

Rockenbach, Bettina; Engel, Christoph

Conference Paper

We Are Not Alone: The Impact of Externalities on Public Good Provision

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Voluntary Cooperation, No. E15-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Rockenbach, Bettina; Engel, Christoph (2010) : We Are Not Alone: The Impact of Externalities on Public Good Provision, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Voluntary Cooperation, No. E15-V3, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37529>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

We Are Not Alone: The Impact of Externalities on Public Good Provision

Christoph Engel and Bettina Rockenbach*

August 2009

Abstract:

Providing public goods is hard, because providers are best off free-riding. Is it even harder if one group's public good is a public bad for another group or, conversely, gives the latter a windfall profit? We experimentally study public goods provision embedded in a social context and find that in the absence of explicit norms externalities have almost no effect. With an endogenously formed provision norm positive as well as negative externalities dampen provision as compared to no externalities. We explain the surprisingly low provision under positive externalities by the providers' increased risk of inequity and stress the importance of institutions sustaining conditional cooperation.

Keywords: Public Good, Externality, Conditional Cooperation, Inequity Aversion, Norms

JEL Classification: C91, D03, D43, D62, H23, H41, L13

* We thank Armin Falk, Özgür Gürek, Martin Hellwig, Wieland Müller, Andreas Nicklisch, Karim Sadrieh, Christian Traxler, Gari Walkowitz and seminar audiences at the University of Tilburg, the Gruter Institute, the University of Cologne, the University of Bonn, the University of Erfurt for valuable comments. Financial support by the Deutsche Forschungsgemeinschaft through RO3071 is gratefully acknowledged.

Authors addresses:

Christoph Engel, Max Planck Institute for Collective Goods, Bonn and University of Bonn, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany, e-mail engel@coll.mpg.de

Bettina Rockenbach, University of Erfurt and Center for empirical research in economics and behavioral science (CEREB), Nordhäuser Str. 63, 99089 Erfurt, Germany, email: bettina.rockenbach@uni-erfurt.de

1. Introduction

The essence of many social problems is the temptation to free ride on others' contribution to the provision of a public good. This essence is not only theoretically well understood (Cornes and Sandler 1996) but has also been backed by a rich experimental literature (Ledyard 1995; Fehr and Gächter 2000; van Dijk, Sonnemans et al. 2002; Masclet, Noussair et al. 2003; Page, Putterman et al. 2005; Potters, Sefton et al. 2005) and is corroborated in the field (Ostrom, Dietz et al. 2002; Anderson, Mellor et al. 2004; Andersen, Bulte et al. 2008). Behavioral research has shown that the core of the problem is not naked greed, but a hurt sense of fairness. Experimental populations frequently “conditionally cooperate”. They are happy to make substantial contributions to a joint project as long as they believe a sufficient portion of others to do so as well (Keser and van Winden 2000; Brandts and Schram 2001; Fischbacher, Gächter et al. 2001; Frey and Meier 2004; Croson, Fatas et al. 2005; Fischbacher and Gächter 2009). This is in line with a strong aversion to be betrayed by others (Bohnet, Greig et al. 2008). The important news for policy makers is that it need not be necessary to force everyone to contribute. It may be enough to make sure that the risk of being the sucker is not too strong, or too salient.

Yet unfortunately political reality is often more complex. Providing the public good often also affects people outside the borders of the community. Equatorial countries preserving the rain forest do not only save their national ecosystems, but the world's climate and biodiversity along with it. If a metropolitan area subsidizes the opera house, it attracts visitors from further away who do not pay local taxes. In these examples public goods provision is not only domestically valuable but additionally creates a positive externality for outsiders. On the other hand, the successful provision of a public good may create negative external effects. Take a country close to the source of an international river, building a dam to secure irrigation water and energy for its population. This deprives countries closer to the estuary of the river's benefits. Or, think of a municipality constructing a landfill close to its borders to keep garbage off its streets. This puts the groundwater in the neighboring community at risk. Of course, one of the economically most prominent examples for negative externalities of cooperation is the formation of a cartel. Successful cooperation among suppliers imposes damage on customers.

How does the existence of external effects affect public goods provision? Do positive external effects make provision “easier” while negative external effects decelerate the provision process? And, in which way is the willingness to contribute to a public good altered if a social

norm, originating in the larger social context, demands abstention or, conversely, establishes a duty to contribute? In this paper we tackle these questions experimentally as well as theoretically.

We model a linear public goods game with externalities on bystanders. These bystanders either profit (*positive externality treatment*) or suffer (*negative externality treatment*) from the actors' provisions to the public good. The situation is asymmetric as bystanders have no direct means to influence the providers' payoffs. To study the influence of norms and the interaction of norms with externalities we study the game in three different phases: first in the absence of any explicit provision norm, second after a provision norm has endogenously been created, and third after the existing norm is formally abolished. In the field, norms are both triggered and justified by externalities, making identification of causalities difficult. Through the design of our experiment, we are able to disentangle the effects of externality and normativity, as well as their interaction.

We find that, in the absence of explicit norms, just knowing that the provision causes an externality has almost no effect on providers. Yet norms, although being in the weak form of mere recommendations, influence provision behavior. Endogenous norm formation internalizes the negative externality: provision norms in the negative externality treatment are significantly lower than in the positive externality treatment or in a control without an externality. Interestingly, though, the provision norm in the positive externality treatment is not higher than in the control. This is noteworthy because in this treatment potential welfare gains are highest. With the endogenously formed provision norm, provisions are highest without an externality. Both forms of externality dampen the provision level. While in the negative externality treatment this is a response to the lower norms, the lower provisions in the positive externality treatment result from two effects: first from norms not being higher than without an externality and second from the fact that – *ceteris paribus* – norm compliance in the positive externality treatment is significantly lower than absent an externality.

How can the observed behavior, especially the surprisingly low provision levels in the positive externality treatment, be explained? Neither the maximization of utility from monetary profits nor the incorporation of inequity aversion (with respect to actors as well as to bystanders) into the utility function can directly explain our observations. Nevertheless, we show that payoff comparisons play an important role in explaining provision behavior. Actors not only seem to compare themselves to the other actors, but also to bystanders, and these comparisons have different consequences. Two motives seem to be prevalent. Actors are

predominantly guided by conditional cooperation, but only if this is not in conflict with the second motive, namely not to fall back behind the passive bystanders. While in the negative externality treatment and in the control these two motives align, they clash in the positive externality treatment. Here conditional cooperators not only face the risk of achieving lower payoffs than free-riding actors, but they may also fall back behind the passive bystanders. Additionally, when comparing themselves with the bystanders, actors seem to be simultaneously and independently sensitive to collective and to individual differences.

Our results clearly point to a limitation of self-governance. Conditional cooperators need institutions to “protect” them against the risk of being the sucker, especially with respect to outsiders gaining a windfall profit. Obviously, the institutional environment of the experiment did not provide this protection. In this light, going back to our examples, it becomes understandable why equatorial countries are compensated for preserving the rain forest by being exempted from the obligation to reduce CO₂ emissions; or why municipalities tax secondary residences, using the second home as a proxy for the benefit from local public goods.

One could think of even more general policy implications of our findings. As long as states would strictly maximize the aggregate utility of their citizens, many transnational public goods would be provided. Even if other states receive a windfall profit, the benefit for the nationals of the providing state would often still be large enough. Yet government has to defend higher taxes and onerous regulation vis-à-vis the citizenry. Not so rarely, political support for an otherwise sensible intervention falters if this gives outsiders a free lunch. A striking illustration is defense. Often, if one country disciplines a rogue state, many other countries benefit as well, yet save their soldiers’ lives. The hurt sense of fairness may explain why even large countries like the US press their allies to join them. It becomes understandable why so many armed conflicts go untamed. And one sees why federations like the United States of America and confederations like the European Union have grown so large: under the federal umbrella, beneficiaries cannot so easily escape contributing their fair share.

Of course, all our examples are embedded in a much richer environment than the one we modeled in our experimental game. Yet in all examples, the underlying conflict has the structure of a public goods dilemma for the internals with positive or negative repercussions on externals who neither contribute nor have a say on the contribution level. Our results suggest that policy makers should be concerned that conditional cooperation is hampered and that norms work less effectively when internal cooperation gives outsiders a windfall profit.

In section 2 we discuss the related literature. In section 3 we present the formal model and in section 4 we derive hypotheses from our theory. Section 5 presents the results and section 6 concludes. The appendices provide the instructions and the statistics.

2. Related Literature

To the best of our knowledge no experimental study on public goods provision with externalities on inactive others has been conducted so far. Surprisingly, even in other contexts there are only a few studies which have aspects of externalities. In a wider sense dictator games (Hoffman, McCabe et al. 2008) qualify. The dictator may be said to create a positive externality for the non-acting recipient. Also (Güth and van Damme 1998) may be interpreted as a game with an externality. The proposer proposes how to divide a pie between three players. The division is executed if and only if the responder accepts. Otherwise, all players receive nothing. These two acting players create an externality on the inactive third player (dummy player). Another example is provided by (Bolton and Ockenfels 2008). An actor chooses between lotteries, and a non-actor's payoff depends on the actor's choice. Dependent on the parameterization and the realization of the lottery, the non-actor may receive a lower or a higher payoff than the actor. (Abbink 2005) plays a two-person bribery game in which corruption negatively affects passive workers. He concludes that the reciprocity considerations between briber and official overrule concerns about distributive fairness towards other members of the society.

Studies with effects on active others are more common. For example, Bornstein and colleagues extensively study team competition in various contexts (for an overview see Bornstein 2003) and find that in social dilemmas the competition with another group increases in-group cooperation. (Abbink, Brandts et al. 2009) find that group members punish each other more severely if the group is in conflict with another. The group position may be interpreted as a joint project of the group.

In the framework of our study it is important to recall that contribution rates in public goods provision are sensitive to framing. If the provision problem is framed as a public good, there is more cooperation than if the very same problem is framed as a public bad (Andreoni 1995; Sonnemans, Schram et al. 1998; Vergnaud, Willinger et al. 1999). These studies, however, did not incorporate any outsiders that are either positively or negatively affected. Croson and Marks find that a "recommendation" increases contributions in a linear public good (Croson and Marks 2001).

In an indirect way, the experimental literature on oligopoly provides evidence. In a recent meta-study (Engel 2007) collusion rates in oligopoly experiments between 1959 and 2006 were analyzed. Collusion is significantly lower if the opposite market side is represented by real subjects (collusion rates of about 7%), rather than a computer bidding a predetermined demand function (collusion rates of about 43%). This might indicate that participants shy away from imposing harm on other participants, which would imply that cooperation is lower if it entails a negative externality. Moreover, in the meta-study, collusion is significantly higher if the otherwise identical game is not framed as a market (collusion rates of about 36%), but is framed neutrally (collusion rates of about 57%). With a neutral frame, experimental subjects have no chance to activate their world knowledge about the undesirability of collusion. This might indicate that normativity dampens cooperation when cooperation engenders a negative externality.

3. A Public Goods Game with Externalities

We introduce a linear public goods game in which public goods provision causes externalities to non-actors. The game consists of n_A active players, the *actors*, and n_B passive players, the *bystanders*. Actors are endowed with e_A and may contribute any amount $0 \leq g_i \leq e_A$ to a public good, which benefits all actors. As in a standard public goods game, the sum of all actors' contributions $G = \sum_{k=1}^{n_A} g_k$ is augmented by $a \cdot n_A$ and then equally distributed among the actors. The parameter $\frac{1}{n_A} < a < 1$ is the marginal per capita rate (MPCR) that specifies the marginal individual return each actor receives from her own contribution to the public good. The actors' payoff is given in equation (1):

$$(1) \quad \pi_i^A = e_A - g_i + aG, \quad i = 1, \dots, n_A$$

Bystanders receive an endowment e_B and cannot contribute to the public good. But – dependent on the sign of the parameter b – they either benefit from ($b > 0$), suffer from ($b < 0$), or are unaffected by ($b = 0$) the contributions of the actors. Accordingly, for a given b all bystanders earn an identical payoff which is solely determined by the actors' actions and is out of the bystanders control. The profit function of bystanders is given by equation (2).

$$(2) \quad \pi^B = e_B + bG$$

Experimental Implementation

In our experimental implementation we set the endowments of actors and bystanders to be equal $e_A = e_B = 20$, and keep $a = 0.4$ fixed. For the sake of comparability we choose parameters that are standard in public goods experiments without externalities¹. Our subject groups consist of 7 subjects, $n_A = 4$ actors and $n_B = 3$ bystanders. Subjects play the above game repeatedly over 10 subsequent rounds in fixed groups. We restrict the contribution rates to be either $g = 0$, $g = 10$, or $g = 20$.

