Department of Economics
ISSN number 1441-5429
Discussion number 06/18

Third-Party Punishment: Retribution or Deterrence?

Fangfang Tan[1] and Erte Xiao[2]

Abstract:
We conduct an experiment to examine the role of retribution and deterrence in motivating third party punishment. In particular, we consider how the role of these two motives may differ according to whether a third party is a group or an individual. In a one-shot prisoner's dilemma game with third party punishment, we find groups punish more when the penalty embeds deterrence than when it can only be retributive. In contrast, individual third parties' punishment decisions do not vary on whether the punishment has any deterrent effect. In general, third party groups are less likely to impose punishment than individuals even though the punishment is costless for third parties.

Keywords: Third-party punishment, group decision making, retribution, deterrence, social dilemmas, experiment
JEL Codes: C72, C92, D63, D70

[1] Department of Public Economics, Max Planck Institute for Tax Law and Public Finance.
Email: fangfang.tan@tax.mpg.de
[2] Department of Social and Decision Sciences, Carnegie Mellon University. Email: exiao@andrew.cmu.edu

1. Introduction

Punishment plays an important role in maintaining social order and economic relationships (for literature reviews see Gächter and Herrmann, 2009; Chaudhuri, 2011). Numerous studies point out that societal stability relies not only on second-party punishment in which only an implicated party can punish a wrongdoer, but perhaps more importantly on third-party punishment, in which those who sanction norm violators do not directly benefit from their decisions (Fehr and Fischbacher, 2004).

Recent biological evidence on animals indicates that third-party punishment is unique to humans. Even our closest relatives, chimpanzees, do not engage in third-party punishment (Riedl et al., 2012). In addition, recent empirical findings on antisocial punishment and retaliatory punishment also highlight the importance of third-party punishment as an effective means to defend justice and avoid the escalation of violence that might lead to feuds.[1] Studies on second-party punishment suggest that it can lead to the escalation of retaliation and unjustifiable punishment. It is therefore efficient to centralize the punishment in the hands of third parties in large societies, as their decisions are less selfishly motivated (Herrmann et al., 2008, Carpenter and Matthews, 2010; Jensen, 2010).

In contrast to the large literature devoted to understand second-party punishment (see, e.g., Carpenter and Matthews, 2009; Casari and Luini, 2012; Xiao and Houser, 2005), what motivates third-party punishment remains poorly understood. Moreover, the few current studies focus exclusively on individual third party[2]. In reality, however, a majority of third party punishment decisions are made by a group such as a committee or a jury instead of by individuals. In this paper, we examine how retribution and deterrence motivate third-party punishment. More importantly, we investigate how large role these two motives play depending on whether the

[1] Recent studies show that unrestricted second-party peer punishment can lead to antisocial punishments, feuds, or inequality (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Hermann et al., 2008; Houser and Xiao, 2011; Nikiforakis, 2008; Nikiforakis and Engelmann, 2011; and Nikiforakis et al., 2012). These studies draw attention to the importance of designing restricted punishment mechanisms. In addition to the intervention of a third party, peer punishment can be regulated by predetermined rules (Andreoni and Gee, 2012; Xiao and Houser, 2011; Xiao and Kunreuther, forthcoming) or voting mechanisms (Ertan et al., 2009; Kosfeld et al., 2009; Noussair and Tan, 2011; Tyran and Feld, 2006) .

[2] See, e.g., Almenberg et al., 2011; Chavez and Bicchieri, 2013; Fehr and Fischbacher, 2004; Walker and Halloran, 2004; Kurzban et al., 2007; Ottone et al., 2008; Marlowe et al., 2008; Carpenter and Matthews, 2010, 2012; Ouss and Peysakhovich, 2012; Casari and Luini, 2012. Some law literature (MacCoun and Kerr, 1988) studied mock jury decision making and identified several biases in group judgment such as the defendant bias (being less willing to convict a defendant). Our paper focuses on uncovering the underlying motivation whether certain behavior deserves punishment or not, rather than assessing the likelihood that a defendant commits a crime.

third-party punishers are individuals or unitary groups (i.e., a group that makes a joint decision). From a policy-maker's perspective, understanding what motivates people's punishment decisions can help in designing regulations, since people are more willing to comply with regulations and rules that are in line with their judgment of what should be done (Tyler, 2006).

The literature in philosophy, law and social psychology reasons for imposing punishment has pointed to two relevant motivations to economists: retribution and deterrence (Carlsmith et al., 2002; Durkheim, 1973; Kurzban and DeScioli, 2013; Woods, 2006).[3] Retribution is non-utilitarian, in that it is triggered by the desire to give the wrongdoers their just deserts for the harm they inflict on others.[4] In contrast, deterrence is utilitarian, since punishment is used as a tool to prevent future norm violations. Differentiating between these two motives helps explain and predict the conditions under which people punish. Deterrence-driven punishment is correlated with the probability of future violation and hence should not happen if future violation is no longer possible. In contrast, retributive punishment, correlated with the harm inflicted on the victims, will occur even if the violator no longer has any chance to inflict any harm on others.

This paper attempts to answer two questions: (1) To what extent do third parties use punishment as retribution or to deter defectors? (2) How does the answer to the first question differ according to whether the third party is a group or an individual?

We study these questions using controlled laboratory experiments based on a framework of special interest to economists: a one-shot prisoner's dilemma (PD) game. A third party can decide whether or not to *approve* any punishment decision proposed by players in the PD game on their counterparts. The punishment proposer, rather than the third party, will pay the cost of punishment if the third party approves it. As we elaborate later in details, one advantage of using this design is that it provides clean evidence for understanding the motivation of punishment decisions. It also captures two features of punishment institutions in civilized societies. First,

[3] There are many ways to classify the motives underlying third-party punishment behavior. For example, Kriss et al. (2013) examine the motives of punishment based on whether people intrinsically enjoy punishment or feel obligated to do so due to image concerns. In this paper, we classify the motives according to the goal of the punishment outcome: Is it to deter future violations or restore justice?

[4] Although retributive punishment is inconsistent with the homoeconomics assumption, it could be driven by negative emotion towards norm violators (Xiao and Houser, 2005; Hopfensitz and Reuben, 2009; Casari and Luini, 2012), moral concerns (Cubitt et al., 2011) or the relative earnings comparison (Dawes et al., 2007; Houser and Xiao, 2010; Bosch-Rosa, 2012). Numerous experimental studies show it occurs in social and economic relationships: Victims or even third parties often punish the wrongdoers in a one-shot scenario with no future interaction (Baldassarri and Grossman, 2011; Fehr and Gächter, 2000, 2002; Fehr and Fischbacher, 2004; Güth, 1995; Henrich et al, 2006, Kurzban et al, 2007).

punishment often will not occur if victims do not make an accusation. One example of this is the U.S. criminal procedure in which a jury composed of members randomly selected from the eligible population has the opportunity to ascertain the guilt or innocence of a defendant when the defendant is accused by a victim. Second, to ensure the justice of punishment, the decisions of third parties (e.g., committee members, jury, or judges) often do not have any significant direct impact on their payoffs (Babcock and Loewenstein, 1997; Xiao, 2013).

The experiment has a two-by-two design, resulting in four treatments. A third party could either be an individual or a group (Individual vs. Group). The decisions of a third party could be announced either before or after the two players have decided whether or not to cooperate in a PD game and proposed punishment after knowing the counterpart's cooperation decision (Ex-ante vs. Ex-post). The treatment difference between Ex-ante and Ex-post measures the extent to which third-party punishment is motivated by deterrence.

Our main finding is that groups are more likely to punish instrumentally than individuals: Groups are more likely to approve punishment toward defectors when the punishment has a deterrence effect (in the Ex-ante treatment) than when the punishment can only be retributive (in the Ex-post treatment). In contrast, individual third parties punish both in the Ex-ante and Ex-post treatments in similar ways. Furthermore, consistent with previous studies such as Tan and Xiao (2012), a significant amount of proposed punishment is negated by the third party. However, a new result of this study is that when the third party is a group, the punishment disapproval rate is even higher, even though it is a costless decision to punish.

Our study contributes to the understanding of both third-party punishment behavior and group decision making. First, previous studies show that, although third parties are willing to incur costs to punish wrongdoers, the frequency of costly third-party punishment is often much lower than second-party punishment (e.g., Fehr and Fischbacher, 2004; Kurzban et al., 2007; Carpenter and Matthews, 2012; Pederson et al., 2013). In all of the studies, punishment is costly for the third party. In addition, implicated parties in the game cannot communicate with their third party whether they think punishment should be imposed. In principle, the low frequency of punishment can be attributed to the cost of punishment (Ottone et al., 2008) and the uncertainty surrounding whether the implicated parties themselves would like the punishment to happen. Our study excludes these possibilities and suggests that many third parties are simply not willing to

punish even when the implicated parties have explicitly expressed the desire to punish the wrongdoer at their own cost.

Second, in the literature regarding group decision-making on punishment, a well-documented finding is that groups are less willing to impose costly punishment than individuals (Bornstein and Yaniv, 1998; Robert and Carnevale, 1997; Auerswald et al, 2013).[5] Since decision-makers in those studies have a stake when they make decisions, it is therefore unclear what contributes to the observed difference in group and individual decision-making. Are groups more likely to seek to maximize profits? Or, it is the judgment of what behaviors deserve to be punished changes in the group decision process? The data from our experiment suggest that, even when the decision is payoff-independent, groups are less likely to punish than individuals. Moreover, groups' decisions are more likely to be consistent with utilitarianism than individual decisions. We analyze chat messages to help understand how groups make decisions.

The rest of the paper is organized as follows. Section 2 describes the experimental design and procedures. Section 3 constructs a model framework and outlines our testing hypotheses. Section 4 presents the results from the experiment. Section 5 discusses the results and section 6 concludes.

2. Experiment design and procedures

To provide clean evidence to understand the motivation of third-party punishment decisions, we design our experiment based on a simple one-shot simultaneous prisoner's dilemma (PD) game. Person A and B simultaneously choose between the action of cooperation and defection (see Table 1).[6] Then, upon observing each other's choices, A and B independently decide whether or not to propose earnings deductions to each other.

A third party (Person C) could decide whether to approve punishment proposals and, in case of approval, how many tokens to deduct from the earnings of the punishee. We use the strategy method to elicit third-party decisions in each of seven scenarios where at least one implicated party proposed certain punishment. Since we are mostly interested in the punishment decisions toward defectors, the strategy method allows us to obtain more observations for this case than a

[5] An exception is Keck (2013) who shows that groups punish more than individuals if they are deceived by other players, even though groups cheat more if given this opportunity. Engel (2010), Kugler et al. (2012) and Charness and Sutter (2012) offer excellent reviews of this literature.

