Leadership with Individual Rewards and Punishments

Gürerk, Özgür and Lauer, Thomas and Scheuermann, Martin

RWTH Aachen University, University of Cologne

17 July 2015

Leadership with Individual Rewards and Punishments

Özgür Gürerk*, Thomas Lauer[§], Martin Scheuermann*

*RWTH Aachen University, [§]University of Cologne

March 21, 2018

Abstract: Leading by example is considered an important means to influence followers. In most organizations, however, to influence followers, leaders use a variety of instruments. Most frequently, leaders possess the power to administer rewards or punishments to team members. Do individual rewards or punishments reinforce the impact of leading by example on team members' contributions? Because of confounding factors, it is difficult to research leading by example using field data. Here, we investigate the effects of leading by example and the effect of rewards or punishments on contributions in controlled lab experiments. We find that both rewards and punishments are more effective in fostering contributions than leading by example as such. When leading by example comes together with reward or punishment power, it does not improve on the effects of rewards or punishments as such in increasing contributions.

Keywords: Leadership; Incentives; Punishment; Reward; Leading by Example; Public Goods

JEL classification: C9, H4, M5

***Corresponding Author:** Özgür Gürerk (e-mail: ozgur.gurerk@rwth-aachen.de): School of Business and Economics, RWTH Aachen University, Templergraben 64, D-52056 Aachen, Germany
Thomas Lauer (e-mail: thomas.lauer@uni-koeln.de): Experimental and Behavioral Economics, University of Cologne, Universitätsstraße 22a, D-50937 Köln, Germany
Martin Scheuermann (e-mail: martin.scheuermann@rwth-aachen.de): School of Business and Economics, RWTH Aachen University, Templergraben 64, D-52056 Aachen, Germany

Acknowledgments: We thank participants of the 2014 Economic Science Association Conference in Prague, and the 2014 meeting of the Gesellschaft für experimentelle Wirtschaftsforschung in Passau for helpful comments. Financial support granted by the Deutsche Forschungsgemeinschaft (DFG) through GU 954/3 and LA 3372/1, is gratefully acknowledged.

1. Introduction

"The most important thing I learned is that soldiers watch what their leaders do. You can give them classes and lecture them forever, but it is your personal example they will follow."

General Colin Powell[1]

''The [. . .] most obvious way to bring about cooperation between employees [. . .] is to pay for cooperation or to punish uncooperative behaviour."

Edward P. Lazear[2]

Leaders may induce cooperative behavior by influencing their followers and coordinating them on efficient outcomes (Foss 2001). According to conventional wisdom, one possibility to exert such influence is through leading by example. If leaders exert high efforts first, followers can be expected to mimic them. The commander of a platoon, the captain of a football team, the head of a political party; all are expected to engage themselves in a particular way and serve as role models to motivate soldiers, fellow teammates, or supporters.

In addition to leading by example, leaders usually use a variety of other instruments simultaneously. Often, leaders are equipped with the power to reward or punish their subordinates individually. Rewards may be monetary, such as bonus payments or a salary increase, or they may be non-monetary, like the provision of a better work environment, or a representative car. On the other hand, leaders can exert individual punishments, like ending a job contract, assigning the employee less preferable tasks, or withholding expected bonuses. In the military, leaders may even use a disciplinary "non-judicial-punishment" and send their subordinates to jail, up to several weeks.

Given that leaders often can administer individual rewards or punishments to their followers, it is surprising that no experimental study yet systematically investigated the combined effect of leading by example and rewards or punishments. In this paper, we fill this gap by examining the following research question: Do individual rewards or punishments reinforce leading by example's assumed cooperation-enhancing effect?

Since many other confounding factors may play a role, investigating leading by example in the field is not easy. The advantage of controlled lab experiments is that we can disentangle the

[1] Taken from http://www.themilitaryleader.com/resource-recommendations/quotes/ on August 29, 2016.
[2] Lazear (1998, pp. 269–270).

effect of individual rewards and punishments from the effect of leading by example. To do this, we conduct 2x3 factorial between-subject treatments, with and without leaders.

We have two main contributions to the existing literature. First, we inquire into the leader's influence on contributions in the absence of any rewards or punishments. Since some previous studies experimentally investigated this question, we contribute to the literature with a robustness check. Second, we ask whether rewards or punishments reinforce leading by example with respect to increasing contributions. By comparing different treatments, we identify the effects of leading by example and individual rewards and punishments. The second point is our novel contribution to the literature.

Our main findings are as follows: First, in the absence of any rewards or punishments, leader-free teams achieve similar contribution levels as teams with leaders. Second, rewards and punishments are helpful to increase contributions, both in teams with and without leaders. We find, however, that setting an example does not further increase the impact of rewards and punishments on contributions. We conclude that in our setting, not leading by example, but the use of rewards or punishments seems crucial for achieving high contributions.

2. Related Literature

2.1. The Mechanics of Leading by Example

Leading by example may work via (social) preferences. Empirical evidence shows that many people are conditional cooperators, who prefer to match others' contributions (Fischbacher et al. 2001). Thus, theoretically, in the presence of conditional cooperators, leaders' high efforts can influence the followers to exert high efforts, too. Experimental evidence, however, shows that many conditional cooperators slightly undercut others' efforts (Fischbacher and Gächter 2010), which often leads to a decline of average contributions over time.

Leading by example could also work by shaping the followers' beliefs. Cartwright and Patel (2010) theoretically show that an individual may prefer to contribute first in a sequential public goods game if she believes that a sufficient number of other players would imitate her contribution. Related to this, Gächter and Renner (2014) experimentally show, in a repeated setting, that the initial behavior of leaders seems decisive. In the subsequent rounds, however, followers put more weight on the fellow teammates' behavior than the leader's efforts.

Leading by example may work even better "in the dark" (Weber 2015) if followers do not have full information on the productivity of the team project. In an asymmetric information model,

Hermalin (1998) considers a team where only one single member knows about the team's overall productivity. Hermalin (1998) shows that the team can achieve a higher welfare if the informed team member exerts effort – as a leader – before her teammates do. With her costly commitment, the leader can credibly signal being in the preferred high-productivity state and motivate her teammates to exert high effort, too. This model nicely demonstrates that the existence of a leader and her revelation of private information may result in high cooperation in teams. Komai et al. (2007) extend Hermalin (1998), by showing that some cases exist, where preventing full revelation of the state may induce followers to exert higher efforts, compared to if the leader revealed the state fully with her action (as in Hermalin 1998).

In a fundraising context, Vesterlund (2003) presents a similar theoretical prediction as Hermalin (1998), when there is uncertainty about the quality of a charity (the value of the public good). To increase overall contributions, fundraisers of high-quality charities should publicly announce the first contribution. This model again provides a rationale for the effectiveness of having a leader for increasing contributions.

In their model, Huck and Rey-Biel (2006) consider a team with two inequality averse players. If a leader exerts effort first, team output and payoffs are higher, compared to simultaneous play. Huck and Rey-Biel (2006) also show how leadership can arise endogenously. In their setting, to maximize team output, the least productive member should be selected as the leader.[3]

2.2. Leading by Example Experiments without Rewards or Punishments

Leading by example is found to be beneficial in coordination experiments (Cartwright et al. 2013), as well as in settings with asymmetric information on the value of the team project (Potters et al. 2007, Komai et al. 2011). In this paper, we focus on leading by example settings with symmetric information on the value of the team project.

Sequential public goods games emerged as the workhorse for the experimental investigation of leaders' influence on team cooperation. In these experiments, the leader contributes first. Since other teammates, also denoted as followers, are informed about the leader's contribution before contributing themselves simultaneously, the leader can set an example.

In a public bad experiment, Moxnes and van der Heijden (2003) show that the presence of a leader improves the group outcome to a small extent, but significantly so (followers invest 13% less in the public bad). In a public goods game, Güth et al. (2007) observe higher overall

[3] To our best knowledge, there are no theoretical studies investigating leading by example in combination with reward or punishment power.

contributions if a leader is present, compared to a no-leader setting. Levati et al. (2007) find a positive effect of a leader's existence on contributions when endowments are heterogeneous, but the effect is less pronounced for the case of homogenous endowments. Pogrebna et al. (2011) also find that contributions are higher when a subject makes a binding announcement before fellow group members contribute – compared to the simultaneous decision setting.

There are, however, at least as much studies reporting no significant leader effect (Sturm and Weimann 2007, Haigner and Wakolbinger 2010, Gächter and Renner 2014, Sahin et al. 2015). In a field experiment conducted in Bolivia, Jack and Recalde (2015) do not find a significant leader effect, if the leader is elected randomly compared to baseline simultaneous treatment without a leader. Hence, for experiments with symmetric information on the value of the public good, the literature is not conclusive on leaders' influence on overall contributions.

In general, leaders' and followers' efforts are highly correlated (see, e.g., Gächter and Renner 2014). The followers, however, persistently undercut the leader's contributions, which in turn makes the leader lower her contribution in the following round. Gächter and Renner (2014) argue that a leader, after realizing that the followers contributed less than herself, wants to avoid the feeling of being exploited again and reduces her contribution in the next round. Overall, this behavior leads to a decay of contributions over time, similar to what one observes in simultaneous contribution experiments without a leader.

2.3. Experiments with Centralized Rewards or Punishments in Simultaneous settings

Ample evidence from public goods experiments shows that decentralized (peer) rewards or punishments can increase contributions (see, e.g., Balliet et al. 2011, Milinski and Rockenbach 2012). The absence of a coordination device, however, often hinders the efficient use of rewards or punishments.