In our two treatments, we vary the way in which bystanders are influenced by the contributions of actors, i.e. we vary b . In the *positive externality treatment PE*, we set $b = 0.2$. In the *negative externality treatment NE*, we set $b = -0.2$. In addition to the two treatments we run a *control* with no externality, i.e. $b = 0$. Table 1 summarizes the experimental parameters.

Table 1: Experimental parameters

Treatments	endowment $e_A = e_B$	MPCR of actors a	MPCR to bystanders b
positive externality	20	0.4	0.2
negative externality	20	0.4	-0.2
control	20	0.4	0

Normativity

Previous experiments have already shown that subjects are influenced by the framing of the experimental setup, e.g. as a public goods or a public bads experiment. Is this the consequence of an experimenter's demand effect, i.e. do participants feel obliged to cooperate in public goods experiments while, in oligopoly experiments, they feel obliged not to collude? Or do they (at least partially) follow an (implicit) norm not to harm others? To tackle these questions we study norm formation and norm abidance, but we deliberately refrain from introducing normativity exogenously and instead endogenize norm formation. To do so, we subdivide the experiment into three phases of 10 rounds each.

Phase 1: no *contribution norm*

Phase 2: *contribution norm*

Phase 3: *contribution norm* removed

¹ In the meta-study by (Zelmer 2003) on 27 studies with 711 distinct groups, the mean MCPR was 0.404 and the average number of periods was 10.4.

In the first phase no contribution norm is salient. At the beginning of phase 1 subjects are randomly allocated to the roles of actors and bystanders, and play the public goods game with fixed roles. At the beginning of phase 2, all 7 players of a group decide on a desirable contribution level for phase 2. After voting is completed roles of actors and bystanders are randomly assigned for phase 2. We have deferred voting until the second phase to make sure that participants had made experiences with the cooperation problem in the respective treatment. Voting takes place in the “entire society” and under the veil of ignorance to reflect a fundamental tenet of rule of law and of democracy: rules should not be ad hoc, but should be “general”. Their field of application should transcend the case that triggered their introduction. For that reason, statutes define conflicts and solutions in abstract terms. Since participants did not know future roles, we did also not have to worry about the formation of coalitions along predefined interests. In an alternative with fixed roles in which only actors vote, the voting process would resemble a form a pre-play communication among the actors, which is not in the focus of our present research.

It is made clear in the instructions that the norm is meant to be a non-binding guideline, and that there is no enforcement mechanism. Specifically all players vote on one of the three possible contribution rates (0, 10, or 20) as a “recommendation” for the later actors. The *contribution norm* is determined by absolute majority vote. If no contribution level receives an absolute majority in the first voting round, a run-off ballot between the two contribution levels with the highest number of votes is conducted. Phase 2 allows to study how cooperation rates change if a norm is newly introduced, and how norm compliance interacts with the externality. We finally want to know whether a norm even helps subjects coordinate once it is no longer “in force”. To that end, we again randomly assign roles at the beginning of phase 3 to the members of the same group of 7. We announce that the previous norm has been “abolished”, and have them play another 10 rounds. The three phases allow us to investigate the interaction between externalities and normativity in a voluntary contribution game: if no norm is salient (phase 1); when an endogenously created norm is present (phase 2); and when this norm is removed (phase 3).

The exact phase structure was unexpected for the subjects. Prior to phase 1, subjects received the instructions for that phase (see Appendix A.1) and were informed that, after completion of this phase, the experiment would continue. They were told that they would receive new instructions for the continuation and that their phase-1-behavior had no consequences for their strategic position in the continuation. After the termination of phase 1, subjects received the

instructions for phase 2 (see Appendix A.2) and were told that, after completion of this second phase, the experiment would continue and that they would receive new instructions for the continuation. Neither their phase-1-behavior nor their phase-2-behavior would have any consequences for their strategic position in the continuation. After the termination of phase 2 subjects received the instructions for phase 3 (see Appendix A.3) and were told that after completion of this third phase the experiment would end. The accumulated payoff of all three phases was paid out in cash immediately after the completion of phase 3.

Conduct of the Experiment

The experiment was run at the University of Erfurt (*elab*) in June 2008 with a computerized interaction using z-Tree (Fischbacher 2007). Subjects were invited using ORSEE (Greiner 2004). Each subject played in one of these three parameter variations (two treatments and control) and no subject played in more than one. We collected nine independent observations in each of both treatments and in the control, adding up to 27 independent observations with a total of 189 subjects. Each session lasted about one hour and subjects earned on average 14.19 € in the control sessions (14.29 € for actors, and 14.06 € for bystanders), 14.79 € in the *PE* sessions (14.76 € for actors, and 14.84 € for bystanders), and 13 € in the *NE* sessions (13.29 € for actors, and 12.61 € for bystanders).

4. Theoretical considerations and hypotheses

Contributions

As can be easily seen from equation (1), actors solely motivated by the maximization of their own monetary gains are completely unaffected by the presence of bystanders and free-ride on the public good provision (i.e. choose $g_i = 0$, which implies $G=0$). In that case the actors' as well as the bystanders' payoff is the initial endowment e_A and e_B , respectively. In case all actors are solely motivated by payoff-maximization, contributions of zero prescribe the unique Nash equilibrium of the stage game.

What happens when we relax the assumption of pure money maximization? A growing family of models formalizes deviations from money maximization in the direction of social preferences (for a survey see Fehr and Schmidt 2002) capturing, among others, sensitivity to group efficiency (Charness and Rabin 2002), maximin preferences (Engelmann and Strobel 2004) or esteem (Ellingsen and Johannesson 2008). Since this is what we are introducing, for

our purposes models of inequity aversion are best suited. We prefer (Fehr and Schmidt 1999) (henceforth FS) over (Bolton and Ockenfels 2000), since the former allows for different weights for advantageous and disadvantageous inequality. This permits to generate separate hypotheses for treatments with positive and negative externalities. Since both income and wealth are symmetric in our setting, we have no reason to discuss these potential qualifications (Buckley and Croson 2006).

In the FS model, actors gain disutility both from having a payoff disadvantage and from having a payoff advantage in comparison to others, where the model assumes the disutility from the disadvantage to be greater than the disutility from a payoff advantage:

$$\begin{aligned}
 (3) \quad u_i^A = & \pi_i^A - \\
 & - \alpha_i \frac{1}{n_A + n_B - 1} \left[\sum_{\substack{j=1 \\ j \neq i}}^{n_A} \max\{\pi_j^A - \pi_i^A, 0\} + n_B \max\{\pi^B - \pi_i^A, 0\} \right] \\
 & - \beta_i \frac{1}{n_A + n_B - 1} \left[\sum_{\substack{j=1 \\ j \neq i}}^{n_A} \max\{\pi_i^A - \pi_j^A, 0\} + n_B \max\{\pi_i^A - \pi^B, 0\} \right] \quad \beta_i \leq \alpha_i, 0 \leq \beta_i < 1
 \end{aligned}$$

Thus, the utility of actor i is composed of the actor's monetary payoff π_i^A , reduced by the utility loss from disadvantageous payoff differences, weighted by α_i (second line in (3)), and the utility loss from advantageous payoff differences, weighted by β_i (third line in (3)). In measuring the payoff differences, the actor compares her payoff to the payoffs of all other players, both actors and bystanders.

The unique Nash equilibrium under the assumption of monetary payoff maximization, i.e. zero contributions of all actors, is also an equilibrium in the FS model. In the Nash equilibrium there is neither inequity with respect to the other actors nor with respect to the bystanders. All actors and all bystanders have a payoff of 20.² A unilateral deviation to a strictly positive contribution would create both a lower monetary payoff and disutility from disadvantageous inequity of the deviator, both with respect to the other actors and with

² This is a consequence of the identical endowments of actors and bystanders. Assuming different endowments would create inequity in the Nash equilibrium and might thus lead to different predictions in the Nash equilibrium and the FS equilibrium. To keep the model simple and to align theoretical predictions, we chose symmetric endowments.

respect to the bystanders³. Therefore, for all parameters α and β in the admissible range complete free-riding is also an equilibrium under inequity aversion as modeled by FS. Are there any additional equilibria? The answer is yes, but only for very extreme values of the inequity parameters. Full contribution of all actors (20, 20, 20, 20) may be an equilibrium in both treatments and the control, in case all actors extremely suffer from advantageous inequality, i.e. have very high β values. Then the monetary advantage of deviating to a lower contribution is more than eaten up by the disutility of having a higher payoff. In *PE* this requires $\frac{6}{7} < \beta_i < 1$ and $\beta_i \leq \alpha_i < 8\beta_i - 6$ for all actors i . This prescribes a narrow parameter range, which would for example be satisfied by $\beta = 0.9$ and $\alpha = 1.1$. In this example, the utility of sticking to the contribution of 20 would be 29.8, while the utility of unilaterally deviating to zero would be 29.6. In *NE* an equilibrium with full contributions is possible if all actors have inequity parameters satisfying $\frac{6}{7} < \beta_i < 1$ and $\beta_i \leq \alpha_i$. And, finally in the control treatment full contribution of all actors is in equilibrium if $\frac{3}{4} < \beta_i < 1$ and $\beta_i \leq \alpha_i$ is satisfied for all actors i . Actually, in the control we also find the asymmetric contribution profile (20, 20, 20, 10) to be in equilibrium if $\frac{36}{38} < \beta_i < 1$ and $\beta_i \leq \alpha_i < \frac{1}{5}(43\beta_i - 36)$ holds for the actors contributing 20 and $\frac{3}{4} < \beta_i < 1$ and $\beta_i \leq \alpha_i$ holds for the actor contributing 10. In *PE* and *NE* there is no equilibrium with asymmetric contributions. Nevertheless, according to the empirical estimations of the inequity parameters (e.g. (Fehr and Schmidt 1999:844) and (Blanco, Engelmann et al. 2006)) observing such high β values for all actors, i.e. having a group of actors in which each actor is so upset by having a higher payoff than the others, is almost impossible and thus we expect complete free-riding, even under the assumption of inequity aversion:

H₁ Contributions: Neither in one of the treatments nor in the control are contributions significantly different from free-riding. This is true, even if we allow for other-regarding preferences in the form of inequity aversion as modeled by FS.

Joint payoff maximization

Our public goods game with externalities shares the social dilemma character of the public goods games in which externalities remain unconsidered that – dependent on the parameterization of the game – the prescriptions of individual and joint payoff maximization may be opposed. Suppose that all actors contribute an identical amount $g_i = g$ to the public

³ Notice that this is also true in *NE*. A unilateral deviation to a contribution of 10 (20) leads to a payoff of 14 (8) for the deviator, a payoff of 24 (28) for the other actors and a payoff of 18 (16) for the bystanders. Under complete free-riding all actors and bystanders have a payoff of 20.

good. Then the joint payoff of all actors is $\Pi^A = n_A[e_A - g + n_A ag]$ and the joint payoff of all bystanders is $\Pi^B = n_B[e_B + n_A bg]$. Thus, the joint profit of actors and bystanders is $\Pi = \Pi^A + \Pi^B = n_A e_A + n_B e_B + n_A g[n_A a + n_B b - 1]$. The term in brackets ($E = [n_A a + n_B b - 1]$) determines the marginal productivity of each unit contributed to the entire society of actors and bystanders by adding the marginal return it provides to the n_A actors ($n_A a$) and the marginal return it provides to the n_B bystanders ($n_B b$), minus the marginal cost of provision (i.e. 1). Obviously, the sign of E determines which contribution g^{SO} maximizes the joint profit Π of all actors and bystanders. It is either one of the boundary values ($g^{SO} = e_A$ for $E > 0$ and $g^{SO} = 0$ for $E < 0$), or, in case of $E = 0$, any contribution leads to the same joint profit. In our parameterization joint payoff maximization demands full contributions of all actors in all treatments, independent of whether the actors strive for maximizing the joint profit of actors alone Π^A or the joint profit of actors and bystanders Π .

