Third – party punishment: hot or cold?
Peter Lewisch (University of Wien)
Stefania Ottone (EconomEtica and University of Eastern Piedmont)
Ferruccio Ponzano (University of Eastern Piedmont)

Introduction
The role of punishment (in the broad sense) for the stability of social order has been the object of intense economic research, both analytical and empirical, in recent years. This research has shown that individuals are willing to carry out punishment even in the absence of any instrumental purpose. It is often claimed that such an “altruistically motivated punishment” (in particular, punishment for the sake of satisfying one’s own desire for retributive justice) is the missing “golden headstone” in an encompassing explanation of social order. The aforementioned literature has, however, typically confined its analysis to the study of (altruistic) punishment by one single potential punisher. Little attention has been devoted to punishment patterns in group settings. The aforementioned literature seems to implicitly assume that altruistic punishment would not generate spill over effects to other potential punishers in the group (as would be the case if an actor derives retributive utility from seeing a wrongdoer punished by someone else). If any retributive reward of punishment is, as assumed, in fact always and categorically confined to the punisher him-/herself, no spill over effects to other potential punishers and, hence, no free riding on altruistic punishment can occur. Once we relax this – crucial, but by no means self-evident – limitation, punishment patterns become much more complex; and so does the question of social order: altruistic punishment then carries “public good properties”, free riding on altruistic punishment becomes possible, and the aggregate amount of punishment is undetermined, possibly resulting both in under- and over-punishment.

Given the unclear nature of punishment, we trust that an empirical – namely experimental – study can contribute to a better understanding of the aforementioned set of issues. In this research project, we want to clarify the public good nature of altruistic punishment by comparing the amount of individually provided punishment in group settings (where more than one potential punisher carry out punishment by separate choice) to the stand alone case (with only one single potential punisher). This approach also allows us to study the extent of free riding, namely whether potential punishers in a group would behave as conditional cooperators when providing punishment, and – as a related question – the amount of aggregated punishment levels (under- versus over-enforcement).
**Expected results and the related literature**

The study of punishment in experimental economics is large and increasing (Fehr and Gacther, 2000, Nikiforakis, 2008). In particular, the common question in these works is if a costly punishment is used by individuals and what is its impact in economic choices. If we consider a population of purely selfish individuals, our expected result is trivial: no one will punish others in any case. Obviously, the literature shows a completely different scenario where people are ready to punish if they perceive a non-fair situation.

As we briefly argued in the introduction, we want to add the answers of three questions in a third party punishment game. In order to elicit these results, we modify the experiment by Fehr and Fischbacher (2004) allowing comparing a one observer scenarios and multi-observers scenario. First of all, we want to analyse if punishment is an ordinary good or not. Secondly, we investigate if punishment is a private good or a public good and, then, if the second hypothesis is true, we will analyse the presence of conditional cooperation or not. Analyses about these three issues are formerly preformed in contexts of second party punishment. According to the presence of ordinary good, Anderson and Putterman (2006) and Carpenter (2007) note that as the price of punishment increases. This leads to assuming that punishment could be considered an ordinary good.

The second problem addresses to the public or private nature of the good punishment. In particular, we define punishment as a private good if the presence of a second punisher does not influence the level of contribution of a subject. A previous experiment of Casari and Luini (2005) based on a second party punishment shows that punishment could be approximated to a private good. We want to check if it is true also in third party punishment scenario.

If it is not the case and we obtain that punishment can be considered as a public good, our third aim is to check if punishers act as free riders or not. In particular, our question about cooperation concerns the fact that punishment is conditional or not. A further results we elicit using this analysis is a comparison of the level of punishment when we have only a potential punisher and when the number of potential punishers is higher than one. If we assume that player free-rides the total level of punishment should be lower in the multi-potential-punishers scenario with respect to the one-punisher situation.

