Dipartimento di Politiche Pubbliche e Scelte Collettive – POLIS
Department of Public Policy and Public Choice – POLIS

Working paper n. 159

January 2010

Free-riding on altruistic punishment? An experimental comparison of third-party-punishment in a stand-alone and in an in-group environment

Peter Lewisch, Stefania Ottone and Ferruccio Ponzano

UNIVERSITA' DEL PIEMONTE ORIENTALE "*Amedeo Avogadro*" ALESSANDRIA

Periodico mensile on-line "POLIS Working Papers" - Iscrizione n.591 del 12/05/2006 - Tribunale di Alessandria

Free-Riding on Altruistic Punishment?

An Experimental Comparison of Third-Party-Punishment in a Stand-Alone and in an In-Group Environment

Peter Lewisch (University of Vienna)
Stefania Ottone (University of Milano - Bicocca)
Ferruccio Ponzano (University of Eastern Piedmont)

Abstract. This paper deals with the subject of third-party punishment. The paper compares, by means of an economic experiment, punishment by a third party (Stand-Alone case) with punishment by third parties (In-Group environment). This deliberate introduction of a second potential punisher is neither subtle nor marginal. Shifting punishment choices into this "enlarged environment" allows us to study, in a systematic way, the complex relationship between the punisher's expectations about her/his peer's punishment decisions and her/his own punishment choices. In particular, we aim to examine whether, on average, individual punishment is systematically lower in an In-Group environment compared with the Stand-Alone case.

JEL classification: C91, C92, K42

Keywords: Third-Party Punishment; Collective Punishment;

1. Introduction

This paper deals with the subject of third-party punishment. The paper compares, by means of an economic experiment, punishment by a third party (Stand-Alone case) with punishment by third parties (In-Group environment). This deliberate introduction of a second potential punisher is neither subtle nor marginal. Shifting punishment choices into this "enlarged environment" allows us to study, in a systematic way, the complex relationship between the punisher's expectations about her/his peer's punishment decisions and her/his own punishment choices. In particular, we aim to examine whether, on average, individual punishment is systematically lower in an In-Group environment compared with the Stand-Alone case.

Our starting point is the empirical finding that, within several economically relevant contexts, subjects are seemingly willing to punish others even when they have to incur the cost of doing so and regardless of any foreseeable economic benefits. An already established body of experimental literature deals with such attitude in two different environments: "second-party punishment" (SPP) and "third-party punishment" (TPP).

In cases of second-party punishment, the potential punisher is the "victim" - typically the other party in an "exchange relationship". The seminal work on SPP is Güth et al. (1982). This article examines the phenomenon of "irrational punishment" in the framework of what is called the "Ultimatum Game" (UG), in which a Proposer decides on how to divide her/his initial endowment with a Receiver. Whereas it is the Proposer´s task to propose the respective split, the Receiver may decide whether or not to accept a positive offer. If s/he rejects the proposed split, neither party receives anything. Empirical evidence shows that the Receivers prefer to reject positive offers when they consider them unfair, thus sacrificing the value of their share for "punishing" the Proposers for their unfair offers. Another variant of this "victim-as-punisher-approach" is Fehr and Gächter (2000), who analyze costly individual punishment in the framework of an exchange relationship in public good settings. They show that the presence of a punishment mechanism generates high cooperation levels in these public goods game settings. Subsequent articles (including, among others, Anderson and Putterman, 2006; Carpenter, 2007; Casari and Luini, 2008; Nikiforakis, 2008) provide a more detailed understanding of the characteristics of sanctions (when applied among peers). The sanctions applied increase with perceived unfairness, they show the characteristics of standard ordinary goods (their demand diminishes when the cost increases), and they are comparable to private goods due to their emotional nature. However, as argued in Fehr and

Fischbacher (2004), the application of SPP is by definition limited, because it exclusively focuses on punishment among subjects that are directly involved in a given situation. Indeed, "if only second parties imposed sanctions, a very limited number of social norms could be enforced because norm violations often do not directly hurt other people" (p. 64). Typically, the stability of a social norm requires enforcement by outsiders, i.e., the intervention of subjects that are not directly involved in the game. This motivates the study of TPP, which analyses the altruistic behaviour ("altruistic punishment") by third-party subjects who are willing to carry out costly punishment in order to sanction norm violators, even when this violation does not cause any direct loss to them. Following Fehr and Fischbacher (2004), subsequent articles (Charness et al., 2007, Bernhard et al., 2006, Ottone, 2008) provide evidence regarding the effectiveness of TPP in different socio-cultural situations. In fact, many contributions in the literature view such "altruistic" provision of costly punishment as the "golden cornerstone" for social stability, namely as a device to ensure compliance with social norms. Or so it would seem.

The literature, so far, does not address the question as to whether such an "altruistic punishment" is robust with respect to the presence of a further potential punisher. In our paper, we want to pursue the aforementioned line of research regarding TPP by examining whether and to what extent such punishment would also occur if we increase the number of potential punishers (allowing for a second potential punisher).

A study of this set of issues requires "solid ground", and we seek to establish this solid ground first by replicating experimentally the existence of TPP. Moreover, we want to examine whether this TPP shows the characteristics of an ordinary good such that its amount decreases with its cost. On this basis, we proceed with examining our central research question, which concerns the impact of a second potential punisher on the amount of individually provided punishment. In particular, we want to examine whether TPP is a private good or a public good - in the sense that there is a mutual influence of punishment choices in the In-Group scenario. The consequences of such a move from a Stand-Alone scenario to the In-Group case are complex. The introduction of a second punisher may have (i) no effect (because of an in-person nature of punishment that makes possible punishers uninterested in the amount of punishment meted out by their peers); (ii) a strong effect because the presence of a second punisher would allow free-riding so that TTP may even collapse; (iii) some effect if punishment comprises elements of both the aforementioned cases. Eventually, we examine whether punishment in the In-Group scenario leads to "over-punishment" in the sense that the

aggregated amount of a punisher's own punishment plus the expected punishment of her/his peer exceeds the amount of punishment by the same individual in the Stand-Alone case.

