

Glöckner, Andreas; Kube, Sebastian; Nicklisch, Andreas

## Conference Paper

# The benefits of latent payback in social dilemmas

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Voluntary Cooperation, No. E15-V2

### Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Glöckner, Andreas; Kube, Sebastian; Nicklisch, Andreas (2010) : The benefits of latent payback in social dilemmas, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Voluntary Cooperation, No. E15-V2, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37280>

### 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.*

# The benefits of latent payback in social dilemmas

Andreas Glöckner\*

Sebastian Kube\* ¶

Andreas Nicklisch\*

\* *Max Planck Institute for Research on Collective Goods,  
Kurt-Schumacher-Str. 10, 53113 Bonn, Germany*

¶ *Department of Economics, University of Bonn, Adenauerallee  
24-42, 53113 Bonn, Germany*

email: [surname]@coll.mpg.de

November 2, 2009

## Abstract

Latent payback is a natural element of social interactions: Non-cooperators face substantial threats of not being supported in situations of dire need, or of being punished in seemingly unrelated situations. In the controlled environment of the laboratory, we experimentally explore the effects of latent sanctions on cooperation *per se*, as well as on the effectiveness of immediate sanctions. Our data indicate that latent sanctions alone are able to sustain cooperation in groups. However, the real power of latent sanctions is unleashed when combined with immediate sanctions. These situations enormously enhance cooperation within groups – and, do so moreover, with less intense sanctioning. This surprising result provides a potential explanation for the co-existence of latent and immediate sanctions in social interactions.

*Keywords: Decentralized Punishment, Latent Sanctions, Public Goods, Sanctioning Effectiveness, Voluntary Contribution Mechanism, Laboratory Experiment*

JEL: C92, H40, H41, K00

A key element in societies is the maintenance and stabilization of cooperation. Mechanisms that support cooperation even in the absence of centralized institutions have been subject to intense research in social sciences (e.g., Trivers, 1971, Axelrod & Hamilton, 1981, Gintis, 2000, Gürerck et al., 2006). Sanctioning mechanisms in the private provision of public goods have drawn special interest (e.g., Ostrom, et al., 1992, Fehr & Gächter, 2000, 2002, Masclet et al., 2003, Nikiforakis & Normann, 2008, Herrmann et al. 2008). Here, persons form a group and choose to contribute to a joint project (the *public good*), with the total amount of contributions being increased by a certain factor and equally divided so that that persons face a social dilemma: although full contribution to the joint project is socially efficient, each individual has a (monetary) incentive not to contribute but to free-ride on the others' contributions. Contributions thus serve as a proxy for cooperativeness in the group. In these situations, where social and self-interest are in conflict, cooperation tends to break down over time. When persons are given the opportunity to apply costly sanctions, subsequent punishment of free-riders regularly stabilizes cooperation (Herrmann et al., 2008). This is not only true in humans. Animals also use punishment to enforce cooperative behavior. For example, chimpanzees form supportive coalitions to gain access to resources and attack allies that do not support them (de Waal, 1982), and superb fairy-wrens punish designated helpers who do not support the group in nurturing the offspring (Mulder & Langmore, 1993).

The main motive for the use of punishment, besides venting one's anger (Vyrastekova et al., 2008), is to train of defectors (de Quervain et al., 2004). One might expect sanctions to be most effective in achieving educational aims when punished individuals instantaneously learn about their received sanctions; i.e., if consequences *immediately* follow the behavior (Skinner 1938, 1969). Therefore, research on decentralized sanctioning has focused on mechanisms that provide *immediate* feedback about punishment (hereafter: *immediate sanctions*). Interestingly, however, many situations involve *latent* feedback about punishment (hereafter: *latent sanctions*). In these cases, the punished person only is subject to the severe consequences of sanctions after a time delay: e.g., group members promise not to support a defector in the future; neighbors spread rumors which later ruin the defector's reputation in the neighborhood; colleagues might start workplace bullying as a reaction to peers' free-riding. Furthermore, punishment and its consequences in animals are often widely spaced in time (Clutton-Brock & Parker, 1995). But why do such latent sanctions exist? As we will demonstrate in this article, latent sanctions seem to be an essential ingredient for the stabilization of cooperation in groups. In particular, as our results suggest, the combination of

immediate and latent sanctions leads to an exceptional increase in cooperation and efficiency rates.

**METHOD** We conducted a series of laboratory experiments to clearly isolate the effects of latent sanctions on cooperation behavior. Players in our design are anonymously matched in groups of four and play ten consecutive periods of a voluntary contribution game with punishment options. Each period consists of two stages. In stage one, players are endowed with 20 tokens each and simultaneously decide how many of the tokens to transfer to a group account. The sum of contributions to the group account are multiplied by 0.4 and provided to each group member. Hence, each player profits equally from the group account, independent of his or her contribution. The total sum of income within the group is maximized if all group members contribute fully, but given the other players' contribution decisions, each player could increase his or her individual payoff by withholding his or her own contribution. Thus, the game reflects exactly the kind of social dilemma that is present in many situations in everyday life.