[6] In the experiment, we use natural language by labeling cooperation as "Option I" and defection as "Option II".

direct response method[7]. The scenarios include four possible outcomes that both players defect, both cooperate, and one defects and the other cooperates. In each scenario, we distinguish the case in which one person proposes to punish the other, versus the case in which two persons mutually propose to punish each other. All seven scenarios are displayed at once and in a fixed order on the third party's decision screen across all treatments (see a screen shot for groups labeled as Figure 1). The maximum amount of punishment is 40 tokens, which is exactly the earnings of a unilateral defector. Earnings are reset to be zero if they are negative as a result of punishment.

Table 1: Payoff table of the prisoner's dilemma game

		Person B	
		Option I(Cooperate)	Option II (Defect)
Person A	Option I (Cooperate)	30 / 30	40 / 15
	Option II (Defect)	40 / 15	20 / 20

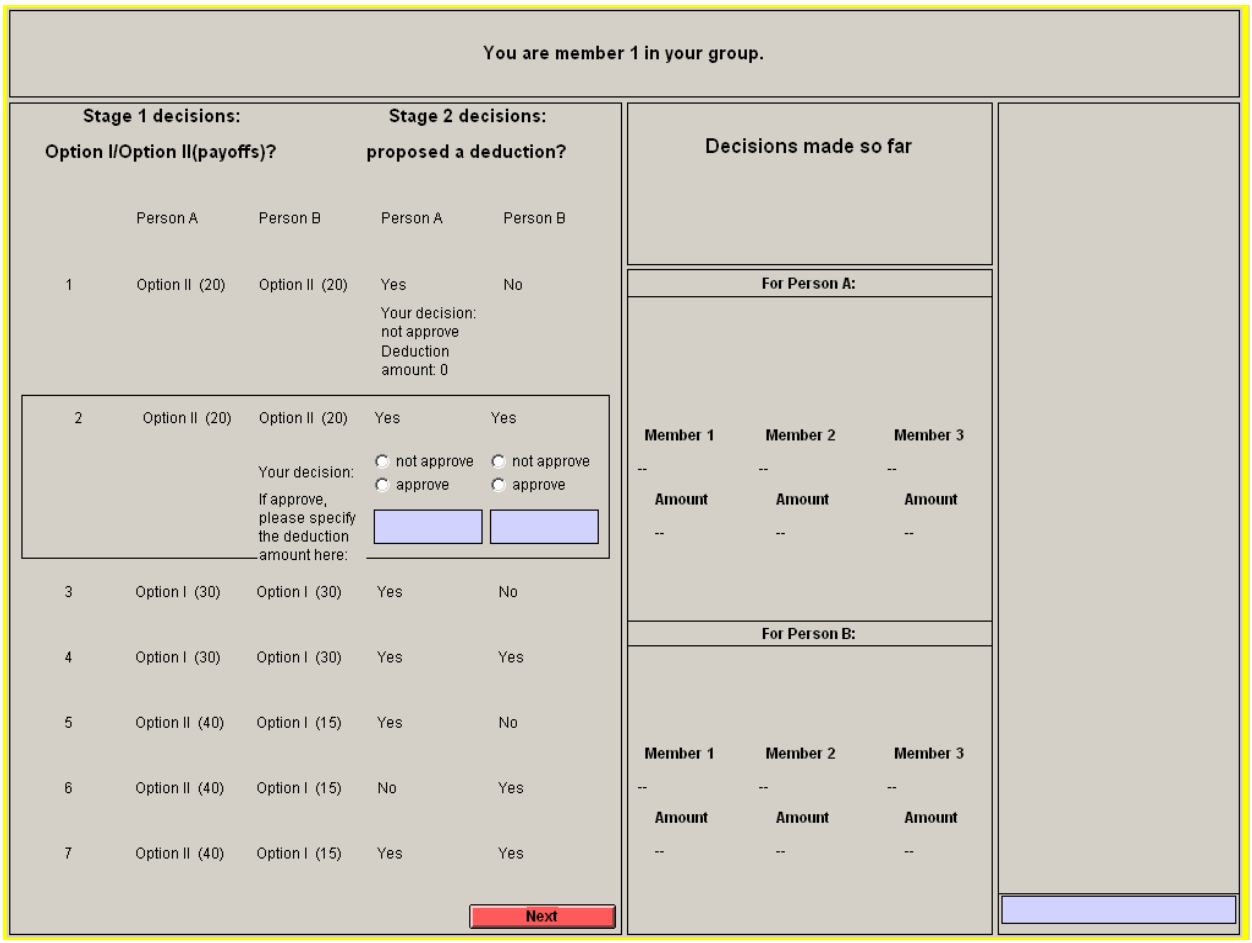
Note: The exchange rate is 5 tokens to $1.

The third party's earnings are not affected by whether he approves the punishment proposal.[8] If the third party decides to approve the proposal, the punishment proposer (Person A or B) will incur a fixed cost of 5 points and the deduction amount does not go to the earnings of any player. All these are common knowledge. As we mentioned before, the profit-independence of punishment decisions in our experiment captures the nature of many third punishment mechanisms in natural occurring environments. More importantly, this feature allows us to learn the true punishment preference of third parties as their own profit no longer plays any role in the decisions.(Anderson and Putterman, 2006; Carpenter, 2007). These features of the third party punishment mechanism are applied to all treatments.

[7] It has been argued that the main difference between the strategy and the direct response methods is that "hot" emotion is more likely to play a role in decision using the latter method (Brandt and Charness, 2009). Strategy method thus better suits the purpose of our study as it minimizes the emotion differences among treatments.

[8] This feature distinguishes our paper from other studies in which the solitary punisher also profits from the game (e.g., DevlinFoltz and Lim, 2008; O'Gorman et al., 2009).

Figure 1: Screen shot in the group treatments



Our study follows a two-by-two design. Table 2 displays the design and the number of independent observations in each treatment. The first treatment variable is the timing in displaying the punishment strategies of a third party. In the Ex-post treatments, a third party is told that his decision in each scenario will be revealed to Persons A and B only after they have decided whether to cooperate and whether to propose to punish the counterpart. Thus, it is clear to the third party that his decision will not affect their behavior. In the Ex-ante treatments, the third party's punishment decisions in all scenarios are revealed to Person A and B before they make any choices in the prisoner's dilemma game. Thus the punishment decisions can influence A's and B's decision. This design also captures the feature of the punishment mechanism in

reality, where and laws rules are set out and potential norm-breakers can know the consequences of their behavior before making decisions.

Table 2: Treatment design

	Individuals	Groups
Ex – post	IndExpost: 31	GrpExpost: 31 (31 × 3 = 93 players)
Ex – ante	IndExante: 31	GrpExante: 29 (29 × 3 = 87 players)

Note: The number of independent observations in each treatment is in each cell.

The second treatment variable is whether a third party is an individual or a group of three individuals. When a third party is a group, the three members can exchange messages freely with each other via an electronic chat box during the decision-making stage until they reach joint agreements.[9]

Figure 1 displays the decision screen in the group treatments. All seven scenarios are presented to third parties on the decision screen at once in a fixed order. Each group will discuss what decision to make in each scenario one by one. Once the group finishes making a decision in one scenario, they could move on to the next but could no longer change the previous decisions anymore. We do not impose any particular decision rules on the group. That is, group members could first discuss among several possible solutions and agree upon one, or one member could start making one proposal, and other members express their opinions whether or not to agree. The decisions of all previous scenarios remain on the screen when third parties make decisions in a new scenario.

It is well established that decisions can be influenced by relative payoff comparisons (Fehr and Fischbacher, 2004; Dawes et al. 2007; Leibbrandt and Lopez-Perez, 2011,2012). In our experiment, third party's decisions can be potentially affected by earnings comparisons between himself and Person A/Person B in the PD game. We designed our experiment similar to Xiao (2013) to minimize such confounds: The third party's payoff is determined by a random number

[9] In many judicial systems, the jury's requirement to reach unanimity before convicting a defendant is legislated. For example, in criminal law jury trials, a guilty verdict is often required to be unanimous in many commonwealth countries such as Canada (see Criminal Code of Canada, Part XX: Jury Trials). Another reason we choose the unanimity rule is to facilitate result comparisons with the existing literature, most of which use the unanimity rule for groups.

from the prisoner's dilemma payoff matrix (15, 20, 30, 40) with equal probabilities. Moreover, a third party knows his or her earnings only at the end of the experiment. Neither Person A nor Person B knows the earnings of the third party throughout the experiment. All these are common knowledge.

We expect group third parties to take a much longer time to finish decisions than the individual third parties. If we conduct the PD game in the same session as as the punishment stage, participants in the PD game will have to wait for a much longer time in the group treatments than the individual treatments. To minimize the differences among treatments, we decided to first collect decision data of third parties for all of the four treatments in our experiment.[10] After that, we ran the corresponding one-shot PD game with new subjects as Persons A and B. We randomly selected 16 third parties from each treatment and matched their decisions with the punishment requests by Persons A and B in the PD game. Each third party, if selected, is matched with one pair of Persons A and B only. When someone proposes to punish, the computer program searches for the corresponding decisions of the third party in the database and implements the decisions accordingly. All of the above information is common knowledge to all players (i.e., Persons A and B, and the third party).

To better understand decisions of Persons A and B, we also design a survey to elicit beliefs. In both treatments, Persons A and B are asked to guess each other's decisions after submitting their choices: "How likely (on a scale from 0 to 100) that your counterpart is going to choose to cooperate?" In the Ex-post treatments, we also asked Persons A and B to guess the amount of punishment their matched third party has assigned in each scenario after they have submitted their choices. One of the answers is randomly selected and subjects earn an extra $1 if their answer is correct.

We conducted the experiment at the Pittsburgh Experimental Economics Laboratory in April 2012. A total number of 370 subjects from various academic backgrounds participated in the study, among which 242 are in the role of third parties (Person C), and 128 are in the roles of Persons A and B. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). On average, an experiment session lasted about 50 minutes, including the

[10] One concern here is of course that participants may not believe that the PD game will be conducted later. However, the Pittsburgh Experimental Economics Laboratory where we conducted our experiment strictly prohibits deception. We also sent emails to third-party subjects if their decisions are chosen to be matched with players in the PD game. Thus, we think this concern is minor.

instructions and payment stage. Subjects earned an average of $10.20 in the role of Person C, and $11.62 in the role of Persons A and B in the prisoner's dilemma game, including the show-up payment.

3. Measures of deterrence vs. retribution

The comparison of third-party punishment in Ex-ante and Ex-post conditions allows us to learn the relative importance of retribution and deterrence motives. In this paper, we focus on the punishment decisions for defectors. Since punishment in our experiment does not affect the third party's payoff, we can assume that a third party makes punishment decisions to best perform the function of punishment that he values: retribution or deterrence or both. Some third parties are non-punishers meaning they do not care about either of the motives and thus do not approve any punishment in either Ex-ante or Ex-post treatments.

