JENA ECONOMIC RESEARCH PAPERS

2012 – 046

Motives of Sanctioning: Equity and Emotions in a Public Good Experiment with Punishment

by

**Paolo Crosetto
Werner Güth
Luigi Mittone
Matteo Ploner**

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Motives of Sanctioning: Equity and Emotions in a Public Good Experiment with Punishment

Paolo Crosetto^{a,*}, Werner Güth^a, Luigi Mittone^b, Matteo Ploner^b

^aMax Planck Institute of Economics, Strategic Interaction Group, Jena, Germany

^bDepartment of Economics - CEEL, University of Trento, Italy

Abstract

We study conditional cooperation based on a sequential two-person linear public good game in which a trusting first contributor can be exploited by a second contributor. After playing this game the first contributor is allowed to punish the second contributor. The consequences of sanctioning depend on the treatment: whereas punishment can reduce inequality in one treatment, it only creates another inequality in the other. To capture the effect of delay on punishment both treatments are run once with immediate and once with delayed punishment. Moreover, to investigate the effect of pure voice, all four treatments are also run in a *virtual* condition with no monetary consequences of punishment. Results show the emergence across all conditions of a strong norm of conditional cooperation. Punishment is generally low, it is higher when not delayed and it is not used to reduce inequality in payoffs. The main motive of sanctioning appears to be the need to punish a violation of the reciprocity norm, irrespective of monetary consequences.

JEL classification: C70; C72; C92; H41

Keywords: Public good games; Punishment; Experiments; Conditional cooperation

*Corresponding author. Max Planck Institute of Economics, Strategic Interaction Group, Kahlaische Straße 10, 07745 Jena, Germany Tel. +49.3641.686.627. E-mail address: crosetto@econ.mpg.de

1 Introduction

Punishment can be an equilibrium aspect, e.g., in the form of tit-for-tat behavior in repeated games (Aumann, 1992; Axelrod, 1984)), but usually not an equilibrium outcome.¹ This has long been known, as illustrated by the name “Folk Theorems” for theoretical analyses of such equilibrium threats (see also Ostrom (2000) for evidence of threats in recurrent field settings). Nonequilibrium threats are the main finding of ultimatum experiments where rejection of a positive offer hurts both the proposer and the responder. Whereas in the ultimatum experiment threat efficiency, as measured by the ratio of what the proposer and the responder lose by rejection, is endogenously determined by the offer, this ratio is constant and exogenously given in many recent experimental studies (see, e.g., Fehr and Gächter, 2002). Here the evidence of nonequilibrium punishment is strong, also with respect to its prevention effect; it is, however, weak when threat efficiency is low,² e.g. close to 1 (Ahlert et al., 2001).

The observation of no or little punishment when threat efficiency is close to 1 suggests that threat efficient punishment is used as a means to reduce inequality in earnings. By comparison, the purpose of voicing one’s anger by punishing seems minor since, even with low threat efficiency, one can voice one’s anger. Actually, one should voice anger by small punishment amounts when punishment efficiency is low. Thus the prediction is that:

1. there is no or little punishment when threat efficiency is low (not higher than 1), and
2. large punishments are mainly observed when threat efficiency is high (larger than 1, e.g., 2 or 3).

Our innovative attempt to compare the two ex post purposes³ of punishment, namely “inequality reduction” and “voicing one’s anger” is based on the following two-person sequential public good game:

1. first Player X chooses his contribution about which Player Y is informed, before
2. choosing his contribution as well. Thus Player X can “lead” by providing a good example to which Player Y, however, can react by free riding. This might upset
3. Player X who, knowing Y’s contribution, can finally punish Y.

According to prediction 1 above, one would expect little or no punishment if threat efficiency is low. To avoid this we have set threat efficiency constantly equal to 2, i.e., Player Y’s loss is twice as high as Player X’s cost of punishment.

We nonetheless distinguish:

1. a treatment where punishing does not reduce inequality in payoffs [1] and

¹In the case of trembles, in which deviating from equilibrium behavior can be non-intentional, one will have to punish with positive probability to discourage hiding behind trembles. This would render punishment not only an equilibrium phenomenon but also an equilibrium outcome.

²Threat efficiency can be lower than 1, e.g., in the impunity variant of the ultimatum game where the responder’s rejection destroys the responder’s but not the proposer’s payoff (see Bolton and Zwick (1995) and Güth et al. (2001) for experimental studies). In the spite variant of the ultimatum game, in which the responder’s rejection destroys only the proposer’s but not his own payoff, threat efficiency is infinitely large, inducing proposers to be very fair (Güth and Huck, 1997)

³The main ex ante purpose of sanctioning is to prevent the need to punish.

2. a treatment where, in line with most of the literature, it does [2].

We capture [1] by letting the consequences of punishment depend on a binary unbiased chance move: parties earn what they gained in the public good game (before punishment) for one chance event, whereas they receive their equal endowments minus punishment costs in case of the other chance event. This definitely implements [1] since gains, and thereby inequality in gains, from the public good game cannot be reduced when they are actually paid out and since positive punishment only creates an additional inequality in earnings in case of the other chance event. There may, of course, be an inclination of the punisher to counterbalance one inequality by an opposite other inequality.⁴ This, however, is not in line with the popular view of inequity averse social utilities (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). Treatment [2] features the usual additivity of earnings from the public good game and from punishment. Since both treatments rely on the same chance move, they yield for each choice constellation the same expected payoffs for both players, although their payoff implications differ for the two chance events.⁵

Inspired by the ultimatum experiment of Grimm and Mengel (2011), showing that delay in response cools down a frustrated responder and makes him more willing to accept, both treatments are implemented once with immediate (*hot*) and once with delayed (*cold*) punishment choice. We expect less punishment:

1. when this does not reduce payoff inequality and
2. when the choice is delayed.

Finally, to further disentangle the pure “voicing of one’s anger” via punishment from that with additional harming, we run all the treatments in a *real* condition, in which punishment has monetary consequences according to the treatments above, and a *virtual* condition, in which punishment follows the rules detailed above, especially concerning feedback information although punishment is only virtual. We expect in the *virtual* condition

1. low to no punishment and
2. high punishment in the few cases with punishment to voice one’s anger more clearly.