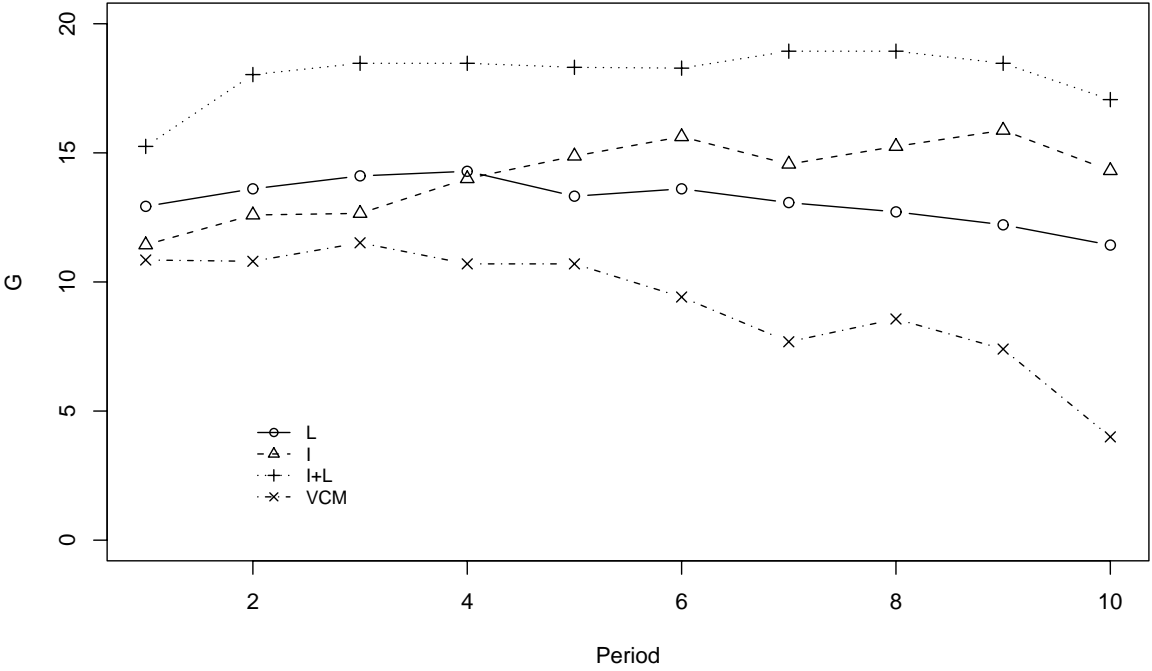
In stage two, after having observed the contribution decisions of all group members, each player can assign up to ten sanctioning points to any other group member. Each point costs one token for the player who assigns it, while reducing the earnings of the player who receives the point by three tokens. Finally, all players are informed about their own income in this period, respectively about their total earnings after the last period. Assigning points provides a second-order public good for the other group members: all group members benefit from disciplining the non-cooperator, but the player who actually disciplines bears the full costs alone. (Not only) economists have frequently pointed out that selfish players should not be expected to punish in the finitely repeated version of this game. Consequently, players anticipating this should also be reluctant to contribute to the initial public good in stage one for the same reasons.

Our first treatment replicates the existing standard sanctioning mechanism with immediate feedback over received punishment (I) (cp. Fehr and Gächter, 2002). Here, each player immediately learns about his received sanctioning points at the end of each period. The second treatment modifies the sanctioning mechanism so that received points are accrued over all periods and are deducted (and observable) only after the final period of the experiment (L). This treatment enables us to test whether the exclusive use of latent sanctions is able to maintain cooperation. Finally, the third treatment features both mechanisms (I+L), where players can choose in each period how many sanctioning points they want to carry out

immediately and how many they want to apply latently. This treatment allows us to test how the co-existence of immediate and latent sanctions affects cooperation.

In total, we ran 9 sessions with 23 groups (92 subjects), resulting in 960 punishment decisions and 8 independent group observations in the I and I+L treatments and 840 punishment decisions and 7 independent group observations in the L treatment. The appendix contains details on the experimental procedure. The computerized experiments were conducted in Bonn in 2008 (Greiner, 2004; Fischbacher, 2007). For comparison, we also include the data of a voluntary contribution mechanism experiment without punishment published in Herrmann et al. (2008), which also took place in Bonn and used exactly the same set of parameters (denoted as VCM).

Figure 1: Average contributions over periods and treatment conditions



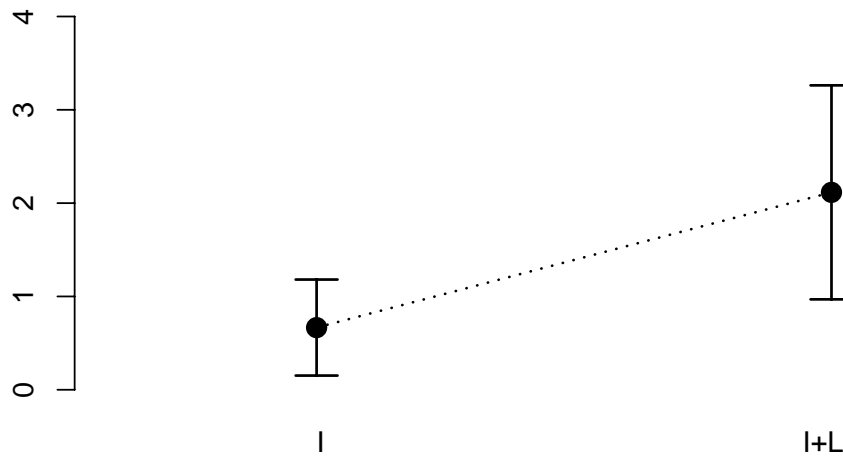
**RESULTS** Figure 1 illustrates the average contributions over time for the three treatment conditions. While contributions in the absence of punishment exhibit the decline over time usually observed, the three sanctioning mechanisms foster cooperation. Average distributions in all three punishment conditions are higher than in VCM [exact rank-sum test, here and in the following with group averages over all ten periods as independent observation, two-sided; comparing VCM to I+L, I, and L;  $p < 0.001$ ,  $p = 0.03$ , and  $p = 0.07$ ]. Contributions in I rise over time and are maintained over a finitely repeated number of games. A similar (though less

distinct) effect is observed in L, and there is no significant difference between average contribution levels in treatments I and L [exact rank-sum test;  $p=0.69$ ]. Even if we focus on the first five or on the last five periods, there is no significant difference [exact rank-sum test;  $p=0.87$ , and  $p=0.23$ ]. Yet, comparisons between I+L and I as well as between I+L and L reveal significant differences [exact rank-sum test;  $p=0.005$  and  $p=0.015$ ]. Moreover, contributions in the I+L treatment are already higher in the first period than in the other treatments [exact rank-sum test;  $p=0.006$  and  $p=0.087$ ], indicating the correct anticipation that the combination of both mechanisms allowing an extremely effective discipline of free-riders, meaning that players increase their contributions from the outset.

We observe no significant differences between the number of sanctioning points distributed in I (on average 0.43 points) and L (on average 0.5) [exact rank-sum test;  $p=0.61$ ]. In I+L, where both types of points are available, subjects assign on average 0.13 immediate sanctioning points and 0.09 latent sanctioning points. The types of points are more often substitutes for one another rather than a complementary means, however, because only 16% of all punishment decisions in I+L use both type of points simultaneously.