Table 2: Model Predictions under Symmetric Contributions⁴

		Symmetric contribution g		
		0	10	20
1. Actors' monetary payoff		20	26	32
2. Bystanders' monetary payoff				
2.1. PE ($b = 0.2$)		20	28	36
2.2. Control ($b = 0$)		20	20	20
2.3. NE ($b = -0.2$)		20	12	4
3. Joint monetary payoff of actors	$\Pi^A = 80 + 2.4g$	80	104	128
4. Joint monetary payoff of actors and bystanders				
4.1. PE ($b = 0.2$)	$\Pi = 140 + 4.8g$	140	188	236
4.2. Control ($b = 0$)	$\Pi = 140 + 2.4g$	140	164	188
4.3. NE ($b = -0.2$)	$\Pi = 140$	140	140	140
5. Actors' utility (according to FS)				
5.1. PE ($b = 0.2$)	$\pi^A - 0.1\alpha g$	20	$26 - \alpha$	$32 - 2\alpha$
5.2. Control ($b = 0$)	$\pi^A - 0.3\beta g$	20	$26 - 3\beta$	$32 - 6\beta$
5.3. NE ($b = -0.2$)	$\pi^A - 0.7\beta g$	20	$26 - 7\beta$	$32 - 14\beta$

⁴ In case of symmetric contributions $\pi^A = e_A - g + aG$, $\pi^B = e_B + bG$ and

$$u_A = \pi^A - \alpha \frac{1}{n_A + n_B - 1} [n_B \cdot \max\{\pi^B - \pi^A, 0\}] - \beta \frac{1}{n_A + n_B - 1} [n_B \cdot \max\{\pi^A - \pi^B, 0\}]$$

Table 2 summarizes the individual and joint monetary payoffs and the actor's utility in the special cases of symmetric contributions.

Conditional cooperation and the risk of being the sucker

Numerous past experiments of finitely repeated public goods games have shown that subjects initially significantly deviate from the equilibrium prescription of free-riding, with average contributions of about 40-60% of the endowment, but that contributions decay over time towards free-riding (Ledyard 1995; Zelmer 2003). This indicates that the equilibrium is not instantaneously reached, but results of a dynamic process. A well established explanation of the dynamics is that the subject population contains a majority of conditional cooperators. Conditional cooperators are happy to cooperate – presumably in order to achieve higher (joint) payoffs – as long as they are sufficiently optimistic that the others will do too. However, due to the nature of the social dilemma, contributors earn lower payoffs than those that free-ride by deviating from cooperation. The unease of being exploited (“being the sucker”) drives contributions down. Every bad experience of low payoffs induces each of the conditional cooperators to individually update her beliefs about the number of cooperators downwards, which leads to a vicious cycle (Keser and van Winden 2000; Brandts and Schram 2001; Fischbacher, Gächter et al. 2001). Recent findings show that the effect is exacerbated by the fact that most conditional cooperators are imperfect in the sense that they have a preference for contributing positively, but less than others, conditional on others contributing (Fischbacher and Gächter 2009). Moreover they only imperfectly update their beliefs in the light of experiences made in previous periods (Fischbacher and Gächter 2009).⁵ These considerations lead to

H₂ Development of contributions over time: Average contributions in initial rounds are above 0 and decline over time.

The different roles of actors and bystanders

Up to now we have assumed that, when actors compare themselves to other players, they do not distinguish between actors and bystanders. This neglects role differences. The considerations presented so far may well be modified by assigning different weights to the payoffs and utilities of other actors and bystanders. Because we are not aware of any

⁵ This is in line with the general finding that experimental subjects have a conservatism bias in updating beliefs (Huck and Weizsäcker 2002).

quantitative measures or formal models for this, we will discuss this problem qualitatively. Actors might perceive themselves as an in-group and bystanders as an out-group. Although the two groups do not compete for resources, in the spirit of the minimal group paradigm (Brewer 1979; Diehl 1990) this could help actors to coordinate internally. This is indeed what has been found in prisoner's dilemma games (Insko, Schopler et al. 1994; Goren and Bornstein 2000; Goren 2001), and in public goods games in particular (Rapoport and Bornstein 1989). The effect can be explained as an instance of group polarization (Doise 1969; Moscovici and Zavalloni 1969; Myers and Lamm 1976; Isenberg 1986). Another effect of acknowledging different roles may be to not (only) compare the collective performance of the community of actors with the collective performance of the community of bystanders, but (also) to individually strive for a higher payoff than bystanders.

In *PE*, collectively actors can at best break even with bystanders. This requires that they all contribute nothing, and forego any internal gains from cooperation. In *NE* and in the control, breaking even with bystanders is the worst possible collective outcome. It happens if neither actor contributes anything. In *PE*, there is a trade off between internal cooperation and the worsening of the external competitive position. In the control and in *NE*, internal cooperation has a double dividend. In *NE*, the externality gives the double dividend extra leverage.

Even if actors care about a competitive advantage over the bystanders, this might not play itself out the same way in all treatments. Only in *PE* and *NE* are actors in a position to actively influence the competitive position of bystanders. This might make the ingroup-outgroup effect more salient, and thereby more pronounced.

H₃ Cooperation in the light of the different roles of actors and bystanders: Actors motivated by achieving a competitive advantage over bystanders should contribute in the control and in *NE*. Contributions should be highest in *NE*.

Voting Behavior

At the beginning of phase 2, all subjects vote for the recommended contribution in phase 2. Since the rule is cheap talk, rational agents are indifferent with respect to its contents. If, however, we slightly relax the rationality assumption and assume that agents believe that some players might abide by the rule, voting gains a strategic component. This provides an incentive to vote for the contribution that yields the highest expected payoff under the veil of uncertainty, i.e. maximizes $\pi^{VU} = \frac{4}{7}\pi^A + \frac{3}{7}\pi^B = 20 + \frac{12}{7}g(0.2 + b)$. Because in *PE* both actors

and bystanders profit from high contributions, it is in their interest to establish a norm of high contributions (see table 3, 1.1.).

Table 3: Payoffs and utilities under the veil of uncertainty

	Symmetric contribution g		
	0	10	20
1. Expected payoff under the veil of uncertainty			
1.1. PE ($b = 0.2$) $\pi^{VU} = 20 + 24g/35$	20	188/7	236/7
1.2. Control ($b = 0$) $\pi^{VU} = 20 + 12g/35$	20	164/7	188/7
1.3. NE ($b = -0.2$) $\pi^{VU} = 20$	20	20	20
2. Expected utility under the veil of uncertainty			
2.1. PE ($b = 0.2$) $u = 20 + \frac{2}{35}g(12 - \alpha)$	20	$\frac{1}{7}(188 - 4\alpha)$	$\frac{1}{7}(236 - 8\alpha)$
2.2. Control ($b = 0$) $u = 20 + \frac{6}{35}g(2 - \beta)$	20	$\frac{1}{7}(164 - 12\beta)$	$\frac{1}{7}(188 - 24\beta)$
2.3. NE ($b = -0.2$) $u = 20 - \frac{2}{5}\beta g$	20	$20 - 4\beta$	$20 - 8\beta$

Hence we expect a high proportion of voting outcomes of 20 in *PE*. In the *NE* treatment, actors benefit from high contributions, while bystanders dislike contributions to the public good. However, all three possible contribution norms yield the same expected payoff (see table 3, 1.3.). In the control, actors profit from high contributions, while bystanders are unaffected. The expected payoff under the veil of uncertainty increases in contributions (see table 3, 1.2.), which is why we expect a high proportion of voting outcomes 20 in the control.

If we again assume that subjects are inequity averse, they do not solely decide on the basis of the expected payoff, but of the expected utility, which incorporates the disutility out of advantageous and disadvantageous inequity. Under the veil of uncertainty a player has to consider the potential utility loss or gain out of inequality when becoming an actor (with respect to the bystanders) and the potential utility loss or gain out of inequality when becoming a bystander (with respect to the actors) and weight these losses and gains with the corresponding utilities, i.e. $u(\pi^{VU}) = \frac{4}{7}u_A + \frac{3}{7}u_B$. Results are presented in table 3, part 2. Under the usual assumptions on the parameters α and β , a contribution norm of 20 yields the highest expected utility in *PE* (for $\alpha < 12$) and in the control (for $\beta < 2$), whereas a contribution norm of 0 yields the highest expected utility in *NE* (for $\beta > 0$).

Instead of assuming that subjects maximize their expected payoff or utility, subjects could reason that, when becoming a bystander in *NE*, they would want the contribution to be 0, and when playing as an actor they would want the contribution to be 20. Weighting the vote of 0 with the likelihood of becoming a bystander and the vote 20 with the likelihood of becoming an actor results in a vote of 10 ($\frac{4}{7} \cdot 20 + \frac{3}{7} \cdot 0 = 11.43$). In *PE* this reasoning would obviously result in a vote of 20. In the control treatment, the voting outcome is irrelevant for future bystanders. Future actors benefit from a high voting outcome. Since bystanders are indifferent, this rule predicts a clear vote of 20. Under this rule, there is no difference in votes between the positive externality and the control treatment.

H₄ Voting behavior: In the positive externality treatment and in the control we predict votes of 20. In the negative externality treatment, we predict votes of 0 or 10.

Contributions in phase 2

In the experimental instructions, we made it absolutely clear that the norm is neither binding nor enforceable. Money maximizing actors should therefore not make any contribution, whatever the norm is. The same holds true for actors with FS preferences. From the perspective of money maximizing actors, the “norm” is just cheap talk. Experimentally, in multiple player public goods cheap talk has been shown to be even counterproductive (Wilson and Sell 1997) or ineffective in some settings (Bochet, Page et al. 2006), and highly elusive (Güth, Levati et al. 2007) or volatile in others (Palfrey and Rosenthal 1991), while it has been effective in two player coordination games (Clark, Kay et al. 2001). Face to face communication as well as verbal communication in a computerized chat room has a positive effect though (Cason and Khan 1999; Belianin and Novarese 2005; Bochet, Page et al. 2006).

Our setting only allows for the expression of desired contribution levels, and four actors interact which – in light of the previous literature – speaks against a positive effect of communication on contributions. It seems, however, plausible to assume that at least some actors, at least in early rounds, interpret the norm as a social expectation, and interpret their participation in the norm creation process as a commitment. Moreover, at least some actors are likely to aim at being consistent with their individual votes. Hence our predictions for votes and voting outcomes should matter for contributions in phase 2. Comparing with phase 1, contributions should accordingly be higher in the *PE*, possibly also in the control treatment, but they should be lower than phase 1 contributions in the *NE* treatment.

H₅ Contributions in phase 2: In the control and in the *PE* treatment contributions in phase 2 are higher than in phase 1; they are lower in the *NE* treatment.

The fact that the norm is neither binding nor enforceable should make norm-abiding imperfect. The temptation to cheat should be the higher the more costly norm abiding behavior is for the respective actor. We hence expect norm-abiding behavior to be more pronounced if the norm is 10, rather than 20; if the population has voted for 0, obviously only positive norm violations are conceivable. These primary effects should entail a secondary effect on conditional cooperators. Actors should change their beliefs about cooperation rates in the population (Fischbacher and Gächter 2009). To the extent that norm-abiding is driven by this secondary effect, it decays over time. If the decay mainly results from this secondary effect, it should be similar in all three treatments. These considerations lead to

H₆ Norm abidance in phase 2: The higher the norm is, the higher are contributions. However, if the norm is 20 or 10, norm-abidance is imperfect and it is the lower the higher the norm is. Norm abidance decays over time and decay is similar in all treatments.

Contributions in phase 3

In a society of money maximizing actors, the norm is irrelevant in the first place and the fact that the norm is “abolished” in phase 3 should therefore not change behavior. The same holds for actors with FS preferences. If there is a commitment effect (either of the voting outcome or of the individual vote) in phase 2, it should vanish or it should at least be substantially weakened in phase 3. This primary effect of our phase manipulation should entail a secondary effect in conditional cooperators. They should expect less cooperation. We accordingly expect contributions in phase 3 to be lower than in phase 2, and since we again expect conditional cooperation to matter, contributions should also decay over time.

H₇ Contributions in phase 3, compared to phase 2: In phase 3, contributions should not be influenced by the voting outcome in phase 2. Within phase 3, contributions decay over time.

5. Results

The analysis of our data will be based both on non-parametric tests over the 27 independent subject groups (9 in each of both treatments and 9 in the control) and on regression analyses. The model requirements for the regression analyses are explicated in box B.1 in Appendix B.

Contributions in phase 1

We start our data analysis by looking at the contributions in phase 1. Figure 1 displays the average contributions per period and treatment and shows significantly positive contributions in all periods.⁶ Looking at treatment differences in phase 1, *NE* seems to induce higher contributions in the first 5 rounds, whereas *PE* and control look alike. There is indeed a weakly significant difference between *PE* and *NE* and between the control and *NE* in the first 5 rounds⁷ whereas we find no difference between the treatments when taking the data of the entire first phase⁸. Both treatments and the control display a decreasing trend in contributions. In fact, a regression analysis displays a significant negative effect of the period on the contribution (see table B.1, for details). Thus, the observed dynamic behavior is well in line with what was observed in numerous previous experiments, supporting our hypothesis H_2 (while rejecting H_1).