The paper is organized as follows. After this introduction (section 1), we introduce the experimental design and procedure in section two. The third section is devoted to the expected results, while the fourth section presents the results. Section five concludes.

2. The experimental design

The Game. This experiment is designed such that it allows for a comparison of third-party punishment in the Stand-Alone case (i.e., one single potential punisher) and the In-Group case (i.e., two potential punishers). The underlying scenario is that of a Third-Party Punishment Game (Fehr and Fischbacher, 2004). A Dictator Game is played by the Dictators and the Receivers (As and Bs). The potential punishers (the Cs) do not themselves participate in the Dictator Game, but may inflict punishment on the Dictator (the As) following the latter's decision as to how to share the initial endowment with their "game partner" (the Receivers). If punished, the Dictators suffer monetary loss. Punishment is costly both for the As (suffering the punishment) and for the Cs (carrying out this punishment). The Cs have to invest real resources, namely a certain amount of their own financial endowment, for the goal of imposing punishment (in terms of a multiple of the punishment cost) on A; details are given below.

FIGURE 1 HERE

Technically, the tool (see Figure 1) is a variant of the Third-Party Punishment Game. 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. A assigns a certain portion of his original endowment to B; B is entirely passive: s/he receives what s/he is given. 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. C gets the opportunity to impose a monetary disutility ("punishment") on A. This punishment, available in discrete units, has to be "purchased" by C. It inflicts a pure monetary loss on A (i.e., there is no transfer of the deducted amount to the punisher). Before making her/his punishment choices, each player C is informed that s/he has ½ probability of being the only player C in the group (Stand-

Alone scenario) and ½ probability of being in a group together with another player C (In-Group scenario). In the latter case, where we have two Cs, we call them C1 and C2. The strategy method is implemented at this level. Each player C is asked to declare her/his level of punishment[1] for both cases (Stand-Alone and In-Group) and for each level of transfer from A to B.

The Treatments. In our experimental design, we also introduce differences in the cost of punishment to study the impact of "different price-levels" both for individual punishment choices and for In-Group interaction. The experimental design consists of three treatments: 1) the High-Cost Treatment (HCT) where all player Cs are C_H types; the Low-Cost Treatment (LCT) where all player Cs are C_L types; 3) the Mixed-Cost Treatments (MCT), where in the Stand-Alone scenario player Cs are a mix of C_H and C_L types, while in the In-Group scenario one punisher is a C_H type and the other one a C_L type. After the Cs have made their choices according to the strategy method for all cases, their effective role is randomly drawn and the relative decision implemented. At the end of the game, first order beliefs of players C regarding the expected punishment level by his/her peer (the other punisher) player C if s/he is in a group of four are elicited according to the standard procedure.

The endowment and the payoffs. 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. Player A can choose her/his transfer to Player B among seven options – 0, 5, 10, 15, 20, 25 or 30 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.[2] In the MCT, in groups of four subjects one of the two players C bears the high cost while the other player C bears the low one. Each token value is 0.10 Euro.

In all treatments: 1) subject A's payoff is the remaining sum after her/his transfer to B and the punishment received by player C(s); 2) subject B's payoff is the sum of her/his initial endowment and the amount received by subject A; player C's payoff is the difference between her/his initial endowment and the sum spent to punish player A. More specifically:

$P_A = 100 - t - s$

$P_B = 50 + t$

[1] Our study focuses on monetary sanctions. Consequently, player C's punishment consists of reduction of player A's payoff.

$P_C = 75 - c$

where:

t = number of tokens transferred from A to B

c = number of tokens spent by C to punish A

s = sanction received by player A

The way s is computed depends both on punishment return (equal to 3 if player C's type is C_L and equal to 2 if player C's type is C_H) and on the number of player Cs in the group. In particular:

s = 3*c in the LCT – Stand-Alone scenario

s = 2*c in the HCT - Stand-Alone scenario

s = 3*c in the MCT - Stand-Alone scenario when C = C_L

s = 2*c in the MCT - Stand-Alone scenario when C = C_H

s = 3*(c_1 + c_2) in the LCT - In – Group scenario

s = 2*(c_1 + c_2) in the HCT - In – Group scenario

s = 3*c_1 + 2*c_2 in the MCT - In – Group scenario when $C_1 = C_L$ and $C_2 = C_H$

s = 2*c_1 + 3*c_2 in the MCT - In – Group scenario when $C_1 = C_H$ and $C_2 = C_L$

The Procedure. The experiment has been run at 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). Participants were all undergraduate students of the University of Trento. In each session, participants were split into four groups of 3 subjects (A, B and C) and two groups of 4 subjects (A, B, C1 and C2). Overall, eight one-shot sessions have been run, with a total of 160 participants; namely two sessions of 20 participants in the HCT and in the HCT, and four sessions of 20 subjects in the MCT. The instructions were read by participants on their computer screen while an experimenter read them out loud. After reading the instructions, but before playing the game subjects had to answer some control questions to ensure that they understood the rules of the experiment. At the end of each session, subjects were asked to fill out a brief survey to check for socio-demographic data. Each subject participated in one session only and partners' identities were unknown even after the experiment. To guarantee anonymity subjects received their payment after the experiment. No communication among the participant was allowed. All information concerning the payoff function and the rules of the game where common knowledge. Each session lasted about 60 minutes. Each subject earned on average 7 Euros.

[2] Only transfers of entire tokens are allowed and no participant can earn a negative payoff.

3. Analytical framework: Research questions and behavioural assumptions

a. Research Questions

In this experimental study, we examine three questions. Firstly, we analyse whether the Cs punish at all and how the level of punishment by the Cs varies as the transfer from A to B increases (so that punishment choices are affected by the perceived level of "unfairness" regarding A's behaviour). Secondly, we examine whether third-party punishment is an ordinary good. This would be the case if the amount of punishment decreased with its cost increased. Thirdly, and most importantly, we analyse the impact of an increase in the number of potential punishers for the amount of individually provided punishment.

b. Behavioural Foundations

Let us turn now to the behavioural foundations regarding the voluntary individual provision of costly punishment. According to the classical economic approach, as based on the *Homo Oeconomicus* model, potential punishers will not punish. Any positive level of punishment would imply a monetary cost without conceivable monetary benefit (in terms of deterrence or whatever instrumental purpose) in an one-shot scenario. The predicted absence of punishment would hold true in either scenario (Stand-Alone and In-Group). In turn, based on the findings of the existing experimental literature, one would expect some positive degree of punishment. In light of these findings, one would expect "altruistic punishment" to actually occur and to be not infrequent. Since punishment would not buy these individuals anything in terms of monetary benefits, actual punishment choices are a matter of taste. One would expect these "preferences for punishment" to differ between subjects, similar to any other preference.