Interestingly, our results suggest that although the cooperation is strikingly higher in I+L, this increased cooperation is associated with less intense sanctioning. Specifically, Probit regressions indicate a significantly lower probability that players assign either immediate or latent sanctioning points (irrespective of the number of points). Furthermore, regression results show that the number of both types of points assigned decreases significantly in I+L compared to the other treatment conditions, even when controlling for the differences in contributions between the treatments [see the supplementary material for regression details]. This finding is also reflected in the sanctioning effectiveness of immediate sanctioning points. By sanctioning effectiveness, we mean the change in players' contribution (between the period where they were punished and the subsequent period) per immediate sanctioning point. Average sanctioning effectiveness in conditions I and I+L are shown in Figure 2. We find an average sanctioning effectiveness of 0.67 in the I condition, i.e. that the sanctioned player increases his or her average contribution by 0.67 tokens in the subsequent period. In contrast, we find a significantly higher average sanctioning effectiveness of 2.12 in the I+L condition [exact rank-sum test,  $p=0.04$ ]. This means that the effect of punishment on contributions is more than tripled when immediate sanctions are accompanied by latent sanctions, making immediate sanctions highly productive in the I+L condition compared to the I condition.

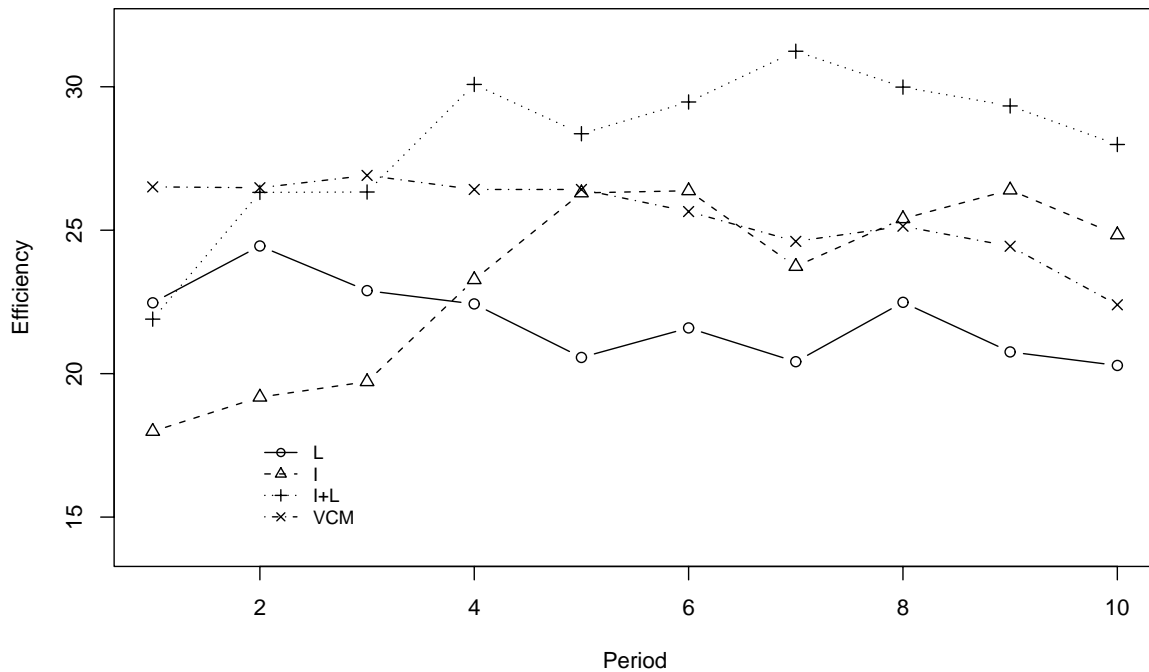
Figure 2: Average sanctioning effectiveness of immediate sanctioning points



Note: Bars indicate the 95% confidence intervals

Figure 3 shows the development of efficiency – defined as players’ average payoff – over time. On average, players earn 28.1 tokens (out of a maximum of 32) in I+L, 21.8 in L, 23.3 in I, and 25.5 in VCM. Thus, average efficiency is highest in the I+L condition [exact rank-sum test, comparing I+L to I, L and VCM;  $p=0.02$ ,  $p=0.01$ , and  $p=0.03$ ]. In contrast, the sanctioning mechanisms when isolated do not lead to better efficiency rates when compared to the voluntary contribution mechanism without punishment [exact rank-sum test, comparing VCM to I and L;  $p = 0.27$ , and  $p = 0.12$ ]; we might expect, however, the immediate sanctioning mechanism to enhance efficiency if the number of periods were sufficiently large (Gächter et al., 2008). Still, the mere fact that both sanctioning mechanisms are available together tremendously increases the efficiency of group cooperation in this dilemma situation within a few periods and furthermore does so *without* the substantial short-run efficiency losses due to punishment. In our view, this ultimately provides a possible explanation why the combination of latent and immediate sanctions is present in many everyday situations.

Figure 3: Average efficiency over periods and treatment conditions



**DISCUSSION** Previous evidence has shown that the availability of immediate sanctioning mechanisms leads to stable cooperation in social dilemmas, but usually at the cost of efficiency losses due to punishment activities in the short run. We find that latent sanctions also serve this purpose and lead to a comparable level of cooperation. Most interestingly, however, our findings demonstrate that the co-existence of both potential threats leads to higher cooperation and yields tremendous efficiency gains. The mere existence of latent sanctions more than triples the sanctioning effectiveness of immediate sanctions. Moreover, both the number of immediate and latent sanctioning points are significantly reduced.

Our results might provide a potential explanation why a combination of both sanctions is available in many real-world situations. Persons may potentially view immediate sanctions as warning signs of harsh latent consequences and therefore adapt their behavior and cooperate. One might hypothesize that this behavioral mechanism relies on some kind of deeply rooted experience to avoid provoking latent payback (e.g., a later passionate burst of anger).

In this spirit, our experiment shows the beneficial effects of the co-existence of immediate and latent sanctioning mechanisms in social dilemmas. Due to its appealing simplicity and practicability, one should consider those means as important self-reinforcing elements that stabilize cooperation in societies.



