward and low/high leverage.

previous experiments have established that punishment is more effective than reward.[7]

Participants played the stage game for 30 rounds in the same groups.[8] The partner design creates an environment conducive for developing nuanced cooperation dynamics such as the ones we want to test. To provide participants with multiple opportunities to (re-)initiate complex cooperation, we introduced a (known in advance) restart after each of the first two blocks of ten rounds (Cookson, 2000). Since the session consists of 30 rounds, the game is over after the third block of ten rounds. To strengthen the restart effect, one block was randomly selected to determine the final payoff in the experiment.

The communication was implemented by ninety seconds of electronic chat at the beginning of every 5-round sub-block.[9] All three members of the group could send free-form messages to the group. Participants were not allowed to reveal any identifying information, but were otherwise not restricted in the content of the messages. In the no-communication treatments, there was a brief waiting phase after each 5 rounds instead of the communication phase. Table 1 summarizes the experimental design and the number of independent observations in each treatment.

The experiment was programmed in z-Tree (Fischbacher, 2007) and conducted in the Negev Experimental Economics Laboratory at Ben-Gurion University of the Negev. To avoid negative payoffs, participants performed an unrelated task at the end of the experiment, for which they received an additional amount. The payoff rule for the additional task was explained only after the main part of the experiment was finished. The average total payoff was 54.70 NIS (approximately 15 USD).

2.1 Content analysis

Three independent raters rated the contents of the discussions in the communication treatment on several dimensions. First, the raters counted the number of

[7]In the exogenous treatments of Sutter et al. (2010), for example, rewards have no significant effect on contributions whereas punishment has a strong positive effect. The increase in contributions overcomes the social costs of punishment if the leverage is high and the time horizon is long enough (Gächter et al., 2008; Nikiforakis and Normann, 2008). The difference in leverage should thus outweigh the effect of having identified players across rounds, an environment in which reward is generally more effective (Rand et al., 2009).

[8]As in the standard game with all-positive gains from cooperation, the stage game has a unique equilibrium in dominant strategies in which no player contributes and no player punishes or rewards. By backward induction this is the only subgame perfect equilibrium of the repeated game.

[9]Allowing for six chat periods allowed participants ample opportunities to communicate without overly lengthening the session duration.

Table 2: Contributions.

	Standard VCM		Punishment		Reward	
	Majority	Minority	Majority	Minority	Majority	Minority
No communication						
Mean	10.97	1.02	9.53	2.11	8.29	2.18
Std. Dev.	7.65	3.55	6.32	5.25	6.25	3.42
N	660	330	540	270	540	270
Communication						
Mean	12.17	3.62	11.50	2.53	13.36	2.26
Std. Dev.	7.74	6.78	7.55	4.95	6.40	3.28
N	660	330	540	270	540	270

total, purely social (e.g., 'Hello!'), and minority messages in each communication period. Next, the raters assessed whether the communication included any messages recommending other players to contribute, not contribute, not contribute if you're the minority player, avoid punishment, punish contributing players, punish non-contributing players, reward contributing players, reward non-contributing players, reward the minority player, reward the minority player only if she contributes, and messages indicating satisfaction with past behavior. Three more ratings estimated the subjects' understanding of the rules as reflected in the messages, however inter-rater reliability for those items was low.

The interclass correlation for the first items (number of total, social, and minority messages) was high, 0.996, 0.937, and 0.879, respectively ($p < 0.001$ for all three). Fleiss' κ for the binary indicators was fairly high,[10] between 0.498 and 0.860 ($p < 0.001$ for all) with the exception of the indicator for messages calling to punish or reward those who don't contribute, which all three raters agreed never occurred. We aggregated the three ratings by averaging (and rounding) the cardinal variables and taking the majority in cases of non-unanimity for the binary variables.

3 Results

Figure 1 shows contributions by majority and minority players over time, separated by communication treatment and institution. Table 2 presents summary

[10]Fleiss' κ (Fleiss, 1971) is a measure of agreement on categorical ratings between multiple raters, calculated as the degree of agreement beyond that expected by chance.

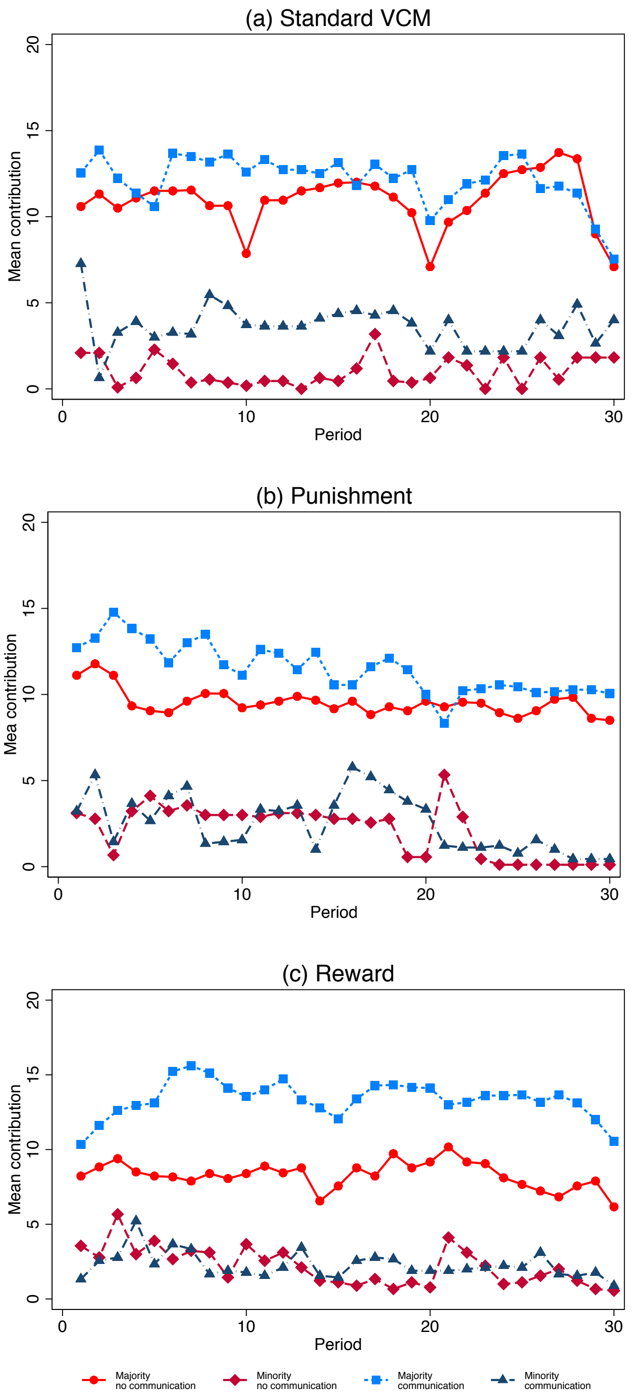

Figure 1: Contributions by treatment, role, and period.