It seems conceptually helpful to distinguish between two cases of "altruistic punishment", namely, "instrumental" and "emotional" punishment. We consider punishment an "instrumental good" if the utility of an Observer (player *i*) is (also) influenced by the level of punishment of the other Observer (player *j*). In such a case, player *i* is interested in seeing the Dictator punished, but this very interest is also satisfied if the punishment is carried out by player *j*; *i* can thus derive some utility from the fact that someone else carries out punishment. The utility of punishment can also be enjoyed by a bystander: punishment "spills over" from one punisher to the other. In contrast, we consider punishment an "emotional good" if no such influence exists. In this case, the utility of punishment (in terms, say, of the punisher's own emotional satisfaction) can only be enjoyed when one carries it out oneself. Punishment by the peer does not matter.

We use a simple utility function that reflects the aforementioned framework:

$$U_i = f\left(x_i, p_i(t), p_j(t)\right) \tag{1}$$

where x_i is the payoff of the Observer i, p_i is the effect of punishment of observer i and p_j is the effect of punishment of Observer j and t is the transfer from A to B.

In our simultaneous one-shot experiments, players do not know the real level of p_j but they have only expectations about it. Then, formula (1) becomes:

$$U_i = f\left(x_i, p_i(t), E\left(p_j(t)\right)\right) \tag{2}$$

where E (p_j) represents the expected value for player i of the effect of punishment of player j.
Then, we explicit (for each level of t) our utility function using quasi-linear preferences as in Casari and Luini (2008). Then, Equation (2) becomes:

$$U_i = x_i + a\ln\left(p_i + bE\left(p_j\right)\right) \tag{3}$$

with a budget constraint:

$$x_i + qp_i = y\,. \tag{4}$$

where a and b are two parameters, q is the cost of punishment for player i and y is her/his revenue.
Maximizing our utility function subject to the given budget constraint by means of a Lagrange substitution, we establish that the optimal value of p_i is:

$$p_i = \frac{a}{q} - bE\left(p_j\right)^3 \tag{5}$$

c. Assumptions

Now to our assumptions. Assumption 1 concerns a general behavioural assumption regarding the relationship between the Dictator's transfer and the respective punishment for all cases of "altruistic

[3] The lowest level of p_i is obviously 0.

punishment". Assumptions 2-6 distinguish between different categories of punishers according to their "punishment characteristics". All of these assumptions provide experimental hypotheses for our experiments.

Assumption 1. *The level of third-party punishment decreases as the transfer from the Dictator to the Receiver increases.*

This first assumption concerns, presupposing the empirical existence of some positive level of punishment, the responsiveness of its incidence to the degree of its underlying "event/act". Assuming the perceived unfairness by the Dictator to be the determining factor of punishment, one would predict that increases in the Dictator's transfer to the Receiver will, all other things being equal, decrease the punishment by the Observer. Such a prediction is also based on the experimental literature (see, for instance, Fehr and Fischbacher, 2004; Ottone, 2008) which shows that an increase in the level of transfer reduces the level of punishment. In fact, as the level of *t* increases, we expect a reduction in the level of utility received by punishment. In particular, if a player thinks that the level of transfer is fair, the utility of punishment is obviously equal to 0.

The subsequent assumptions concern different behavioural categories of punishers and specify the respective experimental hypotheses.

Assumption 2. *If third-party punishment is a good with* $a > 0$[4] *in equation (5), it is an ordinary good.*

There is no theoretical reason to assume that third-party punishment is not an ordinary good. Even if one remained agnostic as to the precise motivation of TTP, one would expect less punishment to be carried out if the cost of doing so increases. Equation (5) shows that the level of punishment decreases as the cost of punishment increases. Empirical results along these lines are common in the literature dealing with second-party punishment (Anderson and Putterman, 2006; Carpenter, 2007).

[4] If a = 0, player *i* does not receive utility from punishment, then the level of p_i is always equal to 0. This is the classical *Homo Oeconomicus*.

Assumption 3. *If third-party punishment is a good with $a > 0$ in equation (5), the level of punishment chosen by subject i in the Stand-Alone scenario $p_i = \frac{a}{q}$ is her/his desired level of punishment.*

In the Stand-Alone scenario, Observer *i* is the only player in the group who is allowed to punish. Expected punishment of Observer *j* does not exist. Consequently, her/his chosen level of punishment corresponds to her/his desired level of punishment – namely, her/his desired effect of punishment on the Dictator given the transfer to the Receiver.

Assumption 4. *If third-party punishment is an emotional good with $b = 0$ in equation (5), the punishment of player j does not influence the utility of player i. In particular:*

1) *the level of p_i in the Stand-Alone and in the In-Group scenario will be the same;*
2) *the sum of p_i and E (p_j) in the In-Group scenario is higher than or equal to p_i in the Stand-Alone scenario.*

If third-party punishment is an emotional good, player *i*'s desired level of punishment corresponds to her/his "in persona" punishment. The expected punishment of Observer *j* – whatever the amount – is irrelevant. A clear consequence is that in the In-Group scenario over-punishment may occur.

Assumption 5. *If third-party punishment is an instrumental good with $b = 1$ in equation (5), the punishment of player j does influence the utility of player i. In particular:*

1) *the level of p_i in the Stand-Alone scenario is higher than in the In-Group scenario when subject i expects a positive level of punishment coming from the other punisher*[5]*;*
2) *the sum of p_i and E (p_j) in the In-Group scenario is equal to p_i in the Stand-Alone scenario.*

If third-party punishment is an instrumental good, player *i*'s desired level of punishment corresponds to her/his desired effect of punishment on the Dictator. The expected punishment of Observer *j* is now relevant. It provides a perfect substitute for Observer *i*'s sanction. Observer *i* is only interested in the result (i.e., the fact that the Dictator is punished to the appropriate extent), the person/identity of the punisher does not matter.