We can classify the punishers into three types based on the motives: retribution-driven punisher (R); deterrence-driven punisher (D); retribution and deterrence-driven-punisher (RD). We assume that the distribution of the three types of punishers is the same in both Ex-post and Ex-ante conditions.

R type punishes only for retribution. According to Carlsmith et al. (2002), the primary concept of the retribution concern is that "punishment should be proportionate to the harm" (p. 285). Suppose they believe the penalty amount P^*_r should be imposed on the defectors to restore justice. They will approve the punishment proposal in both treatments and the magnitude of punishment assigned will be the same in the Ex-ante as in the Ex-post treatment: $P_R^{Ex\text{-}ante} = P_R^{Ex\text{-}post} = P^*_r > 0$.

D type punishes only for deterrence. Suppose they believe to deter defection, the amount of punishment should be no less than $P^*_d > 0$. They will approve the punishment proposal and assign $P_D^{Ex\text{-}ante} \geq P^*_d$ in the Ex-ante treatment and reject the punishment proposal in Ex-post treatment. Thus $P_D^{Ex\text{-}ante} > P_D^{Ex\text{-}post} = 0$. Note that unlike retribution-driven punishment where the amount should exactly justify serving as retribution, the deterrence-driven punishment can be any amount higher than or equal to P^*_d.

RD type punishes for both retribution and deterrence. Suppose they believe a penalty of $P^*_r > 0$ should be imposed on the defectors to restore the justice and a penalty of $P^*_d > 0$ is sufficient to deter defection. The RD type will approve the punishment proposal in both Ex-ante

and Ex-post treatments. However, the amount they assign in each treatment may differ depending on the weight of retribution and deterrence motives and the relationship between P^*_r and P^*_d:

If $P^*_r \geq P^*_d$: The third party will assign P^*_r in both treatments: $P_{RD}^{Ex\text{-}ante} = P_{RD}^{Ex\text{-}post} = P^*_r$. The reason is that in the Ex-ante treatment, a punishment magnitude of P^*_d is not enough for retribution purpose, even though it is sufficient to deter defection. To assign P^*_r in the Ex-ante treatment can serve both deterrence and retribution purpose.

If $P^*_r < P^*_d$: Under the Ex-post treatment, the third party will assign $P_{RD}^{Ex\text{-}post} = P^*_r < P^*_d$ just for the retribution purpose. The interesting case is the Ex-ante treatment. On one hand, the third party thinks at least P^*_d should be imposed to deter defection. On the other hand, he believes P^*_d is too high to be justified from the perspective of retribution. The more a third party is motived by deterrence, the higher $P_{RD}^{Ex\text{-}ante}$ will be. Similarly, the more he values retribution, the lower $P_{RD}^{Ex\text{-}post}$ will be. Thus, $P^*_r = P_{RD}^{Ex\text{-}post} < P_{RD}^{Ex\text{-}ante}$ and the more a third party value deterrence, the greater the difference between $P_{RD}^{Ex\text{-}post}$ and $P_{RD}^{Ex\text{-}ante}$ is.

Taken all together, RD types will approve punishment in both treatments. The amount they assign in the Ex-ante treatment is at least as high as that in the Ex-post treatment. In particular, when a majority of RD types hold the belief of $P^*_r < P^*_d$ and value deterrence more than retribution, we will observe that the magnitude of approved punishment is significantly greater in Ex-ante treatment than Ex-post treatment.

Comparing the changes in punishment behavior between the Ex-ante and Ex-post treatments informs us the extent to which punishment decision is driven by deterrence and retribution motives. Assume that the population consists of a proportion of p^R of R type, p^D of D type and p^{RD} of RD type and the rest are non-punishers. The aggregated punishment approval rate in the Ex-post treatment $p^{Ex\text{-}post} = p^R + p^{RD}$ and in the Ex-ante treatment $p^{Ex\text{-}ante} = p^R + p^{RD} + p^D$. The average magnitude of *approved* punishment in the Ex-post treatment $P^{Ex\text{-}post} = (p^R P_R^{Ex\text{-}post} + p^{RD} P_{RD}^{Ex\text{-}post}) / (p^R + p^{RD})$ and in the Ex-ante treatment $P^{Ex\text{-}ante} = (p^R P_R^{Ex\text{-}ante} + p^{RD} P_{RD}^{Ex\text{-}ante} + p^D P_D^{Ex\text{-}ante}) / (p^R + p^{RD} + p^D)$. From the previous analysis, we know that 1) $P_{\mathrm{R}}^{\text{Ex-ante}} = P_{\mathrm{R}}^{\text{Ex-post}}$, and 2) $P_{\mathrm{RD}}^{\text{Ex-ante}} \geq P_{\mathrm{RD}}^{\text{Ex-post}}$.

Thus, we derive the following hypotheses (summarized in Table 3):

1) If we observe $p^{Ex\text{-}ante} = p^{Ex\text{-}post}$ and $P^{Ex\text{-}ante} = P^{Ex\text{-}post}$, then it implies that p^D is small and the third party population is dominated by either R type punishers or the RD type punishers who

hold the belief of $P^*_r \geq P^*_d$ or both. That is few third parties are utilitarian and care only about deterrence.

2) If we observe $p^{Ex\text{-}ante} = p^{Ex\text{-}post}$ but $P^{Ex\text{-}ante} > P^{Ex\text{-}post}$, then the third party population is dominated by RD type who holds the belief of $P^*_r < P^*_d$ and is relatively more motivated by deterrence than retribution.

3) If we observe $p^{Ex\text{-}ante} > p^{Ex\text{-}post}$, then a significant proportion of punishers should be D type motivated only by deterrence concern. In this case, the magnitude of *approved* punishment in the Ex-ante treatment can be greater or less or the same as in the Ex-post treatment. [11]

Table 3. Punishment behavior interpretation

Possible outcomes		Hypotheses
Freq.	Magn.	Majority third-parties are…
$p^{Ex\text{-}ante} = p^{Ex\text{-}post}$	$P^{Ex\text{-}ante} = P^{Ex\text{-}post}$	R type and/or RD type with $P^*_r \geq P^*_d$
$p^{Ex\text{-}ante} = p^{Ex\text{-}post}$	$P^{Ex\text{-}ante} > P^{Ex\text{-}post}$	RD type with $P^*_r < P^*_d$
$p^{Ex\text{-}ante} > p^{Ex\text{-}post}$	*any relationship*	D type

A point worth noting is that the three types of punishers could not be directly identified in the experiment due to our between-subject design. [12] Rather, this framework contributes to the understanding how our treatments help to learn the punishment motivation. The assumptions of this theory could be falsified if we observe other results than the three listed above.

In this study, we do not advance any hypothesis regarding the relative importance of retribution and deterrence motives in groups versus individual third parties. However, we could use the difference-in-differences approach to construct comparisons. For example, if we observe less punishment approved in the Ex-post treatment than in the Ex-ante treatment when the

[11] When the proportion of D type is much higher than the other two types, the average amount of approved punishment is close to $P^{Ex\text{-}ante}$. Thus, the difference in the punishment magnitude between the two treatments depends on the comparison between the amount of punishment D type would assign in the Ex-ante treatment and the average punishment amount assigned by the R and RD types in the Ex-post treatment.

[12] We employ a between-subject design instead of a within-subject design to avoid the discussion of a possible order effect. In fact, we suspect that it is highly likely that subjects anchor their decisions from the first treatment to the second one. This is in particularly true if they get tired or bored, and the extrinsic motivation of thinking hard is low in our experiment (i.e., their decisions do not affect their monetary payoffs in any case). Moreover, as it might take a lot of time for groups to reach consensus in labs, a within-subject design might raise time constraint problems.

decision makers are groups but not when they are individuals, we can draw the inference that deterrence motive plays a more important role for groups than for individuals.

4. Results

We use the strategy method to elicit punishment decisions in each of the seven scenarios: four scenarios in which one player proposes punishment (i.e., scenarios 1, 3, 5, 6 in Figure 1) and the remaining three scenarios (i.e., scenarios 2, 4 and 7 in Figure 1) in which both players propose punishment. In total there are ten decisions: five on defectors and the other five on cooperators. We find that a third party's decision whether or not to approve one person's punishment proposal toward a counterpart does not vary according to whether the counterpart has also proposed punishment. [13,14] Therefore, in the data we report below, we do not differentiate these two cases.

Although a third party has to make punishment decisions for defectors (*Pun_Def*) as well as cooperators (*Pun_Coop*), only the former situation is consistent with cooperative norms and third-party punishment in the real world. Not surprisingly, we find both the punishment approval rate and magnitude are significantly higher when punishment is imposed on defectors compared to cooperators (45.57% versus 32.13%; 7.01 versus 2.89, two-sided Wilcoxon sign-rank tests, $p < 0.01$)).[15] Given our interest in the role of retribution and deterrence in punishment decisions, our discussion below focuses on the punishment decisions toward defectors. It includes the cases when a defector proposes punishment toward a defector (*Def_Pun_Def*) and those when a cooperator proposes punishment toward a defector (*Coop_Pun_Def*). Punishment decisions towards cooperators are reported in Appendix A2.

4.1 Third-party punishment decisions

We first compare third-party decisions in different treatments. Then we turn to the cooperation outcomes for Persons A and B in the PD game. In Table 4, we calculate the average

[13] We treat each individual or group as an independent observation. In each treatment, we run Wilcoxon sign-rank tests on both punishment approval frequencies and magnitudes for a given third party, comparing his decision in the case in which only Person A or Person B proposes punishment, versus the case in which both persons A and B propose punishment. None of these tests are significant at the 10% level.

[14] A related study by Güth and Otsubo (2011) also find that a third party favors "equality before the law," meaning that a third party assigns equal punishment if both of them choose to defect, even though there exists a Pareto improvement equilibrium in which the third party punishes one of them.

[15] These numbers are aggregated over individuals and groups, pooling all punishment decisions together.

punishment frequency (i.e., approval rate) and magnitude for each individual or group third party in each category.