statistics on contributions by treatments. Table 4 presents the results of OLS regressions for contributions, efficiency, and punishment/reward behavior on treatments and types.[11] Unless otherwise stated, all marginal effects and p-values reported in the following text are based on the regressions reported in Table 4. Finally, Figure 2 summarizes the punishment and rewards allocated based on the difference in contributions between the allocator and the recipient. In the following, we discuss the results separately for the different institutions.

3.1 *Standard VCM*

The most prominent result, as could be expected, is that majority players contribute much more than minority players (an estimated margin of 9.95 and 8.55 tokens in the No communication and Communication treatments, respectively, $p < 0.001$ for both treatments; Wilcoxon rank-sum $z = 2.871, p < 0.005$ and $z = 3.223, p < 0.005$). The main question is, however, whether and how the minority players affect majority contributions. The *bad apple* conjecture predicts that majority players will decrease their contributions after observing low contributions by minority players—despite the minority's justification for withholding contributions.

Table 3 reports the marginal effect of the (other) majority and minority players' contributions in the previous period on current period contributions of majority and minority players. We find that majority players' contributions significantly increase with the contributions of the other majority player in the preceding period. The OLS regression reported in Column (1) of the table reveals a small yet significant effect of past minority contributions on majority contributions in the No communication treatment, and a positive but non-significant effect in the Communication treatment. We expect the minority player's contributions to be independent of the majority players' behavior, and therefore represent exogenous variation. Indeed, Columns (3) and (4) show that the effect of past majority contributions on the minority contributions is close to zero and not significant. Column (2) adds individual fixed effects to the regression on majority contributions. The fixed effects control for unobserved heterogeneity which is correlated with the lagged regressors, including overall, rather than the dynamic, effect of peers' contributions. The coefficients of the other majority player's past contribution are somewhat smaller compared to the OLS analysis, but still highly significant. The effect of the minority player's past contribution, however, becomes negative and non-significant in both communication treatments. In the appendix we report the

[11] The regressions include the period number across blocks. Including the block number and period within block has virtually no effect on the other coefficients.

Table 3: Marginal effects of others' contributions in $t-1$, Standard VCM.

	(1) Majority	(2) Majority	(3) Minority	(4) Minority
Majority in $t-1$				
No communication	0.715***	0.356***	0.092	-0.022
	(0.067)	(0.075)	(0.081)	(0.024)
Communication	0.645***	0.434***	0.231	-0.014
	(0.096)	(0.085)	(0.210)	(0.055)
Minority in $t-1$				
No communication	0.104**	-0.025		
	(0.046)	(0.053)		
Communication	0.063	-0.107		
	(0.091)	(0.107)		
Individual fixed effects	No	Yes	No	Yes

Notes: Marginal effects based on OLS and fixed effects regressions on contributions in period t by contribution of others in $t-1$, interacted with communication treatment, and period. Robust standard errors clustered on groups. *,** ,*** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

full regressions, as well as appropriate Arellano-Bond (Arellano and Bond, 1991) estimations of dynamic specifications of the model controlling for path dependence, which corroborate the results from the fixed effects models.

Thus, majority players are sensitive to the dynamic contributions of their majority peer, but not to those of the minority player. Nonetheless, the OLS result suggests that, at least in the No communication treatment, majority players do somewhat respond to the minority player's behavior (but not vice versa). We test this conclusion further by using the group level data, taking the minority player's first period contribution as an exogenous factor. The mean contribution by majority players over all periods (excluding the first period, which cannot be influenced by the contemptuous minority contribution) is positively correlated with the minority player's contribution in the first period (Spearman's $\rho = 0.525, p = 0.097, N = 11$ and $\rho = 0.586, p < 0.06, N = 11$ in No communication and Communication, respectively; $\rho = 0.502, p < 0.05, N = 22$ when pooling across communication treatments).[12] Taken together, the evidence lends partial support for Conjecture 1.

Result 1. *We find some evidence that majority players are sensitive to the minority player's overall contribution level in the standard VCM. Although majority players are sensitive to their majority peer's round-to-round changes in contributions, they appear to ignore the minority player's contribution dynamics.*

[12]The difference between the correlations in the two treatments is not significant.

In contrast to the robust finding in standard public goods experiments, we find that, with harmed minorities, communication does not facilitate majority contributions and efficiency.[13] The content analysis complements this result, as we find that 37.9% (25/66) of chats included messages calling for withholding contributions. In fact, all but one of the 11 groups in the Standard VCM with Chat treatment discussed a strategy of not contributing at least once. Such discussions are associated with a reduction of 5.2 points in contributions averaged over the next block of 5 periods (see Table A2 in the appendix for details). Note that this relationship may not be causal. Nonetheless, it indicates a real tendency to refrain from contributions that is reflected in the communication, and counters the robust positive effect of communication on contributions documented in the literature.

Thus, the evidence from the Standard VCM treatment reveals that Potential Pareto Public Goods are indeed perceived differently from universal public goods. Group members that benefit from the public good respond to the effect on the minority player. Without the possibility of compensation and Pareto improvement, group members do not perceive contributions as promoting a team objective. Communication, which has the potential to promote group identity and team reasoning, is ineffective in raising contribution levels.

3.2 Punishment

As can be clearly seen in Figure 1 and Tables 2 and 4, contributions are not significantly higher in the punishment treatment than in the standard VCM. Because punishment is costly, efficiency in the punishment treatments, both with and without communication, is lower than in the other treatments.