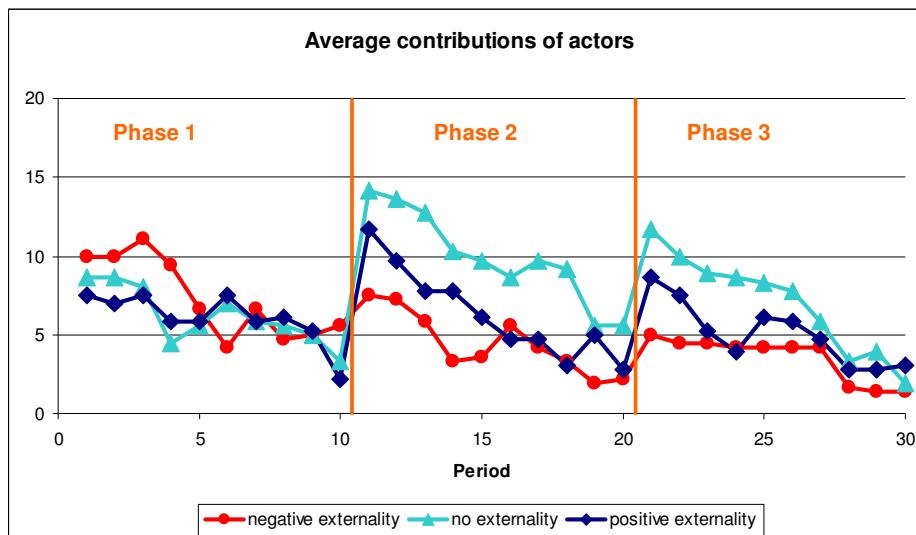


Figure 1 Average contributions

Additionally, the regression shows that contributions are significantly influenced by the treatment. It detects a negative effect of the control treatment, and an even more pronounced negative effect of the *PE* treatment, as compared to the *NE* treatment. This is well in line with

⁶ Since 0 is the lower limit of our support, we cannot directly test the hypothesis that contributions are 0. We can, however, approximate this by testing the hypothesis that contributions are 0.1. In each treatment and phase separately we can reject the hypothesis of (close to) zero contributions (two-sided one-sample sign rank test over the 9 independent subjects groups per treatment, largest p-value .0077).

⁷ Contributions in *NE* are weakly significantly higher than in *PE* (Mann-Whitney u-test over the 18 independent subjects groups, $p=0.083$, two-sided), and higher than in the control ($p = 0.063$).

⁸ A Kruskal Wallis test comparing mean contributions per group in the first phase is indeed insignificant ($N = 27$, $p = .316$). Pairwise Mann-Whitney u-tests between the treatments and the control are also insignificant, even at a significance level of $p < 0.1$.

the group formation argument in **H₃**. From a policy perspective, this is troubling news. Knowing that this hurts outsiders not only does not prevent insiders from collaborating. It even fuels internal cooperation.

Voting Behavior

Table 4 reports the mean votes per treatment, irrespective of the subject's role (total) and separated by the subject's role in phase 1. Pooling over the roles in phase 1, average votes in *NE* are significantly lower than in control⁹ and significantly lower than in *PE*¹⁰. Control and *PE* do not display any significant differences in average votes¹¹. This is in accord with **H₄**. Hence, the endogenous norm formation internalizes the negative externality, but does not seem to internalize the positive externality.

Table 4 Mean votes per treatment and role in phase 1

	control	PE	NE
total	14.921 (.812)	14.286 (.774)	8.571 (.929)
bystander in phase 1	15.185 (1.237)	14.074 (1.224)	7.037 (1.672)
actor in phase 1	14.722 (1.090)	14.444 (1.011)	9.722 (1.015)

(standard errors in parenthesis)

To answer the question whether this effect is mainly driven by the previous bystanders, we separate the votes by the subject's role in phase 1. Figure 2 allows a closer look at the frequency distribution of votes, separated by the role in phase 1. In *PE* as well as in the control, the previous role does not affect voting behavior¹². In *NE* the picture is completely different. Previous bystanders vote significantly lower than previous actors¹³. Does this mean that the whole effect of lower votes in *NE* is driven by the previous bystanders? The answer is no! Comparing *PE* and *NE*, we find that also previous actors vote significantly lower in *NE* than in *PE*¹⁴. This is also true when comparing control to *NE*¹⁵. The lower votes in *NE* are thus not only driven by the experience of receiving a low payoff. Former actors also protect

⁹ Mann Whitney u-test, N = 18, p = 0.0003, two-sided

¹⁰ Mann Whitney u-test, N = 18, p = 0.0005, two-sided

¹¹ Mann Whitney u-test, N = 18, p = 0.5247, two-sided

¹² In both treatments: Fisher's exact test, N = 63, p = 1.000; we use this test, instead of means per group and role in phase 1, since the latter test produces spurious differences: while the means differ, at the level of individual votes, there is no significant difference.

¹³ Fisher's exact test, N = 63, p = 0.001.

¹⁴ Fisher's exact N = 54, p=0.001 for bystanders and N = 72, p=0.006 for actors.

¹⁵ Fisher's exact N = 54, p=0.001 for bystanders and N = 72, p=0.003 for actors.

themselves against the risk of becoming bystanders in the future. Comparing *PE* and control we find that neither previous bystanders nor previous actors vote significantly differently¹⁶. Thus, the positive externality is neither internalized in the votes of previous actors nor in the votes of previous bystanders.

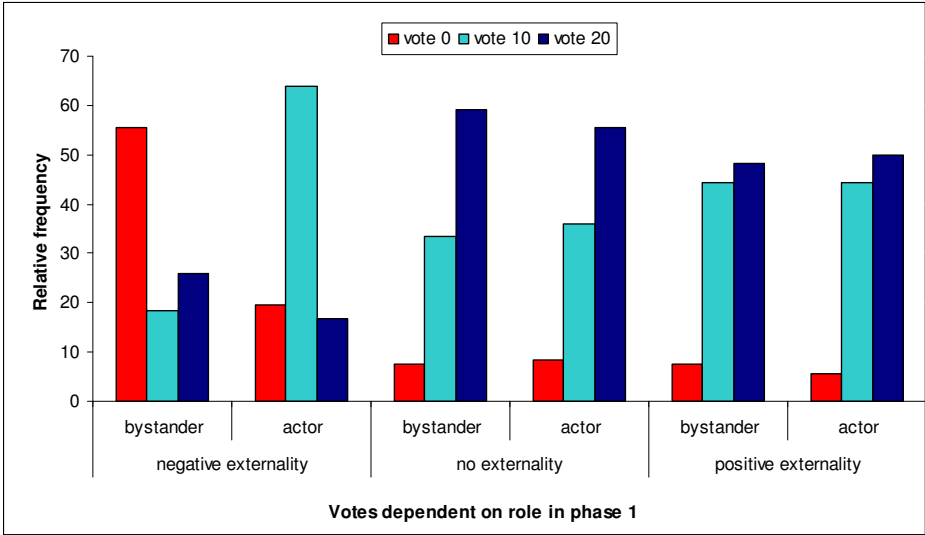


Figure 2 Votes per treatment and role in phase 1

The majority voting rule leads to the voting results as displayed in figure 3. Voting outcomes in *NE* are significantly lower than in control¹⁷ and significantly lower than in *PE*¹⁸. Obviously, there is no difference in the voting results of *PE* and control.

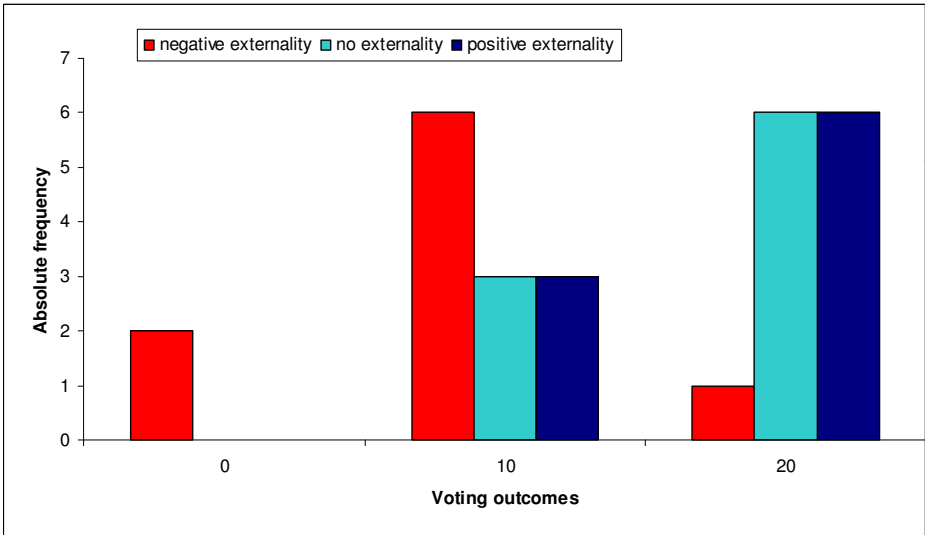


Figure 3 Voting outcomes per treatment

¹⁶ Fisher's exact N = 54, p=0.831 for bystanders and N = 72, p=0.758 for actors

Contributions in Phase 2

Looking at the display of phase 2 contributions in Figure 1, we see that average contributions increase with the introduction of the norm in the control and in *PE*, while average contributions decrease in the *NE* treatment. Comparing just the first five periods of each phase (i.e. periods 1-5 to periods 11-15) the difference is significant in all three cases¹⁹. A comparison of the average contributions in the entire phase 1 to the entire phase 2 (i.e. periods 1-10 to periods 11-20) shows a (weakly) significant difference for the control and for the *NE* treatment, but is insignificant for the *PE* treatment²⁰. Thus the endogenously formed contribution norm significantly increases contributions in the control, significantly dampens contributions in *NE* and only initially increases contributions in *PE*. This is only partial support for **H₅**, because we did not expect the very weak effect of normativity in the positive externality treatment.

The most striking observation, however, is that both forms of externalities lead to lower contribution levels compared to the control and that the contributions in *PE* and *NE* are not significantly different²¹. This is particularly startling since the norm does not help when it would be socially most beneficial, namely when contributions have a positive externality.

Abiding by the norm

Let us dig deeper into the data to analyze the potential causes of this striking observation. On the aggregated level we identify two important causes. The first one is the already discussed observation that the voting outcomes in the positive externality treatment do not differ at all from those of the control, which means that the positive externality was not internalized in the same way as the negative externality. The second one concerns norm compliance. To which degree are contributions influenced by the voting outcome? Regression analysis demonstrates that voting outcome significantly influences contributions in phase 2²². This supports **H₆**, stating that higher norms induce higher contributions. However, adding the treatment effects to the regression²³ one only finds a significant effect of control, while the effect of treatment

¹⁷ Mann Whitney u-test, two-sided, N = 18, p=0.0128

¹⁸ Mann Whitney u-test, two-sided, N = 18, p=0.0128

¹⁹ Two-sided Wilcoxon matched-pairs signed-ranks test: N = 9, control p=0.009; *PE* p=0.0147; *NE* p=0.0118

²⁰ Two-sided Wilcoxon matched-pairs signed-ranks test: N = 9, control p=0.0502; *PE* p=0.3726; *NE* p=0.0502

²¹ Two-sided Mann-Whitney u-test: N = 18, control vs. *NE* p=0.0169; control vs. *PE* p=0.0575; *PE* vs. *NE* p=0.1112. Also in a regression, we do not establish a significant difference between *PE* and *NE* (Table B.2, App. B). Note that *NE* is the reference category, so that the regressor for *PE* directly compares both treatments.

²² See results of the Tobit regression in table B.2 in appendix B, model 1

²³ See results of the Tobit regression in table B.2 in appendix B, model 2

PE, after having controlled for voting outcomes and individual votes, is not significantly different from zero. This shows that the dependence on voting outcomes and individual votes cannot be the complete story because the distribution of voting outcomes and votes in PE and in the control is not significantly different.²⁴ The missing link is norm compliance.