Table 4. Punishment frequencies and magnitudes across treatments (to defectors)

	Third-party punishment decisions					
	Exp		**Freq**		**Magn**	
	Ex-post	Ex-ante	Ex-post	Ex-ante	Ex-post	Ex-ante
a) **Individual third parties**						
			Overall summary			
Pun_Def	8.48	9.85	0.57	0.59	9.39	11.30
	(5.51)	(7.63)	(0.29)	(0.31)	(4.99)	(7.08)
	p = 0.66		*p = 0.44*		*p = 0.43*	
			Conditional on punishers			
Def_Pun_Def	3.33	4.60	0.44	0.46	4.92	7.51
	(5.08)	(6.55)	(0.37)	(0.42)	(5.53)	(6.69)
	p = 0.83		*p = 0.43*		*p = 0.21*	
Coop_Pun_Def	16.19	17.71	0.76	0.77	18.59	20.33
	(12.17)	(12.19)	(0.36)	(0.36)	(11.16)	(10.78)
	p = 0.61		*p = 0.44*		p =0.53	
b) Group third parties						
			Overall summary			
Pun_Def	3.45	6.23	0.24	0.43	8.12	10.04
	(5.04)	(6.02)	(0.32)	(0.40)	(5.14)	(4.38)
	p = 0.08		*p = 0.06*		*p = 0.06*	
			Conditional on punishers			
Def_Pun_Def	2.08	2.62	0.17	0.33	7.15	7.27
	(4.34)	(3.29)	(0.31)	(0.44)	(5.45)	(3.37)
	p = 0.53		*p = 0.07*		*p = 0.76*	
Coop_Pun_Def	5.50	12.59	0.34	0.57	15.52	20.28
	(9.33)	(13.80)	(0.47)	(0.48)	(9.52)	(12.19)
	p = 0.02		*p = 0.04*		*p = 0.32*	

Notes: Standard deviations are reported in parentheses. For "Freq", we count the number of times a third-party approves punishment proposal in the corresponding category, and then divide this number by the total number of decisions need to be made in the category. Specifically, the category "Pun_Def" includes five decisions made in scenarios 1, 2 (two decisions), 6 and 7 in Figure 1. The category "Def_pun_def" includes three decisions made in scenarios 1 and 2 (two decisions). The category "Coop_Pun_Def" includes two decisions in scenarios 6 and 7. For "Exp", we average the amount of punishment in all relevant decisions and treat disapproval as zeros. For "Magn",

we average the amount of punishment conditional on approval in the category. The p-values of the nonparametric test of the variable "Freq" are one-sided proportional tests due to the hypothesis that the punishment frequency in the Ex-ante treatment is never strictly smaller than that in the Ex-post treatment. The p-values of the nonparametric tests for the variables "Exp" and "Magn" are two-sided Mann-Whitney tests.

The column "Freq" reports the average approval rate of a certain proposal. The column "*Magn"* reports the average amount of punishment when a punishment request is approved. The column "*Exp"* reports the expected punishment, taking the average of the product of punishment frequency and magnitude for each third party. In other words, the column "*Exp*" treats the disapproval of a punishment proposal as zero in magnitude, while "*Magn*" excludes these proposals. We compare these variables between the Ex-post and the Ex-ante treatments using proportion tests for punishment frequencies ("Freq") and Mann-Whitney tests for magnitude ("Magn") and expected punishment ("Exp").

4.1.1 Individual third parties

We find no significant difference in the punishment behavior of individual third parties between Ex-post and Ex-ante treatments. This result indicates that retribution is a dominate motive of individual third party punishers.

As shown in Table 4, in the Ex-post treatment, almost 60% of third parties approve punishing defectors. The magnitude of punishment (9.39) is about 25% of the maximum earnings of a defector. Compared to the Ex-post treatment, however, neither the difference in the punishment frequency (57% versus 59%, $p = 0.44$) nor the magnitude (9.39 versus 11.30, $p = 0.43$) is significantly larger in the Ex-ante Treatment. This pattern, based on our discussion in Section 3, suggests that few individual third parties are D type, who are utilitarian and punish only for the purpose of deterrence; rather, retribution is an important motive in their punishment decisions.

We also find that the individual third parties condition their decisions on the actions of punishment proposers. Compared to punishment proposals by defectors (*Def_pun_Def*), proposals by cooperators (*Coop_pun_Def*) are approved at a higher frequency (76% versus 44% in the Ex-post treatment, $p < 0.01$; 77% versus 46% in the Ex-ante treatment, $p < 0.01$), and at a greater magnitude (18.59 versus 4.92 in the Ex-post treatment, $p < 0.01$; 20.33 versus 7.51 in the Ex-ante treatment, $p < 0.01$). This pattern also clearly indicates that individual third-parties do

not punish at random, but rather act consistently with retribution motive that punishment should be in proportion to the harm. [16]

4.1.2 Group third parties

Our data suggest unlike individual third parties, deterrence motive plays a significant role in group punishment decisions.

To see this, the data reported in the lower panel of Table 4 reports that groups are more likely to approve a request in the Ex-ante treatment than in the Ex-post treatment (43% versus 24%, $p = 0.06$). When conditioning punishment on the type of punishers, we find that the results are mostly driven by *Coop_pun_Def* type proposals. A significantly higher percentage of groups approve punishment requests to defectors by cooperators in the Ex-ante than the Ex-post treatment (57% versus 34%, $p = 0.04$). Table 4 also shows that magnitude of punishment is higher in the Ex-ante than the Ex-post treatment although the difference is only marginally significant (10.04 versus 8.12, $p = 0.06$). Based on our discussion in Section 3, we can infer that a significant amount of group third parties are D types who are utilitarian and punish only for the purpose of deterrence.

Comparing individual and group third parties, it is interesting to note that groups approve punishment proposals less frequently than individuals, especially in the Ex-post treatment (24% versus 57% in the Ex-post treatment, with $p < 0.01$ and 43% versus 59% in the Ex-ante treatment, $p < 0.1$). However, the punishment magnitude conditional on approval do not differ significantly (Ex-post: 8.12 versus 9.39, with $p = 0.67$, and Ex-ante: 10.04 versus 11.30, with $p = 0.98$). This also implies that significantly more group third-parties behave like D-type compared to individual third-parties.

[16] One could argue that third parties might also think of the deterrence motive even under the Ex-post condition. To examine this argument, we study the correlation between how they assign punishment and their reported predictions regarding the outcome of the PD game. In the experiment, third parties need to answer a question, "What is the most likely outcome of the PD game?" after they have made punishment decisions in all scenarios. The answers could be "both Persons A and B choose to cooperate," "only one chooses to cooperate," or "neither Persons A and B choose to cooperate - The labels "cooperation" and "defection" are only added later to facilitate the readers. If the argument holds, we should observe a positive correlation between their punishment assignment and beliefs. That is, the heavier the punishment towards defectors, for instance, the more likely they think Persons A and B will cooperate. We find no significant correlation between punishment decisions and beliefs only exist under the Ex-ante condition.

4.1.3 Regression analysis

To provide econometric evidence on the relative importance of the two motives in group third-party punishment decisions as compared with individual ones, we run the following hurdle regression. The hurdle model is a generalization of the Tobit model in which the decision to punish and the amount given are determined by two separate stochastic processes. Hence, the likelihood function for the hurdle model is the manipulation of the likelihood that a third party decides to punish (estimated from a standard Probit model) and the conditional likelihood of the punishment amount when punishment is approved (a truncated linear regression). The two parts of the hurdle model are estimated separately (McDowell, 2003). Again, we focus on norm-enforcing punishment, that is, punishment proposals on defectors.

Table 5. Average treatment effect analysis

Dependent variable: Punishment of defectors (Yes = 1 for Probability (Prob.))						
	Pun_Def		Def_Pun_Def		Coop_Pun_Def	
	(1)		(2)		(3)	
	Prob	Magn	Prob	Magn	Prob	Magn
β_1: Ex-post Individual	0.36***	0.91	0.30***	4.50	0.36***	5.10**
	(0.08)	(2.03)	(0.10)	(2.89)	(0.08)	(2.57)
β_2: Ex-ante Group	0.22***	3.61	0.20*	2.05	0.22**	5.88**
	(0.10)	(2.40)	(0.12)	(4.34)	(0.12)	(2.89)
β_3: Ex-ante Individual	0.38***	2.84	0.32***	2.10	0.38***	6.59***
	(0.08)	(2. 05)	(0.10)	(3.00)	(0.04)	(2.54)
β_4: Proposer cooperate	0.28***	12.64***	--	--	--	--
	(0.04)	(1.31)				
β_5: Other player propose	0.03	- 0.57	0.08	- 0.19	- 0.03	- 0.71
	(0.03)	(1.31)	(0.05)	(1.76)	(0.04)	(1.81)
β_0: Constant	--	7.60***	--	12.19***	--	16.61***
		(2.52)		(3.18)		(2.20)
R^2	0.11	0.29	0.05	0.05	0.11	0.38
# Obs.	610	278	366	129	244	149

Notes: Baseline is punishment behavior for group third parties in the Ex-post treatment. "Prob" reports the marginal effects from a Probit regression calculated at the mean; "Magn" is a truncated-linear regression; standard errors are in parentheses and are clustered at the individual level;*** significant at the 1% level, ** at 5% level; * at 10% level.

The hurdle model allows punishment to depend on whether the punishers are groups or individuals, whether third-party decisions are displayed ex-post or ex-ante, whether the punishment proposer is a defector or cooperator, and whether the other player also proposed punishment. We first run a specification by pooling all data together (Regression (1)), and then two specifications conditional on the punisher's action in the PD game (defector: Regression (2) or cooperator: Regression (3)).

Results in Table 5 report consistent findings with the nonparametric statistics. We first focus on approval probability. Individuals do not react to the timing of punishment decisions, since a two-sided chi-squared test could not reject the null hypothesis that the coefficients β_3 and β_1 are equal for all three specifications.[17] Groups, on the other hand, are significantly more likely to approve punishment in the Ex-ante treatment than in the Ex-post treatment (as indicated by the significantly positive coefficient β_2 in all three specifications). Moreover, the approval rate changes of group third parties between the Ex-ante and Ex-post treatments are marginally significantly larger than those of the individual third parties (β_2 versus $\beta_3 - \beta_1$, one-sided chi-squared test, $p = 0.07$), suggesting retribution plays a less important role for groups than for individuals. Combined with the above results, our data suggest that not only are individual third parties less likely to be driven by deterrence only (D type) than groups (β_1 is significant), but they are also more retributive than group third parties ($\beta_3 - \beta_1 > \beta_2$).