**The experimental design**

The experimental design consists of three treatments: the High-Cost Treatment (HCT), the Low-Cost Treatment (LCT) and the Mixed-Cost Treatments (MCT). In all the treatments the tool (see Figure 1) is a variant of the Third-Party Punishment Game (Fehr and Fischbacher, 2004). At the beginning of each session in each treatment subjects are randomly assigned a role (A, B and C). In the first stage participant A (the Dictator) and participant B (the Receiver) are paired and play a Dictator Game. In the second stage, participants C enter the game and are assigned to a couple
(Player A + Player B) who played the Dictator Game in the first stage. At this point each player C is informed that s/he has \( \frac{1}{2} \) probability of being the only player C in the group and \( \frac{1}{2} \) probability of being in a group where another player C has entered the game. In the latter case, we call C1 and C2 the two players C. The strategy method is implemented at this level. Each player C is asked to declare her/his level of punishment in both cases and for each level of transfer from A to B. After that, their effective role is randomly drawn and the relative decision implemented. At the end of the game, first order beliefs of players C on the punishment level of the other player C if s/he is in a group of four are elicited according to the standard procedure.

In each treatment, A’s initial endowment is 100 tokens, C’s initial endowment is 75 tokens while B’s initial endowment is 50 tokens. In the HCT, the cost for all participants C to punish participant A for the amount of 2 tokens, is 1 token. In the LCT, the cost for all participants C to punish participant A for the amount of 3 tokens, is 1 token.\(^1\) In the MCT, players C1 and C2 bear the high/low cost to punish player A, while the other players C bear the low/high cost. Each token’s value is 0.10 Euro.

**The experimental procedure**

The experiment has been run in the Experimental Economics Laboratory (CEEL) at the University of Trento, Italy. The experiment has been programmed and conducted with the software z-Tree (Fischbacher, 2007). Overall, 8 sessions has bee run, with a total of 160 participants (40 participants in the HCT, 40 in the HCT and 80 in the MCT). At the end of each session, subjects have been asked to fill in a brief survey to check for socio-demographic data. Each subject participated in one session only and partners’ identities have been unknown even when the experiment was over. Each session lasted for about 60 minutes. Each subject earned on average 7 Euros.

**Preliminary results**

In this section we present our main findings. In particular, we analyse the Observers’ transfer with respect to: 1) the Dictators’ transfer to the Receivers; 2) the number of Observers in the group; 3) the cost of punishing. Consequently, we study eight subgroups of punishment: LT-alone, HT-alone, LT-MIX-alone, HT-MIX-alone, LT-not-alone, HT- not-alone, LT-MIX- not-alone, HT-MIX- not-alone. The former four subgroups refer to the level of punishment implemented by Observers in groups of three participants in the LCT, in the HCT and in the MCT when they bear the low cost and the high cost of punishing respectively. The latter four subgroups describe the same series of treatment but when the Observer is in a group of four.

Figures and Tables provides data of this subgroups. In particular, the level of punishment of each subgroup for each possible transfer from the Dictator to the Receiver.

\(^1\) Only transfers of entire tokens are allowed and no participant can earn a negative payoff.
Result 1. In line with the existing experimental literature, the Observer’s level of punishment decreases as the Dictator’s transfers increase.

Description of result: This result (see Figure 2) is perfectly in line with the experimental evidence obtained in previous works (see for instance Fehr and Fischbacher, 2004; Bernhard, 2005; Ottone, 2005, 2008). A random-effects Tobit regression of the Observer’s punishment on the Dictator’s transfer to the Receiver confirms that a positive relation exists between the level of punishment and the degree of unfairness (AGIVES, \( p = 0.000 \)).

Interpretation: The results described here are in fact twofold; either one confirms findings of the existing literature. The first finding is simply that individuals carry out costly punishment at all in a one shot interaction with no material (instrumental) benefit to gain: these individuals, in principle “passive observers”, impose a certain disutility on someone else because of this person’s unfair treatment of the respective recipient, and they do so irrespective of the fact that this punishment comes at a cost without any conceivable benefit in terms of deterrence. Whereas this finding does not fit economic textbook assumptions regarding self-interested behaviour, it replicates the results of the pertinent experimental literature on punishment.\(^2\) It is worthwhile to note, however, that this “altruistic punishment” takes place in the seemingly trivial scenario of the above described experiments, which are unlikely to cause strong feelings of anger or revenge. Still, these individuals “punish for the sake of punishment”; they satisfy their desire for punishment and “purchase”, in some real sense, “retributive justice”.\(^3\)