## References

- Axelrod, R. & W.D. Hamilton (1981). The evolution of cooperation. *Science* 211, 1390-1396.
- Clutton-Brock, T.H. & G.A. Parker (1995). Punishment in animal societies. *Nature* 373, 209-216.
- de Quervain, D.J.-F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck & E. Fehr (2004). The neural basis of altruistic punishment. *Science* 305, 1254-1258.
- de Waal, F.B.M. (1982). *Chimpanzee Politics*. London: Allen & Unwin.
- Fehr, E. & S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980-994.
- Fehr, E. & S. Gächter (2002). Altruistic punishment in humans. *Nature* 415, 137-140.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171-178.
- Gächter, S., E. Renner & M. Sefton (2008). The long-run benefits of punishment. *Science* 322, 1510.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206, 169-179.
- Greiner, B. (2004). An online recruitment system for economic experiments. In: *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, edited by K. Kremer and V. Macho, 79-93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.
- Güerke, Ö., B. Irlenbusch & B. Rockenbach (2006). The competitive advantage of sanctioning institutions. *Science* 312, 108-111.
- Herrmann, B., C. Thöni & S. Gächter (2008). Antisocial punishment across societies. *Science* 319, 1362-1367.
- Masclot, D, C. Noussair, S. Tucker & M.C. Villeval. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review* 93, 366-380.
- Mulder, R.A. & N.E. Langmore (1993). Dominant males punish helpers for temporary defection in superb fairy-wrens. *Animal Behaviour* 45, 830-833.
- Nikiforakis, N. & H.T. Normann. (2008). A comparative-statics analysis of punishment in public-good experiments. *Experimental Economics* 11, 358-369.
- Ostrom, E., J. Walker & R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86, 404-417.
- Skinner, B.F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Skinner, B.F. (1969). *Contingencies of reinforcement*. East Norwalk, CT: Appleton-Century-Crofts.
- Trivers, R.L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology* 46, 35-57.
- Vyrastekova, J., Y. Funaki & A. Takeuchi (2008). Strategic vs. non-strategic motivations of sanctioning. Tilburg University Discussion paper.

# The benefits of latent payback in social dilemmas: Supplementary Materials

The supplementary material in Appendix A provides a detailed description of the experiment and the instructions. We report a detailed analysis of sanctions in Appendix B, and we display the groups' mean contribution levels over time in Appendix C.

## Appendix A: Experimental method

At the beginning of each session, participants had to draw lots, in order to assign each of them to a carrel, where we asked them to take their seats immediately. Once all subjects were seated, instructions were distributed and read out aloud. Afterwards, participants could pose clarifying questions to the experiment supervisor in private. After questions were answered individually, participants had to answer a set of control questions to ensure that everybody had understood the game.<sup>1</sup> Control questions were corrected individually, and wrong answers were explained privately. Participants were randomly and anonymously matched in groups of four players each. The composition of the group remained constant throughout the entire 10 periods of the experiment. To prevent subjects from identifying each other across periods, each received a new identification number between 1 and 4 at the beginning of each period; the number only remained in force for that period.

Altogether, 92 subjects, mostly students from the University of Bonn majoring in various fields, participated in the experiment (10 percent were non-students). Mean age was 24.5 years (standard deviation 5.5 years), 62 percent were females. Each subject participated only once in the experiment. None of the participants had participated in a public good experiment before. A session lasted for about 60 minutes. Tokens earned were accrued over all periods and converted at an exchange rate of 3 Euro per 100 tokens. Participants were paid out individually to ensure their anonymity. They earned on average 13.86 Euro<sup>2</sup> (standard deviation 1.45 Euro), including a show-up fee of 5 Euro.<sup>3</sup>

---

<sup>1</sup> Questions are almost identical to the control questions of Herrmann et al. (2008). In one of the original questions, the sum of punishment points exceeds ten which was not possible in our study. We adjusted numbers in such a way that they sum up to a value below ten (i.e., nine).

<sup>2</sup> Corresponds to \$20.40 (situation November 2009)

<sup>3</sup> Notice that since actual period payoffs could be negative due to costs for deduction points or the punishment of deduction points (which rarely occurred), all players received an additional endowment of 50 tokens at the

## Instructions for the I+L condition<sup>4</sup>

### ***General explanations for participants***

You are taking part in an economic science experiment. If you read the following explanations closely, then you can earn a rather significant sum of money, depending on the decisions you make. It is therefore very important that you pay attention to the following points.

The instructions you have received from us are intended solely for your private information. *During the experiment, you will not be allowed to communicate with anyone.* Should you have any questions, please direct them directly to us. Not abiding by this rule will lead to exclusion from the experiment and from any payments.

In this experiment, we calculate in Taler, rather than in euro. Your entire income will therefore initially be calculated in Taler. The total sum of Taler will later be calculated in euro as follows:

$$1 \text{ Taler} = 3 \text{ Euro cent}$$

The Euro you will have accumulated will be paid to you *in cash* at the end of the experiment.

The experiment is divided into separate periods. It consists of a total of 10 periods. Participants are randomly assigned into groups of four. Each group, thus, has three further members, apart from you. During these 10 periods, the constellation of your group of four will remain unaltered. *For 10 periods you will therefore be in the same group.* Please note that the identification number assigned to you and the other members of the group changes randomly in each period. Group members can therefore not be identified as the periods progress.

Each participant will receive from us one instalment of 50 Taler, with which you will be able to counterbalance potential losses.

---

beginning of the experiment. However, no player accrued an overall negative payoff at the end of the experiment.

<sup>4</sup> Differences to the (L), resp. to the {I} condition, are marked with brackets, resp. with braces. Screens differed accordingly.

The following pages outline the exact procedure of the experiment.

### ***Information on the exact procedure of the experiment***

#### ***Step 1***

At the beginning of each period, each participant is allotted *20 Taler*, which we shall henceforth refer to as his *endowment*. The player's job is now to make a decision with regard to using his endowment. You have to decide how many of the 20 Taler you wish to pay into a *project* and how many you wish to keep for yourself. The consequences of your decision are explained in greater detail below.

At the beginning you will see the following contribution screen:



Period 1 Remaining time (sec): 107

In this period you are group member 1

Step 1

You receive in this period an endowment of 20 Taler.

Out of these 20 Taler you pay ...