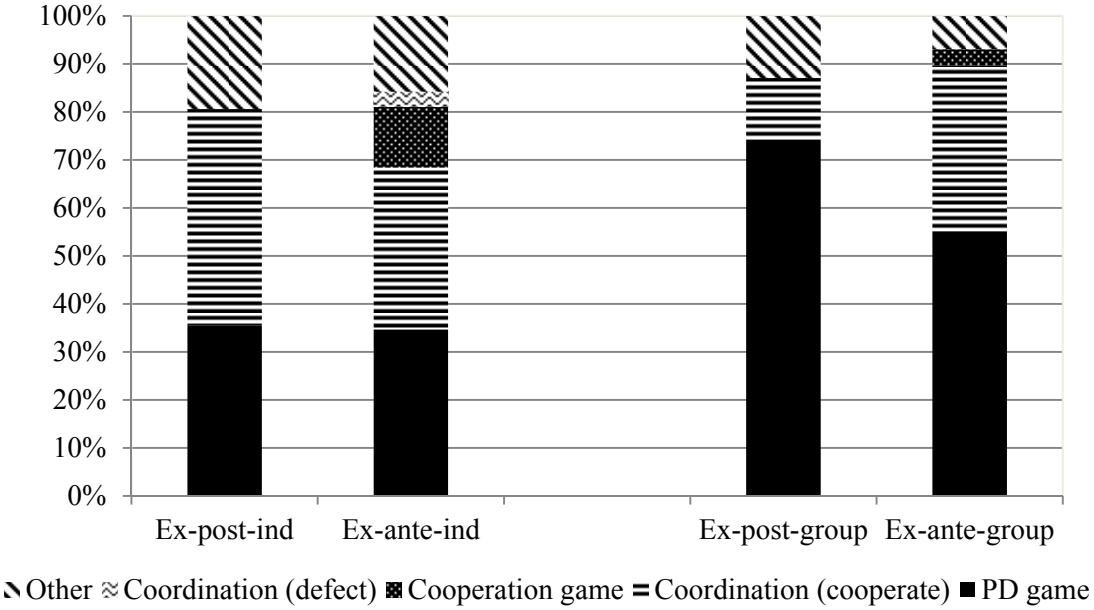
Second, the regression on punishment magnitude if the punishment proposal is approved shows similar treatment comparisons as the punishment frequency. For individual third parties, the difference in the punishment magnitude between the Ex-ante and the Ex-post treatments is not statistically significant in all three specifications, the null hypothesis $\beta_3 = \beta_1$ could not be rejected. For groups, we find the difference between Ex-post and Ex-ante conditions to be particularly strong when cooperators punish defectors (β_2 of column 3), but not so when defectors punish defectors (β_2 of column 2). In summary, based on our discussion in Section 3, the results from the regression analysis further support our main finding that a significant amount of third party punishment decisions by groups, rather than individuals, are driven by deterrence motives only.

[17] The p-values are 0.72, 0.77, and 0.83 in these three specifications.

4.2 The effects of third-party punishment

To evaluate the effectiveness of punishment strategies in each treatment, we recalculate the payoff matrix of the one-shot prisoner's dilemma game based on the third party's punishment decisions. Specifically, for each third party, we average out punishment assigned in each of the seven situations. We treat the amount of punishment as zero for the cases when punishment proposals are not approved. We assume that Persons A and B in the PD game have rational expectations on the third party's decisions, even in the Ex-post treatments when they do not know the third party's decision.

Figure 2: Nature of the game adjusted for punishment



Notes: Coordination (defect) means that a PD game is modified as a coordination game. (cooperate, cooperate) and (defect, defect) are the two pure-strategy equilibria, but (defect, defect) is payoff dominant. Coordination (cooperate) follows a similar definition, except that (cooperate, cooperate) is payoff dominant. Cooperation game means (cooperate, cooperate) is the unique Nash equilibrium. PD game means (defect, defect) is the unique Nash equilibrium.

Depending on the punishment decisions by a third party, the one-shot prisoner's dilemma game could become a coordination game (payoff-dominant strategy could either be cooperation or defection), or a cooperation game (the dominant strategy is to cooperate), or remain as a

prisoner's dilemma game (the dominant strategy is to defect). To change a PD game to a cooperation game, a third party needs to assign heavy punishment to defectors, regardless of whether the opponents are defectors. To change a PD game to a coordination game, by contrast, only requires defectors to be punished sufficiently heavily when they defect unilaterally. Section A.1 in the Appendix presents the analysis in detail. Figure 2 presents the types of games after the intervention of third parties in each treatment.

As shown in Figure 2, nearly half of the individual punishers turn a PD game into a coordination game in the Ex-post treatment. In the Ex-ante treatment, this percentage decreases although not significantly (45.16% versus 34.48%, a two-sided proportional test, $p = 0.42$). This implies that these individuals promote conditional cooperation. That is, they assign heavy punishment to players who defect unilaterally. When their counterparts also defect, on the other hand, punishment becomes not sufficiently deterrent. Even in the Ex-ante treatment, only approximately 12.90% (4 out of 31) of individual third parties punish defectors heavily regardless of the decisions of their counterparts, making cooperation a dominant strategy. Under both the Ex-post and Ex-ante conditions, 35.48% of the individual third parties punish too little such that a PD game remains as a PD game.

In contrast, the game distributions after the intervention of group third parties significantly differ compared with those of individuals, both in the Ex-post and the Ex-ante treatments ($p < 0.01$, two-sided Fisher exact tests). In the Ex-post treatment, most of the games remain as PD games, which is also significantly higher than the individuals (74.19% versus 35.48%, $p < 0.01$, a two-sided proportional test). Compared with the Ex-post treatment, significantly fewer groups in the Ex-ante treatment keep a PD game as it is (55.17% versus 74.19%, $p = 0.09$, a two-sided proportional test). Instead, more groups change the PD game into a coordination game (12.90% versus 34.48%, a two-sided proportional test, $p = 0.05$).

The above results are consistent with our main finding that compared with individuals, groups are more likely to be utilitarian and driven only by deterrence motive. Nevertheless, neither groups nor individuals punish defectors hard enough to ensure that cooperation will be a dominant strategy. The cooperation rates are exactly identical when the third party is a group (25% for both conditions). In the individual treatments, cooperation under both conditions slightly increases to above 30% (31% and 32%). However, the differences in the cooperation rate between group and individual third party treatments are not statistically significant at any level.

We do similar analysis on final earnings after punishment. The average earnings are slightly higher in the presence of group third parties than individual third parties (24.06 versus 22.97 in the Ex-post treatments, and 22.82 versus 21.31 in the Ex-ante treatments). However, none of the differences are statistically significant. As we discuss next, one explanation is that the level of punishment is insufficient to render defection unprofitable.[18]

5. Discussion

Our main finding is that group third parties assign punishment based on whether punishment decisions are announced before agents make their decisions, while individual third parties' decisions do not vary on the timing of the decision. This implies that few individuals are driven only by deterrence motive, while a significant proportion of groups perceive punishment only as an instrument to elicit cooperation, and suppress its use when this function is absent.

Previous research on group decision making suggests that the way group members make collective decisions depends on what they consider to be socially desirable and justifiable in this situation. In our experiment, socially desirable punishment behavior may change based on whether punishment decisions are displayed ex-post or ex-ante, which leads to distinctive patterns in the group decision process. According to an established theory in social psychology called the social comparison theory (SCT) (see, e.g., Burnstein et al. 1973), people are motivated both to perceive others and to present themselves in a socially desirable way. Hence, during group deliberation on whether and how much to punish, they could adjust their initial proposal to be more in line with what they think is socially desirable. For instance, it is possible that some retribution-driven punishers think others in their group are deterrence-driven, and thus adjust their punishment proposals in order to be perceived as socially desirable. [19] In view of this, we examine how groups reach decisions through their electronic chats in Table 6. We find

[18] One could argue that it might be too complicated for the subjects to calculate the ex-post earnings in each cell given such a short period of time, even though all decisions of third parties are displayed ex-ante. We fully acknowledge that point. On the other hand, however, this is also an indication that punishment to defectors is on average not strong enough that subjects could quickly identify cooperation as the only profitable strategy.

[19] A companion theory is called Persuasive Argument Theory (PAT), which emphasizes the importance of communication and information influence, so that group members' opinions are influenced by the persuasiveness of an argument (see, e.g., Myers et al, 1980). In the Ex-ante treatment, group members are more likely to increase punishment if deterrence-driven punishers convince others that punishing defectors heavily is a legitimate and convincing argument. Unfortunately, our current design does not allow us to directly test and compare SCT and PAT.

supporting evidence that the social comparison theory even seems to hold when subjects' identities are not publically reveal. Subjects' chat record suggests that punishment is perceived to be less socially desirable in the Ex-post treatment, making it harder for a group to reach this consensus.

Table 6. Chat analysis for group third parties' punishment decisions toward defectors

	Def_Pun_Def		Coop_Pun_Def	
Section a. Summary statistics				
Treatment	Ex-post	Ex-ante	Ex-post	Ex-ante
(No. obs)	(62)	(58)	(62)	(58)
Group chat freq.	51.61%	67.24%	30.65%	36.21%
	(32)	(39)	(19)	(21)
Section b. For the groups that do not chat				
Approval freq.	10%	21.05%	20.93%	37.84%
	(3/30)	(4/19)	(9/43)	(14/37)
Punishment magn.	2.4	3.5	16.00	15.71
Section c. For the groups that chat				
Approval freq.	25%	35.90%	63.16%	90.48%
	(8/32)	(14/39)	(12/19)	(19/21)
Start with disapproval end up with disapproval	100% (17/17)	100% (17/17)	100% (3/3)	66.67% (2/3)
Start with approval end up with approval	70% (7/10)	80% (12/15)	85.71% (12/14)	100% (14/14)
Initial proposal magn.	6.43	10	17.5	26.67
Final punishment magn.	9.13	9.43	16.42	26.84

Notes: Group members could coordinate their decisions in two ways - either by sending chat messages to each other, or simply typing the proposals in a decision display panel visible to all other members in a group. Section a of Table 6 reports the total number of relevant cases in each treatment ("No. obs") and frequency that groups send chat messages explicitly ("Group chat freq."). Section b records the approval frequencies ("Approval freq."), for each treatment and each situation in which groups need to decide punishment on defectors. Section b also records punishment magnitude for groups did not leave any chat messages. Section c documents statistics for groups which chat. We further record the first magnitude proposal via chat within a group ("Initial proposal magn."), and the final amount submitted as a group's collective decision ("Final punishment magn."). Again we focus on the punishment decisions toward defectors.

The first evidence is the proportion of groups reaching tacit agreement to punish without chats. This is true especially for the case when a cooperator proposes punishing a defector: When groups do not chat, only one-fifth of them approve punishment in the Ex-post treatment, but this percentage almost doubles in the Ex-ante treatment (20.93% versus 37.84%, a two-sided proportional test, $p = 0.09$).[20] In other words, we observe more need for discussion to agree on punishing the defectors in Ex-post treatment. For the groups that chat, the percentage agreeing to punish is significantly lower under the Ex-post condition (63.16% versus 90.48%, a two-sided proportional test, $p = 0.04$).[21]

Moreover, not only are groups less likely to agree upon punishment absent of discussion, but the initial punishment amount proposed in a group is also smaller in the Ex-post than the Ex-ante treatment (17.5 vs. 26.67, a two-sided Mann Whitney test, $p= 0.03$). The difference in the initial punishment amount also contributes to the difference in the final assigned punishment amount between the two treatments, because roughly 90% of the time the initial proposal is the same as the final approved amount in both treatments. To a less extent, the differences in chats also apply to the cases where defectors propose to punish defectors.