A number of studies investigated the centralized use of rewards or punishments when team members contribute simultaneously. O'Gorman et al. (2009) find that centralized execution of punishment does not increase contributions more than peer punishment. Under perfect information, Fischer et al. (2016) do not find a difference in aggregate behavior between peer and central punishment, either. Nosenzo and Sefton (2014) report a similar result in a setting with both rewards and punishments. Grieco et al. (2017) compare several centralized punishment mechanisms to a decentralized peer punishment setting. Similar to O'Gorman et al. (2009), if only one randomly selected group member can punish, contribution levels are not different from the decentralized setting. Contributions are significantly higher, however, if the

group member with the highest contribution of the current round has the power to punish. In a within-subject design, Harrell and Simpson (2016) investigate whether the introduction of punishment leads to higher contributions when only the leader has punishment power compared to peer punishment. They do not find a significant contribution difference between the leader and the peer punishment treatments. Teams with prosocial leaders, however, contribute more than teams with proself leaders, but no more than in the peer punishment treatment. In Gürerk et al. (2009), leaders can choose between a positive and a negative incentive scheme. The selected scheme is used for ten consecutive periods. While initially 19 out of 20 leaders prefer the reward scheme, many of them switch to the punishment scheme after observing decreasing contributions in their group. In the final and third phase, the majority of leaders choose the punishment scheme, which generates higher contribution levels than the reward scheme.

In contrast to a central authority in simultaneous settings as above, the leader in our *sequential* setting is possibly more effective in increasing contributions, since she can set a clear example. If teammates do not follow, she can use rewards or punishments to motivate them.

2.4. Experiments with Centralized Rewards or Punishments in Sequential Settings

To our best knowledge, only two studies investigate a (sequential) leading by example setting in combination with a reward or punishment power of the leader. In Güth et al. (2007), the team leader can punish just one team member, by temporarily excluding her from the team and the team output. When leaders have such exclusion power, contributions increase. Sutter and Rivas (2014) report a treatment in which the leader can reward just one single follower. In this setting, contributions are higher compared to a control treatment without a leader.

Our study is different from the works mentioned above in some important ways: In the present study, unlike in Güth et al. (2007), leaders cannot exclude players from the group but can reward (or punish) teammates by increasing (or reducing) their payoffs. Different from Sutter and Rivas (2014), where leaders may only reward or punish one group member, we give the leaders the possibility to reward or punish each teammate individually.

3. Experimental Design

3.1. Treatments

We explain our experimental design with respect to treatments in three steps. First, to investigate the leader's influence in the absence of rewards or punishments, we conduct two treatments. While in the leader treatment L, the leader contributes first, in the leader-free treatment P, peers contribute simultaneously. Second, to investigate the impact of rewards or punishments in the absence of a leader, we conduct the treatments P-REW, and P-PUN, and contrast them to the P treatment. Third, to inquire into the leader's influence in the presence of individual rewards or punishments, we conduct two more treatments: the treatment L-REW with reward possibilities for the leader, and the treatment L-PUN with punishment possibilities for the leader. To find out the combined effect of leading by example and rewards or punishments, we compare these treatments to the peer treatments P-REW and P-PUN.

3.2. The Game

In the treatments L and P, we deploy a one-stage public goods game with voluntary contribution mechanism. In the other four treatments, there is an additional stage with the possibility to allocate rewards or punishments. The game is played for 20 periods, in groups of four, with identical group composition over the rounds.

Public Goods Stage: In each period, each player receives an endowment of $e = 20$ and decides on her contribution $0 \leq c_i \leq e$ to the public good. The sum of contributions $C = \sum_{i=1}^{4} c_i$ is multiplied by 1.6, and is equally split among all group members. Thus, the marginal per capita return *MPCR* amounts to 0.4. After each period, each player is informed about individual contributions and payoffs. In the leader treatments, in each group, one player is randomly chosen to serve as the leader for all periods. Leaders contribute first, the followers after receiving feedback about the leader's contribution.[4]

Reward and Punishment Stage: In the leader treatments L-REW and L-PUN, the leader is equipped with additional 20 points that she can either keep for the private account or use for individual rewards or punishments. In the peer treatments P-REW and P-PUN, each subject is

[4] We deliberately decided to use a random selection mechanism in order to avoid confounds between our variables of interest and the potential effect from selection procedures, and to keep our design comparable to the previous studies mentioned in section 2. We are also aware of potential effects of the selection modus (random) of the leader on contributions. Previous studies show a rotating leader (Güth et al. 2007) who is chosen exogenously does not significantly affect leader contributions or the followers' propensity to follow the leader's example. On the other hand, if a randomly selected group member voluntarily decides to lead, team contributions are higher compared to teams with involuntary leaders (Haigner und Wakolbinger 2010). The focus of this study, however, is on the effectiveness of leading by example as such, and on the additional effects of individual rewards or punishments.

given 5 additional points which she can either keep in the private account or use to reward (or to punish) other teammates individually.[5] In all treatments with rewards or punishments, every allocated reward point increases the receiver's payoff by 3 points, and every punishment point decreases the receiver's payoff by 3 points. In all treatments, players keep the "unused" points in their private accounts.

Payoffs: The payoff π_i for player *i* for each treatment is given below. The sum of reward or punishment points that a player *i* allocates to others is given by r_i and p_i, with r_{-i} and p_{-i} being the sum of the points that player *i* receives from other players.

(1) L and P: $\pi_i = 20 - c_i + 0.4C$

(2) L-REW: $\pi_i^{Leader} = 20 - c_i + 0.4C + 20 - r_i,\ \pi_i^{Follower} = 20 - c_i + 0.4C + 3\, r_{-i}$

(3) L-PUN: $\pi_i^{Leader} = 20 - c_i + 0.4C + 20 - p_i,\ \pi_i^{Follower} = 20 - c_i + 0.4C - 3\, p_{-i}$

(4) P-REW: $\pi_i = 20 - c_i + 0.4C + 5 - r_i + 3r_{-i}$

(5) P-PUN: $\pi_i = 20 - c_i + 0.4C + 5 - p_i - 3p_{-i}$

3.3. The Procedures

We programmed the experiment with z-Tree (Fischbacher 2007) and conducted it at the RWTH Aachen University using ORSEE (Greiner 2015) for subject recruitment. 288 subjects participated in 72 independent observations (12 per treatment). After the experimenter read the instructions (see Appendix) aloud, subjects could privately ask clarifying questions. Sessions lasted about 60 minutes. Each subject was paid privately. The average payoff was €12.70.

4. Hypotheses

Assuming money-maximizing actors with self-centered preferences, and applying backward-induction, it is straightforward to see that no leader or follower should engage in costly punishment or reward. Following the same rationale, it is also obvious that no player contributes to the public good.

[5] To clearly separate contribution behavior and the use of rewards or punishments, we provide additional tokens for the reward and punishment stage. We do this to keep the reward and punishment mechanisms comparable, and to ensure that leaders and peers face the same costs for a given level of reward or punishment. A number of studies apply a similar design (Gürerk et al. 2009, Choi and Ahn 2013, Nosenzo and Sefton 2014).

In contrast, models assuming social preferences (e.g., Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) show that players may contribute to public goods and forego parts of their earnings to punish uncooperative players.

Based on the evidence from the experimental literature discussed in Section 2 we state the following three hypotheses corresponding to our three main research questions.

Hypothesis 1: *In the absence of rewards or punishments, teams with leaders achieve different levels of contributions than leader-free teams* (Comparison of L and P treatments).

Hypothesis 2: *In the absence of leaders, teams with reward or punishment options achieve higher contributions than teams without reward or punishment possibilities* (Comparison of P-REW and P, and P-PUN and P treatments, respectively).

Hypothesis 3: *In the presence of rewards or punishments, teams with leaders achieve different levels of contributions than leader-free teams* (Comparison of L-REW and P-REW, and L-PUN and P-PUN treatments, respectively).

5. Results

We first focus on the test of our three hypotheses (Results 1-3). After that, we report results about the evolution of contributions and individual contribution behavior. We then investigate the use of rewards and punishments, and teammates' reactions in terms of contribution changes. Finally, we analyze the payoffs.

5.1. The Effect of Leadership on Contributions in the Absence of Rewards or Punishments

Table 1 shows the average contributions, payoffs, and the average rewards and punishments for all treatments. What is the effect of leadership on contributions in the absence of any punishment and reward possibilities? Note that, in our setting, being a leader always means to contribute first. As can be seen, over all periods, the leader treatment L induces lower contributions than the peer treatment P, the difference, however, is not statistically significant (9.8 and 11.3 points, Mann-Whitney-U test (MWU), two-sided, $p = 0.564$).[6]

[6] This result may be considered as somewhat surprising since some of the previous studies (mentioned in Section 2.2.1.) report that groups with leaders tend to have higher or at least similar contribution levels as treatments without leader. We are not the first to find insignificantly lower average contributions in a leader setting compared to a no-leader environment (Gächter and Renner 2014: no-leader treatment 10.24, leader treatment 9.64, $p = 0.603$).

Result 1. *In the absence of rewards or punishments, teams with leaders do not achieve different levels of contributions than leader-free teams.*

As also can be seen from Table 1, the followers in the L treatment contribute significantly less than the leaders do (9.3 and 11.2 points, Wilcoxon-matched-pairs test (WMP), two-sided, p = 0.021). This result is in line with previous studies. We do not find a difference between the followers' contributions in the leader treatment L and the teammates' contributions in the peers treatment P (9.3 and 11.3, MWU, two-sided, p = 0.453).

Table 1: Treatment averages over all periods. Standard deviations in parentheses.