Punishment can only be effective if it is used against low contributors (labeled *pro-social punishment* by Grechenig et al., 2010). We conjectured that the existence of a harmed minority undermines the efficacy of the punishment institutions through antisocial punishment used by the minority player to subdue contributions, and by generating ambiguity as to the proper target. Indeed, the lowest (unique) contribution is made by the minority player in 93% (69%) of all cases.

Figure 2 lends support to both conjectures. The bars in the figure show the average allocated amount of punishment points (top panels) or reward (bottom panels). The bars are conditioned on the difference in contributions between the recipient of the punishment or reward and the one who distributes the punishment or reward. The numbers on top of each bar represent frequencies. First, the top right panel depicts a clear increase in punishment by minority players as a

[13]See non-significant coefficient on COM in Table 4.

Table 4: OLS regressions results.

	(1) Contributions	(2) Efficiency	(3) Received	(4) Distributed
COM	2.597	0.998	-0.007	-0.204
	(1.827)	(1.113)	(0.201)	(0.224)
PUN	1.083	-5.025**		
	(1.443)	(1.503)		
REW	1.157	-0.840	1.378**	0.926*
	(0.924)	(0.914)	(0.426)	(0.397)
COM × PUN	-2.171	1.103		
	(2.395)	(2.484)		
COM × REW	-2.519	1.047	4.500**	1.296
	(2.055)	(1.306)	(1.357)	(0.801)
Majority	9.947***		0.072	-0.272
	(1.610)		(0.168)	(0.227)
COM × Majority	-1.402		-0.143	0.152
	(2.335)		(0.199)	(0.232)
PUN × Majority	-2.521			
	(2.596)			
REW × Majority	-3.836		0.576	1.254*
	(1.973)		(0.506)	(0.495)
COM × PUN × Majority	2.940			
	(3.658)			
COM × REW × Majority	6.396*		-3.846*	0.959
	(2.960)		(1.517)	(0.720)
Period	-0.051	0.062	-0.039***	-0.039***
	(0.027)	(0.031)	(0.009)	(0.009)
Constant	1.819*	63.631***	0.953***	1.183***
	(0.699)	(0.888)	(0.198)	(0.248)
Observations	5,220	1,740	3,240	3,240
R-squared	0.320	0.139	0.396	0.429

Notes: OLS regressions. Robust standard errors clustered on groups in parentheses. *,** ,*** indicate significance at the 0.05, 0.01, and 0.001 levels, respectively.

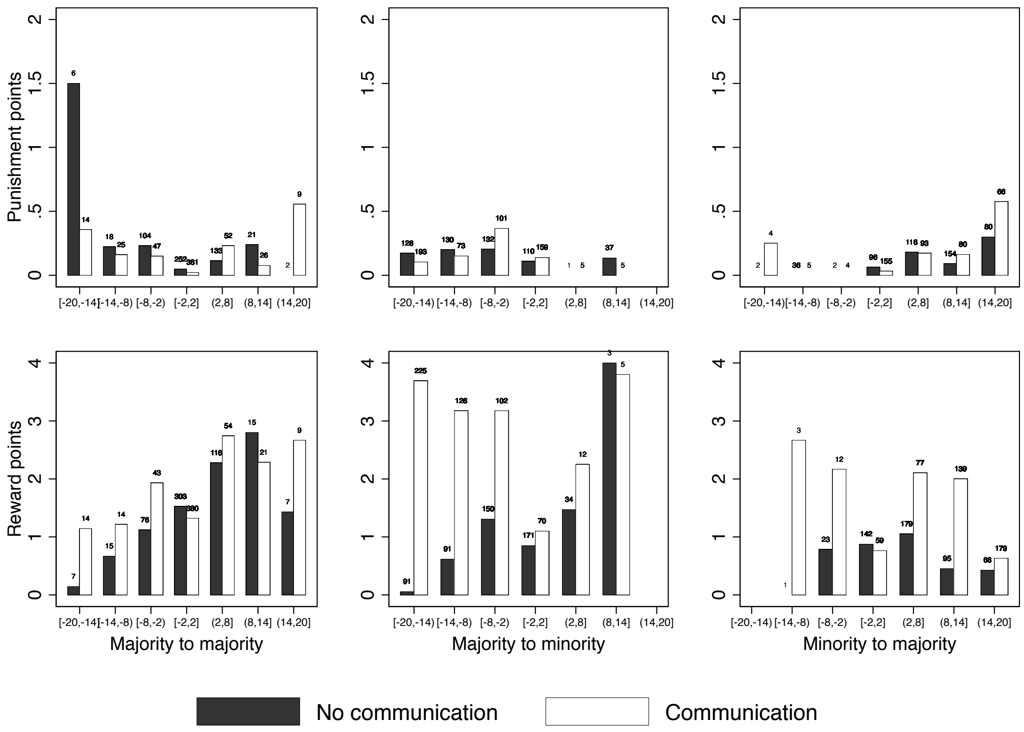

Figure 2: Punishment and reward by player types and difference in contributions.

function of the punished party's relative contribution. Second, the within-majority punishment depicted in the top left panel does not follow the typical pattern found in public goods with punishment experiments. For the majority of cases, there is almost as much punishment of positive deviators as there is for negative deviators, which is at odds with the standard finding that there is much more pro- than anti-social punishment. The only evidence we find for strong pro-social punishment are six observations in the treatment without communication, where the punisher contributed her full endowment and the punished participant contributed less than 6 tokens (zero in four of the cases). [14]

Columns (1) and (2) of Table 5 report mixed effects regressions analyzing the effect of being punished on contributions in the next period. This analysis, too, reveals that punishment is ineffective in increasing contributions. The only effect that approaches significance is that of being punished by the minority player. Since minority players punish those who contribute, this more likely reflects a desire to counter-punish the minority player through contributions.

The communication analysis provides complementing evidence for minority players using the punishment institution to curb majority contributions, as the minority players' active participation in the communication—measured by number of minority player messages—is negatively correlated with the mean contributions in the subsequent block of five periods. Each additional message by the minority player was associated with a decrease of around 0.9 points in the average contributions majority players (see Table A3 in the appendix).