The second, somewhat more specific finding, still in line with the existing literature, is that levels of punishment vary with the degree of perceived unfairness. This result suggests, in line with everyday intuitions, that individuals perceive unfairness not as a binary category (“fair”, “unfair”), but as a continuum; the response to such unfair behaviour mirrors the graveness of the offence. It should be noted that this result can be observed both in the 3-person and the 4-person environment. In a nutshell one may characterize this finding that “punishment is demand dependent”: the higher the (subjectively experienced) interest in seeing the wrongdoer punished (i.e. the higher the potential punisher’s demand), the higher the willingness to pay for such punishment.

Result 2. Third-party punishment is sensitive to its cost

Description of result: The higher the cost, the lower is the level of punishment at each level of transfer from the Dictator to the Receiver. It is clear from Figure 2 where it turns out that the level

\(^2\) With respect to the subsequent results of the paper, it is good to have “the right basis” as a starting point, namely a replication of “altruistic punishment”.

\(^3\) In a situation, in which they see their own concepts of fairness (which seemingly command a fair treatment of the receiving party) violated, they are willing to spend real resources to satisfy their own concepts of justice.
of punishment is significantly lower when the Observer bears the high cost of punishing. A random-effects tobit confirms this result (HT and MXT, \( p \leq 0.05 \)).

Interpretation: This second result is in line with economic predictions. This finding captures, what may be called the “supply-side” of punishment: The higher the price for a certain service, the less of it will be consumed. This finding can again be observed both in the 3-person and the 4-person environment. Whereas this finding does not come as a surprise, it should be noted that there is relatively scarce and relatively recent experimental literature on this subject matter.\(^4\) Insofar, this result confirms the findings reported in the aforementioned literature

**Result 3. Third-party punishment is a public good**

Description of result: If we compare the level of punishment of the Observers when they are alone and when they are in a group where another Observer is interacting, it is clear that, in all treatments, punishment is higher in groups of three. See Figure 2 and the random-effects tobit regression (ALONE, \( p = 0.000 \)).

Interpretation: Result 3 describes a main finding of this paper. The comparison of individual punishment levels as performed either on a stand-alone basis or in a group-setting has not yet been tested out systematically in the experimental literature. The results show that individuals punish significantly less, when a second potential punisher is present, as opposed to a setting wherein they are the only punishers. The presence of a further potential punisher, hence, generates spill-over effects on an individual’s punishment choices. If the personal interest of a potential punisher in punishing a wrongdoer was “in his/her flesh” (“in persona”), the presence of a further party would not matter and could not lead to any modification of behaviour.

**Result 4. As in a typical public good scenario, most players behave as conditional cooperators**

Description of result: As we found out that third-party punishment is a public good, we want to check whether the well-known pattern of conditional cooperation emerges. In a punishment scenario this would be represented by a situation where, in groups of four subjects, the Observers’ level of punishment is correlated to the expected punishment of the other third party. Visual inspection (see Figure 3) and statistical analyses (Spearman correlation test; \( p < 0.005 \))\(^5\) confirm our hypothesis: most Observers conditionally punish.

Interpretation: Result 4 embodies another main finding of the paper. In the presence of a second punisher, individual punishers behave as “quasi conditional cooperators”. If they expect the other

---

\(^4\) See Fehr/Fischbacher (2004) p 189: “From an economic viewpoint, one would expert that as the costs of sanctions increase for the sanctioning subject, there will be fewer and weaker sanctions. Recent research has shown that this is indie the case (C.M. Anderson and L. Putterman, unpublished; J.P. Carpenter, unpublished)”.

\(^5\) We ran this test for each subgroup. The H0 is independence of the two variables.
punisher to punish severely, they increase their (costly) punishment and vice versa. We are not aware of similar findings in the literature.