			Contributions			
Treatment	**L**	**P**	**L-REW**	**P-REW**	**L-PUN**	**P-PUN**
All	9.8 (5.7)	11.3 (4.6)	14.5 (5.0)	15.9 (4.2)	13.8 (6.8)	16.9 (4.4)
Leader	11.2 (5.6)	-	13.1 (5.9)	-	14.6 (6.5)	-
Follower	9.3 (5.9)	-	15.0 (5.4)	-	13.5 (7.2)	-
	Payoffs from the public goods stage excluding rewards and punishment					
Treatment	**L**	**P**	**L-REW**	**P-REW**	**L-PUN**	**P-PUN**
All	25.9 (3.4)	26.8 (2.8)	28.7 (3.0)	29.5 (2.5)	28.3 (4.1)	30.1 (2.6)
Leader	24.4 (4.2)	-	30.1 (5.2)	-	27.5 (5.7)	-
Follower	26.3 (3.4)	-	28.2 (3.1)	-	28.5 (3.9)	-
	Net Payoffs including rewards and punishments					
Treatment	**L**	**P**	**L-REW**	**P-REW**	**L-PUN**	**P-PUN**
All	25.9 (3.4)	26.8 (2.8)	38.8 (5.1)	40.4 (4.9)	31.6 (4.4)	32.6 (4.1)
Leader	24.4 (4.2)	-	40.0 (1.9)	-	45.8 (5.9)	-
Follower	26.3 (3.4)	-	38.4 (6.9)	-	26.9 (4.3)	-
	Average allocated rewards and punishments (average of the team's total)					
Treatment	**L**	**P**	**L-REW**	**P-REW**	**L-PUN**	**P-PUN**
All	-	-	-	11.7 (5.1)	-	2.5 (2.3)
Leader	-	-	10.2 (5.1)	-	1.6 (1.4)	-

5.2. The Effect of Rewards and Punishments on Contributions in the Absence of Leaders

To see, in the absence of leaders, whether teams with reward or punishment options achieve higher contributions than teams without, we compare P to P-REW, and P to P-PUN, respectively. As can be seen in Table 1, teams with peer rewarding possibilities (P-REW) achieve significantly higher contributions than the teams in the peer treatment P (15.9 and 11.3,

MWU, two-sided, $p = 0.043$). Similarly, teams with peer punishment options achieve higher contributions than the teams in P (16.9 and 11.3, MWU, two-sided, $p = 0.006$).

Result 2. *In the absence of leaders, teams with reward or punishment options achieve higher contributions than teams without reward or punishment possibilities.*

5.3. The Effect of Leaders on Contributions in the Presence of Rewards or Punishments

To investigate whether leaders influence contributions in the presence of reward or punishment possibilities, we compare the average contributions of the L-REW treatment to P-REW, and the contributions in L-PUN treatment to P-PUN, respectively. Based on our finding that leading by example does not increase contributions in the absence of reward or punishment opportunities (Result 1), we argue that any potential difference in contribution levels between L-REW and P-REW or L-PUN and P-PUN should come from the combined effect of having a leader with the power to reward or punish. Over all 20 periods, we observe no significant differences between the treatments L-REW and P-REW (14.5. and 15.9 points, MWU, two-sided, $p = 0.237$), and between L-PUN and P-PUN (13.8 and 16.9 points, MWU, two-sided, $p = 0.106$). In fact, the average contributions even tend to be higher in the peer treatments (see Table 1).

To separate the (pure) effect from having a leader from the effect of having a leader with reward or punishment power we run two panel regressions. The results are presented in Table 2. In column (I), we compare individual contributions in treatments L, P, L-PUN and P-PUN to see whether a leader can increase contributions in a punishment environment. Neither the dummy that indicates whether the groups have a leader (1), nor the independent variable for the combined effect from having a leader with punishment power (3) is significantly different from zero. In contrast to that, the dummy for being in a punishment treatment (2), is highly significant. Column (II) compares the treatments L, P, L-REW and P-REW, and shows a very similar result for the reward environment. There is no effect from having a leader (1) or from having a leader with reward power (3). Being in a reward environment (2), however, results in significantly higher contributions. In both regressions, the other followers' average contribution in the previous period (4) has a positive and significant effect on contributions in the current period.

Result 3. *In the presence of rewards or punishments, leaders have no (additional) effect in increasing contributions.*

Table 2: Treatment differences - the effects of leadership, punishment, and rewards.

Dependent variable: Teammate's contribution	(I) L-PUN, P-PUN, L & P	(II) L-REW, P-REW, L & P
(1) Leader dummy	−0.553 (0.560)	−0.634 (0.685)
(2) Punishment (I) / reward (II) treatment dummy	2.111*** (0.558)	1.677*** (0.595)
(3) Punishment (I) / reward (II) treatment dummy x Leader dummy	−0.543 (0.878)	0.526 (0.931)
(4) Other followers' (peers') average contribution in t-1	0.732*** (0.031)	0.692*** (0.037)
Constant	2.596*** (0.505)	3.065*** (0.598)
Observations	3192	3192
R^2 overall	0.615	0.491

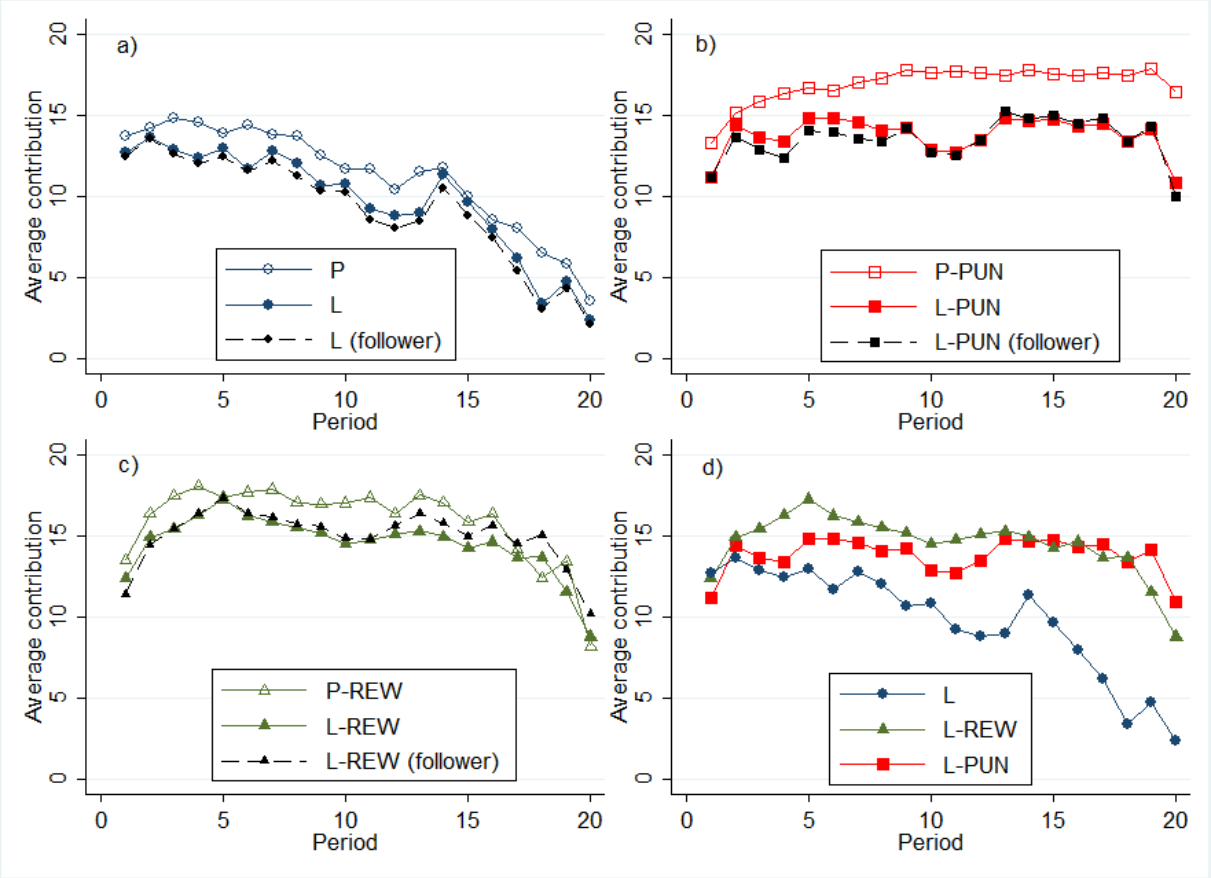
*** Significant at 1%. Panel regression with cluster robust standard errors in parentheses (clustered on group level). Note: To check for potential collinearity problems we run the same specifications without (4). The results remain qualitatively the same."

5.4. Evolution of Contributions and Individual Contribution Behavior

Evolution of Contributions

Panel a) of Figure 1 shows the evolution of contributions in treatments L and P. Note, for the leader treatment L, we plot the overall team average, and additionally, the average contribution for the followers. Over time, contributions both in L and P treatments show a decreasing trend. Compared to the first half of the experiment (periods 1-10), average contributions both in L and P are significantly lower in the second half (average of the periods 1-10 in L: 12.2 points, in periods 11-20: 7.3 points, WMP, two-sided, $p = 0.007$), (P treatment: 13.7, and 8.8, WMP, two-sided, $p = 0.006$). Both in L-REW and P-REW treatments (see Figure 1, panel c), contributions in the second half are also significantly lower than in the first half (L-REW: 15.3 points and 13.6, WMP, two-sided, $p = 0.023$, P-REW: 16.9 and 14.8, WMP, two-sided, $p = 0.034$). Interestingly, this decreasing trend is absent in both punishment treatments (see Figure 1, panel b). While we observe a non-decreasing trend in the L-PUN treatment, we even see a clearly increasing trend in the P-PUN treatment: We find no difference in average contributions in L-PUN (13.8 in periods 1-10 and 13.8 in periods 11-20, WMP, two-sided, $p = 0.969$), whereas contributions are significantly higher in the second half in P-PUN (16.3 and 17.5, WMP, two-sided, $p = 0.050$).

Figure 1: Average contributions.



Individual Contribution Behavior: Does Leading by Example Work?

To analyze teammates' individual contribution behaviors, we ran a series of panel regressions that we present in Table 3. In column (I), we compare the behavior in treatments L and P to see whether leaders make a difference in the absence of any rewards or punishments. In column (II) and (III), we compare L-REW and P-REW, and L-PUN and P-PUN, respectively. Column (IV) looks at the three leader treatments aiming at disentangling the effects of leaders' contributions and the effect of their use of reward or punishment. Column (V) displays the effects of a reward or punishment setting in the absence of the leader.