The low punishment levels (see Table 6) are also consistent with the conjecture that majority players are reluctant to punish. Astoundingly, none of the 54 communications in the punishment with chat treatment included messages referring to punishment of non-contributors. Thus, the data support Conjecture 2.

Result 2. *In the presence of a harmed minority, punishment is not used to deter free riding, and is even used by the minority to deter contributions. Consequently, the punishment institution does not increase contribution levels, while reducing efficiency.*

3.3 Reward

Without communication, rewards between the two majority players is, on average, double the rewards directed at the minority player. Figure 2 shows that rewards received are positively correlated with contributions. Notwithstanding, majority players allocate positive amounts to players who contributed less then themselves.

[14]In the Communication treatment, on the other hand, we have nine observations with relatively high average punishment of high contributors by low contributors.

Table 5: Marginal effects of receiving punishment and reward on contributions.

	(1) Majority, punishment	(2) Minority, punishment	(3) Majority, reward	(4) Minority, reward
Received from majority in $t-1$				
No communication	-0.087	-0.386	0.878**	0.485***
	(0.332)	(0.285)	(0.371)	(0.159)
Communication	-1.100	-0.053	0.357	0.193
	(1.036)	(0.133)	(0.244)	(0.128)
Received from minority in $t-1$				
No communication	0.494*		-0.092	
	(0.292)		(0.395)	
Communication	0.759		0.677	
	(0.518)		(0.827)	
N	1044	522	1044	522

Notes: Marginal effects based on mixed effects regressions on contributions in period t by received punishment or reward in $t-1$, interacted with communication treatment, and period with random effects for individuals nested in groups. Robust standard errors clustered on groups. *,**,*** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table 6: Allocated punishment and reward by treatment and type.

		Punishment			
		No Communication		Communication	
		To		To	
		Majority	Minority	Majority	Minority
From	Majority	0.13 6.5%	0.17 9.3%	0.08 3.3%	0.17 5.6%
	Minority	0.29 15.6%	—	0.18 10.4%	—
		Reward			
		No Communication		Communication	
		To		To	
		Majority	Minority	Majority	Minority
From	Majority	1.62 54.4%	0.86 46.9%	1.58 63.1%	3.11 78.3%
	Minority	0.75 52.2%	—	1.29 58.9%	—

Note: Means of, and proportions of players engaging in, punishment and reward.

Overall contribution levels are somewhat lower than in the baseline treatment, however the difference is not significant ($\beta = 0.528, p = 0.630$ for the marginal effect of the reward treatment on majority contributions based in the regression reported in Column (1) of Table 4).

Communication alters this behavioral pattern significantly. As can be seen in Table 6, most majority players reward the minority players. In total, the minority player receives more than 6 reward points on average. Columns (3) and (4) of Table 4 confirm that majority players significantly increase the amount of points they reward by 2.204 points ($p < 0.05$) and direct rewards toward the minority player, who receives an estimated amount of 4.493 reward points more from each majority player as a result of the communication ($p < 0.005$). Thus, communication facilitates compensation of the harmed minority. Furthermore, as can be seen from Columns (3) and (4) of Table 5, receiving punishment from a majority player leads to higher contributions in the next period—only without communication. With communication, rewards are no longer used to encourage contributions, but used as a compensation channel.

The communication content provides further evidence on the interaction of communication and rewards. As discussed above, the existence of the harmed minority leads many groups to discuss not contributing. Such messages appear in 37.9% and 38.9% of all communications in the Standard VCM and Punishment treatment, respectively. In contrast, when the reward institution enables compensation, contribution is no longer viewed as negative, and messages advocating not contributing appear in only 9.3% of communications in the Reward treatment ($\chi^2 = 15.21, p < 0.001$).[15]

Contrary to our conjecture, minorities do not significantly increase their contributions when compensation is feasible. As can be seen in Figure 2, when communication is available, minority players receive compensation even when they contribute much less than the compensating majority player. Our content analysis complements this finding. In many case, the majority players acknowledge the harm incurred by the minority player and accept that she should refrain from contributing and be compensated.[16]

Compensation, however, is accompanied by an increase in contributions by

[15]This result is robust to controlling for repeated observations. Only 3 of 9 groups discussed not contributing in the Reward treatment compared to 8 of 11 and 8 of 9 in the Standard VCM and Punishment treatments, respectively ($\chi^2 = 6.56, p < 0.05$). A logistic regression of discussing not contributing by institution also results in significant differences between the Reward and the other two treatments ($p < 0.005$ for both pairwise comparisons).

[16]Communication and rewards may facilitate minority contributions if they can be compensated for the harm incurred by a full contributions profile, which is excluded by the experimental parameters. The communication content suggests, however, that participants tend to perceive self-inflicted harm as undesirable, even if the harm can be redressed.

the majority players, who, in the reward treatment, contribute 5.072 tokens more with communication than without communication ($p < 0.005$). Consequently, efficiency is slightly but significantly higher with communication (2.044, $p < 0.005$). This result suggests that majority players are reluctant to contribute to a public good that is not universally and unequivocally good. This tentative conclusion, however, is in contrast with the results of Engel and Rockenbach (2011, 2014). Whether negative externalities play a different psychological role when imposed within or without the active group requires a systematic study.

The behavioral pattern observed in the reward treatment is summarized in the following result.

Result 3. *Communication is a necessary condition for compensation. When rewards are available, communication leads to compensation of the harmed minority, increased majority contributions, and higher efficiency.*

4 Conclusion

Potential Pareto Public Goods are someone's public bad. We find that when the public good harms a member of the group, contribution is no longer perceived as unequivocally socially desirable.[17] Group communication, which typically leads to high contributions (Bochet et al., 2006; Frohlich and Oppenheimer, 1998; Koukoumelis et al., 2011) is often used to deter contributions, and is consequently not effective in enhancing cooperation. Furthermore, the existence of a minority player who contributes little may pull down the contributions of majority players. This finding complements and extends the recent finding by Gangadharan et al. (2015) that the efficiency-enhancing effect of communication is diminished if group members draw differential benefits from the public good.