Altogether, we thus rely on a $2 \times 2 \times 2$ factorial design whose detailed protocols are described in section 3 after introducing the games more formally in section 2. Section 4 reports the main findings, and section 5 tests the hypotheses indicated above. Section 6 discusses the results and concludes.

2 The game models

Let X be the first contributor, choosing c_x to which Y , the second contributor, reacts by choosing $c_y(c_x)$, with $c_i \in [0, e]$ for $e = 20$ and $i = X, Y$. Contributions c_i as well as punishment

⁴Thus we do not claim to rule out reciprocal inclinations which, however, can impact in similar ways in both treatments.

⁵Thus the two treatments are equivalent from an a priori but not ex post perspective.

levels $p \in [0, P]$ with $P = 3$ are restricted to integers, meaning that one can choose only between $p = 0$ (no punishment), $p = 1$ (weak punishment), $p = 2$ (mild punishment), and $p = 3$ (strong punishment). Since both endowments are e , equal payoff can be achieved by equal contributions and no punishment since the marginal productivity of contributions is the same for both players. After his initial choice of c_x and Y 's reaction $c_y(c_x)$ to it, X is informed about c_y and finally chooses $p \in [0, P]$, with threat efficiency set to 2. The game ends with an unbiased chance move of "Head" or "Tail".

We present the payoffs in a way illustrating that both game models imply the same expected earnings for all choice constellations $(c_x, c_y(c_x), p)$. In the additive treatment [2] "Add" payoffs for X and Y , respectively, are:

$$\pi_x = \begin{cases} e - c_x + \alpha [c_x + c_y(c_x)] + e - p & \text{if chance selects "Head"} \\ 0 & \text{if chance selects "Tail"} \end{cases}$$

and

$$\pi_y = \begin{cases} e - c_y(c_x) + \alpha [c_x + c_y(c_x)] + e - 2p & \text{if chance selects "Head"} \\ 0 & \text{if chance selects "Tail"} \end{cases}$$

All that differs in treatment [1] "Alt" are the consequences of the chance move, namely:

$$\pi_x = \begin{cases} e - c_x + \alpha [c_x + c_y(c_x)] & \text{if chance selects "Head"} \\ e - p & \text{if chance selects "Tail"} \end{cases}$$

and

$$\pi_y = \begin{cases} e - c_y(c_x) + \alpha [c_x + c_y(c_x)] & \text{if chance selects "Head"} \\ e - 2p & \text{if chance selects "Tail"} \end{cases}$$

Clearly, the expected payoffs are:

$$E[\pi_x] = \frac{1}{2} [2e - c_x + \alpha [c_x + c_y(c_x)] - p]$$

and

$$E[\pi_y] = \frac{1}{2} [2e - c_y(c_x) + \alpha [c_x + c_y(c_x)] - 2p]$$

in both treatments. The parameter $\alpha = \frac{3}{4}$ satisfies the usual conditions $0 < \alpha < 1 < 2\alpha$ implying that, for given choices by the other, one personally gains by lowering one's own contribution; this, however, reduces the sum of payoffs. The main difference between treatments "Add" and "Alt" is that:

- in case of "Head" any inequality of π_x and π_y cannot be avoided or reduced by positive punishment p in treatment "Alt", whereas this, due to threat efficiency being 2, is possible in treatment "Add".
- in case of "Tail" equal zero payoffs are imposed in treatment "Add", whereas in treatment "Alt" equality of payoffs with level e endogenously results only if $p = 0$ is chosen.

Thus positive punishment p can possibly reduce inequality of payoffs in treatment “Add”, namely if “Head” occurs, whereas it would only create another inequality in treatment “Alt”, namely in case of “Tail”.

Assuming common (and known) material opportunism allows to solve the two game models by backward induction. Obviously, the optimal punishment is $p^* = 0$ regardless of c_x and $c_y(c_x)$ since X suffers from $p > 0$ with probability $\frac{1}{2}$. But then $c_y^*(c_x) = 0$ is best for Y for all possible choices c_x by X . Finally, anticipating all the consequences, X had better choose $c_x^* = 0$.

For common (and known) efficiency seeking players $p^+ = 0$ is also best, whereas due to $2\alpha > 1$ the efficiency minding contribution choices are $c_y^+(c_x) = e$ for all c_x and $c_x^+ = e$. If Y is known to be a conditional cooperator in the sense of $c_y^+(c_x) = c_x$ for all c_x , this would render $c_x^+ = e$ optimal. This illustrates that “leading by example” ($c_x^+ = e$) and “conditional cooperation” ($c_y^+(c_x) = c_x$) in the sense of following the example (see, e.g., Güth et al., 2007) can also lead to an efficient conflict settlement with $p^+ = 0$. Such conditional cooperation can be stabilized by off-play punishment threats such as “choose $p = 3$ whenever $c_y(c_x) < c_x$.” For $c_x = e$, for example, this suffices to overcome Y ’s free-riding incentives of $e(1 - \alpha) = 5$. Thus $(c_x = e, c_y(c_x) \equiv c_x, p = 0 \text{ for } c_y(c_x) = c_x, p = 3 \text{ otherwise})$ qualifies as an equilibrium, however one in weakly dominated strategies, that is not being subgame perfect.⁶

3 Experimental protocols

We rely on a $2 \times 2 \times 2$ factorial between subjects design.

One variation concerns the payoff function, which can be either “Add” or “Alt”, as detailed above.

A second variation concerns whether decision makers in role X can immediately react with punishment p after learning the reaction $c_y(c_x)$ chosen by Y or else are engaged in a filling task before choosing p : we call this “Hot” or “Cold”. To keep the changes between “Hot” and “Cold” minimal, in all treatments all (X and Y) subjects are shown the same tasks and spend the same amount of time fulfilling them:

- in “Hot” the punishment decision is taken immediately by X , and only after the choice of p the X participants can proceed with the filling task;
- in “Cold” the task takes place in between learning Y ’s reaction $c_y(c_x)$ and the possibility to punish via $p \in \{0, 1, 2, 3\}$.