The independent variable "leader dummy" (1) indicates whether the group has a leader or not. Interestingly, the mere existence of a leader has no significant effect on teammates' contributions, in any of the comparisons. In all leader treatments, however, the leader's contribution (2) does have a significant influence on followers' contributions. The magnitude of this effect, however, is not different from the effect of the other followers' contributions (3)

which is also significant.[7] The interaction between these two influencing factors (2) and (3) shows that the existence of a leader significantly reduces the impact from teammates' average contribution (4).

Table 3: Determinants of contributions (all periods).

Dependent variable: Teammate's contribution	(I) L & P	(II) L-REW & P-REW	(III) L-PUN & P-PUN	(IV) L-REW, L-PUN & L	(V) P-REW, P-PUN & P
(1) Leader dummy	−0.952 (0.608)	1.423 (2.639)	−0.904 (2.316)		
(2) Leader contribution	0.441*** (0.034)	0.142** (0.058)	0.212*** (0.082)	0.404*** (0.036)	
(3) Other followers' (peers') average contribution in t-1	0.818*** (0.032)	0.645*** (0.090)	0.680*** (0.078)	0.430*** (0.045)	0.738*** (0.036)
(4) Leader x Other followers' (peers') average contribution in t-1	−0.447*** (0.067)	−0.225* (0.121)	−0.255** (0.105)		
(5) Reward treatment dummy				6.208*** (1.777)	1.463*** (0.511)
(6) Reward treatment dummy x Leader contribution				−0.255*** (0.082)	
(7) Punishment treatment dummy				3.724** (1.752)	2.079*** (0.555)
(8) Punishment treatment dummy x Leader contribution				−0.164* (0.085)	
Constant	1.585*** (0.376)	5.501*** (1.584)	5.578*** (1.613)	0.472 (0.481)	2.524*** (0.543)
Observations	1596	1596	1596	2052	2736
R^2 overall	0.549	0.341	0.650	0.580	0.514

* Significant at 10%, ** at 5%, *** at 1%. Panel regression with cluster robust standard errors in parentheses (clustered on group level). In column (IV) we only look at the followers´ contributions (3 per group), in column (V) we look at the contributions of all team members (4 per group), hence the different number of observations (2052*4/3=2736).

A closer look at the relation of leaders' contributions and the availability of rewards (6) and punishments (8) reveals that the mere possibility to punish or to reward reduces the impact of leaders' contributions. It seems that leading by example becomes less important as soon as the leader is equipped with additional means to influence followers' contributions. The following correlations also support this claim. While leaders' and followers' contributions are highly correlated (average of Spearman's rho = 0.746) in the L treatment, the leaders' impact is much

[7] We tested whether the coefficients (2) and (3) are different. To do so, we calculated the chi-square for (2)-(3)-(4)=0. The p-value is 0.405 (Wald test).

smaller in L-PUN (average Spearman's rho 0.504), and even less so in L-REW (average Spearman's rho 0.205). The average correlation coefficients (rho) obtained in the L treatment are significantly higher than the coefficients in the L-PUN and L-REW (L and L-PUN: MWU, one-sided, $p = 0.002$, L and L-REW: MWU, one-sided, $p = 0.001$). Comparing the impact of the mere possibility of reward (5) and punishment (7) shows that a centralized use of reward or punishment power by the leader (IV) affects contribution behavior much stronger than the same possibilities in a decentralized environment (V) (Wald-test, rewards: $p = 0.003$, punishment: $p = 0.002$).

Result 4. *The effect of leading by example (measured as the correlation between leader and follower contributions) is less strong if the leader has rewards or punishments at hand.*

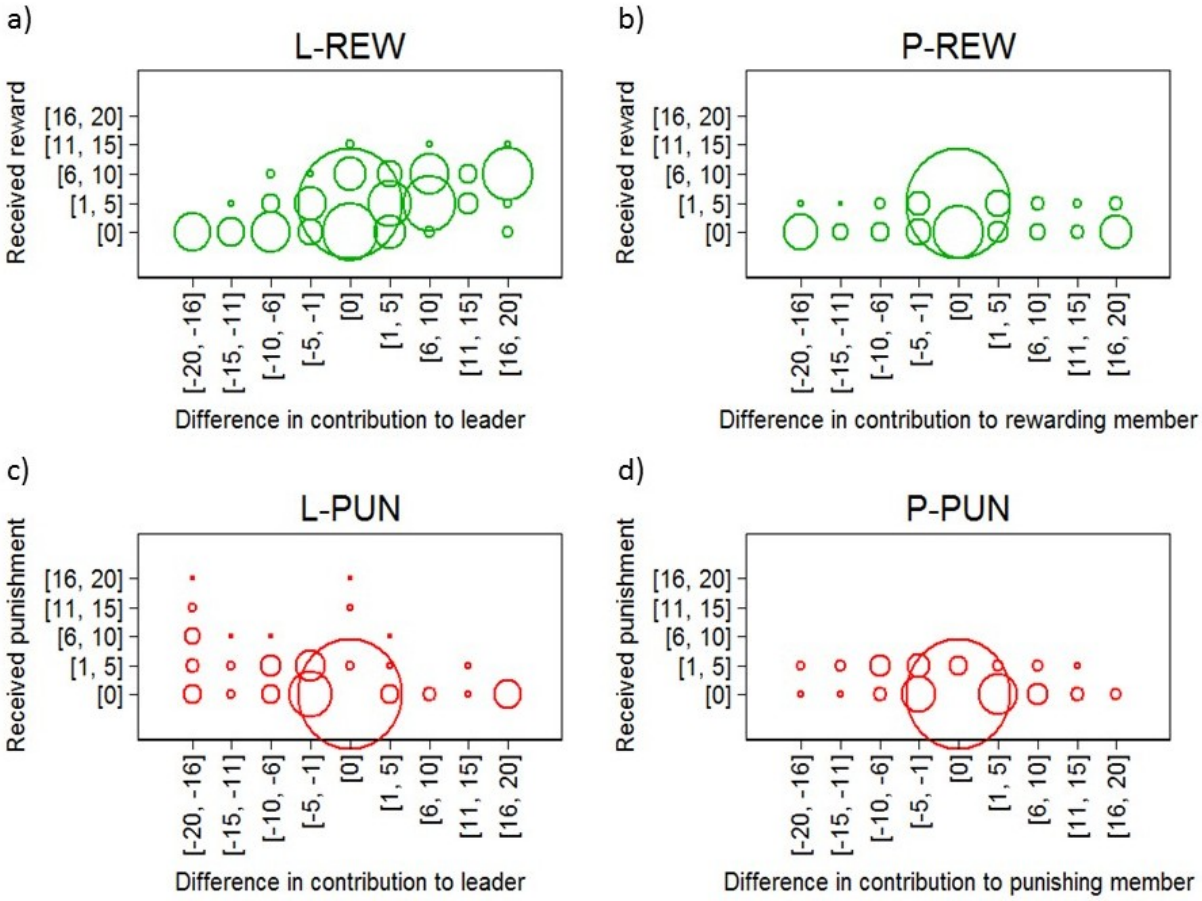
5.5. The Use of Rewards and Punishments

As can be seen in Table 1, leaders, as well as peers, clearly allocate more reward points than punishment points (MWU, two-sided, $p < 0.001$ in both cases). On average, leaders do not allocate more points than peers do, neither rewards (MWU, two-sided, $p = 0.544$), nor punishments (MWU, two-sided, $p = 0.299$).[8]

Figure 2 shows the average points a teammate received, dependent on the contribution difference between her and the rewarding/punishing leader; in the peers treatments, the difference between her and the rewarding/punishing team member, respectively (averages include zeros if no reward or punishment point was assigned).[9] Let us first look at the rewarding behavior. In L-REW, leaders allocate the highest average rewards (4.76 points) to those teammates who contributed more than the leaders themselves. Interestingly, all leaders rewarded (at least one time) followers who contributed the same amount (3.32 points) or even less (0.72 points). In the P-REW treatment, teammates on average allocate the most reward points to peers who contributed the same amount as they did. On average they allocated 1.20 points to those who contributed the same amount, 0.68 to those who contributed more, and 0.31 to those who contributed less than they did.

[8] If we only consider cases with positive reward and punishments, still there are no differences between L-REW and P-REW (11.9 and 12.2, MWU, two-sided, $p = 0.71$), nor between L-PUN and P-PUN (5.9 and 5.6, MWU, two-sided, $p = 0.85$). For a figure showing the evolution of rewards and punishments over time, see Appendix A.4.
[9] Recall that in the leader treatments, the leader may assign a maximum of 20 points, in a discretionary way, to one single teammate or to several teammates. In the peer treatments, each teammate may allocate up to 5 points.

Figure 2: Average received reward or punishment.



Based on single rewards and punishments, not the totals a subject might have received in a period. As peers only have 5 points to allocate, in the panels on the right-hand side, we only see bubbles for the received reward/punishment categories zero and the interval [1, 5]. The diameter of a bubble reflects the relative frequency of the respective reward or punishment level.

Now, let us look at punishments. In both PUN treatments, on average, those subjects who contributed less than the person allocating the points were punished more heavily (L-PUN 1.58 points, P-PUN 0.94 points), compared to teammates who contributed equal (L-PUN 0.11 points, P-PUN 0.04 points) or more (L-PUN 0.45 points, P-PUN 0.19 points). Teammates punish others more harshly, the greater the contribution differential between the punisher and the teammate is. Nevertheless, some subjects also punish teammates who contributed more than they themselves did. Interestingly, this "anti-social" punishment is significantly less frequent in L-PUN than in P-PUN (3.3% and 13.3% of all punishments, MWU, two-sided, $p = 0.046$).