Since contributing to the Potential Pareto Public Good is not unequivocally good for the group, groups don't use the punishment institution effectively to boost cooperation. We attribute part of the inefficacy of punishment to the fact that a majority free rider is likely to still contribute no less than the minority player, creating ambiguity with regard to the proper target of punishment. Such ambiguity is known to undermine the efficacy of punishment institutions, for ex-

[17] Delaney and Jacobson (2015) studied a public bad that harms three players in a group of four, however the fourth player draws a high benefit from the public bad, making it selfishly dominant for her to contribute. Such privileged players contributed less than their full endowment if they knew that some of the other group members would not get a chance to be in the privileged role in any of eight rounds played in the group. Unlike our setting, here the public bad is never desirable from a social point of view, and the negative feature of the situation is outside the stage game. Notwithstanding, the result mirrors our findings in that people are less likely to contribute to a project that is, in some sense, unfair to their peers.

ample when there is exogenous uncertainty regarding relative contributions (Ambrus and Greiner, 2012; Bornstein and Weisel, 2010; Grechenig et al., 2010).

This is reflected in the group communication. Group members advocate not contributing as much as they do without the possibility of punishing, and often suggest avoiding punishment altogether. Messages calling to punish free riders did not occur even once in the experiment. Accordingly, punishment levels are very low, and majority players are not punished more by their peers when they contribute less, unless the difference in contributions between the punisher and punishee is extremely high.[18] Punishment levied by minority players on contributing majority players further erodes cooperation under punishment, indirectly by effectively being a form of antisocial punishment—which is known to undermine the potential for punishment to increase contributions—and by deterring contributions directly.

The ability to reward has the potential to drastically alter the perception of the public good. When the minority player can be compensated, contributions once more become unequivocally socially desirable. We find, however, that this only happens when the group members can communicate. Without communication, majority players use the reward mechanism to send money to each other. As there is no efficiency gain from rewarding, this is no more than a symbolic trust-building exercise.

With communication, on the other hand, majority players increase their contributions and redistribute the generated surplus to the minority players. The analysis of the messages reveals that indeed the public good is no longer perceived negatively, as the rate of discussions in which players advocate not contributing drops from above 35% in the other treatments to less than 10% in the reward treatment. Whereas without communication minority players only receive compensation when they contribute, with communication the majority players compensate the minority indiscriminately.

Why is it crucial for the public good to be unequivocally desirable? An answer can be found in the notion of *Team Reasoning* (see Gold and Sugden, 2007, for a comprehensive overview). In its basic form, team reasoning stipulates that, rather than reasoning about 'what should *I* do?', individuals pose the question 'what should *we* do?', with respect to some group payoff function U. If the group objective points at outcome x, each individual who identifies with the group proceeds to perform his part of the strategy profile that results in x (Bacharach, 1999,

[18]This result mirrors findings by Reuben and Riedl (2009), who found that the introduction of high benefit players that personally benefit from contributing diminishes the positive effect of punishment on contributions. In both studies, the existence of a player with a clear strategy—be it to contribute as in Reuben and Riedl (2009) or not to contribute as in the current study—erodes the efficacy of punishment.

2006; Sugden, 1993, 2000). In social dilemma games, such as symmetric linear public goods games, team reasoning will lead individuals to cooperate.[19]

The effect of harmed minorities in Potential Pareto Public Goods is related to the two main issues that arise in theories of team reasoning, namely *what is a team?* and *what is the team's objective?* (Bacharach, 2006; Sugden, 2000). We argue that the existence of harmed minorities undermines team reasoning both by hindering group identification—which is a necessary condition for individuals to team reason—and by eliminating cooperation as a common goal.

Group identification can arise from elements external to the game. Nonetheless, Bacharach (2006) hypothesized that some game features facilitate group identification. Specifically, his *interdependence hypothesis* states that group identification is stimulated if the group members perceive that there is some outcome s^* that they can bring about through joint action, which all members rank higher than some other outcome s, which can be obtained as a game-theoretic solution of the corresponding game between individually-minded players. This condition naturally holds in the standard symmetric public goods game, but is absent from the game with harmed minorities.

What determines the group payoff function U? Other than assuming Paretian ordering, most treatments take this as exogenously fixed. In a recent paper, Sugden (2015) defined a principle of *mutually beneficial practice* as a necessary feature of the group objective. A mutually beneficial practice is one that, if all group members follow it, guarantees to every member a payoff strictly higher than what she can unilaterally secure (formally, her maximin payoff).[20] Whereas in the standard public goods game, cooperation is a mutually beneficial practice, this is not the case in Potential Pareto Public Goods.

Thus, the existence of harmed minorities both undermine group identification, and removes the common goal. This is reflected in our analysis showing that cooperation is not viewed as unequivocally desirable. Two additional observations are in line with this interpretation of the effect of harmed minorities. First, communication typically facilitates cooperation by boosting group identity (Dawes et al., 1988), and by allowing for "public acts of joint commitment" that provide the necessary assurance that the group members are committed to team reasoning (Sugden, 2003, p. 178). As harmed minorities undermine both aspects, the lack of communication effects become clear when interpreted as mediated by team

[19] The number of times an individual used the word "I" was negatively correlated with her mean contribution over all periods ($r(87) = -0.272, p = 0.011$). The use of the word "we" was positively correlated with contributions ($r(87) = 0.225, p = 0.037$). However, the correlations were not significant at the group level.

[20] A second condition, that each group member contributes to the others by playing out his part, is less relevant to the current discussion.

reasoning.[21] Second, the reward mechanism reinstates interdependence in the Bacharach (2006) sense, allowing for group identification to arise from the game features. It also reinstates cooperation as a common goal satisfying the conditions of a mutually beneficial practice. Hence, group communication leads to higher contributions, and no longer involves prevalent calls to refrain from contributing.

In sum, this paper focuses on Potential Pareto Public Goods, which are harmful to some members of the community. We believe that this hitherto largely neglected aspect of public good provision is potentially relevant across many instances of public goods. We explore the implications of a harmed minority under different institutions. We conclude that the existence of a harmed minority dramatically affects the way in which the public good is perceived by the group, and consequently players' considerations and the efficacy of different institutions designed to facilitate public good provision.