[5] In our design we are not able to distinguish about between the emotional or instrumental nature of punishment if and only if the level of punishment does not change in the two stages and E (p_j) is equal to zero. As we will show in the results, only two players show this kind of behaviour.

Assumption 6. *If third-party punishment is a mixed good with* $0 < b < 1$ *in equation (5), the punishment of player j does influence the utility of player i. In particular:*

1) *the level of* p_i *in the Stand-Alone scenario is higher than in the In-Group scenario when subject i expects a positive level of punishment coming from the other punisher;*
2) *the sum of* p_i *and* $E(p_j)$ *in the In-Group scenario is higher than* p_i *in the Stand-Alone scenario.*

If third-party punishment is a mixed good player *i*'s desired level of punishment corresponds to her/his desired effect of punishment on the Dictator, but the identity of the Observer is not completely irrelevant. Observer *i* takes account of the expected punishment of Observer *j,* but not with the same intensity. That is, Observer *j*'s sanction provides a positive utility to Observer *i*, even if the benefit coming from an "in persona" punishment is still higher.

4. Experimental 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 cost of punishing; 3) the number of Observers in the group. Consequently, we study eight subgroups of punishment: LT-Stand-Alone, HT-Stand-Alone, LT-MCT- Stand-Alone, HT-MCT-Stand-Alone, LT-In-Group, HT-In-Group, LT-MCT-In-Group, HT-MCT-In-Group. The former four subgroups show the level of punishment implemented by Observers in groups of three participants in the LCT, the HCT and in the MCT when they bear the low cost and the high cost of punishing respectively. The latter four subgroups cover the same treatments, when the Observer is in a group of four.

Result 1. Third-party punishment takes place and the Observer's level of punishment decreases as the Dictator's transfers increase.

Description of result. Our experiment shows positive levels of third-party punishment for all given scenarios, with the average punishment level that declines with increased transfers from the Dictator to the Receiver. Both results - namely the positive average levels of punishment in terms of tokens spent and the decline of punishment when transfer levels increase - are graphically presented in Figure 2. Since actual punishment (in terms of tokens deducted from the Dictator's money) is always a multiple of the Observers' punishment investment, the aforementioned relationship can also (indirectly) be seen, in a numeric presentation, in Table 1. 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, $p = 0.000$).

FIGURE 2 AND TABLE 1 HERE

Interpretation. Both of the two aforementioned results are perfectly in line with the experimental evidence obtained in previous works (see, for instance, Fehr and Fischbacher, 2004; Bernhard, 2005; Ottone, 2005, 2008). The first finding is simply that individuals carry out costly punishment at all in a one-shot interaction without gaining any material benefit. This positive amount of punishment shows that individuals are willing to impose a certain disutility on someone else because of this person's unfair treatment of the respective Receiver, and they do so irrespective of the fact that this punishment comes at a cost without any conceivable benefit in terms of deterrence. In a sense, these individuals "punish for the sake of punishment"; they satisfy their desire for punishment and "purchase", in some real sense, "retributive justice".[6]

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.[7] It confirms our first experimental assumption. This result suggests, in line with everyday intuitions, that individuals perceive unfairness not as a binary category ("fair", "unfair"), but as a continuum and that the response by potential punishers 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 in terms of a "demand-dependence" of altruistic punishment: 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 of punishment, the lower its demand at each level of transfer from the Dictator to the Receiver is. Table 1 illustrates the respective findings numerically. For each level of transfer by the Dictator punishment is higher in the LC scenario. In other words, when the Observer bears the high cost of punishing, the level of punishment is lower for each subgroup in both scenarios. A random-effects Tobit confirms this result (HT and HT-MCT, $p \leq 0.05$).

[6] 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.

[7] We adopt the notion of unfairness presented by Fehr and Schmidt (1999) – the departure from the "equal split". Therefore we assume that for inequity-averse subjects any transfer from Player A different from 25 is unfair.

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.[8] In this respect, this result confirms the findings reported in the aforementioned experimental literature on SPP.

Result 3. Third-party punishment nature is multifaceted

Overview. This result is threefold. (a) The introduction of an additional potential punisher matters: it reduces, on average, the amount of individually provided punishment. Punishment in the Stand-Alone case is always higher than in the In-Group scenario. (b) Aggregated punishment in groups (as the sum of the punisher's own punishment and the expected punishment of his peer) is, on average, always higher than the respective Observer's punishment in the Stand-Alone case. (c) Individual punishers are heterogeneous as to their individual punishment characteristics and to the degree by which the presence of a second punisher affects their choices.

Description of Result 3a. On average, individually provided punishment is lower in the In-Group environment. A comparison between the Stand-Alone scenario and the In-Group scenario shows that, in all treatments, punishment is lower in the presence of a second potential punisher. Figure 2 provides a graphic illustration of this main finding in terms of average tokens spent on punishment: this graph shows that the line depicting the average number of tokens spent on punishment in the Stand-alone case line is always above the respective In-Group line. Table 1 presents the same result numerically: average punishment is, for all possible cases, lower in the In-Group column (denoted with "I") compared with the respective Stand-Alone case. A random-effects Tobit regression confirms this finding (ALONE, $p = 0.000$).

Interpretation. Result 3a allows for the conclusion that punishment is not a purely emotional matter. If an Observer cared only about the intrinsic benefits of her/his own altruistic punishment, s/he would disregard the presence of a further potential punisher. That is, the expected punishment by her/his peer would not matter. The empirical findings give an inverse result, showing that the presence of a second punisher is relevant for punishment choices.

Description of Result 3b. "Aggregated" punishment levels (in terms of an individual Observer's own punishment plus her/his expected punishment by the peer in the In-Group scenario) exceed punishment in the Stand-Alone case.[9] The potential number of tokens subtracted from the Dictator's final payoff in the In-Group scenario, the sum of p_i and $E\ (p_j)$, is significantly higher (Mann-Whitney test, $p < 0.03$)[10] than the number of tokens subtracted in the Stand-Alone scenario (p_i). These findings are depicted graphically in Figure 3 showing that the "aggregated" In-Group line is always above the Stand-Alone line.