Result 5. *Leaders exert anti-social punishment less-frequently than peers.*

In public goods experiments with peer punishment, a non-negligible fraction of peers punishes those who contributed more than themselves (see, e.g., Nikiforakis 2008). One possible

motivation for anti-social punishment is retaliation punishment. The fact that followers cannot retaliate after the leader has punished them may at least partly explain why we observe less anti-social punishment in L-PUN.

Do Leaders Use Rewards and Punishments Differently than Peers Do?

To investigate how reward and punishment are used on individual level, and whether there are differences in behavior between leaders and peers, we run a series of hurdle regressions. In the probit columns (I)-(IV) depicted in Table 4, we estimate how the probability to get a reward or punishment depends on a teammate's contribution. The Tobit columns (V)-(VIII) provide information on the magnitude of the received points by the teammate, given she received a positive amount of reward or punishment.

Let us look first at the probit columns (I) and (III) to see which variables increase the chance to get punished. In L-PUN, a higher contribution decreases the chance to be punished (1), but in P-PUN, the effect is not significant. In both punishment treatments, both the relative contribution compared to others (3), and that to the punishing teammate (2) increase the chances of getting a punishment significantly. Interestingly, for a leader's decision to punish a follower, it is more important how much a follower contributes compared to the other followers than compared to the leader herself, as the respective coefficient (3) is considerably higher than (2). The Tobit columns (V) and (VII) show, in both punishment treatments, the amount of punishment decreases with the contribution level (1) by a similar magnitude, but only significantly so in the P-PUN treatment. In both PUN treatments, teammates get less punishment, the more they contributed compared to the other two teammates (5).

Table 4: Determinants of punishment and reward decisions (all periods).

Dependent variable:	Probit: probability of subject *i*'s to be punished (rewarded)				Tobit: the amount of punishment (reward) subject *i* receives			
	(I) L-PUN	(II) L-REW	(III) P-PUN	(IV) P-REW	(V) L-PUN	(VI) L-REW	(VII) P-PUN	(VIII) P-REW
(1) *i*'s contribution	−0.030*	0.151***	−0.020	0.136***	−0.153	0.287***	−0.155*	0.159***
	(0.018)	(0.017)	(0.037)	(0.019)	(0.113)	(0.036)	(0.086)	(0.027)
(2) *i* contributes less (REW: more) than the punishing (rewarding) subject (0/1, binary)	0.803**	0.598***	0.846***	-1.093***				
	(0.338)	(0.225)	(0.231)	(0.205)				
(3) *i*'s contribution is lower (REW: higher) than the average contribution of the other two subjects in the group (0/1)	1.713***	−0.250	0.594***	−0.002				
	(0.292)	(0.340)	(0.203)	(0.153)				
(4) Difference between *i*'s and the punishing (rewarding) subject´s contribution					−0.130	0.104***	0.014	−0.103***
					(0.167)	(0.025)	(0.085)	(0.012)
(5) Difference between *i*'s and the average contribution of the other two subjects in the group					−0.690***	0.026*	−0.257***	0.076***
					(0.213)	(0.049)	(0.077)	(0.016)
Constant	−1.980***	−1.457***	−1.372*	−1.591***	−5.033*	−1.723***	−1.196	−1.992***
	(0.371)	(0.166)	(0.704)	(0.264)	(2.853)	(0.520)	(1.859)	(0.495)
Observations	720	720	2880	2880	720	720	2880	2880
Pseudo R^2	0.513	0.518	0.242	0.368	0.222	0.242	0.152	0.192

* Significant at 10%, ** at 5%, *** at 1%. Since in P-treatments, each subject can reward/punish the peers, we have four times more observations than in the L-treatment. Cluster robust standard errors in parentheses (clustered on group level).

Now let us turn to rewards. As the columns (II) and (IV) show, in both reward treatments, the own contribution level increases the probability to be rewarded (1). In L-REW, contributing more than the leader (2) increases the chances for a reward significantly, but not the relative contribution compared to other teammates. Interestingly, in P-REW, contributing more than others is decreasing the chances to get a reward. As the Tobit columns (VI) and (VIII) show, in both treatments, if rewarded, the magnitude of reward increases with the own contribution (1). In L-REW, the reward increases with the contribution difference between the reward-receiving teammate and the rewarding teammate (leader) while in P-REW, the reward is decreasing with the contribution differential between the rewarding and the rewarded peer (4). It seems that peers use rewards to encourage those who contributed less than themselves, while leaders reward to compensate those who already contributed a lot. In P-REW, the reward increases with the contribution difference between the rewarded and the other two teammates (5). Overall, we find a larger effect (in terms of magnitude) from falling behind the other two team members' average contribution in PUN treatments (coefficients (5) L-PUN: -0.690, P-PUN: -0.257) compared to the REW treatments (coefficients (5) L-REW: 0.026, P-REW: 0.076).

Leaders seem to have different reference contributions for evaluation when allocating rewards and punishments. While in case of punishment, the relative contribution of the teammate to other followers is decisive, when giving rewards, leaders evaluate a teammate's contribution to her own (the leader's) contribution. In particular, a teammate is more likely to be punished if she contributed less than other followers. In case of rewards, a teammate who contributes more than the leader is more likely to get a reward.

Teammates' Reaction to Rewards and Punishments (in the next round)

To what extent do teammates change their contribution in period t after being rewarded or punished in period t-1 (given they received positive rewards or punishments)? To answer this question, we run regressions as depicted in Table 5, with the contribution differential (contribution in t – contribution in t – 1) being the dependent variable. We ran separate regressions for each treatment. As the columns (I) and (II) show, perhaps unsurprisingly, receiving a reward does not significantly change a teammate's contribution, neither in L-REW, nor in P-REW. On the other hand, as the columns (III) and (IV) show, punishments have a positive and highly significant effect on the contribution change of the punished teammate. In the leader treatment L-PUN, the effect is much larger (Wald test $p = 0.015$) than in P-PUN.

Result 6. **(a)** *While rewards do not have a significant effect on contribution adjustments, punishments do.* **(b)** *Punishment from leaders has a much higher impact on contribution adjustments compared to punishment from peers.*

A leader's reward or punishment decisions may have higher acceptance by the other members, since by contributing first, the leader exposes herself to the risk of being exploited. The higher potential acceptance, in turn, may lead followers to stronger contribution reactions to punishments.

Table 5: Contribution change after receiving reward or punishment.

Dependent variable: Contribution change (Contribution in t minus contribution in t-1)	(I) L-REW	(II) P-REW	(III) L-PUN	(IV) P-PUN
Leader contribution	0.048		0.211	
	(0.047)		(0.181)	
Received reward points in t-1	-0.037	-0.145		
	(0.194)	(0.145)		
Received punishment points in t-1			1.172***	0.350***
			(0.199)	(0.088)
Period	-0.146***	-0.296***	0.087	-0.079*
	(0.050)	(0.045)	(0.112)	(0.041)
Constant	-0.505	2.280***	-4.687	1.610**
	(1.554)	(0.736)	(3.564)	(0.747)
Observations	515	764	90	204
R^2 overall	0.028	0.072	0.221	0.107

* Significant at 10%, ** at 5%, *** at 1%. OLS with cluster robust standard errors in parentheses (clustered on group level). We only looked at the team members' contribution change after they received a reward or punishment. Since the frequency of rewards or punishments executions varies across treatments, the number of observations is different in each treatment.

We also observe that except L-PUN, there is a significant negative trend of the contribution change over the periods, as the respective coefficients indicate. In other words, over time, the magnitude of adjustment from the previous round to the current round decreases.

Appendix A.2 and A.3 show group level data of contributions and the use of reward and punishment points. In the majority of the groups in L-REW, the level of reward points assigned by the leader is rather stable, while in P-REW in most groups the amount of reward points issued by the peers decreases over time. In treatments with the possibility to punish, leaders as well as peers in most cases decide not to punish at all. In P-PUN, we observe that peers often use punishment in the first rounds until every group member contributes the maximum amount of 20 points, and then refrain from punishment in the following rounds. If someone deviates from the maximum contribution (usually in the last period), punishment is used again.

5.6. Payoffs

We focus on net payoffs, i.e., payoffs including rewards and punishments.[10] Not surprisingly, due to the payoff increasing reward mechanism (recall each allocated reward point generates two more points), net payoffs in L-REW and P-REW are higher than in L-PUN and P-PUN, respectively (MWU, two-sided, $p = 0.002$, $p = 0.005$). We do not observe significant differences, neither between L-REW and P-REW (MWU, two-sided, $p = 0.453$), nor between L-PUN and P-PUN (MWU, two-sided, $p = 0.564$). Payoffs in L are significantly lower compared to L-REW (L: 25.9 and L-REW: 38.8 points, MWU, two-sided, $p < 0.001$), and compared to L-PUN (L: 25.9 and L-PUN: 31.6 points, MWU, two-sided, $p = 0.004$), perhaps not surprisingly, since in L-REW and in L-PUN, leaders have an additional endowment.

Leaders in L-REW use about half of their 20 additional points to increase the payoffs of the followers. This leads to rather similar net payoffs for the leaders (40.0 points) and the followers (38.4, WMP, two-sided, $p = 0.908$). In L-PUN, by allocating points, leaders cannot increase followers' payoffs. They rather can decrease the payoffs of their subordinates. Indeed, leaders in L-PUN receive much higher payoffs than followers do (45.8 and 26.9, WMP, two-sided, $p < 0.001$).

Which mechanism pays off more for leaders, and for followers? As the numbers above indicate, leaders with punishment possibilities obtain significantly higher payoffs than leaders with rewards (MWU, two-sided, $p = 0.028$), since leaders in the reward treatments invest a higher portion of their budget for rewards. Due to the payoff increasing reward mechanism, it is not surprising that followers in the reward treatment L-REW have much higher incomes than the followers in the L-PUN treatment do (MWU, two-sided, $p < 0.001$). In other words, as a leader, it pays off to be in the punishment setting. For a follower, the rewards setting is more advantageous.