We fill the “cooling-off” period with one of two unincentivized questionnaires⁷: the Domain-Specific Risk Scale (DOSPERT, Blais, 2006) questionnaire on risk attitudes or the Aggression Questionnaire (AQ, Buss and Perry, 1992), whose results we used as controls in our statistical analysis. Of course, it might depend on the filling tasks how effective (in the sense of reducing

⁶This is similar to the multiplicity of equilibria in ultimatum games of which, however, only the backward induction solution occurs in weakly undominated strategy (or subgame perfect).

⁷Both questionnaires are validated, widely used in the psychological literature, and are meant to measure risk and anger attitudes.

p) the cooling-off period will be. We did not explore this in the study at hand, however. We imposed the same filling tasks for both roles (X and Y) to avoid possible demand effects of task differences.

The third variation consists of playing the above treatments as detailed above, with monetary consequences, or in a pure voice scenario, in which the punishment choices have no monetary consequences. We call this variation “Real” and “Virtual”.

We let the players play the game twice. Hence each treatment of the $2 \times 2 \times 2$ factorial design is made up of two rounds of a basic 2-phase design (see the timeline in Fig. 1). In *Phase 1* of the first round, agents start out without knowing the role, X or Y , that they are going to be assigned. Using the strategy vector method each agent $i = 1, 2$ chooses

- her action $c_x^i \in \{0, 5, 10, 15, 20\}$ and
- her reaction $c_y^i(\cdot) = (c_y^i(c_x))_{c_x \in \{0, 5, 10, 15, 20\}}$; $c_y^i \in [0, 20]$.

After agents have made their choices, chance selects actor X and reactor Y , $X \neq Y$, with equal probabilities for both possibilities. Once the roles are fixed, the players enter *Phase 2*: first, actor X is reminded of her own choice c_x and is informed about the actual choice $c_y(c_x)$ of actor Y ; actor Y receives no feedback. At this point, in the “Cold” treatments all actors are asked to go through the 29 or 30 questions of the AQ or DOSPERT questionnaires; in the “Hot” treatments, instead, subject X can immediately decide whether to punish Y or not by choosing $p_x \in \{0, 1, 2, 3\}$ and fill in the questionnaire afterwards. Players receive no further feedback at the end of round 1.

These two phases are then repeated once, without any further feedback, in a perfect stranger matching. The repetition differs from the first round only insofar as actors now know their role already in *Phase 1*; all else stays the same. In particular, the strategy method choice elicitation of *Phase 1* now has a different meaning and different labels: instead of being an alternative-specific strategy, the phase now collects beliefs about the random partner’s behavior in the opposite role. Moreover, subjects who went through the DOSPERT now go through the AQ, and vice versa. To control for order effects, half of the subjects were given first the DOSPERT and then, in the second round, the AQ; the other half first received the AQ and then the DOSPERT.

[Insert Figure 1 about here]

After the repetition one of the two rounds is randomly chosen for payment. The actual payment is randomly determined by the choice of the round and the outcome of a virtual coin toss. Both random draws are group specific, and for practical reasons carried out directly by the software.

3.1 Session details

The experiment was run at the laboratory of the Max Planck Institute Jena, Germany. Three hundred and sixty-eight subjects took part in 12 experimental sessions attended by 28 to 32

subjects each. The experiment was computerized using zTree (Fischbacher, 2007) and took around one hour.⁸

Choices during the experiment were made using Experimental Currency Units (ECU). At the end of the experiment, 3 ECUs were exchanged for 1 euro and earnings thus computed were paid in cash.

The distribution across treatments of the participants and their average earnings in euro, including the €2.50 show-up fee, are summarized in Table 1.

[Insert Table 1 about here]

Table 1 shows that average earnings are quite similar across treatments. However, condition “Add” seems to promote a higher dispersion in individual earnings than condition “Alt”, for which the payoff portfolios are less diversified since both random events yield positive payoffs. Median earnings are considerably higher for treatment “Hot”/“Add” than for other treatments, both in the Virtual and in the Real condition.

4 Results

In the following, we first describe behavior in the Real and in the Virtual conditions separately. Specifically, we analyze conditional contribution choices of Players Y and punishment decisions of Players X. Then we pool together data from the Real and Virtual conditions and present a regression analysis on the determinants of punishment in the experiment.

4.1 Real punishment

Conditional Contributions

Figure 2 shows average reactions of Players Y to a given contribution of X in the Real and Virtual conditions.

[Insert Figure 2 about here]

Figure 2 also shows that reactions of Players Y are generally proportional to the potential contribution of Player X in all experimental conditions and both in Period 1 and Period 2.

To test the proportionality of reactions, we compare each reaction for a positive level of c_x to the reaction to the closest lower level of c_x (e.g., reactions for $c_x = 5$ are compared to reactions for $c_x = 0$, and so on). A series of Wilcoxon Signed Rank Tests (WSRT) shows that for all comparisons Y’s reactions to any positive c_x are bigger than Y’s reactions to the closest smaller c_x (all p-values ≤ 0.029). Thus we observe a very high incidence of conditional cooperation as Players Y seem to strongly condition their cooperation levels on the cooperation levels of their partners.

Result 1: *There is a high incidence of conditional cooperation across treatments and in both rounds.*

⁸The English version of the original German instructions used in the experiment is available at the end of the paper; the experimental software is available upon request.

Let us investigate across treatments which choice c_x would be optimal for X when anticipating such a striking conditional cooperation of Y participants. To explore this, we compute for each $c_x \in \{0, 5, 10, 15, 20\}$ the average expected payoff of X when he is matched with all Y participants in the same Real treatment condition, separately for rounds 1 and 2.

[Insert Figure 2 about here]

According to the upper half of Table 2, full cooperation is always the optimal choice for a risk neutral X , significantly so for all conditions (Wilcoxon Rank Sum Test, p-values always higher than 0.024) but for condition “Cold”-“Alt” (WRST, p-value = 0.394). The optimal choice will generate an average expected payoff ranging from 21.520 to 22.419 ECUs.