FIGURE 3 HERE

Interpretation. Result 3b illustrates that (expected) punishment by her/his peer does not fully "crowd out" the Observer's own punishment. If punishment were purely instrumental, the punisher would reduce his/her own punishment precisely by the expected amount of his/her peer's punishment, so that overall punishment (calculated as the sum of the punisher's own punishment and the peer's expected punishment) would remain constant.

However, over-punishment – a level of punishment higher than the desired one[11] – in the In-Group scenario may be due to different reasons. First of all, such over-punishment may imply that third-party punishment is an emotional good. However, we already know from result 3a that this is not the case, since the level of punishment in the Stand-Alone scenario is lower than in the In-Group context. A second reason for such an over-enforcement may also stem from the fact that, for all punishers, $0 < b < 1$. This means that punishment by the second Observer works as an imperfect substitute. Finally, Result 3b may be the result of a heterogeneity of the punishers ("multifaceted nature of punishment"). The last mentioned possibility requires closer observation, namely an exact breakdown of punishment pattern to the various types of punishers. This breakdown is given in Result 3c below.

Description of Result 3c. Analysis of actual punishment choices on an individual level shows that punishers are in fact heterogeneous. A certain number of Observers never punishes ("*homo*

[8] See Anderson and Putterman, 2006; Carpenter, 2007.
[9] Thus, "aggregated" punishment is, thus, used as a hypothetical concept, reflecting the "potential" number of tokens (the punisher's actual punishment and her/his estimates about her/his peer's punishment) subtracted from the Dictator's final payoff.
[10] We ran this test for each subgroup. The H_0 is the equality of the two variables.
[11] p_i in the Stand-Alone scenario.

oeconomicus"). Among those who do punish, some behave along "instrumental lines" (reducing their own punishment according to expected punishment by the peer), others along "emotional lines" (being unresponsive to the presence of a second potential punisher), others still exhibit mixed characteristics. Table 2 gives the exact breakdown of the pertinent empirical results.

TABLE 2 HERE

The average share of those individuals who never punish is 45%. The percentage of those individuals who – at least to a certain extent – carry out TPP amounts to 48%. The residual 7% makes unclear or undefined choices. Concerning the different types of punishment patterns, it is possible to distinguish heterogeneous preferences for punishment. Overall, punishment is an emotional good for 11% of the subjects - which is around ¼ of the punishers. This 11% ignore the presence of a second punisher such that their punishment remains completely unaffected by the appearance of a second potential punisher. Punishment is an instrumental good for 13% (which is around 1/3 of the punishers). For these individuals, punishment by their peers spills over to their own punishment choices, leading to a full crowding out. Punishment is a mixed good for 24% (which is around ½ of the population of punishers). Their punishment combines elements of an emotional and instrumental good. Consequently, even if there is a spill-over effect, the crowding out is only partial.

Interpretation. Result 3c shows that differences in tastes and preferences also matters for punishment. In general, in our TPP experiment, a slightly higher number of punishers emerges. Only 11% of our individuals punish without regard to the presence of a possible second punisher. In turn, the presence of a second punisher affects punishment choices, at least to some degree, for 37% of our individuals (namely for 13%, for whom punishment is an instrumental good and for 24%, for whom it is a mixed good). This 37% reduce their own punishment efforts in the shadow of a second punisher ("free riding"). "Aggregated" punishment (in terms of the sum of an Observer's actual punishment and his/her estimates as to the punishment by his/her peer), however, exceeds average punishment in the Stand-Alone case.[12] On average, individuals accept an outcome, whereby the appearance of a second punisher (potentially) increases overall punishment levels. In this

[12] The aforementioned results reflect only the two punisher case. They do not cover the more general case of *n* possible punishers, i.e. the effects on punishment when moving to a more encompassing group (with *n* potential punishers). Whereas one would expect a further drop in the individual punishment activities, the overall ("aggregated") effects are unclear.

respect, therefore, even based on the individual punisher's own evaluations, the In-Group scenario holds the potential for excessive punishment.

5. Overall interpretation of experimental findings and application to social and legal norms

In this last section, we discuss implications and applications.

Implications. They concern TPP in its role for the provision of social and legal stability. We consider two basic implications.

The first one is fundamental as to the ability of TPP to provide the crucial enforcement device for social (and legal) stability, i.e. the "cement of social stability". In this respect, our contribution cautions the optimism of the recent literature. The praise in the literature of TPP as the "golden cornerstone" of spontaneous social order seems to be motivated by the mere existence if the (important) empirical phenomenon of "altruistic punishment". Implicitly, this literature seems not to take into account the potential instrumental nature of such a punishment. In our experiments, four times as many individuals as emotional punishers have not punished at all (genuine *homo oeconomicus*). The impact of altruistic punishment, hence, depends to a considerable extent on the behaviour of the residual part of the population (in our results 37%). For these latter individuals, TPP is either a purely instrumental or a mixed good. For either group, the presence of additional punishers erodes, at least partly, the preparedness to engage in punishment. For these individuals, hence, there exists the potential for free riding on punishment. It seems as if this potential for free riding and the instability of TPP therefrom arising has not yet been fully developed in the literature. Our experimental findings, to be sure, do not suggest an overall breakdown of TPP, but they emphasize the potential fragility of the social and legal stability. Whereas TPP does play an important role in the provision of this social stability, it risks to be itself unstable. This point deserves further theoretical and empirical inquiry.