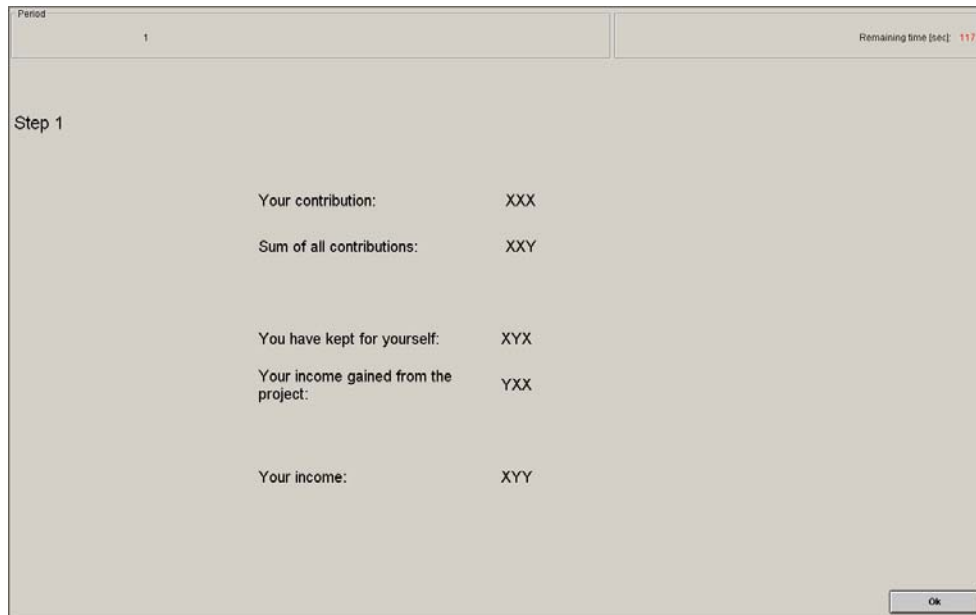
... into the project:

OK

In the left upper corner of the screen you will find the *period number*. In the right upper corner you will find the *remaining time* for your decision *in seconds*.

Your *endowment* is, thus, *20 Taler in each period*. You make a decision on your project contribution by typing any one whole number between 0 and 20 into the appropriate field on your screen. This field can be accessed using the mouse. As soon as you have determined your contribution, you have also decided on how many Taler to keep for yourself, i.e., *20 – your contribution*. Once you have typed in your contribution, please click on *Continue*, again using the mouse. Once you have done this, your decision for this period is irreversible.

Once all members of the group have made their decisions, you will be told how high the total sum of contributions from *all* group members (including your own) to the project is. In addition, you are informed about your own contribution and the number of Taler kept by you; you are also told how many Taler you have earned in total during Step 1.



*Your income* therefore consists of two parts, namely:

- (1) the Taler you have kept for yourself
- (2) the "*income gained from the project*". Your income from the project is calculated as follows:

$$\begin{aligned} & \text{Income from the project} \\ &= .4 \times \text{total sum of all contributions to the project} \end{aligned}$$

Your *income in Taler* in each period equals

$$(20 - \text{Your contribution to the project}) + .4 * (\text{total sum of contributions to the project})$$

The total income at the end of Step 1, in Taler, is calculated according to the same formula for each member of the group. If, for example, the sum of the contributions from all group members adds up to 60 Taler, you and all other members each receive a project income of .4x

$60 = 24$  Taler. If the group members have contributed a total of 9 Taler to the project, you and all other members each receive an income of  $.4 \times 9 = 3.6$  Taler from the project.

For each Taler you keep for yourself, you earn an income of 1 Taler. If, on the other hand, you contribute one Taler from your endowment to your group's project instead, the sum of the contributions to the project increases by one Taler and your income from the project increases by  $.4 \times 1 = .4$  Taler. However, the income of each individual group member also increases by  $.4$  Taler, so that the group's total income increases by  $.4 \times 4 = 1.6$  Taler. The other group members thereby also profit from your contributions from the project. In turn, you profit from other members' contributions to the project. For each Taler contributed to the project by another group member, you earn  $.4 \times 1 = .4$  Taler.

## **Step 2**

In Step 2, you can *decrease, or leave as it is*, the income of each individual group member by giving *points* [*mediate points*]. You have the opportunity to assign two different types of points, immediate and mediate points. The income reduction through immediate points takes place at the end of each period. The income reduction through mediate points takes place only after the end of the experiment. This means that mediate points you have received throughout the experiments will be accumulated and deducted from our total income at the end of the experiment. [The income reduction through mediate points takes place only after the end of the experiment. This means that mediate points you have received throughout the experiments will be accumulated and deducted from your total income at the end of the experiment.] {The income reduction through points takes place at the end of each period.} All other group members are allowed to decrease *your* income, too, if they so wish. You will see this when considering the input screen of the second step.

You will be shown on the screen, along the number of periods and the remaining time, how many Taler each individual group member has contributed to the project. Your contribution will be shown in the row "You", while the contributions of all other three group members will be shown in randomly changing rows over periods.

Period 1 Remaining time [sec]: 110

Step 2

Groupmember	Contribution	Immediate points	Mediate points
You	XXX		
Group member 2	YYY	<input type="text"/>	<input type="text"/>
Group member 3	YXY	<input type="text"/>	<input type="text"/>
Group member 4	YYX	<input type="text"/>	<input type="text"/>

Your income in Taler from step 1 is: XYY

OK

You now have to decide for every group member about the combination of two types of points [points] {points} you wish to distribute to them. It is compulsory to enter a number at this stage. If you do not wish to alter a certain group member's income, please insert 0. If you want to assign points you have to choose a number greater than 0. You can operate within the fields by using the tab key (→|) or the mouse.

When distributing points, you incur costs in Taler which depend on the number of points you distribute to the individual players. *Distributed points are numbers between 0 and 10.* This means that the sum of immediate and mediate points per group member and period need not to exceed 10. [*Distributed points are numbers between 0 and 10.*] {*Distributed points are numbers between 0 and 10.*} The more points you give an individual player, the higher your costs are. The total costs in Taler are calculated as the sum of the costs of all points distributed to all other group members. The following formula shows the connection between the points distributed to an individual group member and the costs of such distribution:

$$\text{Costs for assigned points} = \text{sum of immediate and mediate [mediate] \{ \} points (in Taler)}$$

Each distributed point costs you 1 Taler. For example, if you have allocated 2 points to one member, your cost is 2 Taler; if, in addition, you give 9 points to another group member, your cost is 9 Taler; if you give the final group member 0 points, you have no costs. The *total cost* to you is therefore 11 Taler (2+9+0). As long as you have not yet clicked on *Continue*, you may still change your decision.

If you choose 0 points for a certain group member, you do not alter this group member's income. If you allocate 1 point (choosing 1) to a group member, you *decrease* this particular group member's Taler income from Step 1 by 3 Taler. If you allocate 2 points to a group member (choosing 2), you decrease his income by 6 Taler etc. *Each point allocated by you to a particular group member reduces the group member's Taler income from Step 1 by 3 Taler.*

By how much a group member's income from Step 1 is reduced *overall* depends on the total number of points received. If, for instance, one member receives *a total of 3 points* from all other members, the income in Step 1 is reduced by 9 Taler. If a member receives a total of 4 points, the income in Step 1 is reduced by 12 Taler.