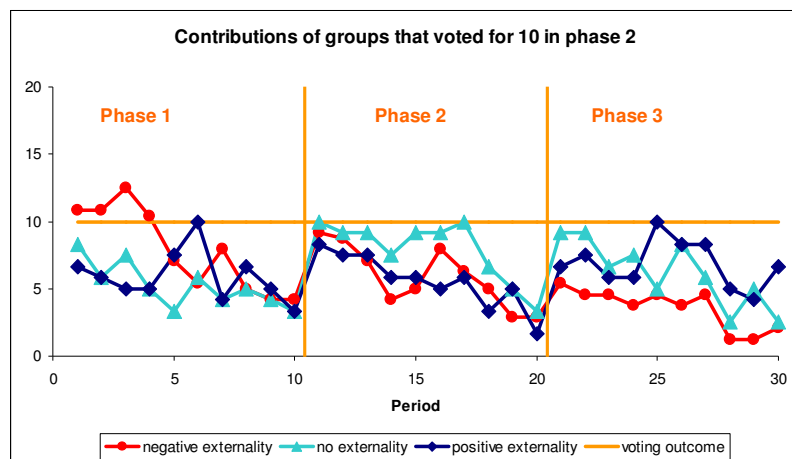


Figure 4 Norm compliance for contribution norm 10

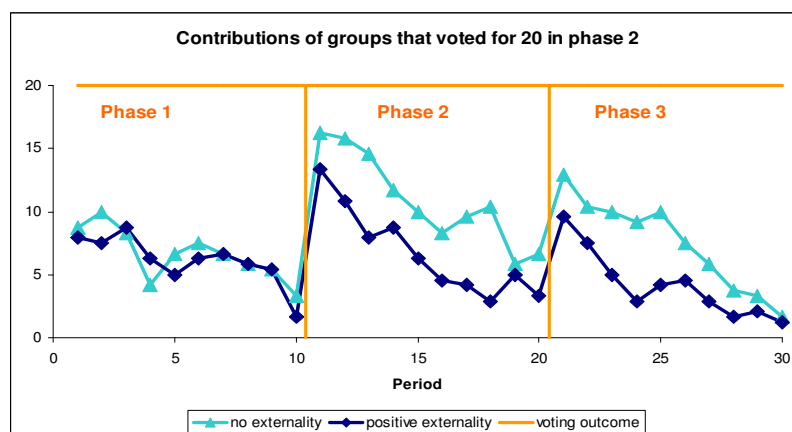


Figure 5 Norm compliance for contribution norm 20

Figure 4 and figure 5 display the contributions of those groups with voting outcomes of 10 and 20, respectively.²⁵ The figures show that, on average, groups contribute less than pledged²⁶. Norm abidance is best if the voting outcome is 0, worse if the outcome is 10, and

²⁴ Model 3 of the regression in table B.2 shows that the player's contribution in phase 2 is not significantly explained by the player's role in phase 1.

²⁵ Since only a single group voted 20 in NE, we drop this observation from the graph.

²⁶ Wilcoxon matched-pairs signed-ranks test for voting outcome 10, testing H_0 that there is no deviation from the voting outcome: $N = 12$, $p = 0.0022$; Wilcoxon matched-pairs signed-ranks test for voting outcome 20: $N = 13$, $p = 0.0015$

the worst if it is 20²⁷. But this is not yet the complete story. Holding the voting outcome constant we also find treatment differences for the contribution norm 20²⁸. In the control, participants obeyed this norm significantly better than in *PE*. The immediate impression that in both cases norm abidance decays over time can also be supported from the significant negative effect of period in both regressions. Additionally, we can show that the pattern of decay does not differ between treatments²⁹.

The analyses presented so far provided a better understanding of why phase 2 contributions in both externality treatments are lower than in the control. In *NE* subjects (previous actors as well as previous bystanders) vote for lower contribution norms, as compared to *PE* and to the control. Since the abidance of lower norms turns out to be quite high, contributions lower than those of the control result. But what happened in *PE*? We learned that subjects casted votes for slightly lower norms than in the control and that voting outcomes are identical to those in the control. The high contribution norm 20 has a low compliance rate, which is even lower if it is achieved in *PE*. These findings nicely round up the picture. Yet, the nature of the individual behavior leading to these processes remains opaque and its study will be the focus of the following analyses.

Contribution dynamics, conditional cooperation and the aversion of being the sucker

How to explain the individual changes in contributions? Conditionally cooperative actors update their belief about the number of other (conditional) cooperators and adjust their contributions accordingly. In our public goods game the own payoff is directly related to the contributions, such that the most basic form of conditional cooperation would predict to increase the own contribution in case past profit was high, and to decrease it otherwise. We test this conjecture in a regression³⁰ and find that the past period's profit indeed has a highly significant positive effect on contribution change (cf. Table B.4.2 regression model 1). The higher the actor's past profit, the more the actor increases her contribution. High actor profit indicates that the other actors have made substantial contributions. This prompts a

²⁷ Jonckhere Terpstra, N = 27, p <0.001

²⁸ See the regression in Table B.3 in appendix B

²⁹ Two-sided Mann Whitney u-test *PE* v. *NE*, N= 18, p = .2660; control v. *NE* p = .8243; control v. *PE* p = .6896. The dependent variable is normabiding in period t minus normabiding in period t+1. The tests use means of this variable, over periods 11 – 19, per group. Tests remain insignificant if we use all available observations per comparison, i.e. if N = 648.

³⁰ See table B.4 in appendix B. The table contains three regressions, one for each of the three phases. Remarkably, the regression coefficients as well as the significance levels are very similar in all phases. Indeed, a regression combining all three phases yields also very similar parameters.

conditionally cooperative actor to adjust her belief about the number of (conditional) cooperators upwards and to increase her contribution.

The same dependence can be found if the past period's utility (in the sense of FS, with average parameters proposed in this paper, i.e. with $\alpha = 0.85$ and $\beta = 0.315$) instead of the past period's profit is used as an independent variable (cf. regression model 2). The substantially higher R^2 shows that experienced FS utility explains the dynamics even better than past profit. The explanation with the past period's utility not only enriches the payoff comparison with respect to the other actors (in regression model 1) with a consideration of disutility from inequity with respect to other actors, but also incorporates a comparison to the bystanders as an additional source of disutility from inequity.

But, how relevant, if at all, is a comparison to the bystanders' profit? For actors solely motivated by conditional cooperation, non-acting outsiders should be irrelevant. They adjust their cooperation level to their belief about the other actors' cooperation level and do not care about outsiders. To study the relevance of the payoff comparison with respect to bystanders, we separate the payoff comparison into the comparison with respect to the other actors and with respect to the bystanders (cf. regression model 3). The result is striking. Both comparisons have a significant influence on contribution change, however, with different signs. Thus it seems that actors compare themselves to the other actors as well as to the bystanders, but their reactions point into opposite directions. The higher the actor's past profit in comparison to the other actors, the more the actor increases her contribution. This reflects the conditionally cooperative relation within the group of actors, as described in model 1. In contrast, however, the difference between the own and the bystanders' profit has a negative effect. If in the subsequent period an actor was better off than the bystanders, the actor reduces her contributions. If the actor was worse off, she increases her contributions.

This finding is best understood if one focuses on the interactions of the variables measuring payoff differences with treatments (c.f. regression model 4). Main effects basically stay unchanged. They capture behavior in *NE* (the reference category) and in the control³¹. In both cases, the more actors cooperate, the bigger the payoff advantage of actors over bystanders. Thus, actors seem torn between two motives. They want to conditionally cooperate within the group of actors, but at the same time they hesitate to increase the inequity in relation to

³¹ In phase 2 the control is not significantly different from *NE*. In phases 1 and 3 both main effects are even increased through the interaction effects with the control.

bystanders. The former effect is, however, more than twice as strong as the latter (e.g. phase 2 for *NE*: $1.126 - 0.437 = 0.692$), such that conditional cooperation remains the dominant force.

The results for *PE* are even more revealing. In this treatment, differences between individual actor profit and bystander profit are in most cases negative. They range from -16 to 12, with an average of -1.346. The sum of the negative regressor for the payoff difference between this actor and bystanders and the positive interaction effect is strongly positive (for example in phase 2: -0.437 [main effect] + 3.029 [interaction effect] = 2.592). This coefficient has to be multiplied with the – in most cases negative – difference between individual actor profit and bystander profit. Thus, the regression essentially suggests that, in *PE*, actors are strongly induced to reduce the windfall profit for bystanders. Actually this desire seems to be so powerful that it trumps conditional cooperation: in *PE* the overall effect of the lagged difference between own and other actors' profit is negative (e.g. in phase 2: 1.126 [main effect] – 2.168 [interaction effect] = -1.042). This means that in case of a positive externality – different from what has been observed in *NE* and in control – conditional cooperation among the actor group does not work. Note, however, that this does not mean that actors free-ride. Free-riders would not be influenced by the difference between their own profit and other actors' profits in the previous period. The interaction effect should neutralize the main effect, i.e. it should be about as strong. Actually, the model predicts that actors substantially *reduce* their contributions if they had made more profit than the other actors.

How can we explain the “break-down” of conditional cooperation among actors when there is a positive externality? The reason is that the windfall profits of the bystanders adumbrate the actors' relationship. High actor profits are accompanied by even higher bystander profits and the consequences of “being the sucker” are most fatal, as the following situation exemplifies. Consider the situation that three actors contribute fully, while one actor free-rides. This gives the free-rider a payoff advantage over the other actors which is treatment independent (see table 5). The payoff relation with respect to the bystanders, however, depends on the treatment. In *PE* the contributing actors not only have a lower payoff than the free-rider but also than the bystanders. This is not only true in monetary payoffs, but even more so in utility, if one applies the FS model with the proposed parameters $\alpha = 0.85$ and $\beta = 0.315$ (see table 5). Actually, both forms of externality decrease the sucker's utility compared to the control.

Table 5: Cooperation and deviation

	<u>Scenario 1:</u> all actors contribute 20	<u>Scenario 2:</u> 3 actors contribute 20 (the <i>suckers</i>) and one actor free-rides by contributing 0		
1. Payoffs	actors	free-rider	contributors	bystanders
PE	32	44	24	32
control	32	44	24	20
NE	32	44	24	8
2. Utilities	actors	free-rider	contributors	bystanders
PE	30.30	38.96	17.77	29.04
control	30.11	37.07	20.54	14.90
NE	27.59	35.18	18.65	-3.90

But what are the crucial aspects in the actors' comparison with respect to the bystanders? Do actors compare the whole group of actors to the whole group of bystanders, or do they compare their individual payoff to that of bystanders? Regression model 5 shows that both effects are indeed significantly present. More interestingly even, if we do not control for treatment differences, actors do not care about collective performance. If one controls for the difference between individual profit and bystander profit, the more actors have collectively outperformed bystanders, the more individual actors *reduce* their contributions in the subsequent period. If they can, actors free-ride on other actors' efforts to improve the group's collective position. In the second phase, as long as the individual difference between this one actor and bystanders has been less than 1.38 times as large as the collective difference³², the regression model predicts that the actor reduces her contribution.

Regression model 6 confirms this picture both for *NE* and for the control. In this model, main effects of both difference regressors are as in regression model 5. *NE* is the reference category. The interactions between individual and collective differences on the one hand and control are insignificant. Seemingly though, *PE* is different. The interaction between collective differences and *PE* is significant. It has the opposite sign of the main effect, and it is about twice as strong as the main effect. However, to appreciate this finding, one has to take the range of the independent variable into account. In *PE*, collective differences are bound to be negative or, at best, zero. In phase 2, they range from -3.5 to 0, with mean -1.292. Taken together, the main effect of collective differences (-1.343) and its interaction with *PE*

³² 1.115 [lagged difference between own and bystander profit] * 1.38 - 1.535 [lagged difference between average actor and bystander profit] = 0

(2.814) result in a positive regressor that is even stronger than the main effect (1.471). Given that the independent variable is always negative in this treatment, we learn that in *PE* actors do not care about collective differences either. Actually they do even free-ride on other actors' efforts more strongly than in *NE* and the control.

Remarkably, the contribution dynamics is almost identical in all three phases. Neither the magnitude of the parameters nor their significance differs considerably between phases.³³ This means that the introduction of a contribution norm yields treatment dependent level effects but does not influence the dynamics of the contribution process. The forces driving the contribution process are payoff comparisons with the remaining actors and with bystanders, both at the group level and at the individual level.

Contributions in Phase 3

Finally, let us investigate contributions in the final phase. The instructions for phase 3 formulated that the recommended contribution of phase 2 is no longer in force. Did it nonetheless guide behavior? Was there an indirect influence through the experiences subjects made in phase 2 while the norm was "in force"? In a regression we find a significant positive effect of the previous norm, if we control for treatment effects³⁴. Interestingly, the individual vote in phase 2 is insignificant for contributions in phase 3. Phase 3 actors who have been bystanders in phase 2 are not significantly influenced in their phase 3 contributions by phase 2 experiences, neither in terms of this player's own profit, nor in terms of average contributions in her group. However, both matter strongly for those who were actors in phase 2 and who are again actors in phase 3 (regression models 5 and 6). Interestingly, the coefficient for the group's average contributions is positive, while the coefficient for the own previous profit is negative. Apparently, previous profit says more about a (greedy) type than about the lasting effects of experiences. Note that in a regression model that includes both measures for phase 2 features voting outcome is no longer significant (regression model 6). This demonstrates that there is indeed no spillover of normativity. The abolished norm only matters indirectly since it has induced actors to make favorable experiences. This is support in favor of **H7**.