6. Discussion and Conclusion

In this study, we disentangle the effects of individual rewards and punishments on contributions from the pure influence of leadership. In the absence of rewards and punishments, teams with a leader do not achieve higher contributions than leader-free teams (Result 1). Thus, in our setting, leading by example as such is not effective in increasing contributions. This result is not in line with the early influential experimental literature on leading by example that reports

[10] In Table 1, we also depict payoffs from the public goods stage, which are directly proportional to the contributions.

a positive effect (Moxnes and van der Heijden 2003, Güth et al. 2007, and Levati et al. 2007). Other studies, however, report a non-significant result, (Sturm and Weimann 2007, Haigner and Wakolbinger 2010, Gächter and Renner 2014, Sahin et al. 2015). Why are these discrepancies between the early papers and the followers? Let us look at the results of the early studies in more detail. Moxnes and van der Heijden investigate a "within subjects" design, they report a small but significant effect. In Güth et al. (2007), teams with leadership contribute only weakly significantly more than the control teams without leadership ($p = 0.08$, two-sided). And, lastly, Levati et al. (2007) report a less strong effect of leadership in case of homogenous endowments. Thus, the effects found in these studies are rather small (Moxnes and van der Heijden 2003, Levati et al. 2007), or only weakly significant (Güth et al. 2007). Taken together, in cases of randomly chosen leaders in symmetric settings, the positive effects of leading by example on contributions might not be as robust as commonly believed. This conclusion renders our first important contribution to the literature.

In the absence of leaders, teams with rewards or punishments possibilities achieve higher contributions than teams without these options (Result 2). Thus, in our setting rewards and punishments are effective in raising contributions. This result supports previous findings from the literature on the effectiveness of rewards or punishment options in increasing contributions (see, e.g., Milinski and Rockenbach 2012).

Finally, if reward or punishment possibilities are available, teams with leaders do not achieve a higher level of contributions than leader-free teams (Result 3). In other words, in our setting, powerful leaders who not only lead by example but can administer individual rewards or punishments cannot make an (additional) improvement with respect to increasing contributions. This finding is our novel contribution to the literature.

What are the implications of the above results summarized above? They indicate, to effectively increase contributions, leaders need some additional power. In our setting, individual rewards or punishments are the driving force in increasing contributions, and not the leader's example, as the contrast of our L-PUN and L-REW treatments to the corresponding P-treatments show. We observe, however, that contributions in teams with leaders who can reward or punish are not higher than the contributions in leader-free teams with peer rewarding or punishment. Thus, our results indicate that in small teams with mutual monitoring possibilities, to increase contributions, a leader may not be necessary. The teammates' willingness to invest in costly bilateral rewards and punishments may suffice for maintaining high contributions. Indeed, some real organizations are relying on flat hierarchies, some even with temporary or no leader at all.

One of the most prominent examples is *Valve*, a major game developer and digital distribution company, which is organized in self-managing "boss-free" teams.[11] Other examples with similar, non-rigid hierarchies and flexible organizational structure are *W.L. Gore* or *Partake*.[12]

How does our study relate to the large and growing experimental leadership literature? The present study sheds light on leading by example as such, and the allocation of individual rewards or punishments. A growing number of experimental studies investigate the role of other potentially important aspects of leadership on contributions, such as communication (see, e.g., Pogrebna et al. 2011, Koukoumelis et al. 2012), or the modus of leader appointment (see e.g., Haigner and Wakolbinger 2010, Baldassarri and Grossman 2011, Bruttel and Fischbacher 2013), or the role of leader characteristics (see e.g., Gächter et al. 2012). Future work should inquire into the interplay of these and other aspects with leading by example, and with rewards and punishments to obtain a more complete picture.

[11] http://www.valvesoftware.com/company/people.html, retrieved on June 2, 2017.
[12] https://www.gore.com/about/our-beliefs-and-principles, retrieved on June 2, 2017.
http://www.partake.de/en/was-ist-partake/, retrieved on June 2, 2017.

References

Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27), 11023-11027.

Balliet, D., Mulder, L. B., & Van Lange, P. A. (2011). Reward, punishment, and cooperation: a meta-analysis. *Psychological bulletin*, *137*(4), 594.

Bolton, G., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1), 166-193.

Bruttel, L., & Fischbacher, U. (2013). Taking the initiative. What characterizes leaders? *European Economic Review*, 64, 147-168.

Cartwright, E., & Patel, A. (2010). Public goods, social norms, and naive beliefs. *Journal of Public Economic Theory*, 12(2), 199-223.

Cartwright, E., Gillet, J., & Van Vugt, M. (2013). Leadership by Example in the Weak-Link Game. *Economic Inquiry*, *51*(4), 2028-2043.

Choi, J. K., & Ahn, T. K. (2013). Strategic reward and altruistic punishment support cooperation in a public goods game experiment. *Journal of Economic Psychology*, *35*, 17-30.

Fehr E., & Schmidt K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3), 817-868.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404.

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

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 16, 541-556.

Fischer, S., Grechenig, K., & Meier, N. (2016). Monopolizing sanctioning power under noise eliminates perverse punishment but does not increase cooperation. *Frontiers in behavioral neuroscience, 10*.

Foss, N. J. (2001). Leadership, beliefs and coordination: An explorative discussion. *Industrial and corporate change*, 10(2), 357-388.

Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2012). Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-By-Example. *Economic Inquiry*, 50(4), 953-967.

Gächter, S., & Renner, E. (2014). Leaders as role models for the voluntary provision of public goods, *CESifo Working Paper*, No. 5409.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114–125.

Grieco, D., Faillo, M., & Zarri, L. (2017). Enforcing cooperation in public goods games: Is one punisher enough? *Journal of Economic Psychology*, *61*, 55-73.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2009). Motivating Teammates: The Leader's Choice of Positive and Negative Incentives. *Journal of Economic Psychology*, 30, 591-607.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2014). On cooperation in open communities. *Journal of Public Economics*, 120, 220–230.

Güth, W., Levati, M. V., Sutter, M., & Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, *91*, 1023-1042.

Haigner, S. D., & Wakolbinger, F. (2010). To lead or not to lead: Endogenous sequencing in public goods games. *Economics Letters*, 108, 93-95.

Harrell, A., Simpson, B., (2016). The Dynamics of Prosocial Leadership: Power and Influence in Collective Action Groups, *Social Forces*, 94(3), 1283-1308.

Hermalin, B. E. (1998). Toward an economic theory of leadership: Leading by example. *American Economic Review*, 88(5), 1188-1206.

Huck, S., & Rey-Biel, P. (2006). Endogenous leadership in teams. *Journal of Institutional and Theoretical Economics JITE*, *162*(2), 253-261.

Jack, B. K., & Recalde, M. P. (2015). Leadership and the voluntary provision of public goods: Field evidence from Bolivia. *Journal of Public Economics*, 122, 80-93.

Koukoumelis, A., Levati, M. V., & Weisser, J. (2012). Leading by words: A voluntary contribution experiment with one-way communication. *Journal of Economic Behavior & Organization*, 81(2), 379-390.

Komai, M., Stegeman, M. & Hermalin, B.E. (2007). Leadership and Information. *American Economic Review*, 97(3), 944-947.

Komai, M., Grossman, P.J., & Deters, T. (2011). Leadership and Information in a Single-shot Collective Action Game: An Experimental Study. *Managerial and Decision Economics*, 32(2), 119-134.

Lazear, E. (1998). Personnel economics for managers. New York: Wiley.

Levati, M. V., Sutter, M., & Van der Heijden, E. (2007). Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution*, *51*(5), 793-818.

Milinski, M., & Rockenbach, B. (2012). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology*, *299*, 139-143.

Moxnes, E., & Van der Heijden, E. (2003). The effect of leadership in a public bad experiment. *Journal of Conflict Resolution*, *47*(6), 773-795.

Nikiforakis, N. (2008). Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves? *Journal of Public Economics*, 92(1), 91-112.

Nosenzo, D., & Sefton, M. (2014). Promoting Cooperation: The Distribution of Reward and Punishment Power, in: P.A.M. van Lange, B. Rockenbach, & T. Yamagishi (Eds), Social dilemmas: New perspectives on reward and punishment. New York. *Oxford University Press*.

O'Gorman, R., Henrich, J., & Van Vugt, M. (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 276, 323-329.

Pogrebna, G., Krantz, D., Schade, C. & Keser, C. (2011). Words versus Actions as a Means to Influence Cooperation in Social Dilemma Situations. *Theory and Decision*, 71(4), 473-502.

Potters, J., Sefton, M., & Vesterlund, L. (2007). Leading-by-example and signaling in voluntary contribution games: an experimental study. *Economic Theory*, 33(1), 169-182.

Sahin, S. G., Eckel, C., & Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, 1(1), 100-113.

Sturm, B., & Weimann, J. (2007). Unilateral emissions abatement. *Environmental economics, experimental methods*, *10*, 157.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. *Review of Economic Studies*, 77(4), 1540–1566.

Sutter, M., & Rivas, M. F., 2014. Leadership, Reward and Punishment in Sequential Public Goods Experiments, in Reward and Punishment in Social Dilemmas, edited by P. A. M. van Lange, B. Rockenbach, T. Yamagishi. Oxford University Press.

Vesterlund, L. (2003). The informational value of sequential fundraising. *Journal of Public Economics*, *87*(3), 627-657.

Weber, R. (2015). The Economics of Effective Leadership. UBS Center Public Paper #3. UBS International Center of Economics in Society. University of Zurich.

Appendix

A.1 Instructions (L-PUN)

General information

We welcome you to this economics experiment. It is very important for you to read the following instructions carefully. If you have any questions, please direct them to us.

In this experiment you can earn money. The exact amount of your payout depends on your decisions and on the other participants' decisions.

While the experiment is running, it is not allowed to communicate with other participants. Non-compliance leads to the exclusion from the experiment and from all payments. All decisions are anonymous, i.e. no other participant gets to know the identity of the participant who makes a specific decision. Anonymity is also ensured during the payout process, i.e. no participant gets to know the amount of other participants' payouts.