A person who receives immediate points will be informed about the income reduction immediately at the end of each period, without detailed information on the group member who distributed that point. The reduction of income by mediate points will be revealed not after each period, but only after the final period of the experiment. This means that all received mediate points are accumulated over periods and group members who have distributed them and deducted from the total income after the experiment, without detailed information on the period and the group member who has distributed that point. [The reduction of income by mediate points will be revealed not after each period, but only after the final period of the experiment. This means that all received mediate points are accumulated over periods and group members who have distributed them and deducted from the total income after the experiment, without detailed information on the period and the group member who has distributed that point.] {A person who receives immediate points will be informed about the income reduction immediately at the end of each period, without detailed information on the group member.} For your total income at the end of step 2, it follows that:



*Total Taler income at the end of step 2 = Period income*

= Income after step 1  
– 3\*(sum of received immediate points)  
– cost of points distributed by you

[= Income after step 1  
– cost of points distributed by you]  
{= Income after step 1  
– 3\*(sum of received points)  
– cost of points distributed by you}

Please note that your total Taler income at the end of step 2 can be negative if the costs for distributed points exceed the income after step 1 minus the reduction of income due to received immediate points. [Please note that your total Taler income at the end of step 2 can be negative if the costs for distributed points exceed the income after step 1.] {Please note that your total Taler income at the end of step 2 can be negative if the costs for distributed points exceed the income after step 1 minus the reduction of income due to received points.}

Once all members of the group have made their decisions, you will be informed about your period income in the following screen:

Period 1		Remaining time [sec]: 116
Step 2		
Your income from step 1:	XYY	
cost of points distributed (in Taler):	-yyy	
Immediate points you have received:	xxx	
Your Income reduction due to receive immediate points (in Taler):	-xxx	
Your Taler income in this period:	XYX	

Your total income at the end of the experiment equals the sum of all period incomes minus the sum of mediate points: [Your total income at the end of the experiment equals the sum of all

period incomes minus the sum of mediate points:] {Your total income at the end of the experiment equals the sum of all period incomes :}

$$\begin{aligned} & \textit{Total income (in Taler)} \\ & = \text{Total sum of period incomes} \quad (1) \\ & - 3 * (\text{sum of received mediate points}) \quad (2) \end{aligned}$$

(If the deduction (2) is larger than the sum of period incomes (1), your income is *0 Taler.*)

$$\{ = \text{Total sum of period incomes} \quad (1)$$

(If the sum of period incomes (1) is negative, your income is *0 Taler.*)}

Do you have any further questions?

# Appendix B: Analysis of punishment

We focused on the intensity and effectiveness of immediate and latent sanctions in our main text. To supplement the reported findings, this section describes subjects’ sanctioning behavior in more detail.

We report in the paper that the strikingly high degree of cooperation in I+L is obtained by less intense sanctioning: In the absence of additional control, a bare comparison of average sum of points is insignificant [exact rank-sum test, comparing L and the average sum in I+L, resp. comparing I and the average sum in I+L;  $p=0.24$  and  $p=0.12$ ]. However, an important – but often neglected – aspect in comparing the average punishment over all observations is the fact that contribution levels also differ between treatments. The question then is whether sanctions are less intense *ceteris paribus*, i.e., after taking that the high cooperation levels in treatment I+L into account. Our analysis of sanctioning effectiveness already point in this direction. To stress it further, we compare the sanctioning intensity across treatments for a given kind of “conduct” or “norm violation”. Following the previous literature on social sanctions (e.g., Herrmann et al., 2008), we define the conduct by the difference in contributions between the punisher (denoted as  $g_i$ ) and the sanctioned person ( $g_j$ ) – the idea being that the difference in contributions (i.e.,  $g_j - g_i$ ) measures the severity of the norm violation.

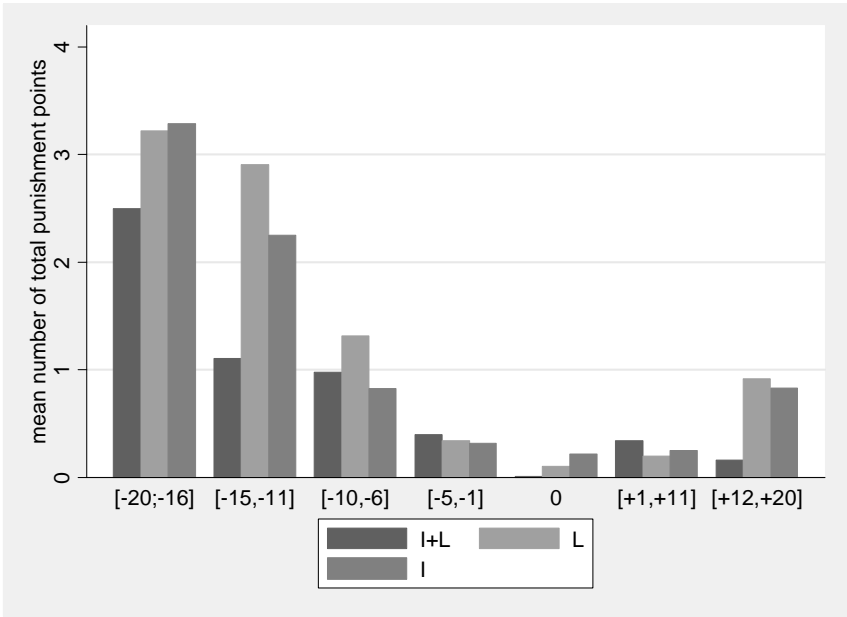


Figure B1: Average amount of punishment (immediate+latent)

Figure B1 compares the punishment intensity for fixed intervals of norm violations between our three treatments. We observe that punishment in I+L is less intense for almost all categories of norm-violations, suggesting that, *ceteris paribus*, less intense sanctioning is required to obtain the cooperation in I+L. As we will see below in the regression analysis, this even holds true after controlling for the other group members' contribution levels.

Before turning to this issue, let us turn to the phenomenon of anti-social punishment as it is visible in the right-hand bars column in Figure B1 above. Earlier research already demonstrated that punishment is sometimes used anti-socially, that is, actions which benefit the society are occasionally sanctioned (e.g., Herrmann et al., 2008). In our game, this happens if a player receives punishment points from someone who contributed less (or the same) than he or she did. We denote this as anti-social punishment, while we refer to pro-social punishment if points are assigned from a player who contributed more than the recipient. Interestingly, the combination of immediate and latent sanctions (treatment I+L) seems to decrease the amount of anti-social punishment in our groups (at least the severe anti-social punishment, i.e., in the rightmost interval of Figure B1). This can also be seen in Table B1 below, where we report descriptive statistics of distributed sanctioning points (as well as sanctioning effectiveness) across treatment conditions, both for pro-social and anti-social instances.