In summary, both the percentage of tacit agreement and higher initial punishment proposals in group chat imply a potential explanation why punishment is higher under ex-ante condition: it is more socially desirable for individuals in groups to propose and agree on larger punishment proposals.[22]

6. Concluding remarks

In this paper, we aim at identifying the motivations (retribution versus deterrence) for punishment decisions by an impartial third party. We find that, in a one-shot prisoner's dilemma game, little individual punishment is driven only by deterrence concerns. This result complements previous studies that show individual third-party punishers are often not

[20] When defectors propose to punish each other, the approval rate is also higher in the Ex-ante treatment, although not significantly.

[21] One plausible explanation could be that the groups that would like to punish heavily are more likely to chat.

[22] Another potential explanation could be that in the Ex-ante treatments, third-parties need to assign heavier punishment since no players will actually propose to punish otherwise. Punishment request data, however, do not support this explanation. Although group third parties assigned heavier punishment in the Ex-ante treatment, only 2 out of 32 players in the PD game requested punishment, compared to 6 players in the Ex-post treatment. When matched with individual third party, more players in the PD game requested punishment in the Ex-ante treatment (22 versus 4), even though the punishment decisions for individual third parties do not vary significantly.

instrumental in settings such as an ultimatum game, a public goods game, or a taking game (Bosch-Rosa, 2012; Casari and Luini, 2012; Carpenter et al.2012; Ouss and Peysakhovich, 2012).

In contrast, we find that groups are dominated by deterrence-driven type punishers. A plausible reason why groups are more reactive to the deterrence concern is that punishment is perceived to be much less socially desirable in the Ex-post treatment. Anticipating this, subjects might be less likely to propose punishment to their group members, even though they would do otherwise if they were given the opportunity to decide *alone*. Furthermore, even though third parties do not bear the cost of punishment, they reject a significant amount of punishment proposals, and the rejection rate is particularly high when the third party is a group. As a result, the amount of punishment is not large enough to deter defectors, especially when the punishment proposer is also a defector.

Our design of the experiment also allows us to learn how the third party's decision may vary according to the behavior of the punishment proposer in addition to the behavior of the punishee. We show the third parties believe a defector deserves much less punishment if the punishment proposer is also a defector. This result is consistent with the findings in Cubitt et al. (2011) that subjects view free riding in a public goods game as less morally wrong if the partner also free rides.

Our experimental design provides future researchers with a framework to study the influences of the effective third-party punishment. The results call for the need to study groups as decision making units, since they offer yet another piece of evidence that groups could punish very differently than individuals, due to the aggregation rule and exposure to peers' opinions. Last but not least, our results provide a first step to potentially inform public policy whether groups appear to be more lenient but utilitarian than individuals in ethical and legal decision making.

References:

Anderson, C. 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*, 51(1), 1–24.

Andreoni, J. and L. Gee (2012), Delegated enforcement and peer punishment in public goods provision, *Journal of Public Economics*, 96, 1036-1046.

Almenberg, J., A. Dreber, C. Apicella, and D. Rand (2011), Third party reward and punishment: Group size, efficiency and public goods. In Psychology of Punishment, Nova Science Publishers. Eds. NM Palmetti et al. ISBN: 978-1-61324-115-8.

Auerswald, H., C. Schmidt, M. Thum and G. Torsvik, Teams punish less, 2013, CESifo working paper series, No.4406.

Babcock, L. and G. Loewenstein (1997), Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, 11(1), 109-126.

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

Brandts, J. and G. Charness, (2011), The strategy versus the direct-response method: a first survey of experimental comparisons, *Experimental Economics*, 14(3), 375-398.

Bornstein, G. and I. Yaniv (1998), Individual and group behavior in the ultimatum game: Are groups more "rational" players?, *Experimental Economics* 1, 101-108.

Burnstein, E., A. Vinokur and Y. Trope (1973), Interpersonal comparison versus persuasive argumentation: A more direct test of alternative explanations for group-induced shifts in individual choice, *Journal of Experimental Social Psychology*, 9(3), 236-245.

Bosch-Rosa, C. (2012), A tale of two tails: Rejection patterns of extreme offers in three-player games, SSRN working paper.

Carlsmith, K., J. Darley and P. Robinson (2002), Why do we punish? Deterrence and just deserts as motives for punishment, *Journal of Personality and Social Psychology*, 83 (2): 284 – 299.

Carpenter, J. (2007), The demand for punishment. *Journal of Economic Behavior and Organization*, 62(4), 522–542.

Carpenter, J. and P. Matthews (2009), What norms trigger punishment? *Experimental Economics*, 12(3): 272-288 (2009).

Carpenter, J. and P. Matthews (2010), Norm enforcement: The role of third parties, *Journal of Institutional and Theoretical Economics*, 166(2), 239-258.

Carpenter, J. and P. Matthews (2012), Norm enforcement: Anger, indignation or reciprocity? *Journal of the European Economic Association*, 10(3): 555–572.

Casari, M. and L. Luini (2012), Peer punishment in teams: expressive or instrumental choice? *Experimental Economics*, 15(2): 241-259.

Charness, G. and M. Sutter (2012), Groups make better self-interested decisions, *Journal of Economic Perspectives*, 26(3), 157-176.

Chavez, A. and C. Bicchieri (2013), Third-party sanctioning and compensation behavior: Findings from the ultimatum game, *Journal of Economic Psychology*, 39, 268-277.

Chaudhuri, A., (2011), Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14, 47-83.

Cubitt, R. P., M. Drouvelis, S. Gächter, and R. Kabalin (2011), Moral judgments in social dilemmas: How bad is free riding? *Journal of Public Economics*, 95, 253-264.

Dawes, C. T., J. H., Fowler, T. Johnson, R. McElreath, and O. Smirnov (2007), Egalitarian motives in humans. *Nature* 446, 794–796.

Denant-Boemont L., D. Masclet and C. Noussair (2007), Punishment, counter-punishment and sanction enforcement in a social dilemma experiment, *Economic Theory* 33, 145-167.

Devlin-Foltz, Z. and K. Lim (2008), Responsibility to punish: Discouraging free-riders in public goods games. *Atlantic Economic Journal*, 36(4), 505-518.

Durkheim, E. (1973), On morality and society, Chicago: University of Chicago Press.

Engel, C. (2010), The behavior of corporate actors: How much can we learn from the experimental literature? *Journal of Institutional Economics*, 6: 445-475.

Ertan, A., T. Page and L. Putterman (2009), Who to punish? Individual decisions and majority rule in mitigating the free rider problem, *European Economic Review*, 53, 495–511.

Fehr, E. and U. Fischbacher (2004), Third-party punishment and social norms, *Evolution and Human Behavior* 25(2), 63-87.

Fehr, E. and S. Gächter (2000), Cooperation and punishment in public goods experiments, *American Economic Review* 90, 980-994.

Fehr, E., and S. Gächter (2002), Altruistic punishment in humans, *Nature* 415, 137-140.

Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics*, 10(2), 171-178.

Guillen, P, C. Schwieren, and G.Staffiero (2007), Why feed the Leviathan?, *Public Choice*, 130, 115–128.

Gächter, S. and B. Herrmann (2009), Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment, *The Royal Society B (Biological Science)* 364, 791-806.

Güth, W. (1995), On ultimatum bargaining experiments - A personal review, *Journal of Economic Behavior & Organization*, 27(3), 329-344.

Güth, W and H. Otsubo (2011), Whom to blame? An experiment of collective harming and punishing, Jena economic research papers, 2011-046.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C.Barett, A.Bolyanatz, J-C., Cardenas, M. Gurven, E.Gwako, N.Henrich, C.Lesorogol,F. Marlowe, D.Tracer and J.Ziker (2006), Costly punishment across human societies, *Science* 312 (5781), 1767-1770.

Herrmann, B., C. Thöni, and S. Gächter (2008), Antisocial punishment across societies. *Science*, 319(5868), 1362-1367.

Hopfensitz, A. and E. Reuben (2009). The Importance of emotions for the effectiveness of social punishment. *Economic Journal* 119, 1534-1559.

Houser, D. and E. Xiao (2010), Inequality seeking punishment, *Economics Letters*, 109(1), 20-23.

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553), 2635-2650.

Keck, S (2013), Group polarization in reactions to dishonesty, mimeo.

Kosfeld, M., A. Okada and A. Riedl (2009), Institution formation in public goods games, *American Economic Review* 99, 1335-1355.

Kriss, P., R. Weber and E. Xiao (2013), Turning a blind eye: On the robustness of costly third party punishment in humans, working paper.

Kugler, T, E. Kausel, and M. Kocher (2012), Are groups more rational than individuals? A review of interactive decision making in groups, WIREs Cognitive Science, 3, 471–482.

Kurzban, R., P. DeScioli, and E. O'Brien (2007), Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28(2), 75-84.

Kurzban, R. and P. DeScioli (in press). Adaptationist punishment in humans. *Journal of Bioeconomics*, 15, 269-279.

Leibbrandt, A. and R. López-Pérez (2011), The dark side of altruistic third-party punishment. *Journal of Conflict Resolution*, 55(5), 761-784.

Leibbrandt, Andreas & López-Pérez, Raúl (2012), An exploration of third and second party punishment in ten simple games, *Journal of Economic Behavior & Organization*, 84(3), 753-766.

Lönnqvist, J., J. Sprenger, M. Verkasalo, G. Walkowitz, P. Wichardt, (2012). Judgment and behaviour in the prisoner's dilemma: The impact of moral and strategic considerations, SSRN working paper.

MacCoun, R. and N. Kerr (1988), Asymmetric influence in mock jury deliberation: jurors' bias for leniency. *Journal of Personality and Social Psychology*, 54(1), 21.

Marlowe, F. W., J. C. Berbesque, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, J. Ensminger, M. Gurven, E. Gwako, J. Henrich, N. Henrich, C. Lesorogol, R. McElreath, D.Tracer (2008), More 'altruistic'punishment in larger societies, *Proceedings of the Royal Society B: Biological Sciences,* 275(1634): 587-592.

McDowell, A. (2003), From the help desk: Hurdle models, *The Stata Journal*, 3(2): 178–184.

Myers, D.G., J.B.Bruggink, R.C Kersting and B.A. Schlosser (1980), Does learning others' opinion change one's opinion? *Personality and Social Psychology Bulletin*, 6, 253-260.

Nikiforakis, N. (2008), Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* 92, 91–112.