References

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic growth 8*(2), 155–194.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature 43*(3), 762–800.

Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics 115*(3), 847–904.

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics 112*(4), 1027–1056.

Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment - an experimental study. *American Economic Review 102*(7), 3317–3332.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies 58*(2), 277–297.

[21] If the majority players conceive of their subgroup as the team, cooperation can be sustained through team reasoning. If this is the case, global communication can even reduce cooperation by extending the team to include the harmed minority player.

Armour, A. M. (1991). The siting of locally unwanted land uses: Towards a cooperative approach. *Progress in Planning 35*, 1–74.

Babcock, R. F. and C. L. Siemon (1985). *The zoning game revisited.* Lincoln Institute of Land Policy.

Bacharach, M. (1999). Interactive team reasoning: a contribution to the theory of co-operation. *Research in economics 53*(2), 117–147.

Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton, NJ: Princeton University Press.

Baland, J.-M. and J.-P. Platteau (1997). Wealth inequality and efficiency in the commons part i: The unregulated case. *Oxford Economic Papers 49*, 451–482.

Balasubramanian, R. and K. N. Selvaraj (2003). Poverty, private property and common pool resources management: The case of irrigation tanks in south india. *South Asian Network for development and Environmental Economics*.

Bardhan, P. and J. Dayton-Johnson (2002). Unequal irrigators: Heterogeneity and commons management in large-scale multivariate research. In *The Drama of the Commons*, pp. 87–112. Washington, DC: National Academies Press.

Bochet, O., T. Page, and L. Putterman (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization 60*(1), 11–26.

Bornstein, G. and O. Weisel (2010). Punishment, cooperation, and cheater detection in "noisy" social exchange. *Games 1*(1), 18–33.

Chambers, R. (1988). *Managing Canal Irrigation: Practical Analysis from South Asia*. Cambridge, UK: Cambridge University Press.

Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics 3*(1), 55–79.

Dawes, R. M., A. J. Van De Kragt, and J. M. Orbell (1988). Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations. *Acta Psychologica 68*(1-3), 83–97.

de Oliveira, A. C., R. T. Croson, and C. Eckel (2015). One bad apple? heterogeneity and information in public good provision. *Experimental Economics 18*, 116–135.

Delaney, J. and S. Jacobson (2014). Those outsiders: How downstream externalities affect public good provision. *Journal of Environmental Economics and Management 67*(3), 340–352.

Delaney, J. and S. Jacobson (2015). The good of the few: Reciprocal acts and the provision of a public bad. *Journal of Behavioral and Experimental Economics 58*, 46–55.

Duflo, E. and R. Pande (2007). Dams. *The Quarterly Journal of Economics*, 601–646.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, 1203–1250.

Engel, C. and B. Rockenbach (2011). We are not alone: the impact of externalities on public good provision. mimeo.

Engel, C. and B. Rockenbach (2014). Give everybody a voice! the power of voting in a public goods experiment with externalities. mimeo.

Engel, C. and L. Zhurakhovska (2014). Conditional cooperation with negative externalities–an experiment. *Journal of Economic Behavior & Organization 108*, 252–260.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature 415*(6868), 137–140.

Fehr, E., S. Gächter, M. Milinski, and B. Rockenbach (2010). Institutions fostering public goods provision. In A. Ockenfels and A. Sadrieh (Eds.), *The Selten School of Behavioral Economics*, pp. 167–184. Berlin: Springer.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischel (2015). *Zoning Rules!* Cambridge, MA: Lincoln Institute of Land Policy.

Fischel, W. A. (1979). Determinants of voting on environmental quality: a study of a new hampshire pulp mill referendum. *Journal of Environmental Economics and Management 6*(2), 107–118.

Fischel, W. A. (1985). *The Economics of Zoning Laws: A Property Rights Approach to American Land Use Controls*. Baltimore, MD: The John Hopkins University Press.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin 76*(5), 378.

Frey, B. S. and F. Oberholzer-Gee (1996). Fair siting procedures: an empirical analysis of their importance and characteristics. *Journal of Policy Analysis and Management 15*(3), 353–376.

Frey, B. S., F. Oberholzer-Gee, and R. Eichenberger (1996). The old lady visits your backyard: A tale of morals and markets. *Journal of political economy*, 1297–1313.

Frohlich, N. and J. Oppenheimer (1998). Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior & Organization 35*(3), 389–403.

Fujiie, M., Y. Hayami, and M. Kikuchi (2005). The conditions of collective action for local commons management: the case of irrigation in the philippines. *Agricultural Economics 33*, 179–189.

Gächter, S., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science 322*(5907), 1510.

Gangadharan, L., N. Nikiforakis, and M. C. Villeval (2015). Equality concerns and the limits of self-governance in heterogeneous populations. *mimeo*.

Gold, N. and R. Sugden (2007). Collective intentions and team agency. *The Journal of Philosophy*, 109–137.

Grechenig, K., A. Nicklisch, and C. Thöni (2010). Punishment despite reasonable doubt: A public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies 7*(4), 847–867.

Güth, W., A. Koukoumelis, and M. V. Levati (2011). One man's meat is another man's poison. *Jena Economic Research Papers*. mimeo.

Güth, W., A. Koukoumelis, M. V. Levati, and M. Ploner (2014). Providing revenue-generating projects under a fair mechanism: An experimental analysis. *Journal of Economic Behavior & Organization 108*, 410–419.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*(5868), 1362–1367.

Koukoumelis, A., M. Levati, and J. Weisser (2011). Leading by words: A voluntary contribution experiment with one-way communication. *Journal of Economic Behavior & Organization*.

Kunreuther, H., P. Kleindorfer, P. J. Knez, and R. Yaksick (1987). A compensation mechanism for siting noxious facilities: theory and experimental design. *Journal of environmental economics and management 14*(4), 371–383.

Lam, W. F. (1998). *Governing irrigation systems in Nepal. Institutions, infrastructure, and collective action.* Oakland, CA: ICS Press.