Table B1: Descriptive statistics of sanctioning points

	L	I	I+L
<i>p</i> (immediate sanctions)		0.43	0.13
<i>p</i> (pro-social)		0.76	0.64
<i>p</i> (anti-social)		0.27	0.03
<i>s</i> (latent sanctions)	0.50		0.09
<i>s</i> (pro-social)	0.90		0.34
<i>s</i> (anti-social)	0.22		0.04
sanctioning effectiveness	0.22 <sup>5</sup>	0.67	2.12

Furthermore, we conducted a regression analysis in which we differentiate between pro-social and anti-social sanctions in greater detail. Our regression results stress that anti-social

---

<sup>5</sup> Notice that this number represents the “anticipated” sanctioning effectiveness of a latent point in L. Of course, this number is purely hypothetical as punished players do not receive feedback on latent points. For I and I+L, we report the sanctioning effectiveness of immediate points.

punishment is present in all three treatment conditions, but that it is less intense in I+L. Interestingly, we also find no evidence that anti-social punishment is predominantly executed by immediate or latent feedback points if both sanctioning mechanisms are applicable (i.e., in treatment I+L).

For our regression analysis, let us define the two dummy variables  $I_{p>0}$  and  $I_{s>0}$  which equal one if player  $i$  assigns immediate, resp. latent, punishment points to player  $j$ , and zero otherwise (to avoid confusing the reader with our treatments identifier, we use  $p$  when referring to immediate punishment points and  $s$  when referring to latent punishment points in our analysis).  $I_{p>0}$  and  $I_{s>0}$  are the dependent variables in two distinct estimations. Further, as independent variables, we introduce the contribution  $g_j$  of the person punished, the contributions of the remaining two group members  $G_{kl}=g_k + g_l$ ,  $k, l \neq i, j$ , and the absolute difference between contributions  $d_{ij}^+ = |\max(g_i - g_j, 0)|$  and  $d_{ij}^- = |\min(g_i - g_j, 0)|$ . We also add a dummy variable  $I_{I+L}$  indicating the I+L condition, and interaction terms  $d_{ij}^+ I_{I+L}$  and  $d_{ij}^- I_{I+L}$ . Therefore,  $g_j$  indicates the effect of the contribution of player  $j$  on the probability of being punished, while  $G_{kl}$  shows the effect of the contributions of other group members, showing whether being in a group of free riders or a group of full contributors affect  $i$ 's punishment decision. The two difference measures allow us to estimate how the (absolute) difference between the

punished player's and punishing player's contributions affects the decision to assign points. We differentiate between positive differences ( $d_{ij}^+$ ) and negative differences ( $d_{ij}^-$ ). Significant positive coefficients for  $d_{ij}^+$  suggest pro-social punishment, whereas significant positive coefficients for  $d_{ij}^-$  suggest anti-social punishment. Finally,  $I_{I+L}$ ,  $d_{ij}^- I_{I+L}$ , and  $d_{ij}^+ I_{I+L}$  show differences between the I+L and the L (for the dependent variable  $I_{s>0}$ ) and between the I+L and the I condition (for the dependent variable  $I_{p>0}$ ), respectively.<sup>6</sup> Table B2 reports the estimation results for the mean marginal effects of the independent variables in a probit regression.<sup>7</sup>

The econometric results underline what we observed above in Figure B1. We find evidence for pro-social and anti-social punishment both with immediate and latent points. The difference between contributions influences the decision whether to punish or not to punish,

---

<sup>6</sup> Of course, the first estimation contains only observations from the I+L and the L conditions, while the second estimation contains only observations from the I+L and the I conditions.

<sup>7</sup> Standard errors are clustered for each group over the entire 10 periods.

as it is indicated by the significant positive marginal effects of  $d_{ij}^-$  and  $d_{ij}^+$ . However, the punished player's absolute contribution level influences the probability that latent and immediate punishment occurs. As one would expect, the probability decreases for higher contributions. Likewise, increasing the contributions of the other players significantly increases the probability that punishment occurs. Most importantly, the significant negative marginal effect of the treatment dummy shows that there is a significantly lower probability in I+L for immediate and latent punishment as soon as we control for the contribution situation, i.e., for the contribution differences across treatments.

Table B2: Mean marginal effects of the Probit estimation

independent	dependent	$I_{p>0}$	$I_{s>0}$
$g_j$		-0.007* (0.004)	-0.010** (0.005)
$d_{ij}^+$		0.020*** (0.007)	0.017** (0.007)
$d_{ij}^-$		0.010*** (0.003)	0.012** (0.005)
$G_{k1}$		0.005** (0.002)	0.007*** (0.002)
$I_{I+L}$		-0.121** (0.047)	-0.120** (0.056)
$d_{ij}^+ I_{I+L}$		-0.019 (0.070)	-0.023 (0.121)
$d_{ij}^- I_{I+L}$		-0.012 (0.045)	-0.011 (0.061)
number of observations		1920	1800
logLik		-545	-509
PseudoR <sup>2</sup>		0.28	0.29
Wald test (7)		619***	295***

Note: Standard errors are reported in parenthesis. \*\*\* indicates significance at a  $p < 0.01$  level, \*\* at a  $p < 0.05$  level and \* at a  $p < 0.1$  level. Marginal effects are evaluated at the means. The constant terms of the models are  $-1.579$ \*\*\* (0.394) and  $-1.584$ \*\* (0.682). The number of observations is reported along the log likelihood (logLik) and the fitness of the estimation by means of the PseudoR<sup>2</sup>. Finally, the Wald test indicates the significance of the estimation's improvement against the null model.