Nikiforakis, N., D. Engelmann (2011), Altruistic punishment and the threat of feuds, *Journal of Economic Behavior and Organization*, 78(3), 319-332.

Nikiforakis, N., C. Noussair and T. Wilkening (2012), Normative conflict and feuds: The limits of self-enforcement, *Journal of Public Economics*, 96(9-10), 797-807.

Noussair, C. and F. Tan (2011), Voting on punishment systems within a heterogeneous group, *Journal of Public Economic Theory*, 13(5), 661-693.

O'Gorman, R., J. Henrich, and M. Van Vugt (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 276(1655), 323-329.

Ottone, S., F. Ponzano, F and L. Zarri, (2008). Moral sentiments and material interests behind altruistic third-party punishment. Universita degli Studi di Verona, Working Paper.

Ouss, A. and A.Peysakhovich (2012). When punishment does not pay: "Cold glow and decisions to punish", working paper.

Pedersen, E., R. Kurzban, and M. McCullough (2012), Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society–B*, 280(1758), 1-8.

Riedl, K., K. Jensen, J. Call, and M. Tomasello (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences*, 109(37), 14824-14829.

Robert, C. and P. Carnevale (1997), Group choice in ultimatum bargaining, *Organizational Behavior and Human Decision Processes* 72 (2), 256-279.

Tan, F. and E. Xiao (2012), Peer punishment with third - party approval in a social dilemma game, *Economics Letters*, 117, 589-591.

Tyler, T. (2006), Why People Obey the Law, Princeton University Press, Princeton (NJ).

Tyran, J and L. Feld (2006), Achieving compliance when legal sanctions are non-deterrent, *Scandinavian Journal of Economics*, 108(1), 135-156.

Walker,J. and M. Halloran (2004), Rewards and sanctions and the provision of public goods in one-shot settings, *Experimental Economics*, 7(3), 235-247.

Woods, A. (2006), Moral judgments and international crimes: The disutility of desert. *Virginia Journal of International Law*, 52(3), pp. 633.

Xiao, E. (2013), Profit seeking punishment corrupts norm obedience, *Games and Economic Behavior*, 77: 321-344.

Xiao, E. and D. Houser (2005), Emotion expression in human punishment behavior, *Proceedings of the National Academy of Sciences*, 102(20), 7398-7401.

Xiao, E. and D. Houser (2011), Punish in Public, *Journal of Public Economics*, 95, 1006–1017.

Xiao, E. and H. Kunreuther (forthcoming), Punishment and cooperation in stochastic prisoner's dilemma game, *Journal of Conflict Resolution*.

Appendix

(Supplemental material, for online publication only)

A.1. Discussion on the modified payoff matrix

Each third party specifies his/her punishment decisions in all of the possible scenarios. Let parameter *a* represents the average punishment of cooperators towards defectors, parameter *b* for the average punishment of cooperators towards other cooperators, parameter *c* for the average punishment of defectors towards other defectors, and parameter *d* for the average punishment of defectors towards cooperators.

The cost of punishment is tricky to model as it is discontinuous, depending on whether a third party approves a punishment proposal. However, results of our experiment indicate that the punishment decisions of third parties do not contingent on the number of punishment proposers. This implies that third parties might not take into account the cost of punishers (5 tokens) when making punishment decisions. Hence, for simplicity, we exclude the cost paid by the punisher conditional on approval of punishment.

		Person B	
		Cooperate	Defect
Person A	Cooperate	30-b / 30-b	15-d / 40-a
	Defect	40-a / 15-d	20-c / 20-c

The following conditions need to be satisfied to ensure that "Defect" is a dominant strategy:

1) $30 - b < 40 - a \rightarrow a - b < 10$
2) $15 - d < 20 - c \rightarrow c - d < 5$

The intuition behind is that the relative punishment of defectors should not be too large relative to that of the cooperators. The following conditions need to be satisfied to ensure that cooperation is a dominant strategy:

3) $30 - b > 40 - a \rightarrow a - b > 10$
4) $15 - d > 20 - c \rightarrow c - d > 5$

The rationale is just the opposite of the previous case. The relative punishment of defectors (compared to the cooperators) should be sufficiently large, no matter whether the other party chooses defect or cooperate.

The conditions under which a game becomes coordination game:

5) $30 - b > 40 - a \rightarrow a - b > 10$
6) $20 - c > 15 - d \rightarrow c - d < 5$
7) $30 - b > 20 - c \rightarrow b - c < 10$ (Cooperation is payoff dominant);
8) $30 - b < 20 - c \rightarrow b - c > 10$ (Defection is payoff dominant);

If the relative punishment of defectors is only sufficiently larger when the punisher cooperates, then the nature of the game turns into a coordination game. If the relative punishment assigned to mutual cooperators is lower than the punishment assigned to mutual defectors by at least 10 points, the payoff dominant strategy is to cooperate. Otherwise the payoff dominant strategy changes to defect.

A.2. Tables and analysis

Table A2.1 Punishment frequencies and magnitudes across treatments (for cooperators)

	Third-party punishment decisions					
	Exp.		**Freq.**		**Magn.**	
	Ex-post	Ex-ante	Ex-post	Ex-ante	Ex-post	Ex-ante
a) Individual third parties						
			Overall summary			
Pun_Coop	3.71	4.57	0.43	0.41	4.79	6.75
	(4.57)	(6.05)	(0.30)	(0.36)	(4.67)	(6.28)
	p = 0.96		*p = 0.54*		*p = 0.33*	
			Conditional on punishers			
	4.02	5.09	0.45	0.46	3.67	5.88
Coop_Pun_Coop	(4.76)	(7.32)	(0.33)	(0.40)	(4.75)	(7.84)
	p = 0.85		*p = 0.47*		*p = 0.66*	
	3.24	3.81	0.39	0.34	5.91	7.87
Def_Pun_Coop	(6.38)	(7.63)	(0.40)	(0.40)	(7.72)	(9.51)
	p = 0.82		*p = 0.65*		*p =0.56*	
b) Group third parties						
			Overall summary			
Pun_Coop	2.02	1.14	0.23	0.21	4.82	3.00
	(3.55)	(1.86)	(0.32)	(0.32)	(2.85)	(1.87)
	p = 0.66		*p = 0.55*		*p = 0.64*	
			Conditional on punishers			
	2.44	1.51	0.27	0.26	4.31	4.37
Coop_Pun_Coop	(4.83)	(2.57)	(0.38)	(0.40)	(6.08)	(2.58)
	p = 0.75		*p = 0.52*		*p = 0.97*	
	1.39	0.59	0.16	0.14	6.14	2.83
Def_Pun_Coop	(4.01)	(1.45)	(0.33)	(0.30)	(6.80)	(2.04)
	p = 0.79		*p = 0.60*		*p = 0.52*	

Notes: Standard errors are reported in parentheses. Although punishment magnitude to cooperators is significantly lower compared to punishment to defectors (7.01 versus 2.89, $p < 0.01$), punishment approval rate (also referred to as the "anti-social punishment") is relatively high (about 40% for individuals and 20% for groups) compared to some previous studies (e.g., Herrmann et al., 2008). A potential reason could that in our experiment punishment is costless for a third party, and that a third-party might feel obliged to approve a request by a second party. More importantly, approximately 70% of punishment magnitude is not larger than the punishment cost (i.e. 5 points). Thus, such decisions could be interpreted as punishment for proposing to punish cooperators. The p-value of the non-parametric test of the variable "Freq" are one-sided proportional test due to the hypothesis that the punishment frequency in the Ex-ante treatment is never strictly smaller than that in the Ex-post treatment. The *p*-value of the non-parametric tests for the variables "Exp" and "Magn" are two-sided Mann-Whitney ranksum tests. * means 10% of significance, ** means 5% of significance and *** means 1% of significance level.

Table A2.2 Probit regression analysis of cooperation decisions of Persons A and B in the PD

Dependent variable: Cooperation (= 1)		
	Ex-post	**Ex-ante**
Beliefs on the other's cooperation propensity	0.019***	0.026***
(100 for cooperation for sure)	(0.007)	(0.009)
Third-party type (1 for groups)	-0.099	-0.541
	(0.368)	(0.406)
Guess third-party punishment on defectors	0.038	--
	(0.028)	
Guess third-party punishment on cooperators	-0.021	--
	(0.338)	
Actual average punishment on cooperators	--	- 0.088*
		(0.054)
Actual average punishment on defectors	--	0.013
		(0.033)
Constant	-1.534***	-1.071***
	(0.467)	(0.054)
Log likelihood	-35.43	- 30.33
Pseudo R^2	0.105	0.221
# Obs.	64	64

Notes: *** significant at the 1% level, ** at 5% level; * at 1% level. The results are robust after controlling for social demographic information (gender, age, nationality, major, political attitude and the Mach IV-scale).

Table A2.2 reports a Probit regression on the cooperation decision of a player's in the prisoner's dilemma. The dependent variable equals to 1 if a player chooses to cooperate and 0 otherwise. We run separate regressions for the Ex-post and Ex-ante treatments. The first explanatory variable measures the subjective belief of a player on the cooperation propensity of his/her counterpart (scale from 0 to 100). The second variable "third-party type" is a dummy equal to 1 if the third party is a group. The remaining variables represent players' subjective beliefs on third-party punishment. Specifically, in the Ex-post treatments, after A and B have made their decisions in the PD game, we ask them to guess the amount their matched third party assigns in each of the seven scenarios (from 0 to 40). Then we take averages of all punishment assigned to defectors and cooperators respectively.

Note that we don't have such data in the Ex-ante treatments, since persons A and B are presented with all decisions of a third-party before they choose their decisions to cooperate. Hence, we replace the belief data with the actual average amount of punishment third parties assign to cooperators and defectors.

In both treatments, the variable having the strongest correlation is the belief of the action one player holds towards his counterpart. The more he believes that the other player will cooperate,

the more likely he will choose to cooperate. The dummy variable representing the group treatment is also not statistically significant, suggesting that, after controlling the (expected) punishment amount, whether the third party is an individual or a group does not affect players' decisions.[23] This is consistent with a finding by Lönnqvist et al. (2012) that regardless of considering mutual cooperation as morally the first-best scenario, their actual cooperation choices in the lab is primarily determined by the subjects' (pessimistic) first order beliefs.

Interestingly, subjective beliefs on third-party punishment do not play a significant role in affecting cooperation decisions in the Ex-post treatments. This result even carries over to the Ex-ante treatments: punishment towards defectors does not affect the propensity to defect significantly, although punishment on cooperators slightly discourages incentives for players to cooperate. This result suggests that cooperation decisions in the one-shot PD game are mainly affected by players' beliefs rather than their perceptions of the third parties. Since beliefs are statistically the same across all treatments (about 40%), the observed cooperation rates turn out to be highly similar.