The second implication concerns "over-punishment". "Over-punishment" is a complex issue, and we approach it only in a few of its aspects from a modest position. In our paper, we do not discuss "over-punishment" in terms of an inefficient duplication of enforcement efforts that may always arise from uncoordinated enforcement activities of various potential enforcers. Rather, we focus on "over-punishment" in terms of excessive punishment. Since the determination of just, fair or

adequate punishment is a difficult, and inherently unsolved matter, we apply an agnostic position. We compare average punishment (in terms of tokens deducted) in the stand alone case with that potential average punishment that results, in the in-group scenario, from the punisher's own punishment and his/her estimates as to the punishment of his/her peers. We examine this aggregated potential punishment to see the extent to which individuals are willing to accept levels of punishment that exceed their own evaluation of adequate punishment in the stand alone case. Our results suggest that the potential for "excessive" punishment in in-group scenarios exists, and it exists, because peer-punishment seemingly is, for a relevant group of the population, not a full substitute for one's own punishment efforts. From an institutional standpoint, these findings suggest room for legal or social rules that would specify the amount of punishment to be inflicted on a wrongdoer in multi-punisher cases. The need for legal intervention may, however, be somewhat smaller than it appears at first hand. In this respect, the multifaceted nature of punishment deserves closer attention. It seems as if the heterogeneity of punishment motivations works as an implicit mitigating device. Social stability in terms of avoidance of excessive enforcement seems to rest on the fact that neither is everybody a *homo oeconomicus* nor an emotional punisher. If the world consisted only of self-righteous enforcers, we would likely run into over-punishment, if it consisted only of narrowly conceived *homines oeconomici*, into breakdown of any enforcement. The population mix, as it exists in a heterogeneous society between *homines oeconomici* who never punish and punishers (of whom some view punishment as an emotional good, others as an instrumental or mixed good) may help to balance out punishment and to avoid over-punishment.

Applications. In a sense, TPP concerns punishment by bystanders. Legally speaking, the pure case of an "unaffected third party" exercising punishment choices is the judge or, more generally, a legal authority. Simple examples range from the more spontaneous the lynching of the crowd to the range of crimes that can be sanctioned at the same time by two authorities (i.e. by a EU court and by a member state; by a civil court and by the antitrust authority). In all these cases, even in different scenarios with different rules, more than one third-party punisher acts and the expected level of punishment imposed by an authority could influence the level of the punishment of the other one.
Our experiment here is of some assistance in better understanding the behavioural framework of lay judges (juries) and legal authorities. In particular, it analyses the relevance of the existence of legal rules when over-punishment may occur. However, this is the first step. In fact, it is rare that lay judges or legal authorities are exposed to uncoordinated punishment choices, because they render

their decisions in an institutionally well defined framework, requiring “coordinated choices”. This point needs further inquiry.

However, we point out that the immediate application of TPP is in the realm of the enforcement of social rules. Even our experimental scenario is not one of an underlying legal offence, but of an act that seemingly is considered unfair by the observer and sanctioned by costly punishment. It is here, where we find a wide field of applications.

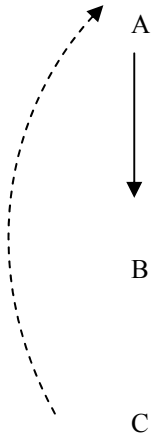
6. Summary

We have conducted experiments comparing individual altruistic punishment in a Stand-Alone case and in a In-Group scenario. The result is fourfold. Firstly, our experiment confirms the empirical phenomenon of third party punishment; punishment levels increase with the perceived unfairness of the underlying behaviour. Secondly, TPP is sensitive to its costs. Thirdly, and most importantly, the introduction of a potential additional punisher reduces, on average, the amount of individually provided punishment. Punishment in the Stand-Alone case is significantly higher than in the In-Group scenario. Fourthly, tastes for punishment differ. A certain number of individuals never punish. Among those who punish, there exists a subgroup of individuals for whom punishment is an emotional good (meaning that their punishment remains unaffected by the appearance of a peer); for a further subgroup punishment is an instrumental good (meaning that expected punishment by the peers crowds out punishment), for the residual subgroup, punishment is a mixed good.

References

- C. M. Anderson and L. Putterman (2006), "Do non-strategic sanctions obey the law of demand? The demand for Punishment in the voluntary contribution mechanism", *Games and Economic Behavior*, 54 (1), 1-24.
- Bernhard (2005), "Third Party Punishment within and across groups – An experimental study in Papua New Guinea, IERE Zurich PWP.
- J. Carpenter (2007), "The demand for punishment", *Journal of Economic Behavior and Organization*, 62 (4), 522-542.
- Casari and Luini (2005), "Group cooperation under alternative peer punishment technologies: An Experiment", LABSI WP 002.
- E. Fehr and U. Fischbacher (2004), "Third Party Punishment and Social Norms", *Evolution and Human Behavior*, 25, 63-87.
- E. Fehr and S. Gachter (2000), "Cooperation and punishment in public goods experiments", *American Economic Review*, 90 (4), 980-994.
- W. M. Landes and R. A. Posner (1975), "The private enforcement of law", *Journal of Legal Studies*, 4 (1), 1-46.
- P. Lewisch, (1995) Punishment, Public Law Enforcement, and the Protective State, Springer.
- N. Nikiforakis (2008), "Punishment and counter-punishment in public good games: Can we really govern ourselves", *Journal of Public Economics*, 92 (1-2), 91-112.
- S. Ottone (2005), "Transfers and altruistic punishments in Solomon's game experiments", POLIS WP 57.
- S. Ottone (2008), "Are people samaritans or avengers", *Economics Bulletin*, 3 (10), 1-8.

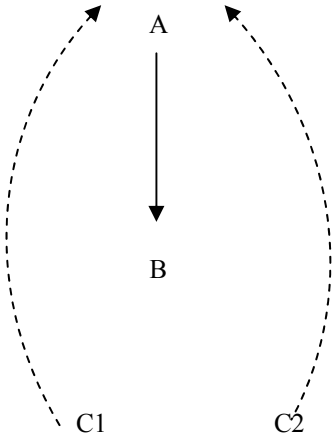
Appendix 1 – Figures and Tables



Third-Party Punishment Game (TPP)

Group of three

Stand-Alone scenario



Third-Party Punishment Game (TPP)