During the experiment your income will be calculated in points. The earned amount of points will be converted to Euro with the following exchange rate:

80 points = 1 Euro.

At the end of this experiment you will receive your payout according to the total number of accumulated points as well as 2.50 Euro for showing up.

In the following we will provide you with a detailed description of the experiment.

Rounds and Groups

- The experiment consists of 20 rounds with each round having the same structure.
- You are a member of a group with 4 members in total. During the experiment the group composition will always stay the same.
    - One group member will randomly be assigned the role of a type A participant, the remaining three members will be type B participants.
    - During the experiment you will maintain your role and only interact with members of your group.
    - Each participant receives a starting capital of 100 points.

Course of the experiment

Each round consists of two stages:

Stage 1: Contributions of the group members

- In every round, each group member receives 20 points.
- Each group member has to decide how many of the 20 points he or she wants to contribute for the group. Points which are not contributed remain with the group member. Possible amounts to contribute are integral numbers from 0 to 20. First, the type A member decides how much to contribute for the group.
- After being informed about the contribution of the type A member, type B members decide on their own contribution.
- The sum of the contributions of all group members (type A and type B) gets multiplied with 1.6 and forms the group result.

  (sum of contributions x 1.6 = group result)
- Each group member (type A and type B) receives a quarter of the group result independently from their own contribution (group result / 4 = individual share of the group result).

Stage 2

- The type A member gets to see how much each group member has contributed.
- The type A member now receives 20 additional points and has to decide if and how many of these 20 points he or she wants to assign to each type B member.
- With each point which the type A member assigns to a type B member, the income of the type B member gets reduced by 3 points.
- Points which are not assigned are kept by the type A member.

Please notice: The order in which type B members are displayed will be determined randomly for each round. Therefore, it is not possible to identify a type B member over the rounds by the position on the displayed lists.

<u>Calculation of you round income</u>

Round income for type A members =

20 (endowment for the round)

- your contribution

+ 1.6 x sum of the contributions of all group members / 4

+ 20 (points which can be assigned to type B members)

- sum of points which are actually assigned to type B members

Round income for type B members =

20 (endowment for the round)

- your contribution

+ 1.6 x sum of the contributions of all group members / 4

- 3 x number of received points

Information at the end of each round

At the end of each round you will be provided with an overview of the group results. For each group member you will get to know: contribution for the group, income after stage 1, assigned or received points, round income.

Please notice: The order in which type B members are displayed will be determined randomly for each round. Therefore, it is not possible to identify a type B member over the rounds by the position on the displayed lists.

Total income

The total income will result from the starting capital of 100 points plus the sum of the earnings from each of the 20 rounds. At the end of the experiment your total income will be paid out with the exchange rate of 1 Euro per 80 points. As already mentioned, you will additionally receive 2.50 Euro for showing up.

A.2 Contributions and Reward Points within Single Observations (Groups).

Figure A2.a: Average contributions and the sum of allocated reward points within single groups of the L-REW treatment.

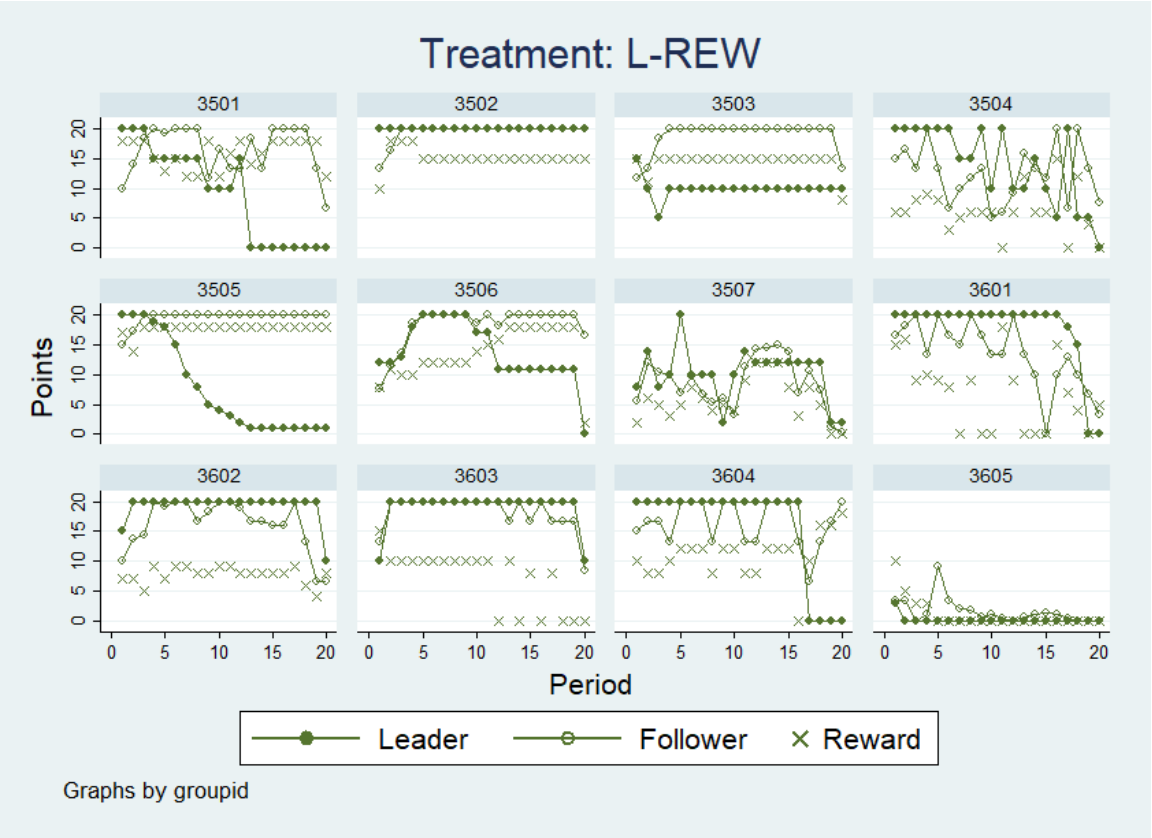
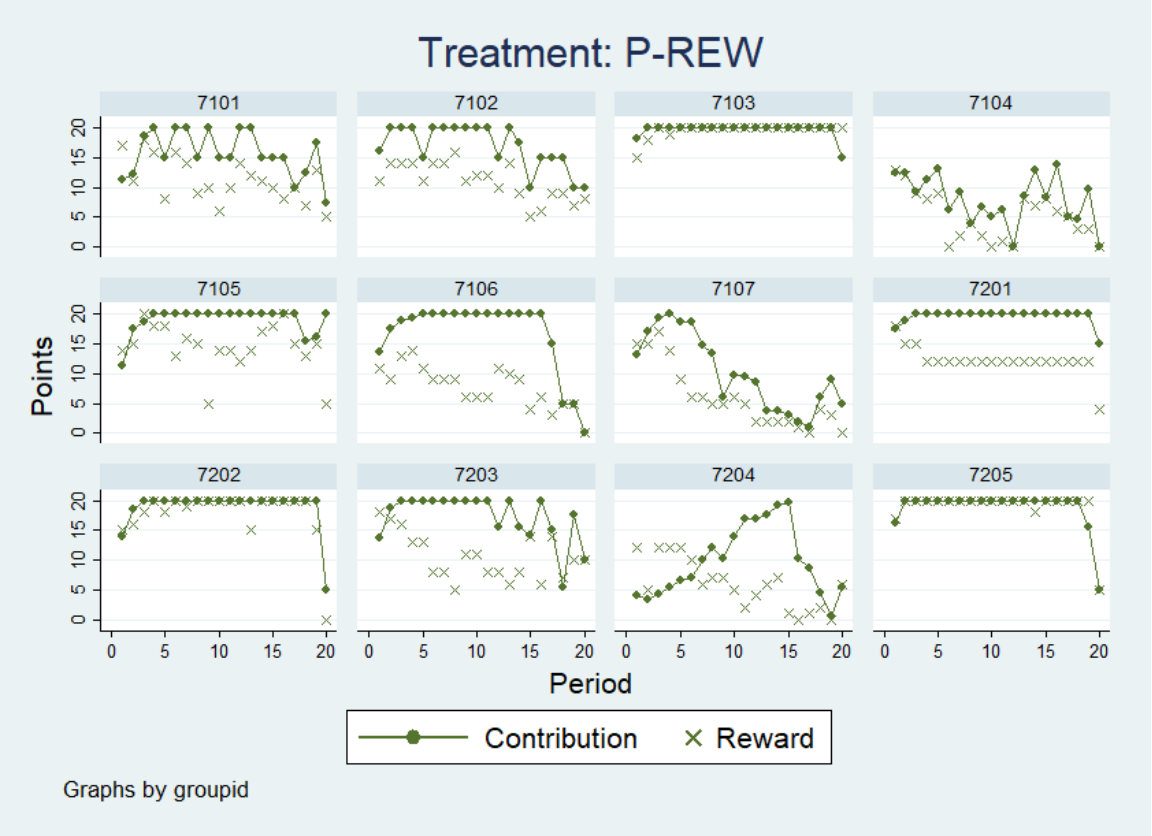


Figure A2.b: Average contributions and the sum of allocated reward points within single groups of the P-REW treatment.



A.3 Contributions and Punishment Points in Single Observations (Groups).

Figure A3.a: Average contributions and the sum of allocated punishment points within single groups of the L-PUN treatment.