[23] Interestingly, when they asked to guess the amount of punishment assigned to a defector if sued by a cooperative counterpart, subjects matched with group third parties think that groups will punish harder than individuals (Mann-Whitney ranksum test, $p < 0.05$).

A.3. Experiment Instructions (Group Ex-ante treatment)

Thank you for coming to the experiment. You will receive 5 dollars for showing up on time. Please read these instructions carefully. If you have a question, please raise your hand, and an experimenter will assist you.

Your task:

At the beginning of the experiment, the computer will randomly match you with two other participants and the three of you will act as a group throughout the experiment. No one will know the identity of his/her group members. In today's experiment, everyone is in the role of Person C and will be given a list of scenarios that will happen in a future experiment (Experiment F) attended by others (NOT anyone in today's experiment). Person C's task in today's experiment is to decide with his or her group members, in each scenario, whether or not to approve any payoff deduction that might be proposed by a participant for his/her counterpart in Experiment F (one of them is in the role of Person A and another is in the role of Person B). If your group approves a payoff deduction proposal, your group will also need to specify the payoff deduction amount that you think should be implemented.

As a group, you will make decisions jointly. That is, the three of you must discuss together and reach an agreement about whether or not to approve payoff deduction proposals and if so, how much to deduct. To facilitate discussion, you could send messages back and forth to each other. Only members in your group could see these messages.

For each pair of Person A and Person B in Experiment F, the computer will randomly select, from today's experiment, one Person C group to match with the pair. The computer will implement this group's decision in the corresponding scenario. Each Person C group in today's experiment is equally likely to be chosen and each Person C group can be matched with no more than one pair.

After Experiment F has been conducted, the experimenter will email each of you to let you know if your group has been randomly chosen to be matched with a pair of Person A and Person B in Experiment F. If you are chosen, you will also see in the email your matched Person A's and Person B's decisions in Experiment F.

Note that Person A and Person B in Experiment F will first see the matched Person C group's decisions in today's experiment and then make their decisions. Also note that no one else, except the experimenter, will know the identity of any decision makers. Nor will the identity information be published or be released to other parties.

Your earnings:

Your earnings are determined by a random process. The computer will randomly assign 15, 20, 30 or 40 tokens, with equal chance. The exchange rate of tokens to dollar is: **5 tokens = 1 dollar.** You will only be informed of your earnings at the end of the experiment. Your decision in this experiment will not affect your own earnings but may affect the earnings of the participants in Experiment F. Note that every member in a group will receive the same earnings randomly assigned by the computer.

To make your decision, you need to first understand Experiment F. A copy of the instructions of Experiment F is attached in the next few pages. Please read them carefully. Again, as you will read from Experiment F instructions, you are referred to as Person C in Experiment F. After you finish reading, you will do some exercises to make sure that you understand the instructions. Then, you and the other two members in your group will be presented the list of scenarios and will be asked whether or not to approve a payoff deduction proposal in each scenario.

A Copy of Experiment F Instructions

(instructions for the implicated stakeholders in PD game)

Thank you for coming to the experiment. You will receive 5 dollars for showing up on time.

Please read these instructions carefully! Talking is not allowed at any time during this experiment. If you have a question, please raise your hand, and an experimenter will assist you.

At the beginning of the experiment, the computer will randomly assign each participant to the role of either Person A or Person B. The computer will also randomly pair a Person A with a Person B at the beginning of the experiment. Person A and Person B interact with each other in a decision task *only once*.

In an earlier experiment, we also recruited some other participants to act in the role of Person C. Person Cs randomly formed a group of three. Each Person C group has made his/her decision on whether or not to approve Person A and/or B's proposal in Stage 2 (see below for details). For each pair of Person A and Person B in today's experiment, the computer will randomly select one Person C group to match with the pair and implement this Person C group's decision to the pair. Each Person C group is equally likely to be chosen and can be matched with no more than one pair.

Person A's and Person B's decision task consists of two stages as described below.

Stage 1:

At the beginning of Stage 1, both Person A and Person B will see all the decisions made by the paired Person C group in each scenario.

Then, Person A and Person B will simultaneously and individually decide to choose either "Option I" or "Option II". The payoffs of Persons A and B are determined as follows: (a) if both Person A and Person B select Option I, each earns 30 tokens; (b) if both Person A and Person B select Option II, each earns 20 tokens; and (c) if one selects Option I and the other selects Option II, the one who selects Option I earns 15 tokens and the one who selects Option II earns 40 tokens. The payoff table below lists all of the possible payoff outcomes for each possible scenario. The number on the left in each cell is Person A's payoff and the number on the right is Person B's payoff. Note that Person A and Person B only make this decision once.

Payoff Table

		Person B	
		Option I	Option II
Person A	Option I	30, 30	15, 40
	Option II	40, 15	20, 20

The exchange rate of tokens to dollar is:

5 tokens = 1 dollar

Stage 2

At the beginning of the second stage, Person A and Person B are informed of the other's decision. Then, each will have an opportunity to propose whether or not to impose whether to pay 5 tokens to impose a payoff deduction for the other.

After Person A and Person B have decided whether or not to propose a payoff deduction for their counterparts, the computer program will search for the matched Person C group's decision under the corresponding scenario to decide whether or not to implement any proposed payoff deduction and, if so, what the amount would be. If neither Person A nor Person B proposed a payoff deduction, Person C's decisions will not affect either Person A's or Person B's earnings.

The following describes how Person Cs form groups and make their decisions.

Person C group's decision

Each Person C in an earlier experiment was given a copy of these instructions. At the beginning of the experiment, Person Cs were randomly assigned into groups of three. The task for the group was to decide jointly, after discussions amongst group member, whether or not to approve any payoff deduction proposal, and if so, how much would be deducted. Each Person C group was asked to specify their decisions for all of the possible scenarios *where at least one person in the pair proposes a deduction for the other.* Each Person C will discuss the decisions with his/her group members and each group will reach an agreement on what to decide and then submit a final decision.

- If a group approved a payoff deduction proposal, the maximum deduction amount the group could impose is 40 tokens. That is, a group could deduct any amount of tokens between 1 and 40. The amount has to be an integer (e.g. 1, 2…). Regardless of the deduction amount a group decided to impose, the person who proposes the deduction will always pay 5 tokens.
- If a group decided not to approve the deduction proposed, the person who proposes the deduction will not pay anything and his/her counterpart will not receive any payoff deduction.

- **Note:** If a Person A's or Person B's final earnings from the decision task are negative, then the final earnings from the decision task will automatically be set to zero and this participant will receive only the $5 show up fee.

Person C group's earnings

Person C group's earnings were determined by a random process. The computer randomly assigned 15, 20, 30 or 40 tokens as his/her earnings. Note that every member in a group will receive the same earnings randomly assigned by the computer. He/she was only informed of his/her earnings at the end of the experiment.

Person C group's feedback

After today's experiment, the experimenter will also email each Person C to let him/her know if the decisions of his/her group have been randomly chosen to be matched with a pair of Person A and Person B in today's experiment. If his/her group is chosen, then in the email, each Person C in that group will also see the matched Person A's and Person B's decisions. Note that no one else, except the experimenter, will know the identity of any decision makers. Nor will the identity information be published or be released to other parties.

Information feedback of Person A/B

At the end of the experiment, Person A/B will see each other's decisions in both stages.

Examples to illustrate payoff calculations

Below are some examples to illustrate how payoffs of Person A, B and C are determined.

Suppose, in Stage 1, Person A chose Option I and Person B chose Option II. Thus, Person A earned 15 tokens and Person B earned 40 tokens in Stage 1. Suppose, in Stage 2, Person A proposed a payoff deduction from Person B's earnings and Person B also proposed a payoff deduction from Person A's earnings. Also, suppose the computer had randomly assigned 20 tokens as the earnings of a matched Person C group. In this case,

- Suppose Person Cs of that group have decided: 1) to approve Person A's proposal and to assign a deduction of 9 tokens to Person B's payoff; and 2) to disapprove Person B's proposal. Each one's earnings (in tokens) are as follows:

  Person A's earnings=15 (earnings in Stage 1)-5 (amount paid to impose the deduction)=10

Person B's earnings=40 (earnings in Stage 1)-9 (amount to be deducted) =31

Each Person C's earnings in that group =20

- Suppose Person Cs of that group have decided: 1) to approve Person A's proposal and to assign a deduction of 9 tokens to Person B's payoff; and 2) to approve Person B's proposal and decided to assign a deduction of 6 tokens to Person A's payoff.

  Each one's earnings (in tokens) are as follows:

  Person A's earnings=15 (earnings in Stage 1) -5 (amount paid to impose the deduction) - 6 (amount to be deducted) = 4

  Person B's earnings=40 (earnings in Stage 1) -5 (amount paid to impose the deduction) -9 (amount to be deducted) =26

  Each Person C's earnings in that group =20

- Suppose Person Cs of that group have decided: 1) to disapprove Person A's proposal; and 2) to disapprove Person B's proposal. In this case, no one's earnings would be changed. Each one's earnings (in tokens) are as follows:

  Person A's earnings=15

  Person B's earnings=40

  Each Person C's earnings in that group =20

  Now suppose, in Stage 2, Person A proposed a payoff deduction from Person B's earnings and Person B did not propose any payoff deduction from Person A's earnings. In this case,

- Suppose Person Cs of that group have decided: 1) to approve Person A's proposal and to assign a deduction of 8 tokens to Person B's payoff.

  Each one's earnings (in tokens) are as follows:

  Person A's earnings=15-5=10

  Person B's earnings=40-8=32

  Each Person C's earnings in that group=20

- Suppose Person Cs of that group have decided to disapprove Person A's proposal. In this case, no one's earnings would be changed. Each one's earnings (in tokens) are as follows:

  Person A's earnings=15

  Person B's earnings=40

  Each Person C's earnings in that group =20

The purpose of the above examples is to illustrate the payoff calculations, rather than provide advice about how to act. You should make the decisions as you wish.

Summary

Person A and Person B interact with each other in a decision task *only once*. This experiment consists of two stages. At the beginning of the first stage, both Person A and Person B will see all the decisions made by the paired Person C in each scenario. Then, each person will decide to choose either Option I or Option II, which will determine each one's earnings in the first stage. In the second stage, after knowing Person A's and Person B's decisions in the first stage, Person A (Person B) will decide whether or not to propose a payoff deduction for the counterpart. The proposed deduction decision will be implemented only if the matched Person C group has approved it in the corresponding scenario. The Person Cs in that group will also decide the deduction amount. Each Person C of a group receives the same earnings randomly determined by the computer. Each Person C will receive an email informing him/her whether his/her group's decisions are implemented. Person A and Person B will make their decisions *after* knowing the matched Person C group's decision.

Please raise your hand if you have any questions at this moment.