The same picture emerges if we look at the number of points rather than at the decision to punish or not. Therefore, for our second regression analysis, let us define the two variables  $p$

and  $s$  which measure the number of immediate (latent) punishment points player  $i$  assigns to player  $j$ . Notice that  $p$  and  $s$  are censored in the interval zero to ten, so that we apply a Tobit regression.  $p$  and  $s$  are our dependent variables in two distinct estimations. As independent variables, we use the same variables as in the first two regressions. Again, the variables  $I_{I+L}$ ,  $d_{ij}^- I_{I+L}$ , and  $d_{ij}^+ I_{I+L}$  indicate differences between the I+L and the L (I) condition. Table B3 reports the estimation results for the mean marginal effects of the independent variables in a robust least square regression.<sup>8</sup>

Results again indicate important treatment differences with respect to the number of immediate and latent punishment points assigned. The significant negative marginal effects for  $I_{I+L}$  show that players assign less immediate and less latent punishment points. Moreover, the weakly significant marginal effect of the interaction  $d_{ij}^+ I_{I+L}$  indicates less pro-social punishment for latent points. Interestingly, there is no evidence that anti-social punishment is mainly executed by latent points if both sanctioning mechanisms are available (i.e., in I+L): the marginal effect of  $d_{ij}^- I_{I+L}$  is neither significantly negative in regression model for immediate points nor significantly positive in the regression model for latent points. Concerning the other independent variables, qualitatively similar results as in the probit regressions are found.

---

<sup>8</sup> Standard errors are clustered for each group over the entire 10 periods.

Table B3: Mean marginal effects of the Tobit regression

independent	dependent	$p$	$S$
$g_j$		-0.119* (0.073)	-0.178** (0.079)
$d_{ij}^+$		0.294*** (0.089)	0.365** (0.181)
$d_{ij}^-$		0.173*** (0.041)	0.230*** (0.083)
$G_{k1}$		0.094*** (0.030)	0.108*** (0.033)
$I_{I+L}$		-2.290*** (0.775)	-2.163** (0.980)
$d_{ij}^+ I_{I+L}$		-0.043 (0.046)	-0.236* (0.127)
$d_{ij}^- I_{I+L}$		-0.129 (0.078)	-0.053 (0.101)
number of observations		1920	1800
logLik		-981	-915
Pseudo R <sup>2</sup>		0.18	0.19
F test (7,number of observations)		26.46***	45.21***

Note: Standard errors are reported in parenthesis. \*\*\* indicates significance at a  $p < 0.01$  level, \*\* at a  $p < 0.05$  level and \* at a  $p < 0.1$  level. The constant terms of the models are  $-4.581^{***}$  (1.164) and  $-4.547^{**}$  (2.177). The number of observations is reported along the log likelihood (logLik) and the fitness of the estimation by means of the PseudoR<sup>2</sup>. Finally, the F-test indicates the significance of the joint coefficients.



# Appendix C: Group's Mean Contribution over Time

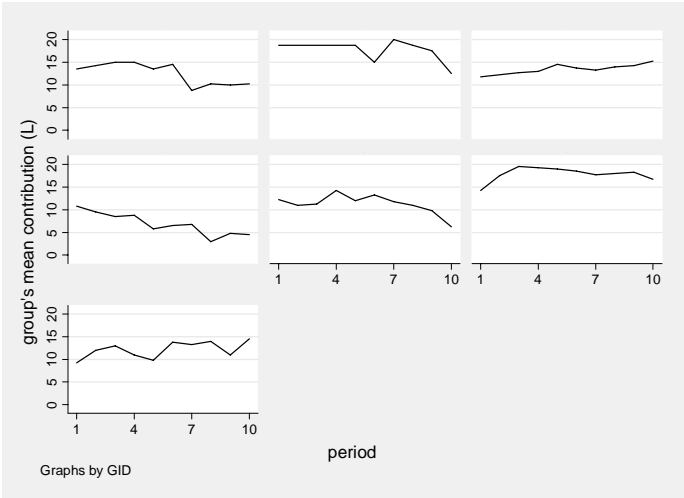


Figure C1: Contribution over time in treatment L

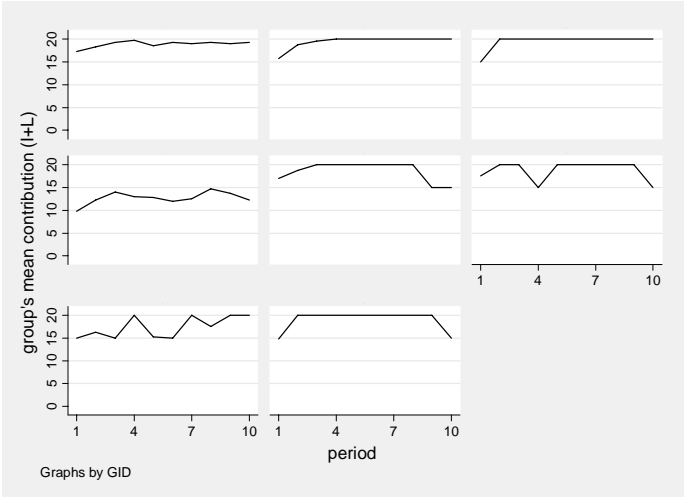


Figure C2: Contribution over time in treatment I+L

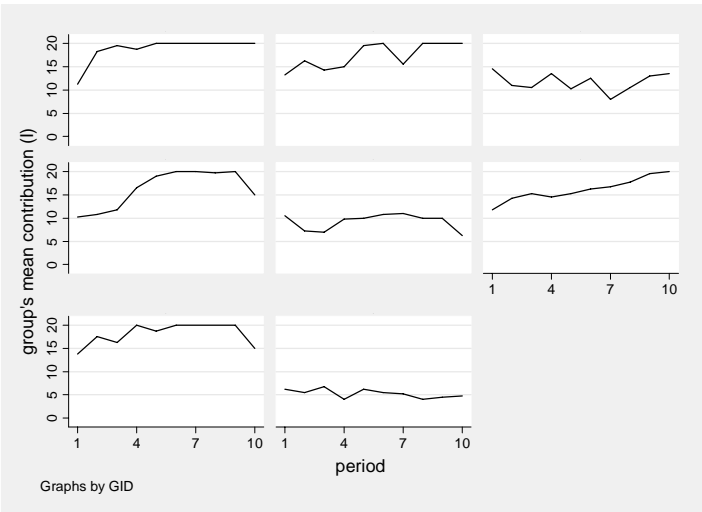


Figure C3: Contribution over time in treatment I

In the main text, we only displayed the development of average contribution over all players in a specific treatment. Figures C1-C3 above include the respective graphs for each single matching group. As can be seen, the fact that average contributions are significantly higher in I+L is not driven by a single group. Rather, all 8 groups in I+L manage to cooperate at very high levels. By contrast, we observe several groups in treatment I and in treatment L which fail to cooperate successfully.