This result deserves closer attention. First, this finding is in line with the bottom line of the overall results in our experiments. Given that empirically punishment is a public good, it should not come as a surprise that individuals behave, in their punishment choices, in a way parallel to the provision of genuine public goods. Second, result 4 is particularly strong, because it is robust even with regard to the – deliberately employed – cost-differentials among potential punishers. Remember that our experiments are designed such that they encompass a high cost and a low cost variant, in which an individual for whom punishment costs are relatively low is paired with a high cost individual. In this environment one could have guessed that individuals take the assignment of either a low-cost or a high-cost technology as a “proxy” for efficient punishment. Whereas it would encourage high cost individuals to economize on (extra-costly) punishment, it would prompt low cost individuals to engage strongly in (relatively cheap) punishment. In the overall, cost-differentials with respect to punishment would facilitate the emergence of efficient enforcement systems. The results show that such behaviour was absent in our experiments. Result 4 prevails in all environments, also in those with high and lost cost Punishers.

**Result 5. The presence of more than one Observer does not result in under-punishment**

Description of result: The last point we want to study is how the nature of punishment influences the final level of third-party intervention when more than one Observer is involved. In principle, the private nature of punishment should lead to over-punishment while a public good scenario should result in under-punishment due to the free-riding phenomenon.

Our results partially disconfirm this hypothesis. If we consider the number of tokens subtracted from the Dictator’s final payoff in groups of three and in groups of four, we observe that in the latter this level is lower but not in a significant way (see Figure 4). A statistical analysis confirms the non-significance of these differences (Mann-Whitney test; \( p > 0.55 \)).

Interpretation: Third-party punishment nature does not allow over-punishment. However, conditional cooperation prevent from under-punishment. This may imply that it may be efficient to assign the task to punish to a third party without risking under-punishment.

**Conclusions**

Our experiment allows finding out some important results about punishment. First of all, we note that punishment decreases as the level of fairness decreases and that punishment is a ordinary good also when we considers third parties. In these two cases we are perfectly in line with the existing literature and third party punishment and second party shows the same patterns.
Differently from second party punishment, third party shows the characteristics of a public good. In terms of Casari and Luini, the emotional aspect of has low impact and rational part is prevalent. Nevertheless, we do not observe free rider behavior and players act as conditional cooperation. In presence of two punishers this avoids under-punishment with respect to a one-punisher scenario.

\footnotetext{6 We ran this test for each subgroup. The H0 is the equality of the two variables.}
References

Appendix 1 – Figures and Tables

Third-Party Punishment Game (TPP)

Group of three

Transfer opportunity

Punishment opportunity

Fig. 1 The Experimental Games

Fig. 2 Punishment Pattern
Fig. 3  Conditional Punishment

Fig. 4  No under-punishment
Appendix 2 – The econometric analysis

\[ P_i = \beta_0 + \beta_1 AGIVES_i + \beta_2 ALONE_i + \beta_3 HT_i + \beta_4 HTMIX_i + \beta_5 LTMIX_i + \beta_6 AGE + \beta_7 GENDER_i + \epsilon_i \]

Dependent variable: punishment \((P_i)\)
Random-effects Tobit regression – censored at the low level (0)

### Variables

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGIVES</td>
<td>transfer from A to B</td>
<td>15</td>
<td>10</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>ALONE</td>
<td>Dummy variable equal to 1 if the Observer is in a group of three</td>
<td>1.73</td>
<td>3.59</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>HT</td>
<td>Dummy variable equal to 1 if the observation belongs to the HCT</td>
<td>0.22</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HTMIX</td>
<td>Dummy variable equal to 1 if the observation belongs to the MCT and the Observer bears the high cost of punishing</td>
<td>0.26</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LTMIX</td>
<td>Dummy variable equal to 1 if the observation belongs to the MCT and the Observer bears the low cost of punishing</td>
<td>0.26</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AGE</td>
<td>Age</td>
<td>21.72</td>
<td>1.97</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>GENDER</td>
<td>Dummy variable equal to 1 if male</td>
<td>0.61</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

n = 62  
T = 14  
N = 868  
Log Likelihood = -991.4223  
Sigma_u = 6.2***  
Sigma_e = 3.3***  

***significance 1%    **significance 5%