Result 2: *Based on rationally expecting the reaction of Y participants, X should choose $c_x = 20$ across all real conditions and in all rounds.*

Punishment

Table 3 summarizes the frequency of punishment for distinct experimental treatments and periods. Moreover, a further dimension is introduced in the table: a Player Y is classified as Perfect Conditional if her contribution is equal or greater than the contribution of the matched Player X , i.e., if $c_y(c_x) \geq c_x$. A Player Y is classified as Imperfect Conditional if her contribution is smaller than the contribution of the matched Player X , $c_y < c_x$. Overall, 68.3% and 64.3% of participants are classified as Perfect Conditional in Period 1 and Period 2, respectively. When punishment is justified by equity considerations, it should emerge only for Imperfect Conditional Players Y and mainly in condition “Add” only.

[Insert Table 3 about here]

Table 3 shows that punishment of Perfect Conditional cooperators is almost absent in the data (4.7% and 0% in Period 1 and Period 2, respectively), while it is more frequent for Imperfect Conditional cooperators (67.5% and 46.7% in Period 1 and Period 2, respectively). The difference in punishment across both categories is highly significant, both in Period 1 and Period 2 (Fisher’s Exact Test (FET), both p-values < 0.001). Thus punishment seems to be triggered by equity considerations.

Focusing on the punishment of imperfect conditional cooperators, no systematic differences across all Real treatments can be identified. A marginally significant difference is registered only for the comparison “Hot”/“Alt” vs “Cold”/“Add” both in Period 1 and Period 2 (FET, p-value= 0.068 and p-value= 0.082, respectively).

Result 3: *In the Real condition punishing imperfect conditional cooperation is frequent ($\geq 28\%$) and observed in both “Add” and “Alt”, although punishment generates another inequality in “Alt” rather than reducing the one resulting from imperfect conditional cooperation.*

4.2 Virtual

Conditional Cooperation

Figure 2 shows that reactions of Players Y are generally proportional to contributions of Players X. When performing the same proportionality test performed for the Real condition, a composite pattern emerges. In Period 1, most differences between reactions to close levels of c_x are statistically significant for conditions “Cold”/“Alt” and “Cold”/“Add” (WSRT, all p-values ≤ 0.012).⁹ In Period 2, many comparisons do not deliver a statistically significant result at the conventional 5% level. In particular, reactions to $c_x = 10$ and $c_x = 15$ are never significantly different (WSRT, all p-values ≥ 0.17). Moreover, reactions to $c_x = 15$ and $c_x = 20$ are significantly different only in condition “Cold”/“Add” (WSRT, p-value=0.007).

The fragmented proportionality pattern emerging from the above analysis suggests different patterns of reactions in the Real and Virtual conditions. However, a statistically significant difference is observed only when comparing average behavior at the individual level in condition “Add”/“Hot”, the condition in which more punishment is expected (Wilcoxon Rank Sum Tests, p-value= 0.086 and p-value= 0.018 for Period 1 and Period 2, respectively).

Table 2 shows that full cooperation is the maximizing choice for a risk neutral X, in most experimental conditions but the HOT condition in round 2 and the “Hot”/“Add” treatment in round 1. The payoffs associated to the optimal choice range from 20.469 to 22.509 ECUs.

All in all, the fraction of conditional cooperators is lower in the Virtual compared to the Real condition; it is still significantly high, though, especially in Period 1.

***Result 4:** Awareness that punishment is virtual weakens conditional cooperation, substantiating the stronger prohibitive effect of real versus virtual punishment threats.*

Punishment

Table 4 summarizes the frequency of punishment for distinct experimental treatments, periods, and types of Player Y in terms of conditional cooperation. Overall, 51.7% and 48.3% are classified as Perfect Conditional cooperators in Period 1 and Period 2, respectively.

[Insert Table 4 about here]

Table 4 shows that Players X tend to punish only Imperfect Conditional cooperators. Overall, the frequency of punishment of Perfect Conditional cooperators is very low (0% and 10.7% in Period 1 and 2, respectively). In contrast, Imperfect Conditional are punished about half of the times (50.0% and 56.7% in Period 1 and 2, respectively). A statistically significant difference between punishment of Perfect Conditional cooperators and Imperfect Conditional cooperators is observed both in periods 1 and 2 (FET, both p-values < 0.010).

The analysis of punishment of Imperfect Conditional cooperators highlights a marginally significant difference between conditions “Cold”/“Alt” and “Hot”/“Add” in Period 1 (FET,

⁹In condition “Hot”/“Alt”, no statistically significant difference is observed when comparing reactions to $c_x = 0$ and $c_x = 5$ (WSRT, p-value= 0.128). In condition “Hot”/“Add”, no statistically significant differences are observed when comparing reactions to $c_x = 10$, $c_x = 15$, $c_x = 15$ and $c_x = 20$ (WSRT, p-value= 0.189 and p-value= 0.101, respectively).

p-value=0.064). In Period 2, significant differences are identified when comparing conditions “Cold”/“Alt” and “Hot”/“Add” (FET, p-value=0.017) and conditions “Cold”/“Alt” and “Hot”/“Alt” (FET, p-value=0.015).

Result 5: In the Virtual condition, punishing imperfect conditional cooperation is frequent, though less frequent than in the Real condition, and is significantly higher in the “Hot” compared to the “Cold” condition.

5 Regression analysis

Table 5 reports the results of a regression analysis carried out to assess the determinants of punishment. Two distinct specifications for the dependent variable are adopted: in column (1), the dependent variable *Punishment* is a dichotomous variable equal to 1 when a Player X chooses a punishment value bigger than zero and equal to 0 otherwise; in column (2), the dependent variable *Punishment* measures the level of punishment and can assume integer values in the interval 0–3. Accordingly, in column (1) the outcomes of a Random-effects logistic regression are reported, and in column (2) the outcomes of a Random-effects Tobit regression are reported. The two regressions provide us with a description of the determinants of the likelihood to punish and of the level of punishment, respectively.