Group of four

In-Group scenario

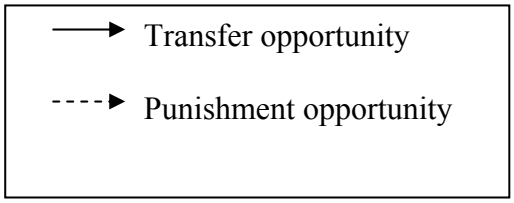


Fig. 1 The Experimental Games

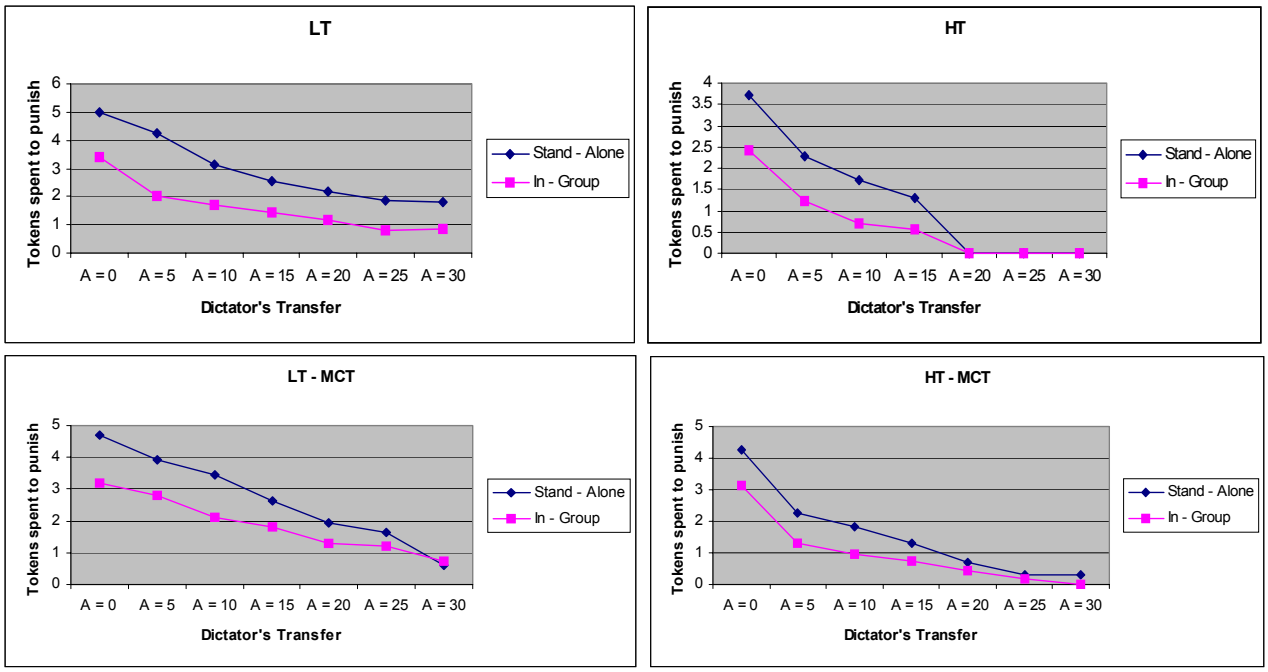


Fig. 2 Punishment Pattern

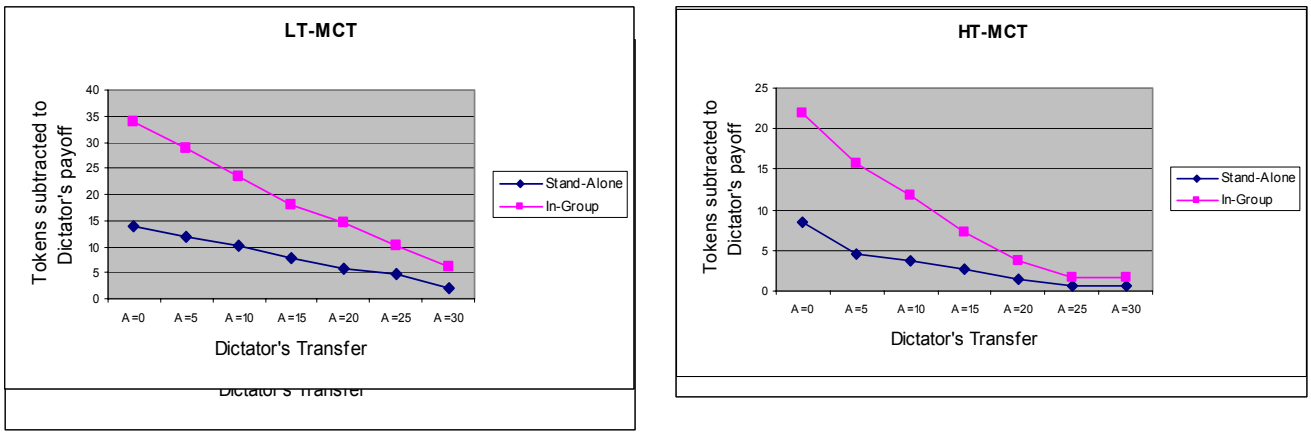


Fig. 3 Potential effect of punishment

Table 1. Average punishment in both scenarios for each Dictator's transfer

	Average punishment (in tokens) when the Dictator's transfer (in tokens) is…													
	…0		…5		…10		…15		…20		…25		…30	
	S[a]	***I***[b]	***S***	***I***	***S***	***I***	***S***	***I***	***S***	***I***	***S***	***I***	***S***	***I***
HT	3.71	2.43	2.29	1.21	1.71	0.71	1.29	0.57	0	0	0	0	0	0
LT	5	3.37	4.25	2	3.12	1.69	2.56	1.44	2.19	1.19	1.87	0.81	1.81	0.87
HT - MCT	4.25	3.12	2.25	1.31	1.81	0.94	1.31	0.75	0.69	0.44	0.31	0.19	0.31	0
LT - MCT	4.69	3.19	3.94	2.81	3.44	2.12	2.62	1.81	1.94	1.31	1.62	1.19	0.62	0.75

a Stand-Alone scenario
b In-Group scenario

Table 2. Third-party punishment nature

	a = 0	**a > 0**			**nd**
		b = 0	**0 < b < 1**	**b = 1**	
HT	65% N = 9	14% N = 2	7% N = 1	14% N = 2	-
LT	31% N = 5	13% N = 2	25% N = 4	25% N = 4	6% N = 1
HT – MCT	56% N = 9	-	38% N = 6	-	6% N = 1
LT - MCT	31% N = 5	18% N = 3	25% N = 4	13% N = 2	13% N = 2
TOTAL	45% N = 28	11% N = 7	24% N = 15	13% N = 8	7% N = 4