³³ The only noticeable differences are in the respective models 4. While the interaction between the lagged difference between own and other actors profit is insignificant in phase 2, it is weakly significant in phase 3, and it is significant at conventional levels in phase 1. Conditional cooperation is thus even more pronounced in the control, in these two phases. Also while the interaction between the lagged difference between own and bystander profit with control is insignificant in phase 2, it is significant in the two remaining phases. Interestingly, if the norm is *not* in force, in the control actors feel a stronger urge to reduce the inequity in relation to the bystanders. Apparently they interpret the norm as justifying the inequity.

³⁴ See models 4 and 5 of Table B.5 in Appendix B. The models 1, 2, and 3, which will not be explicitly discussed here, show that neither treatment, nor voting outcome, nor vote in isolation can explain the effects.

Joint profits and distribution of profits

When looking at the problem from a policy perspective, we adjust our focus from individual contributions to realized joint profit and its distribution. In the control the introduction of the norm significantly increases the efficiency level (i.e. the realized joint profit of all actors and bystanders relative to the maximal achievable level)³⁵ from 82% to 87% (c.f. table 6)³⁶. Since the bystanders' profit is constant, the entire efficiency gain is reaped by the actors and the distribution between actors and bystanders becomes more uneven. In *PE* actors are unable to use the norm for increasing efficiency³⁷ as well as their profit³⁸ or the bystanders' profit³⁹. As a result, in *PE*, the difference between actors' and bystanders' profits is not significantly affected by the norm⁴⁰. The efficiency achieved in *PE* is significantly lower than in the control in all three phases⁴¹. Given our parameters, *NE* is a zero sum game (see Table 2, 4.3), which is why efficiency is meaningless. We can, however, analyze the distribution of profits per group of players. In *NE*, the norm significantly reduces actor profit⁴², increases bystander profit and thus weakly significantly reduces the spread between those⁴³.

Table 6: Average profits and efficiency

treatment	phase	actor profit	bystander profit	joint profit	joint profit in percent of maximal achievable
<i>PE</i>	before	23.63	24.84	24.15	71.64%
	norm	23.80	25.07	24.34	72.20%
	after	23.03	24.04	23.47	69.61%
control	before	23.72	20	22.12	82.38%
	norm	25.95	20	23.40	87.13%
	after	24.22	20	22.41	83.44%
<i>NE</i>	before	24.40	14.13	20	100%
	norm	22.68	16.42	20	100%
	after	22.10	17.20	20	100%

³⁵ Because in *NE* joint profit is always constant and thus efficiency always equals 100%, we will restrict the focus on distributional issues in *NE*.

³⁶ Wilcoxon signed rank test, N = 9, comparing group profit in phases 1 and 2, p = 0.0438

³⁷ Wilcoxon signed rank test, N = 9, comparing group profit in phases 1 and 2, p = 0.3726.

³⁸ Wilcoxon signed rank test, N = 9, comparing actor profit in phases 1 and 2, p = 0.3726.

³⁹ Wilcoxon signed rank test, N = 9, comparing bystander profit in phases 1 and 2, p = 0.3726.

⁴⁰ Wilcoxon signed rank test, comparing the difference of actors and bystanders profits of phases 1 to 2, p=0.3726.

⁴¹ Mann Whitney, the dependent variable is percent of maximally achievable efficiency, N = 18, phase 1: p = 0.0005; phase 2: p = 0.0003; phase 3: p = 0.0009.

⁴² Wilcoxon signed rank test, comparing actor profit in phases 1 and 2, N = 9, p = 0.0502.

⁴³ Wilcoxon signed rank test, comparing group profit: phase 1 vs. 2, p = .0578; phases 1 vs. 3: p = .0077.

6. Conclusions

Public goods frequently spill over to outsiders by either bestowing a windfall profit or by inflicting harm on them. In this paper we show that the interaction of externalities and normativity significantly influences public goods provision. Without an explicit norm, externalities do barely have an effect. An agreed upon contribution norm – although not binding – significantly increases contributions in the case of no externality. Remarkably, however, both forms of externality – positive as well as negative – lead to lower contributions than without an externality. This result might have been expected in the negative externality case; the fact that it is present in the positive externality case is disturbing because social welfare gains would have been highest in this case. Endogenously formed norms internalize the negative externality, but fail to do the same with the positive externality. Moreover, norm compliance is – *ceteris paribus* – much lower in case of a positive externality than in the absence of an externality.

We explain these findings by the interaction of two effects: conditional cooperation and, at the same time, the concern for comparative performance. In the absence of an externality and in the case of a negative externality internal cooperation pays a double dividend in that it additionally gives the group of actors a competitive advantage over the group of bystanders. If however internal cooperation gives outsiders a windfall profit, as is the case in the positive externality, conditional cooperators not only risk falling behind the other actors but also behind the passive bystanders. Individual payoff comparisons with respect to the other actors as well as individual bystanders drive contributions down. Presumably, the anticipation thereof in conjunction with fairness considerations is causal for the relatively low norms in the positive externality treatment. A norm that induces insiders to contribute less when the public good imposes a negative externality on outsiders serves two fairness norms at a time: it protects the needy, and it creates greater equality. By contrast, a norm that induces insiders to contribute more when the public good creates a positive externality for outsiders hurts two fairness norms at a time: it violates the fairness principle of just desert, and it creates greater inequality.

Interestingly, the introduction of the norm has a strong effect on the level of cooperation, in particular in the control and *NE*, but it barely affects interaction dynamics. In all phases, actors respond to experienced cooperativeness of the other actors, but at the same time they aim at reducing the inequity in relation to the bystanders. If this inequity is to their disadvantage, as mostly in *PE*, the latter desire seems so strong that it trumps the beneficial

effect of conditional cooperation. If it is to their advantage, as in *NE* and in the control, it just dampens the willingness to further increase contributions in light of good experiences. These findings reflect the spirit of the Fehr-Schmidt model: players do not only care about detrimental, but also about beneficial inequity; however the sensitivity to detrimental inequity is much stronger.

One must, of course, be cautious when deriving recommendation for institutional design from lab results; context may well suggest otherwise. That said, the implications of our findings differ by the type of externality. Just knowing that internal cooperation inflicts harm on outsiders is not enough. If political intervention made the effect salient, this might even be counterproductive. However if outsiders are given a say on insider behavior, the resulting norms are fairly effective in realigning interests, even absent any enforcement technology. One prominent economic example of a negative externality is collusion in oligopolies. Our results suggest that collusion in oligopolies is not just another instance of cooperation in a public good under a different framing, but strongly influenced by the interaction of the negative externality and normativity. In many markets, one market side is fairly concentrated, while the opposite market side is diffuse, as in consumer markets. Then socially beneficial norms are unlikely to form endogenously. Antitrust intervention is inevitable. Yet our results suggest that the expressive function of antitrust law matters. Even if antitrust is frequently unable to enforce its precepts, it still has some influence since it stigmatizes anticompetitive behavior.

If internal cooperation leaves outsiders unaffected, initially norms are very powerful. However in such situations, cooperation quickly decays. If there is no enforcement mechanism, the norm loses its impact much faster than with a negative externality. This too might have a policy implication. In many areas of law, the probability of enforcement is low. In expected values, law enforcement is far from sufficient to neutralize the incentive to break the law. If initially a substantial portion of the population abides by the norm, such weak forms of intervention might nonetheless suffice. Yet, in such situations mere normativity would be too little. The expressive function of law carries less weight than with negative externalities. Enforcement must at least be frequent and visible enough to make conditional cooperators believe that the norm guides behavior.

While a norm is in force, normativity seems to work on a primary and on a secondary channel. Some individuals seem to feel bound by the norm, or to feel committed to their individual normative statement, i.e. their vote. Others seem happy to cooperate as long as they

expect a sufficiently large portion of the population to do so as well. Once the norm is abolished, the primary effect vanishes. Yet the secondary effect persists. The fact that a group has successfully managed to cooperate while the norm was in force still biases beliefs of conditional cooperators upwards. Here the policy implication is mixed. Initially, such hysteresis is good news. In particular, temporary doubts about validity do not destroy the socially desirable effect on behavior. Yet for the administration of the law, the effect can be dangerous. Those in charge of upholding the norm may be led to erroneously believe that they can lower their efforts. Once beliefs of previous conditional cooperators have changed, intervention must be much heavier handed to restore the expectation that the norm will be widely governing behavior.

Finally, if internal cooperation gives outsiders a windfall profit, normativity is least successful, despite the fact that it would be socially most beneficial. The endogenously generated normative expectation is already below the socially desirable level. Addressees are least likely to abide by the norm. If internals shall be induced to cooperate to the benefit of outsiders, legal intervention must be heavy handed. Internals must be forced to accept a sufficiently demanding norm, and they must be forced to comply with the norm once it has been introduced. The particularly strong need for sovereign intervention may explain why, in legal practice, an alternative solution is more frequent: organization above the level of sovereign states. Federal states like the US, economic zones like the EU, alliances like the NATO, international organizations like the UN, need not replace their members to be justified. It may well suffice, but also be necessary, to prevent the populations of some member states from free riding on other populations' efforts and endurance. If the alliance sufficiently reduces the risk of some to be the sucker, chances are transnational public goods will be provided without direct central intervention. When the provision of public goods spills over to outsiders, inclusion seems to be a most powerful remedy.

References

- ABBINK, KLAUS (2005). Fair Salaries and the Moral Costs of Corruption. Advances in Cognitive Economics. Boicho N. Kokinov. Sofia, NBU Press: ***.
- ABBINK, KLAUS, JORDI BRANDTS, BENEDIKT HERRMANN and HENRIK ORZEN (2009). "Inter-Group Conflict and Intra-Group Punishment in an Experimental Contest Game." American Economic Review **99**: ***.
- ANDERSEN, STEFFEN, ERWIN BULTE, URI GNEEZY and JOHN A. LIST (2008). "Do Women Supply More Public Goods than Men? Preliminary Experimental Evidence from Matrilineal and Patriarchal Societies." American Economic Review: Papers and Proceedings **98**: 376-381.
- ANDERSON, LISA R., JENNIFER M. MELLOR and JEFFREY MILYO (2004). "Social Capital and Contributions in a Public Goods Experiment." American Economic Review: Papers and Proceedings **94**: 373-376.
- ANDREONI, JAMES (1995). "Warm-Glow versus Cold-Prickle. The Effects of Positive and Negative Framing on Cooperation in Experiments." Quarterly Journal of Economics **110**: 1-17.
- BELIANIN, ALEXIS and MARCO NOVARESE (2005). Trust, Communication and Equilibrium Behaviour in Public Goods Game. A Cross-Country Experimental Study <http://129.3.20.41/eps/exp/papers/0506/0506001.pdf>.
- BLANCO, MARIANA, DIRK ENGELMANN and HANS-THEO NORMANN (2006). A Within-Subjects Analysis of Other-Regarding Preferences http://papers.ssrn.com/sol3/papers.cfm?abstract_id=934700.
- BOCHET, OLIVER, TALBOT PAGE and LOUIS PUTTERMAN (2006). "Communication and Punishment in Voluntary Contribution Experiments." Journal of Economic Behavior & Organization **60**: 11-26.
- BOHNET, IRIS, FIONA GREIG, BENEDIKT HERMANN and RICHARD ZECKHAUSER (2008). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States." American Economic Review **98**: 294-301.
- BOLTON, GARY E. and AXEL OCKENFELS (2000). "ERC: A Theory of Equity, Reciprocity and Competition." American Economic Review **90**: 166-193.
- BOLTON, GARY E. and AXEL OCKENFELS (2008). Risk Taking and Social Comparison
- BORNSTEIN, GARY (2003). "Intergroup conflict: Individual, Group, and Collective Interests." Personality and Social Psychology Review **7**: 129-145.
- BRANDTS, JORDI and ARTHUR SCHRAM (2001). "Cooperation and Noise in Public Goods Experiments. Applying the Contribution Function Approach." Journal of Public Economics **79**: 399-427.
- BREWER, MARILYN B. (1979). "In-Group Bias in the Minimal Intergroup Situation. A Cognitive-Motivational Analysis." Psychological Bulletin **86**: 307-324.
- BUCKLEY, EDWARD and RACHEL T.A. CROSON (2006). "Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods." Journal of Public Economics **90**: 935-955.
- CASON, TIMOTHY N. and FEISAL U. KHAN (1999). "A Laboratory Study of Voluntary Public Goods Provision with Imperfect Monitoring and Communication." Journal of Development Economics **58**: 533-552.
- CHARNESS, GARY and MATTHEW RABIN (2002). "Understanding Social Preferences with Simple Tests." Quarterly Journal of Economics **117**: 817-869.
- CLARK, KENNETH, STEPHEN KAY and MARTIN SEFTON (2001). "When are Nash Equilibria Self-Enforcing? An Experimental Analysis." International Journal of Game Theory **29**: 495-515.
- CORNES, RICHARD and TODD SANDLER (1996). The Theory of Externalities, Public Goods and Club Goods. Cambridge, Cambridge University Press.