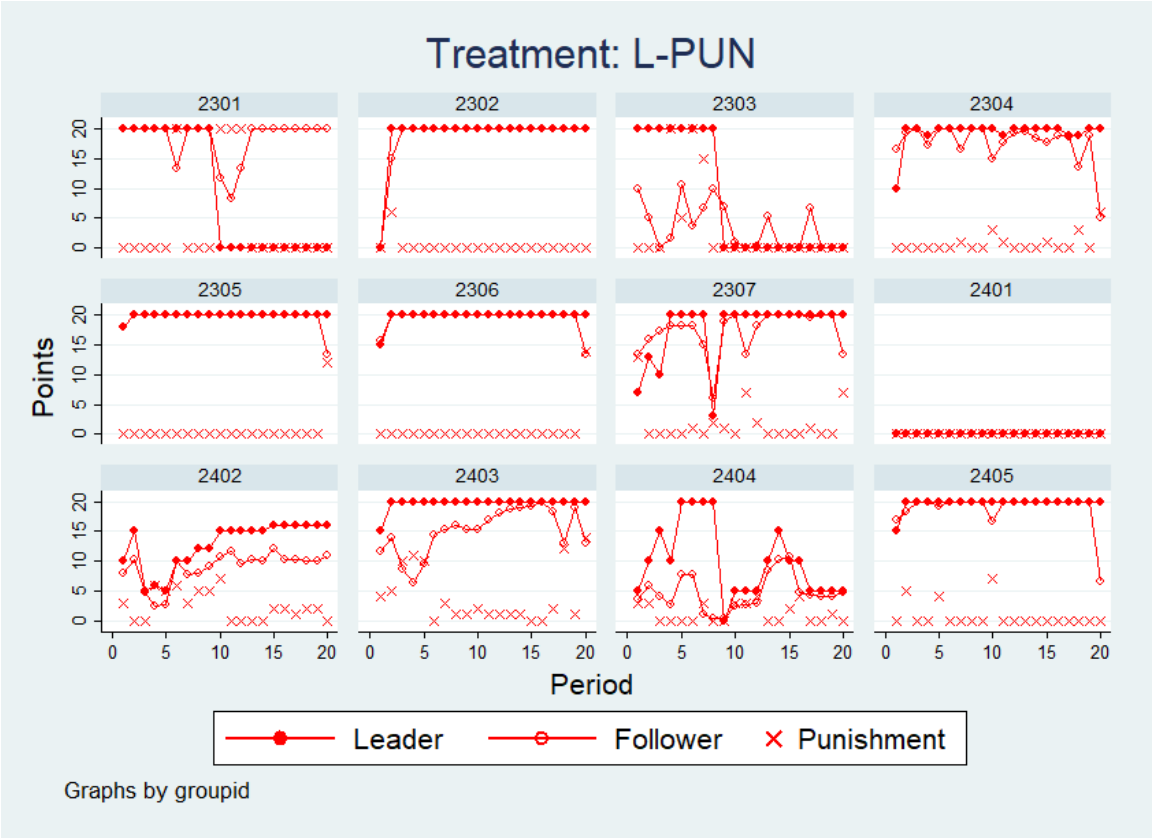
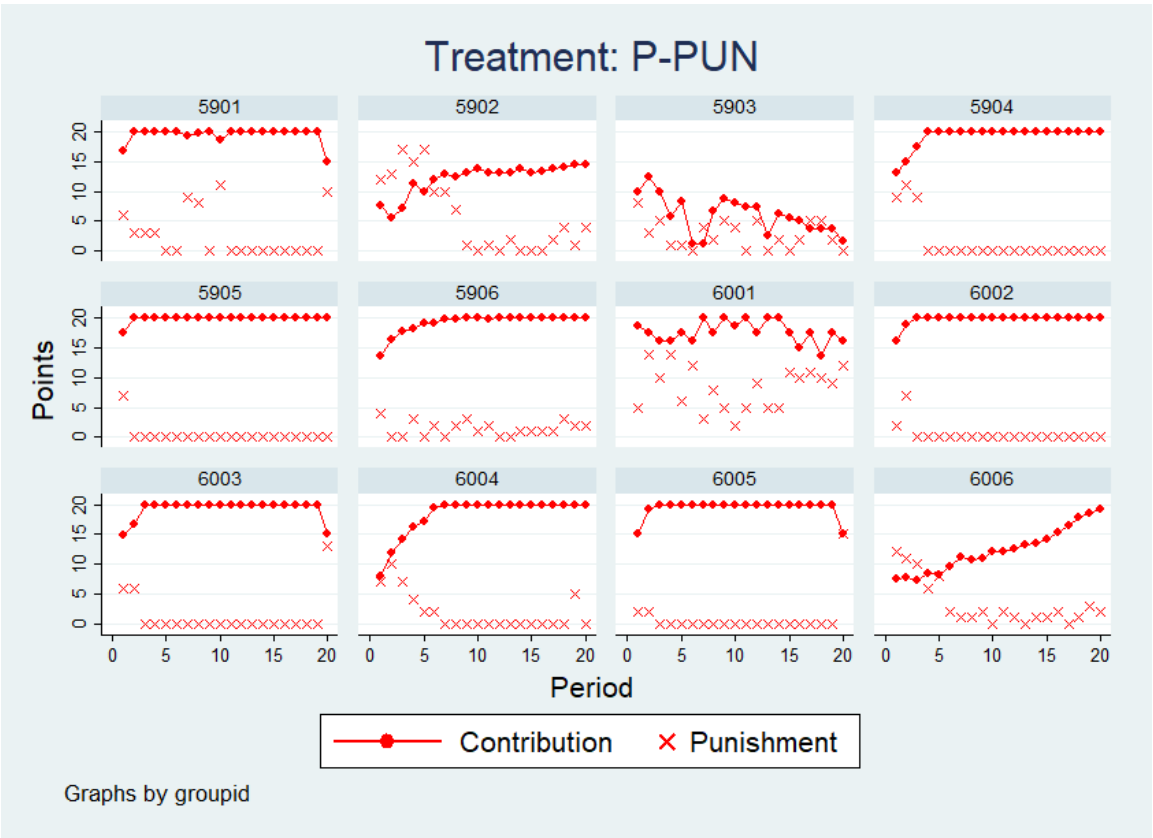
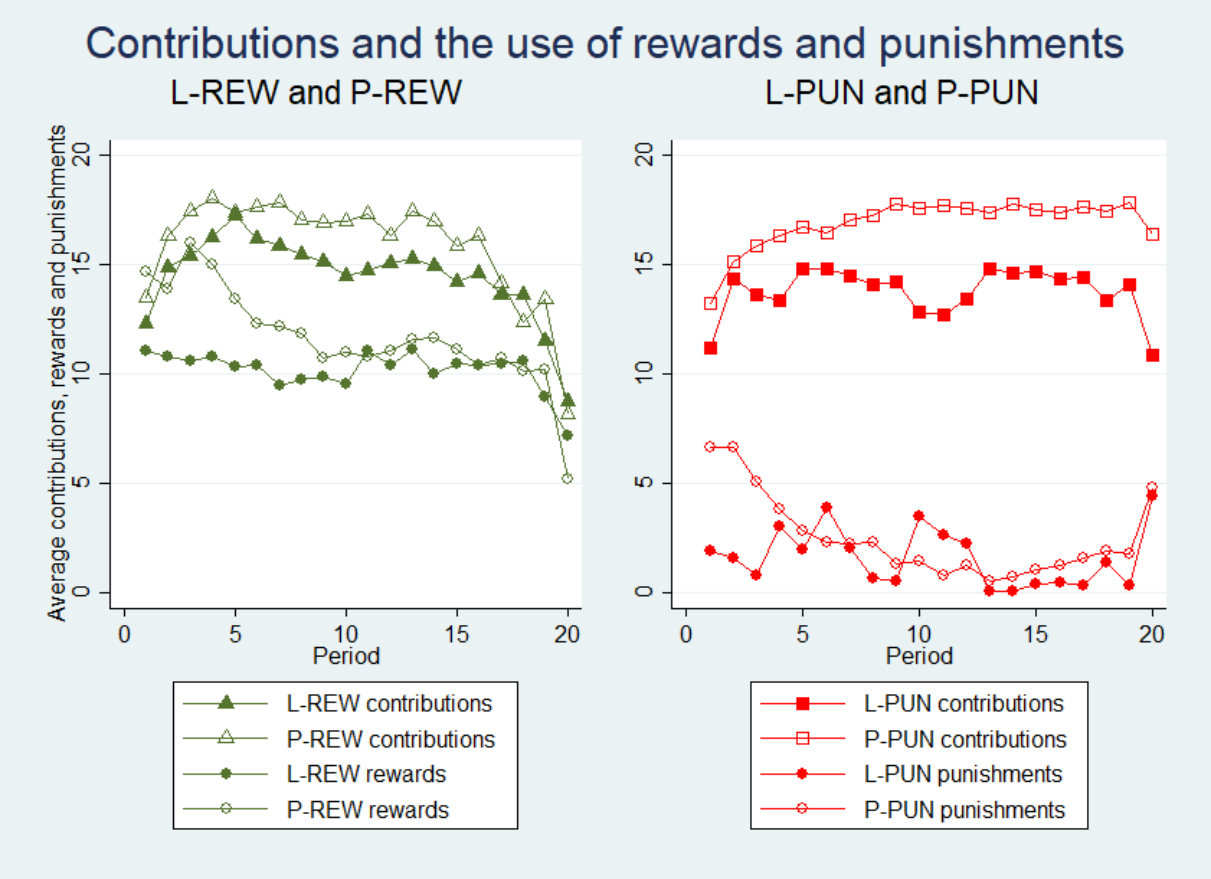


Figure A3.b: Average contributions and the sum of allocated punishment points within single groups of the P-PUN treatment.



A.4 Contributions and Reward or Punishment Points.

Figure A4: Average contributions, and average allocated rewards and punishments.



Rewards treatments show a negative trend in contributions. While in treatments with punishment possibilities, contributions remain stable in the second half (round 11-20) of the experiment compared to the first half (round 1-10), in treatments with rewards, contribution declines significantly. The decline in contributions in reward treatments compared to punishment treatments is in line with findings of other studies (see, e.g., Gürerk et al. 2014, Sutter et al. 2010).