Appendix 2 – The econometric analysis

$$P_i = \beta_0 + \beta_1 AGIVES_i + \beta_2 ALONE_i + \beta_3 HT_i + \beta_4 HTMIX_i + \beta_5 LTMIX_i + \beta_6 AGE + \beta_7 GENDER_i + \varepsilon_i$$

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

Variables	
AGIVES	-.286***
	(0.018)
ALONE	2.01***
	(0.33)
HT	-1.7**
	(0.869)
HT-MCT	-1.667**
	(0.783)
LT-MCT	.36
	(0.817)
AGE	- .032
	(0.087)
GENDER	.411
	(0.507)
Constant	4.897**
	(1.986)
n	62
T	14
N	868
Log Likelihood	-991.4223
Sigma_u	6.2***
Sigma_e	3.3***

*****significance 1% **significance 5%**

Description of the variables used in the regression

Name	Descripton	Mean	Standard Deviation	Min	Max
AGIVES	transfer from A to B	15	10	0	30
ALONE	Dummy variable equal to 1 if the Observer is in a group of three	1.73	3.59	0	30
HT	Dummy variable equal to 1 if the observation belongs to the HCT	0.22	0.42	0	1
HT-MCT	Dummy variable equal to 1 if the observation belongs to the MCT and the Observer bears the high cost of punishing	0.26	0.44	0	1
LT-MCT	Dummy variable equal to 1 if the observation belongs to the MCT and the Observer bears the low cost of punishing	0.26	0.44	0	1
AGE	Age	21.72	1.97	19	28
GENDER	Dummy variable equal to 1 if male	0.61	0.49	0	1

Recent working papers

The complete list of working papers is can be found at http://polis.unipmn.it/pubbl

***Economics Series** ****Political Theory Series** **[ε] Al.Ex Series**
[T]Territories Series **[t]Transitions Series**

2010 n.159[ε] Peter Lewisch, Stefania Ottone and Ferruccio Ponzano: *Free-riding on altruistic punishment? An experimental comparison of third-party-punishment in a stand-alone and in an in-group environment*

2009 n.157* Alessandro Lanteri and Stefania Ottone: *Economia ed etica negli esperimenti*

2009 n.156* Cinzia Di Novi: *Sample selection correction in panel data models when selectivity is due to two sources*

2009 n.155* Michela Martinoia: *European integration, labour market dynamics and migration flows*

2009 n.154* Massimo Pasquariello and GianMarco Chiesi: *Valore aggiunto e tipologia di spesa in Piemonte. Un confronto tra Alessandria e gli altri capoluoghi di provincia*

2009 n.153* Massimo Pasquariello: *Produttività, sistemi locali del lavoro, specializzazione produttiva e scenari futuri. Studio preliminare in preparazione del Piano strategico per il comune di Alessandria*

2009 n.152* Massimo Pasquariello and GianMarco Chiesi: *L'analisi dei determinanti di un'area territoriale. Studio preliminare in preparazione del Piano strategico per il comune di Alessandria*

2009 n.151* Cristina Elisa Orso: *Formal and informal sectors: Interactions between moneylenders and traditional banks in the rural Indian credit market*

2009 n.150* Michele Giuranno: *The logic of party coalitions with political activism and public financing*

2009 n.149* Matteo Migheli: *Sharing the pie: the Lutheran is neither opportunistic nor generous*

2009 n.148* Amedeo Fossati and Marcello Montefiori: *Migrants and mafia as global public goods*

2009 n.147* Alberto Cassone: *L'impatto economico dell'Universita' del Piemonte Orientale "Amedeo Avogadro": un aggiornamento al 2008*

2009 n.146* Daron Acemoglu, Davide Ticchi and Andrea Vindigni: *Persistence of civil wars*

2009 n.145* Daniele Bondonio: *Impact identification strategies for evaluating business incentive programs*

2009 n.144* Barry Eichengreen, Rachita Gullapalli and Ugo Panizza: *Capital account liberalization, financial development and industry growth: a synthetic view*

2009 n.143* Emma Galli and Roberto Ricciuti: *Sulla* political economy *del deficit pubblico nell'Italia liberale*

2009 n.142* Matteo Migheli: *Religiosity and happiness: an ever-winning couple? An answer from India*

2009 n.141** Stefano Parodi: *I media dell'Alessandrino e l'Unione Europea*

2009 n.140* Matteo Migheli: *The two sides of a ghost: Twenty years without the wall*

2009 n.139[ε] Matteo Migheli and Francesco Scacciati: *How does labor supply react to different tax rates? A field enquiry*

2009 n.138[ε] Matteo Migheli and Guido Ortona: *Majority, proportionality, governability and factions*

2009 n.137** Noemi Podestà: *Strumenti di mediazione per la risoluzione di conflitti. L'esperienza dell'Osservatorio per il collegamento ferroviario Torino-Lione*

2009 n.136** Noemi Podestà and Alberto Chiari: *Esperimenti di democrazia deliberativa. Informazioni, preferenze e stili di conduzione in tre giurie di cittadini.*

2009 n.135** Andrea Lanza: *1848 comme reconfiguration des discours politiques.*

2009 n.134* Rongili Biswas, Nicolas Gravel and Rémy Oddou: *The segregative properties of endogenous jurisdictions formation with a welfarist central government*

2009 n.133[ε] Matteo Migheli: *Assessing trust through social capital? A possible experimental answer*

2009 n.132* Piero Cavaleri, Michael Keren, Giovanni B. Ramello and Vittorio Valli: *Publishing an E-journal on a shoe string: is it a sustainable project?*

2009 n.131* Alberto Cassone: *L'impatto economico e sociale dell'Universita' del Piemonte Orientale Amedeo Avogadro*

2009 n.130* Alberto Cassone and Pietro Zaccarella: *Il bilancio sociale delle universita'. Inventario dei problemi e analisi comparata delle esperienze italiane*