- CROSON, RACHEL T.A., ENRIQUE FATAS and TIBOR NEUGEBAUER (2005). "Reciprocity, Matching and Conditional Cooperation in Two Public Goods Games." Economics Letters **87**: 95-101.
- CROSON, RACHEL T.A. and MELANIE MARKS (2001). "The Effect of Recommended Contributions in the Voluntary Provision of Public Goods." Economic Inquiry **39**: 238-249.
- DIEHL, MICHAEL (1990). "The Minimal Group Paradigm. Theoretical Explanations and Empirical Findings." European Review of Social Psychology **1**: 263-292.
- DOISE, WILLEM (1969). "Intergroup Relations and Polarization of Individual and Collective Judgements." Journal of Personality and Social Psychology **12**: 136-143.
- ELLINGSEN, THORE and MAGNUS JOHANNESON (2008). "Pride and Prejudice. The Human Side of Incentive Theory." American Economic Review **98**: 990-1008.
- ENGEL, CHRISTOPH (2007). "How Much Collusion? A Meta-Analysis on Oligopoly Experiments." Journal of Competition Law and Economics **3**: 491-549.
- ENGELMANN, DIRK and MARTIN STROBEL (2004). "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." American Economic Review **94**: 857-869.
- FEHR, ERNST and SIMON GÄCHTER (2000). "Fairness and Retaliation." Journal of Economic Perspectives **14**: 159-181.
- FEHR, ERNST and KLAUS M. SCHMIDT (1999). "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics **114**: 817-868.
- FEHR, ERNST and KLAUS M. SCHMIDT (2002). Theories of Fairness and Reciprocity. Evidence and Economic Applications. Advances in Economics and Econometrics. 8th World Congress, Econometric Society Monographs. Mathias Dewatripont und Stephen J. Turnovsky. Cambridge, Cambridge University Press: 208-257.
- FISCHBACHER, URS (2007). "z-Tree. Zurich Toolbox for Ready-made Economic Experiments." Experimental Economics **10**: 171-178.
- FISCHBACHER, URS and SIMON GÄCHTER (2009). "Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Good Experiments." American Economic Review: ***.
- FISCHBACHER, URS, SIMON GÄCHTER and ERNST FEHR (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Economics Letters **71**: 397-404.
- FREY, BRUNO and STEPHAN MEIER (2004). "Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment." American Economic Review **94**: 1717-1722.
- GOREN, HAREL (2001). "The Effect of Out-group Competition on Individual Behavior and Out-group Perception in the Intergroup Prisoner's Dilemma (IPD) Game." Group Processes & Intergroup Relations **4**: 160-182.
- GOREN, HAREL and GARY BORNSTEIN (2000). "The Effects of Intragroup Communication on Intergroup Cooperation in the Repeated Intergroup Prisoner's Dilemma (IPD) Game." Journal of Conflict Resolution **44**: 700-719.
- GREINER, BEN (2004). An Online Recruiting System for Economic Experiments. Forschung und wissenschaftliches Rechnen 2003. Kurt Kremer und Volker Macho. Göttingen: 79-93.
- GÜTH, WERNER, VITTORIA M. LEVATI and TORSTEN WEILAND (2007). Cheap Talk and Secret Intentions in a Public Goods Experiment <https://people.econ.mpg.de/~levati/Cheap%20Talk.pdf>.
- GÜTH, WERNER and ERIC VAN DAMME (1998). "Information, Strategic Behavior, and Fairness in Ultimatum Bargaining. An Experimental Study." Journal of Mathematical Psychology **42**: 227-247.
- HOFFMAN, ELIZABETH, KEVIN MCCABE and VERNON L. SMITH (2008). Reciprocity in Ultimatum and Dictator Games. An Introduction. Handbook of Experimental Economics Results. Charles R. Plott und Vernon L. Smith. Amsterdam, Elsevier: 411-416.

- HUCK, STEFFEN and GEORG WEIZSÄCKER (2002). "Do Players Correctly Estimate What Others Do? Evidence of Conservatism in Beliefs." Journal of Economic Behavior & Organization **47**: 71-85.
- INSKO, CHESTER A., JOHN SCHOPLER, KENNETH A. GRAETZ, STEPHEN M. DRIGOTAS, DAVID P. CURREY, SHANNON L. SMITH, DONNA BRAZIL and GARY BORNSTEIN (1994). "Interindividual-Intergroup Discontinuity in the Prisoner's Dilemma Game." Journal of Conflict Resolution **38**: 87-116.
- ISENBERG, DANIEL J. (1986). "Group Polarization. A Critical Review and Meta-Analysis." Journal of Personality and Social Psychology **50**: 1141-1151.
- KESER, CLAUDIA and FRANS VAN WINDEN (2000). "Conditional Cooperation and Voluntary Contributions to Public Goods." Scandinavian Journal of Economics **102**: 23-39.
- LEDYARD, JOHN O. (1995). Public Goods. A Survey of Experimental Research. The Handbook of Experimental Economics. J.H. Kagel und A.E. Roth. Princeton, NJ, Princeton University Press: 111-194.
- MASCLET, DAVID, CHARLES NOUSSAIR, STEVEN TUCKER and MARIE-CLAIRE VILLEVAL (2003). "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." American Economic Review **93**: 366-380.
- MOSCOVICI, SERGE and MARISA ZAVALLONI (1969). "The Group as a Polarizer of Attitudes." Journal of Personality and Social Psychology **12**: 125-135.
- MYERS, DAVID G. and HELMUT LAMM (1976). "The Group Polarization Phenomenon." Psychological Bulletin **83**: 602-627.
- OSTROM, ELINOR, THOMAS DIETZ, NIVES DOLSAK, PAUL C. STERN, SUSAN STONICH and ELKE U. WEBER, Eds. (2002). The Drama of the Commons. Washington, National Academy Press.
- PAGE, TALBOT, LOUIS PUTTERMAN and BULENT UNEL (2005). "Voluntary Association in Public Goods Experiments. Reciprocity, Mimicry and Efficiency." Economic Journal **115**: 1032-1053.
- PALFREY, THOMAS R. and HOWARD ROSENTHAL (1991). "Testing for Effects of Cheap Talk in a Public Goods Game with Private Information." Games and Economic Behavior **3**: 183-220.
- POTTERS, JAN, MARTIN SEFTON and LISE VESTERLUND (2005). "After You - Endogenous Sequencing in Voluntary Contribution Games." Journal of Public Economics **89**: 1399-1419.
- RAPOPORT, AMNON and GARY BORNSTEIN (1989). "Solving Public Good Problems in Competition between Equal and Unequal Size Groups." Journal of Conflict Resolution **33**: 460-479.
- SONNEMANS, JOEP, ARTHUR SCHRAM and THEO OFFERMAN (1998). "Public Good Provision and Public Bad Provision. The Effect of Framing." Journal of Economic Behavior & Organization **34**: 143-161.
- VAN DIJK, FRANS, JOEP SONNEMANS and FRANS VAN WINDEN (2002). "Social Ties in a Public Good Experiment." Journal of Public Economics **85**: 275-299.
- VERGNAUD, JEAN-CHRISTOPHE, MARC WILLINGER and ANTHONY ZIEGELMEYER (1999). Mécanisme de Contribution Volontaire et Effet de Contexte. Théorie et Évidence Expérimentelle <http://ideas.repec.org/p/ulp/sbbeta/9907.html>.
- WILSON, RICK K. and JANE SELL (1997). "'Liar, Liar ...'. Cheap Talk and Reputation in Repeated Public Goods Settings." Journal of Conflict Resolution **41**: 695-717.
- ZELMER, JENNIFER (2003). "Linear Public Goods. A Meta-Analysis." Experimental Economics **6**: 299-310.

Appendix A.1 Instructions to the subjects in *PE* for phase 1 (translation of the German original)

Instructions

General Information

- At the beginning of the experiment you will be randomly split into **3 groups of 7 members**. During the whole experiment you will only interact with members of your group.
- The experiment consists of **3 phases**. First you will be informed about phase 1. You will learn about the rules of the next phase as soon as the previous phase has been terminated. **Please note:** The decisions you make in one phase **do not affect** the range of possibilities you have at your disposal in any later phases.

Information for phase 1:

- There are **two types** of players: **active** and **passive players**. There are **4 active players** and **3 passive players**. At the beginning of phase 1 it will be randomly determined whether you are an active or a passive player. Your type will remain unchanged for the whole duration of phase 1.
- You play **10 rounds**, every round will have the same structure.
- Each **active** and each **passive** player receives an **endowment of 20 points in each round**.
- **Active players:** Each active player has to decide how many of the 20 points he/she wants to **contribute** to the public good. All points contributed to the public good will be **multiplied by 1.6** and **equally** split among all 4 active players, i.e. for every point contributed to the public good by an active player, every active player receives 0.4 (=1.6/4). Points not contributed to the public good will stay with the player. More precisely, each active player has to choose one of the following three options:
 - **Contribute** 0 points and keep 20 points,
 - **Contribute** 10 points and keep 10 points or
 - **Contribute** 20 points and keep 0 points
- **Passive players:** Passive players cannot contribute to the public good. The payoff of the passive players depends on the contributions of the active players. For each point contributed to the public good by an active player, each passive player receives **0.2 points**.
- **Payoff per round:**

for active players:	$20 - \text{points contributed} + 0.4 \times \text{sum of the contribution of all active players}$
for passive players:	$20 + 0.2 \times \text{sum of the contribution of all active players}$

Example

If the four active players contribute 0, 10, 10 und 20 (arranged by amount), the sum of contributions by all active players is 40 and each active player receives $0.4 \times 40 = 16$ from the joint project. The individual payoffs per round of the **active players** depend on the amounts contributed and are:

- for the player who contributed 0: $20 - 0 + 16 = 36$
- for the player who contributed 10: $20 - 10 + 16 = 26$ and
- for the player who contributed 20: $20 - 20 + 16 = 16$.

The payoff per round for each **passive player** is $20 + 0.2 \times 40 = 28$.

Payoff

Each player receives a base rate of € 4 once. At the end of the experiment the points will be paid in Euro with the exchange rate: **10 points** are **0.15 €**.

Instructions in *NE* differ in the passage describing the passive player, the passage describing the payoff and the example. These parts read as:

- **Passive players:** Passive players cannot contribute to the public good. The payoff of the passive players depends on the contributions of the active players. For each point contributed to the public good by an active player, each passive player receives a deduction of **0.2 points**.
- **Payoff per round:**

for active players:	$20 - \text{points contributed} + 0.4 \times \text{sum of the contribution of all active players}$
----------------------------	--

for **passive players:** $20 - 0.2 \times \text{sum of the contribution of all active players}$

Example

If the four active players contribute 0, 10, 10 and 20 (arranged by amount), the sum of contributions by all active players is 40 and each active player receives $0.4 \times 40 = 16$ from the joint project. The individual payoffs per round of the **active players** depend on the amounts contributed and are:

- for the player who contributed 0: $20 - 0 + 16 = 36$
- for the player who contributed 10: $20 - 10 + 16 = 26$ and
- for the player who contributed 20: $20 - 20 + 16 = 16$.

The payoff per round for each **passive player** is $20 - 0.2 \times 40 = 12$.

Instructions in the control treatment differ in the passage describing the passive player, the passage describing the payoff and the example. These parts read as:

- **Passive players:** Passive players cannot contribute to the public good. The payoff of the passive players does also not depend on the contributions of the active players.

- **Payoff per round:**

for **active players:** $20 - \text{points contributed} + 0.4 \times \text{sum of the contribution of all active players}$
for **passive players:** 20

Example

If the four active players contribute 0, 10, 10 and 20 (arranged by amount), the sum of contributions by all active players is 40 and each active player receives $0.4 \times 40 = 16$ from the joint project. The individual payoffs per round of the **active players** depend on the amounts contributed and are:

- for the player who contributed 0: $20 - 0 + 16 = 36$
- for the player who contributed 10: $20 - 10 + 16 = 26$ and
- for the player who contributed 20: $20 - 20 + 16 = 16$.

The payoff per round for each **passive player** is 20

Appendix A.2 Instructions to the subjects in *PE* and *NE* for phase 2 (translation of the German original)

Instruction for phase 2:

General Information

- In the second phase you will continue to play in your group with the 7 participants.
- You will again play 10 rounds of the game described in phase 1. Before round 1 and before the determination of the active and passive players, all participants will vote on how many points the active players should contribute to the public good.
- After the voting and before the start of the first round the 4 active and the 3 passive players will be **drawn anew**. Your type will be drawn **randomly and independently of your previous** type.