Meinzen-Dick, R. and M. Bakker (1999). Irrigation systems as multiple-use commons: water use in kirindi oya, sri lanka. *Agriculture and Human Values 16*(3), 281–293.

Meinzen-Dick, R. S. and L. A. Jackson (1997). Multiple uses, multiple users of water resources. In B. M. Swallow, R. S. Meinzen-Dick, L. A. Jackson, T. O. Williams, and T. A. White (Eds.), *Multiple Functions of Common Property Regimes*. Washington, DC: International Food Policy Research Institute.

Nelson, R. H. (1977). *Zoning and property rights: An analysis of the American system of land-use regulation*. Cambridge, MA: MIT Press.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics 92*(1-2), 91–112.

Nikiforakis, N. and H.-T. Normann (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics 11*(4), 358–369.

Noussair, C. N. and F. Tan (2011). Voting on punishment systems within a heterogeneous group. *Journal of Public Economic Theory 13*(5), 661–693.

O'Hare, M. (1977). Not on my block you don't: Facility siting and the strategic importance of compensation. *Public Policy 25*, 409–458.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review 86*(2), 404–417.

Poteete, A. R. and E. Ostrom (2004). Heterogeneity, group size and collective action: The role of institutions in forest management. *Development and change 35*(3), 435–461.

Rand, D. G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak (2009). Positive interactions promote public cooperation. *Science 325*(5945), 1272–1275.

Reuben, E. and A. Riedl (2009). Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution 53*(1), 72–93.

Schively, C. (2007). Understanding the nimby and lulu phenomena: Reassessing our knowledge base and informing future research. *Journal of planning literature 21*(3), 255–266.

Sefton, M., R. Shupp, and J. M. Walker (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry 45*(4), 671–690.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social philosophy and policy 10*(01), 69–89.

Sugden, R. (2000). Team preferences. *Economics and Philosophy 16*(2), 175–204.

Sugden, R. (2003). The logic of team reasoning. *Philosophical explorations 6*(3), 165–181.

Sugden, R. (2015). Team reasoning and intentional cooperation for mutual benefit. *Journal of Social Ontology 1*(1), 143–166.

Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *Review of Economic Studies 77*(4), 1540–1566.

Tan, F. (2008). Punishment in a linear public good game with productivity heterogeneity. *De Economist 156*(3), 269–293.

Tang, S. Y. (1992). *Institutions and collective action: Self-governance in irrigation*. San Francisco, CA: ICS press.

Varughese, G. and E. Ostrom (2001). The contested role of heterogeneity in collective action: Some evidence from community forestry in nepal. *World D 29*(5), 747–765.

Vigdor, J. L. (2004). Community composition and collective action: Analyzing initial mail response to the 2000 census. *Review of Economics and Statistics 86*(1), 303–312.

Walker, J. M. and M. A. Halloran (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics 7*(3), 235–247.

Appendix I: Dynamic regressions

We ran Arrelano Bond estimations of dynamic models

$$y_{it} = \alpha y_{i,t-1} + x_{i,t}\beta + \eta_i + \varepsilon_{it},$$

where y_{it} is the contribution of participant i at period t, and η_i is an unobserved participant effect. To account for the correlation between the lagged contribution and the unobserved participant effect, the estimation used lag 2 to lag 27 as instruments. Table A1 reports the coefficients from these regressions as well as from those reported in table 3.

Table A1: Dynamic regressions, Standard VCM.

	Majority contribution in period t				Minority contribution in period t			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	Fixed effects	Arellano-Bond, No communication	Arellano-Bond, Communication	OLS	Fixed effects	Arellano-Bond, No communication	Arellano-Bond, Communication
Majority in $t-1$	0.715***	0.356***	0.327***		0.092	-0.022	-0.016	-0.041
	(0.067)	(0.075)	(0.080)		(0.079)	(0.024)	(0.031)	(0.045)
Majority in $t-1$ × Communication	-0.070	0.079		0.326***	0.140	0.008		
	(0.117)	(0.113)		(0.077)	(0.225)	(0.060)		
Minority in $t-1$	0.104**	-0.025	-0.014					
	(0.046)	(0.053)	(0.097)					
Minority in $t-1$ × Communication	-0.041	-0.082		-0.155				
	(0.102)	(0.120)		(0.096)				
Own contribution in $t-1$			0.234**	0.098			0.025	0.218
			(0.094)	(0.066)			(0.026)	(0.195)
Communication	1.036				0.666			
	(1.694)				(1.539)			
Period	-0.035*	-0.037	-0.025	-0.076**	0.013		0.041	0.008
	(0.020)	(0.029)	(0.029)	(0.039)	(0.037)		(0.080)	(0.058)
Constant	3.496***	7.710***	5.178***	8.691***	-0.235	2.449***	0.494	3.093**
	(0.987)	(0.770)	(1.154)	(1.579)	(0.813)	(0.364)	(0.579)	(1.225)
N	1276	1276	616	616	638	638	308	308

Notes: *,** ,*** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

Appendix II: Contributions by communication content

Table A2: Contributions by messages to contribute or not to contribute.

	Institution		
	Standard VCM	Punishment	Reward
Contribute	1.177	3.624***	2.779**
	(1.021)	(1.272)	(1.229)
Don't contribute	-5.226***	-3.389***	-1.637
	(1.664)	(1.219)	(2.057)
Block	0.201	-0.175	-0.092
	(0.215)	(0.222)	(0.407)
Observations	174	174	174

Marginal effects based on an omnibus model including interactions with the institution type.
Robust standard errors clustered on groups.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Number of messages and contributions

	(1) All	(2) Non-social messages	(3) Social messages	(4) Minority messages
Non-social messages in…				
Standard VCM	-0.035	0.130		
	(0.194)	(0.105)		
Punishment	0.089	-0.014		
	(0.180)	(0.236)		
Reward	-0.312**	-0.424***		
	(0.143)	(0.153)		
Social messages in…				
Standard VCM	-0.270		-0.252	
	(0.280)		(0.164)	
Punishment	0.128		-0.112	
	(0.157)		(0.150)	
Reward	0.110*		0.136*	
	(0.055)		(0.078)	
Minority messages in…				
Standard VCM	-0.064			-0.256
	(0.366)			(0.242)
Punishment	-0.913***			-0.745***
	(0.176)			(0.111)
Reward	-0.387			-0.361
	(0.229)			(0.288)
Block				
All treatments	-0.168	-0.342*	-0.235	-0.219
	(0.238)	(0.191)	(0.216)	(0.188)
Observations	174	174	174	174

Marginal effects of number of non-social, social and minority messages.