The explanatory variables *ADD*, *HOT*, and *Virtual* capture experimental treatments. *Dist_pos* and *Dist_neg* measure the distance between one’s own contribution and others’ contribution when the latter contribution is bigger or smaller than one’s own contribution, respectively. *Period_1* is a dichotomous variable equal to 1 when choices are made in the first experimental period and equal to zero when choices are made in the second period. The idiosyncratic variables *Age* and *Female* capture age and gender of the respondent, respectively. Variables *HardScience*, *Humanities*, and *Social* identify the field of study of the respondent.¹⁰ Finally, *DOSPRT.score* and *AQ.score* are the scores obtained in the DOSPRT and AQ questionnaires and directly captures risk seekingness and aggression attitudes of the respondent, respectively.

[Insert Table 5 about here]

The two regression analyses summarized in Table 5 provide consistent results. Treatment “Hot” positively affects punishment, while “Add” does not significantly affect it. Moreover, the consequences of punishment in terms of payoff loss seem to not affect the decision to punish, as shown by the estimated coefficient of “Virtual”. The coefficient of *diff_neg* shows that the lower the amount contributed by Y relative to the amount contributed by X, the higher the predicted punishment. In contrast, a higher relative contribution of Y does not

¹⁰The following majors of the Friedrich Schiller University Jena, Germany, were grouped under the label *Hard-Science*: Biology and Pharmacy, Chemical and Earth Sciences, Mathematics and Computer Science, Medicine, and Physics and Astronomy. The following majors were grouped under the label *Humanities*: Philosophy and Theology. The following majors were grouped under the label *Social*: Economics and Business Administration, Law, and Social and Behavioural Sciences. The residual category is given by those majoring in a discipline other than to those listed above.

affect punishment. Finally, both being older and majoring in a discipline belonging to the humanities seems to weakly foster punishment.

Result 6: *The main determinants of punishment are deviation from perfect conditional cooperation and being in the “Hot” condition, while the “Add”-“Alt” variation has no effect both in the Real and Virtual conditions.*

6 Discussion

Altogether, there are four main findings, two of which are as expected, namely:

1. There is a very strong tendency to conditionally cooperate.
2. The “Hot” condition generates higher levels of punishment.

The other two findings are unexpected, namely:

3. No treatment effects along the “Add”-“Alt” variations are observed, although punishing in “Alt” generates another inequality rather than decreasing the inequality due to imperfect conditional cooperation or free riding.
4. Punishment exists and plays a highly similar role in the “Virtual” condition, but with less prohibitive effects suggesting that many X participants just want to voice their anger and not necessarily engage in mutually harming punishment.

Overall, 66.3% and 50% of Y choices can be classified as perfectly conditional in the Real and Virtual conditions, respectively. Concerning punishment, 79.4% and 74.1% of X choices induce no punishment in Real and Virtual conditions respectively (see Tables 3 and 4). Thus Real punishment enhances conditional cooperation; but especially the results of the Virtual condition reveal that a strong reciprocity norm exists in our subject pool, irrespective of the punishment technology.

Hypothesis 1 – *less punishment when this does not reduce payoff inequality* – is not confirmed by the data. Subjects do not respond differently in the “Add” and “Alt” treatments. The fact that punishment creates another inequality in the “Alt” treatment, while it reduces a given inequality in the “Add” treatment, does not change the behavior of our subjects. The number of conditional cooperators and the frequency and size of punishment do not differ significantly across the “Add”-“Alt” dimension, both in the Real and in the Virtual conditions. There is no increased level of punishment when it could decrease inequality.

On the other hand, Hypothesis 2 – *less punishment when the choice is delayed* – is confirmed by the regression analysis of Section 5. Hence the results of Grimm and Mengel (2011) are replicated in our experiment: in condition “Hot”, there is more punishment than in condition “Cold”, and subjects who punish, do so more harshly.

Our data strongly support a norm-based notion of punishment. Subjects do not respond to whether punishment reduces payoff inequalities (“Add” *vs.* “Alt”) nor to whether punishment has real consequences or merely virtual ones (“Real” *vs.* “Virtual”). There is, though, a

strong underlying pattern: punishment strongly and significantly correlates with the distance between one's own and the partner's contribution, once all other factors are controlled for.

Thus subjects punish more on grounds of principle, reacting to a norm violation, than by considering the possible consequences of punishing. This interpretation is consistent with our data along several dimensions.

First, the observed low level of punishment is due to the low overall tendency to violate the implicit norm of conditional cooperation as this high level of compliance is observed even in the Virtual condition.

Second, the norm of reciprocity, on which conditional cooperation is based, can explain the absence of any difference across the "Add" and "Alt" conditions.

Finally, the similar results of the Real and Virtual conditions, documented by the regression results, strongly suggest a pure voice effect, i.e., participants mainly voice their anger mainly in case of no (proper) reciprocation.

Similar to previous studies (e.g., Fehr and Gächter, 2002), we show that punishment promotes cooperation. What is new is that punishment is not used to try to reduce inequality (Fowler et al., 2005) but rather results from a need to sanction a violation of the reciprocity norm. In a sense, punishment behavior does not seem to be driven by its material consequences, but by psychological needs (to "let off steam", contrary but similar to the "warm-glow" effect, see Andreoni (1990)).

References

- Ahlert, M., Crüger, A., Güth, W., 2001. How Paulus Becomes Saulus? – An Experimental Study of Equal Punishment Games. *Homo Oeconomicus* XVIII (2), 303–318.
- Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *The Economic Journal* 100 (401), 464–477.
- Aumann, R., 1992. Irrationality in game theory. *Economic analysis of markets and games: Essays in honor of Frank Hahn*, 214–27.
- Axelrod, R., 1984. *The evolution of cooperation*. Political Science / Science. Basic Books.
- Blais, A.-R. . W. E. U., 2006. A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making* 1, 33–47.
- Bolton, G. E., Ockenfels, A., March 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90 (1), 166–193.
- Bolton, G. E., Zwick, R., 1995. Anonymity versus Punishment in Ultimatum Bargaining. *Games and Economic Behavior* 10 (1), 95–121.
- Buss, A., Perry, M., 1992. The aggression questionnaire. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology* 63 (3), 452.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415 (6868), 137–140.
- Fehr, E., Schmidt, K. M., 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114 (3), 817–868.

- Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.
- Fowler, J., Johnson, T., Smirnov, O., 2005. Egalitarian motive and altruistic punishment. *Nature* 433 (10.1038).
- Grimm, V., Mengel, F., 2011. Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters* 111 (2), 113–115.
- Güth, W., Huck, S., 1997. From ultimatum bargaining to dictatorship—An experimental study of four games varying in veto power. *Metroeconomica* 48 (3), 262–299.
- Güth, W., Huck, S., Miller, W., 2001. The Relevance of Equal Splits in Ultimatum Games. *Games and Economic Behavior* 37 (1), 161–169.
- Güth, W., Levati, M. V., Sutter, M., van der Heijden, E., June 2007. Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics* 91 (5-6), 1023–1042.
- Ostrom, E., 2000. Collective action and the evolution of social norms. *The Journal of Economic Perspectives* 14 (3), 137–158.

A Figures

Figure 1: Timeline of the experiment for “Hot” conditions

Figure 2: Y's Average Reaction

B Tables

	N	Mean	Median	SD
<i>Real</i>				
COLD/ALT	64	9.762	9.167	1.411
COLD/ADD	60	9.172	2.500	7.235
HOT/ALT	64	9.165	9.167	0.783
HOT/ADD	64	10.180	14.170	7.377
<i>Virtual</i>				
COLD/ALT	28	10.150	9.167	1.344
COLD/ADD	28	9.890	8.958	7.598
HOT/ALT	28	9.699	9.167	1.538
HOT/ADD	32	11.590	15.620	7.285

Table 1: Summary statistics of earnings by treatment

<i>Real</i>					
<i>Round 1</i>			<i>Round 2</i>		
	ALT	ADD		ALT	ADD
COLD	20 (22.264)	20 (22.419)	COLD	20 (21.520)	20 (22.312)
HOT	20 (21.689)	20 (22.340)	HOT	20 (21.730)	20 (22.398)
<i>Virtual</i>					
<i>Round 1</i>			<i>Round 2</i>		
	ALT	ADD		ALT	ADD
COLD	20 (22.509)	20 (21.089)	COLD	20 (21.839)	20 (21.679)
HOT	20 (21.397)	10 (21.281)	HOT	0 (20.670)	0 (20.469)

Table 2: The optimal choice c_x of a risk neutral X when anticipating the observed response behavior of Y participants (within parentheses, the associated average expected payoff

$$E[\pi_x] = \frac{1}{2} [2e - c_x + \alpha[c_x + c_y(c_x)]]).$$

Table 3: Punishment (Real)

		Frequency (%)					
		N	0	1	2	3	Median
Period 1	Perfect Conditional						
	COLD/ALT	20 (62.5)	20 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0
	COLD/ADD	21 (70.0)	21 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ALT	23 (71.9)	22 (95.7)	1 (4.4)	(0.0)	(0.0)	0
	HOT/ADD	22 (68.8)	19 (86.4)	(0.0)	(0.0)	3 (13.6)	0
	TOT	86 (68.3)	82 (95.3)	1 (1.2)	0 (0.0)	3 (3.5)	
	Imperfect Conditional						
	COLD/ALT	12 (37.5)	5 (41.7)	1 (8.3)	1 (8.3)	5 (41.7)	1.5
	COLD/ADD	9 (30.0)	3 (33.3)	0 (0.0)	1 (11.1)	5 (55.6)	3
	HOT/ALT	9 (28.1)	2 (22.2)	4 (44.4)	2 (22.2)	1 (11.1)	1
	HOT/ADD	10 (31.2)	3 (30.0)	1 (10.0)	3 (30.0)	3 (30.0)	2
	TOT	40 (31.7)	13 (32.5)	6 (15.0)	7 (17.5)	14 (35.0)	
Period 2	Perfect Conditional						
	COLD/ALT	18 (56.2)	18 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	COLD/ADD	21 (70.0)	21 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ALT	22 (68.8)	22 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ADD	20 (62.5)	20 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	TOT	81 (64.3)	81 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
	Imperfect Conditional						
	COLD/ALT	14 (43.8)	9 (64.3)	0 (0.0)	2 (14.3)	3 (21.4)	0
	COLD/ADD	9 (30.0)	6 (66.7)	1 (11.1)	2 (22.2)	0 (0.0)	0
	HOT/ALT	10 (31.2)	5 (50.0)	1 (10.0)	0 (0.0)	4 (40.0)	(0.5)
	HOT/ADD	12 (37.5)	4 (33.3)	2 (16.7)	2 (16.7)	4 (33.3)	1.5
	TOT	45 (35.7)	24 (53.3)	4 (8.9)	6 (13.3)	11 (24.4)	

Table 4: Punishment (Virtual)