Voting on the amount of the contribution

- Before the determination of active and passive players, all 7 players of the group vote on the **amount of the contribution** in the following 10 rounds. **Please note:** At the time of the voting you do not know yet whether you will be an active or a passive player in phase 2. You do know, however, that the payoff of the passive players in phase 2 depends on the contributions of the active players.
- Each player can vote for one of the three possible amounts (0, 10 or 20).
- After everybody's vote has been casted you will be informed about the result of the voting.
- If one of the amounts obtains the absolute majority, i.e. received 4 or more votes, it is selected (see example 1).
- If there is no amount with an absolute majority, a run-off vote between the two amounts with the highest numbers of votes will be conducted (see example 2). If there are two amounts not with the highest but with an equal number of votes, the amount that will be part of the run-off vote will be randomly drawn from these two (see example 3). The amount of contribution that wins the absolute majority in the run-off vote will be selected.

		Example 1	Example 2	Example 3
Vote no. 1	0 10 20	2 votes 4 votes 1 vote	3 votes 3 votes 1 vote	3 votes 2 votes 2 votes
Result vote no. 1		contribution 10 selected	Run-off vote between contributions 0 and 10	Run-off vote between contributions 0 and 20 (result of random draw between 10 and 20)
Run-off vote	0 10 20	not necessary	5 2 -	1 - 6
Final result of vote		contribution 10 selected	contribution 0 selected	contribution 20 selected

Decision about the contribution of the active players

As in phase 1 only active players can contribute to the public good. They individually decide whether they want to contribute 0, 10 or 20 points. **Please note**, that neither the experimenter nor active nor passive players can enforce that the active players stick to the result of the vote about the amount of the contribution that was previously decided by the whole group.

Payoff

The exchange rate is the same in phase 2: **10 points** are **0.15 €**.

The instructions for the control treatment are identical, besides that the sentence “You do know, however, that the payoff of the passive players in phase 2 depends on the contributions of the active players.” is missing.

Appendix A.3 Instructions to the subjects in *PE*, *NE* and control for phase 3 (translation of the German original)

Instruction for phase 3:

General Information

- In the third phase you will also play in your group with the 7 participants.
- Again you will play 10 rounds of the game described in phase 1, **but the decision taken about the amount of contributions taken in phase 2 does not hold anymore.**
- Before the start of round one of the third phase, the 4 active and 3 passive players will be **drawn anew.** Your type will be drawn **randomly** and **independently of your previous** type.

Payoff

The exchange rate is the same in phase 3: **10 points** are **0.15 €**.

Appendix B Statistics

Box B.1 Model requirements for the regression analysis

Each participant decides every period how much she wants to contribute. Per subject, data is therefore correlated over time. We capture this relatedness by a subject specific error term, i.e. by a random effects model. Moreover contributions are both left and right censored: subjects cannot contribute less than 0, and they cannot contribute more than 20. A Tobit regression is therefore the appropriate functional form. Participants stay together in groups of seven over the entire game, which causes our data to be related within groups. Group dummies capture this relatedness. Finally, we capture effects that are specific to the degree of experience participants have gained by period dummies. We thus work with the following model:

$$\tilde{y}_{it} = \beta_0 + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + \mathbf{p}_t'\boldsymbol{\delta} + \alpha_i + \varepsilon_{it}$$

$$\begin{aligned} y_{it} &= 0 && \text{if } \tilde{y}_{it} \leq 0 \\ y_{it} &= \tilde{y}_{it} && \text{if } \tilde{y}_{it} \in (0,20) \\ y_{it} &= 20 && \text{if } \tilde{y}_{it} \geq 20 \end{aligned}$$

where β_0 is the constant. \mathbf{x} is the vector of explanatory variables we are interested in. These variables are allowed to vary over time, if appropriate. $\boldsymbol{\beta}$ is the vector of regressors. \mathbf{z} is the vector of group dummies, with their (unreported) vector of regressors $\boldsymbol{\gamma}$. Likewise, \mathbf{p} is the vector of period dummies, with their (unreported) vector of regressors $\boldsymbol{\delta}$. α_i is the subject specific error term, while ε_{it} is residual error.

This model still has one limitation. There is no generally acknowledged fixed effects estimator for censored data. Consequently, we are also unable to perform the Hausman test. We must assume that α_i and \mathbf{x}_{it} are uncorrelated. As a double check we run both a random effects and a fixed effects model that ignores the fact that our data is censored, and perform the Hausman test on this mirror model.

Table B.1 Explaining contributions before the introduction of a norm (phase 1)

dependent variable: contribution in phase 1	
control	-10.969*
PE	-17.733**
period	-1.574***
constant	20.851***
N	1080
p model	<.001

In the uncensored mirror models, the Hausman test is insignificant. Random Effects Tobit, group fixed effects and period fixed effects (not reported). *** p <.001, ** p <. 01, * p < .05

Table B.2 Explaining contributions while norm is in force (phase 2)			
dependent variable: contribution in phase 2	model 1	model 2	model 3
control		25.257***	12.943 ⁺
<i>PE</i>		10.427	10.859
voting outcome	1.075**	.554	.566
individual vote	.297*	.297*	.320*
actor in phase 1			-1.580
actor in phase 1 * control			-.594
actor in phase 1 * <i>PE</i>			-.831
period	-2.511***	-2.511***	-2.511***
constant	12.219 ⁺	12.219 ⁺	12.488 ⁺
p model	<.001	<.001	
rho	.213	.213	
Random effects Tobit with group and period fixed effects (not reported) N = 1080; *** p <.001, ** p <. 01, * p < .05, ⁺ p < .1 Hausman test insignificant on uncensored mirror models.			

Table B.3 Explaining the degree of norm abidance while the norm is in force (phase 2)		
dependent variable: norm abidance in phase 2	voting outcome 10	voting outcome 20 (<i>NE</i> omitted)
control	4.012	15.709**
period	-1.590***	-4.307***
constant	14.824***	49.308***
p model	<.001	<.001
N	480	480
Random effects Tobit (lower limit -10 with voting outcome 10, and lower limit -20 with voting outcome 20, upper limit 10 with voting outcome 10, and upper limit 0 with voting outcome 20) with period dummies (not reported). <i>NE</i> excluded with voting outcome 20 since only one group voted that high. *** p <.001, ** p <. 01, * p < .05. Hausman test on mirror models ignoring censoring insignificant.		

Table B.4.1 Explaining dynamics in phase 1						
dependent variable: first differences of contributions in phase 1	model 1	model 2	model 3	model 4	model 5	model 6
lagged profit	.897*** (18.57)					
lagged utility		.835*** (25.07)				
lagged difference between own and average actor profit			1.200*** (13.90)	1.144*** (11.24)		
control* lagged difference between own and average actor profit				.563* (2.31)		
PE*lagged difference between own and average actor profit				-3.471*** (-6.41)		
lagged difference between own and bystander profit			-.455*** (-4.86)	-.405*** (-4.65)	1.145*** (22.57)	1.120*** (16.29)
control* lagged difference between own and bystander profit				-.808** (-3.60)		-.057 (-.45)
PE* lagged difference between own and bystander profit				4.753*** (6.28)		.125 (1.27)
lagged difference between average actor and bystander profit					-1.600*** (-13.90)	-1.525*** (-11.24)
control* lagged difference between average actor and bystander profit						-.751* (-2.31)
PE* lagged difference between average actor and bystander profit						4.628*** (6.41)
cons	-22.190*** (-19.05)	-17.023*** (-25.91)	4.473*** (4.32)	12.416*** (7.27)	4.473*** (4.32)	12.416*** (7.27)
R2 within	.3245	.4123	.4974	.5306	.4974	.5306
R2 between	.1154	.1420	.0394	.0165	.0394	.0165
R2 overall	.2523	.3308	.2705	.1234	.2705	.1234
p model	<.001	<.001	<.001	<.001	<.001	<.001

We estimate a fixed effects model. There is no need for Tobit since the dependent variable is not censored. Since we estimate a linear model, the relatedness of observations within groups can be captured by clustering. For the same reason, White robust standard errors can be used. In most models, the Hausman test is significant. This could be remedied by a Hausman Taylor model (with period as an additional regressor). The results look similar. But the fixed effects model suffices for the research question, and it has the advantage that robust standard errors, clustered for groups, can be used. The N is smaller since observations from the first period cannot be used due to lagged independent variables. T-values are reported in parenthesis. N = 972; *** p <.001, ** p <. 01, * p < .05, + p < .1

Table B 4.2 Explaining Dynamics in Phase 2						
dependent variable: first differences of contributions in phase 2	model 1	model 2	model 3	model 4	model 5	model 6
lagged profit	.820*** (15.73)					
lagged utility		.742*** (19.03)				
lagged difference between own and average actor profit			1.151*** (19.99)	1.126*** (15.26)		
control* lagged difference between own and average actor profit				.127 (.81)		
PE*lagged difference between own and average actor profit				-2.168*** (-6.99)		
lagged difference between own and bystander profit			-.420*** (-6.28)	-.437*** (-6.41)	1.115*** (24.17)	1.064*** (16.90)
control* lagged difference between own and bystander profit				-.195 (-1.22)		-.026 (-.23)
PE* lagged difference between own and bystander profit				3.029*** (7.83)		.139 (1.46)
lagged difference between average actor and bystander profit					-1.535*** (-19.99)	-1.501*** (-15.26)
control* lagged difference between average actor and bystander profit						-.169 (-.81)
PE* lagged difference between average actor and bystander profit						2.890*** (6.99)
cons	-20.816*** (-16.39)	-15.860*** (-20.10)	3.570*** (5.08)	6.804*** (5.29)	3.570*** (5.08)	6.804*** (5.29)
R ² within	.2973	.3739	.4855	.5123	.4855	.5123
R ² between	.0007	.0106	.0293	.0456	.0293	.0456
R ² overall	.1883	.2701	.2238	.2012	.2238	.2012
p model	<.001	<.001	<.001		<.001	<.001

Table B 4.3 Explaining Dynamics in Phase 3						
dependent variable: first differences of contributions in phase 3	model 1	model 2	model 3	model 4	model 5	model 6
lagged profit	.736*** (12.33)					
lagged utility		.740*** (20.01)				
lagged difference between own and average actor profit			1.081*** (13.58)	1.007*** (8.68)		
control* lagged difference between own and average actor profit				.309+ (1.75)		
PE*lagged difference between own and average actor profit				-2.111** (-3.52)		
lagged difference between own and bystander profit			-.427*** (-4.64)	-.372** (-3.80)	1.014*** (16.21)	.971*** (9.61)
control* lagged difference between own and bystander profit				-.460** (-2.82)		-.049 (-.35)
PE* lagged difference between own and bystander profit				3.026*** (4.09)		.212 (1.62)
lagged difference between average actor and bystander profit					-1.441*** (-13.58)	-1.343*** (-8.68)
control* lagged difference between average actor and bystander profit						-.411+ (-1.75)
PE* lagged difference between average actor and bystander profit						2.814** (3.52)
cons	-17.870*** (-12.83)	-15.124*** (-20.98)	3.393** (3.85)	7.695*** (6.89)	3.393** (3.85)	7.695*** (6.89)
R2 within	.2560	.3704	.4340	.4698	.4340	.4698
R2 between	.0450	.1498	.2030	.1810	.2030	.1810
R2 overall	.1741	.2765	.2120	.1355	.2120	.1355
p model	<.001	<.001	<.001	<.001	<.001	<.001

Table B.5 Explaining contributions after norm is no longer “in force”						
dependent variable: contribution in phase 3	model 1	model 2	model 3	model 4	model 5	model 6
control	2.739			19.683**	21.554**	1.176
<i>PE</i>	-11.630			-9.641	-10.285	-12.128 ⁺
voting outcome		-.334	-.366	1.551**	1.618**	.183
individual vote			.233	.232	.165	.078
individual mean profit in phase 2					.463	.632
average contribution in phase 2						-.390
actor in phase 2					34.522*	41.659**
individual mean profit* actor in phase 2					-1.663*	-2.756***
average contribution in phase 2* actor in phase 2						2.565***
period	-2.488***	-2.488***	-2.487***	-2.487***	-2.485***	-2.483***
constant	61.079***	56.138***	53.837***	25.139*	17.196	49.797***
p model	<.001	<.001	<.001	<.001	<.001	<.001
Random effects Tobit with group and period fixed effects (not reported). , N = 1080; *** p <.001, ** p <. 01, * p < .05, + p < .1 Hausman test insignificant on uncensored mirror models.						