Robust standard errors clustered on groups.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix III: Experimental instructions

For Online Publication

General instructions

<Placed on table>

Welcome to the experiment.

Please do not talk to the other participants in the experiment.

During the experiment you will be able to accumulate points according to the decisions you will take.

These points will be converted to money at a rate of **10 points = 1 NIS**.

Additionally you will receive a payment of 10 NIS and a payment for an additional task.

At the end of the experiment please wait for the experimenter to call you for payment.

The instructions will appear on the screen in a minute. Please read them carefully. Please raise your hand if you have any questions.

Instructions for the experiment

<Presented as a pdf document and available throughout the experiment>

These instructions are identical to all the participants.
The experiment is composed of three phases. Each phase includes 10 rounds. **The three phases of the experiment are identical.**
In this experiment you will be randomly and anonymously placed in a group of **three people**, so that None of the group members knows the identity of the other group members. **The composition of the group remains fixed from round to round and from phase to phase,** and will be set by the computer once at the beginning of the first round of the first phase. In the first round of the first phase, each group member will be assigned one of the letters A, B, or C, and will be thus identified. **The identifying letter is fixed throughout the experiment.**

The round

ONLY IN THE COMMUNICATION TREATMENTS

Before the **first and sixth round of every phase** you will have 90 seconds to chat with your fellow group members. When a group member writes a message, all of the group members will see on screen the writer's identifying letter and content of the message. **Important: Do not write anything that could reveal your true identity!** If one of the group members will reveal his identity, the experiment will end for all of the group members and they will receive the minimal payment.
Additionally, each group member will see on this screen the **income from each point in the project** for the group members. This will be explained shortly.

<u>Screen 1</u>
First, each group member will be endowed with **20 points**. Of these points, each group member will choose **how many points to contribute to a project.** The group member keeps the points not contributed. Additionally, the first screen will present your identifying letter and your **"income from each point in the project"**. The income from each point in the project indicates how many points a group member earns **per point contributed to the project account.** One of the three group members will be randomly chosen by the computer. The income from each point in the project for this group member will be 10.4 (minus 0.4). The income from each point in the project for the other group members will be 0.8. The income from each point in the project is **fixed throughout the experiment**.Thereforem the profit from the first screen is calculated thus: if your income from each point in the project is set to 0.8, your first stage earnings are:

$$\boxed{20} - \boxed{\text{The number of points you contributed to the project}} + \boxed{0.8 \times \text{The sum of points the group contributed to the project}} = \boxed{\text{First stage earnings}}$$

If you are the group member whose income from each point in the project is -0.4, your first stage earnings are:

$$\boxed{20} - \boxed{\text{The number of points you contributed to the project}} + \boxed{-0.4 \times \text{The sum of points the group contributed to the project}} = \boxed{\text{First stage earnings}}$$

After you decide how many points to contribute to the project, click on the “continue” button. Once all of the participants have clicked on the “continue” button, the next screen will appear.

<u>Screen 2</u>

On the second screen, each group member will see how many points he contributed to the project, how many points he kept for himself, the sum of contributions to the project in the group, and his profit for the round. This screen will appear for 10 seconds, or until all of the participants clicked on the “continue” button.

ONLY IN THE STANDARD VCM TREATMENTS

<u>Screen 3</u>

On the third screen, each group member will see the contribution and the income from each point in the project of the other group members. This screen will appear for 20 seconds, or until all of the participants clicked on the “continue” button.

ONLY IN THE PUNISHMENT AND REWARD TREATMENTS.
ITALICS DENOTE WHERE THE TREATMENTS DIFFER.

<u>Second stage, Screen 1</u>

On the third screen, each group member will see **the income from each point in the project and the contribution to the project of the other group members.**

Each group member will receive an additional allocation of 8 points. These 8 points can be used to *punish/reward* the other group members as follows. Enter next to each identifying letter how many of the 8 points to allocate to *punishing/rewarding* the group member associated with the identifying letter. **Each point allocated to *punishing/rewarding* another group member will deduct one point from the *punishing/rewarding* group member and three points from the *punished/rewarded* group member.** Therefore, **Your second stage earnings are:**

8	−	The number of points I allocated to *punishing/rewarding* other group members	−/+	3 ×	Sum of points other group members allocated to *punish/reward* me	=	Second stage earnings

And your round earnings are:

First stage earnings	+	Second stage earnings	=	Round earning

After you decide how many points to allocate to *punish/reward* the other group members, click on the "continue" button. Once all of the participants have clicked on the "continue" button, the next screen will appear.

Second stage, Screen 2

At this stage, each group member will be able to see how many points each group member allocated for *punishing/rewarding* and which group members were *punished/rewarded*. Additionally, all participants will be able to see their first stage earnings in the round, their second stage earnings in the round, and their total earnings in the round. **Here is a screenshot of this stage with explanations:**

This screen will appear for 15 seconds, or until all of the participants clicked on the "continue" button.

End of the phase

After ten rounds a new phase will start. Before the beginning of the new phase, each group member will see **How many points were accumulated by the other group members in the previous phase. The composition of the groups, the identifying letters and the income of each group from the project will remain fixed. The experiment is composed of three phases that have 10 rounds each.**

The end of the experiment

At the end of the three phases of 10 rounds each, the experiment will end. You will be asked to remain to participate in an additional task, for which you will also be paid.
The additional task is individual. Your payment for the additional task does not depend on the other participants.

Payment

The computer will randomly select one of the three phases. The payment will be one Shekel per 10 points earned in the selected phase.
If you have any questions, please raise your hand now and the experimenter will approach you.
If you understand the instructions, please switch to the other screen (marked with a leaf) using ALT + TAB and click on "continue".