		Frequency (%)					
		N	0	1	2	3	Median
Period 1	Perfect Conditional						
	COLD/ALT	10 (71.4)	10 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	COLD/ADD	10 (71.4)	10 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ALT	5 (35.7)	10 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ADD	5 (31.2)	10 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	TOT	30 (51.7)	30 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
	Imperfect Conditional						
	COLD/ALT	4 (28.6)	1 (25.0)	0 (0.0)	2 (50.0)	1 (25.0)	2
	COLD/ADD	4 (28.6)	2 (50.0)	1 (25.0)	0 (0.0)	1 (25.0)	0.5
	HOT/ALT	9 (64.3)	3 (33.3)	2 (22.2)	1 (11.1)	3 (33.5)	1
	HOT/ADD	11 (68.8)	8 (72.7)	1 (9.1)	0 (0.0)	2 (18.2)	0
	TOT	28 (48.3)	14 (50.0)	4 (14.3)	3 (10.7)	7 (25.0)	
Period 2	Perfect Conditional						
	COLD/ALT	8 (57.1)	8 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	COLD/ADD	10 (71.4)	9 (90.0)	0 (0.0)	0 (0.0)	1 (10.0)	0
	HOT/ALT	5 (35.7)	4 (80.0)	0 (0.0)	0 (0.0)	1 (20.0)	0
	HOT/ADD	5 (31.2)	4 (80.0)	0 (0.0)	0 (0.0)	1 (20.0)	0
	TOT	28 (48.3)	25 (89.3)	0 (0.0)	0 (0.0)	3 (10.7)	
	Imperfect Conditional						
	COLD/ALT	6 (42.9)	3 (50.0)	0 (0.0)	3 (50.0)	0 (0.0)	1
	COLD/ADD	4 (28.6)	4 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0
	HOT/ALT	9 (64.3)	4 (44.4)	0 (0.0)	0 (0.0)	5 (55.6)	3
	HOT/ADD	11 (68.8)	6 (54.4)	0 (0.0)	0 (0.0)	5 (45.5)	0
	TOT	30 (51.7)	17 (56.7)	0 (0.0)	3 (10.0)	10 (33.3)	

Table 5: Determinants of Punishment

<i>Punishment ~</i>	Coeff (Std. Err.)	
	(1)	(2)
<i>Constant</i>	-16.180(6.403)*	-22.266 (7.394)**
<i>ADD</i>	0.231 (0.955)	0.540 (1.226)
<i>HOT</i>	2.203 (1.096)*	2.638 (1.294)*
<i>Virtual</i>	0.625 (1.020)	1.133 (1.303)
<i>Dist_pos</i>	-0.167 (0.187)	0.031 (0.145)
<i>Dist_neg</i>	0.527 (0.116)***	0.722 (0.127)***
<i>Period_1</i>	0.787 (0.557)	0.580 (0.650)
<i>Age</i>	0.332 (0.196) ^o	0.430 (0.239) ^o
<i>Female</i>	-0.324 (1.076)	-0.529 (1.385)
<i>HardScience</i>	2.634 (2.812)	2.792 (3.375)
<i>Humanities</i>	4.881 (2.920) ^o	6.412 (3.405) ^o
<i>Social</i>	3.019 (2.651)	4.084 (3.146)
<i>DOSPRT.score</i>	-0.006 (0.133)	0.031 (0.169)
<i>AQ.score</i>	0.035 (0.035)	0.051 (0.044)
Obs (groups)	366 (183)	366 (183)
Censoring		lc=284; rc=48

Significance level: *** (0.1%); ** (1%); * (5%); ^o (10%)

(1) Random-effects logistic regression

(2) Random-effects tobit regression

C Instructions

You have been recruited to take part in a computer administered experiment. You receive a 2,5 Euro show up fee for taking part in the experiment. Please read the following instructions carefully. Prior to the experiment, you will have to answer a few questions testing your comprehension of these instructions. Please note that the instructions are written in male gender only for convenience, but refer to both genders equally. Please do not talk and raise your hand if there are any specific questions during the experiment. An experimenter will come to assist you. Please remain silent and switch off your mobile phone. If you violate these rules, we will have to exclude you from the experiment and all payments.

You will be either a Participant X or a Participant Y. Participants will be randomly assigned to role X or to role Y and will keep that role for the rest of the experiment.

The experiment extends over 2 Rounds, each of them made up of 2 Phases. Each participant X will be paired with a different participant Y in the two rounds. Participants will not be informed by us, during or after the experiment, about whom they are matched with.

During the experiment you are going to use ECU (Experimental Currency Units). All participants are given an initial endowment of 20 ECUs. At the end of the experiment, ECU will be converted into Euros at an exchange rate of

$$1 \text{ Euro} = 3 \text{ ECUs.}$$

As an example, 15 ECUs are equivalent to 5 Euro. Your final payment will include the show up fee of 2,5 Euro.

The decisions that you make during the experiment will affect your final payoff. One of the two rounds will be randomly drawn for payment at the end of the experiment.

ROUND 1

Round 1 is made of two phases: Phase 1 and Phase 2.

{*HOT*: Furthermore, after Phase 2 you will be asked to answer some questions. Details about the questions will be given to you directly on the screen.}; {*COLD*: Furthermore, after Phase 1 but before Phase 2 you will be asked to answer some questions. Details about the questions will be given to you directly on the screen.}

You will interact with another participant in a group of two, with one participant in each group being a Participant X and the other being a Participant Y. You will be informed about your role only at the end of Round 1. Thus, in Round 1 you will be asked to choose for both roles, without knowing your actual role.

PHASE 1

In this phase both participant X and Y are given 20 ECUs, and have to decide how to split them between a private account and a group account. In this phase you do not know whether you are a Participant X or a participant Y.

IF YOU ACT AS PARTICIPANT X

If you were chosen to be X, you would have to decide how much to put into the group account. You can choose to put into the account 0, 5, 10, 15, or 20 ECUs. Each ECU that you put into the group account is multiplied by 1,5 and then divided equally between you and Participant Y. Each ECU you do not put into the group account will automatically be put into your private account, without any multiplication.

Participant Y has to make the same choice: each ECU he puts into his private account is kept for himself, and each ECU he puts into the group account is multiplied by 1,5 and then divided equally between him and you. Participant Y can choose to put into the group account any number of ECUs between 0 and 20 (in steps of 1); each ECU he does not put into the group account will be automatically put into his private account, without any multiplication.

Table 1 shows the payoffs for you, if you are a Participant X, for all possible combinations of ECUs that you and the Participant Y put into the group account.

IF YOU ACT AS PARTICIPANT Y

If you were chosen to be Y, you would also have to decide how much to put into the group account. However, as Y you can condition your choice upon the choice of Participant X. Specifically, for each choice of Participant X (0, 5, 10, 15, 20), you are asked to report how many ECUs (between 0 and 20) to put into the group account. Only the choice corresponding to the actual choice of Participant X is going to be employed to define your payoffs as Y in this Phase. Each ECU that you put into the group account is multiplied by 1,5 and then divided equally between you and Participant X. Each ECU you do not put into the group account will automatically be put into your private account, without any multiplication.

Table 2 shows the payoffs for you, if you are a Participant Y, for all possible combinations of ECUs that you and the Participant X put into the group account.

Remember that you will not know, in Phase one, if you will be assigned the role of Participant X or Y. You will be asked what you would do in both roles.

PHASE 2

In this phase both Participant X and Y are given again 20 ECUs.

At the beginning of Part 2 you will be randomly assigned to the role of Participant X or Participant Y, and you will be told of your role. The other member of your group will take the other role.

IF YOU ARE PARTICIPANT X

If you are a Participant X, you will in Phase 2 know the reaction of Y to your contribution to the group account.

{*REAL*: After having learnt this reaction, you can decide whether to punish or not Participant Y. To do so, you can invest in punishment between 0 and 3 ECUs out of your endowment of

20. For each ECU you invest in punishment, Participant Ys endowment will be reduced by 2 ECUs.}

{*VIRTUAL*: After having learnt this reaction, you can decide whether to punish or not Participant Y. Please be aware that punishment is only virtual, that is your choices will not affect yours or Ys payoffs. To punish, you can invest in punishment a virtual amount between 0 and 3 ECUs out of your endowment of 20. For each ECU you invest in punishment, Participant Ys endowment will be virtually reduced by 2 ECUs; Y will be informed only at the end of the experiment of the amount of virtual punishment you have chosen and the corresponding virtual reduction of his payoff. Your endowment for Phase 2 will be equal to your initial endowment of 20 ECUs.}

IF YOU ARE PARTICIPANT Y

If you are a Participant Y, you do not make any choice in Phase 2.

{*REAL*: Participant X will decide to invest or not in punishing you, but you will not be informed about it. Your payoff in this phase will be given by your 20 ECUs minus twice the amount that X invested in punishment.}

{*VIRTUAL*: Participant X will decide to invest or not in virtually punishing you. This has no payoff consequences for you, and you will be informed about it only at the end of the experiment. Your payoff in Phase 2 will be equal to your initial endowment of 20 ECUs.}

ROUND PAYOFFS

In Round 1, the payoffs of Participant X and Y are given by choices in Phase 1 and Phase 2 and by the toss of a coin. The computer tosses a virtual coin.

{*ADD*: If the coin lands on HEADS, the payoffs will be given by the sum of ECUs earned in Phase 1 and Phase 2. If the coin instead lands on TAIL, the payoffs will be equal to zero.}

{*ALT*: If the coin lands on HEADS, the payoffs will be given by ECUs earned in Phase 1. If the coin instead lands on TAIL, the payoffs will be given by ECUs earned in Phase 2.}.

At the end of the round you will be given no feedback about the payoffs. All feedback will be given at the end.

ROUND 2

In Round 2 you will keep the same role you had in Round 1: if you were a Participant X, you will stay X, if you were a Participant Y, you will stay Y. You are asked to make the same choices as in Round 1, and the payoffs are computed with the same rules. Again, the round is made up of two phases and {*HOT*: at the end of phase 2}{*COLD*: between Phase 1 and Phase 2} you will be asked to fill in some questions.

Even if only the choice for your current role will have real consequences, you will be asked to report what you would do if you had been given the other role. This choice has no consequences in terms of earnings in the experiment.

	Your (as X) contribution to the group account				
	0	5	10	15	20
0	20,00	18,75	17,50	16,25	15,00
1	20,75	19,50	18,25	17,00	15,75
2	21,50	20,25	19,00	17,75	16,50
3	22,25	21,00	19,75	18,50	17,25
4	23,00	21,75	20,50	19,25	18,00
5	23,75	22,50	21,25	20,00	18,75
6	24,50	23,25	22,00	20,75	19,50
7	25,25	24,00	22,75	21,50	20,25
8	26,00	24,75	23,50	22,25	21,00
9	26,75	25,50	24,25	23,00	21,75
10	27,50	26,25	25,00	23,75	22,50
11	28,25	27,00	25,75	24,50	23,25
12	29,00	27,75	26,50	25,25	24,00
13	29,75	28,50	27,25	26,00	24,75
14	30,50	29,25	28,00	26,75	25,50
15	31,25	30,00	28,75	27,50	26,25
16	32,00	30,75	29,50	28,25	27,00
17	32,75	31,50	30,25	29,00	27,75
18	33,50	32,25	31,00	29,75	28,50
19	34,25	33,00	31,75	30,50	29,25
20	35,00	33,75	32,50	31,25	30,00

Table 6: Your Payoff if chosen to be X

FINAL PAYMENT

After Round two, the computer will randomly draw one of the two rounds. Your final payment will be given by your payoff in that round.

The ECUs earned in the round thus selected will be exchanged with euro at the conversion rate of 1 euro each 3 tokens.

	X's contribution to the group account				
	0	5	10	15	20
0	20,00	23,75	27,50	31,25	35,00
1	19,75	23,50	27,25	31,00	34,75
2	19,50	23,25	27,00	30,75	34,50
3	19,25	23,00	26,75	30,50	34,25
4	19,00	22,75	26,50	30,25	34,00
5	18,75	22,50	26,25	30,00	33,75
6	18,50	22,25	26,00	29,75	33,50
7	18,25	22,00	25,75	29,50	33,25
8	18,00	21,75	25,50	29,25	33,00
9	17,75	21,50	25,25	29,00	32,75
10	17,50	21,25	25,00	28,75	32,50
11	17,25	21,00	24,75	28,50	32,25
12	17,00	20,75	24,50	28,25	32,00
13	16,75	20,50	24,25	28,00	31,75
14	16,50	20,25	24,00	27,75	31,50
15	16,25	20,00	23,75	27,50	31,25
16	16,00	19,75	23,50	27,25	31,00
17	15,75	19,50	23,25	27,00	30,75
18	15,50	19,25	23,00	26,75	30,50
19	15,25	19,00	22,75	26,50	30,25
20	15,00	18,75	22,50	26,25	30,00

Table 7: Your payoff if chosen to be Y