Teams Do Inflict Costly Third-Party Punishment as Individuals Do: Experimental Evidence

Kenju Kamei

Department of Economics and Finance, Durham University Business School, Mill Hill Lane,
Durham DH1 3LB, UK; kenju.kamei@gmail.com or kenju.kamei@durham.ac.uk

Abstract: Initiated by the seminal work of Fehr and Fischbacher (Evolution and Human Behavior (2004)), a large body of research has shown that people often take punitive actions towards norm violators even when they are not directly involved in transactions. This paper shows in an experimental setting that this behavioral finding extends to a situation where a pair of individuals jointly decides how strong a third-party punishment to impose. It also shows that this punishment behavior is robust to the size of social distance within pairs. These results lend useful insight since decisions in our everyday lives and also in courts are often made by teams.

1. Introduction

A social dilemma is a collective action problem in which agents benefit from peers' cooperative actions, but in which selfish individuals have no incentives to cooperate, the canonical example being the prisoner's dilemma. Because social dilemmas are a pervasive feature of people's interactions, scholars have actively studied people's cooperation behaviors, along with mechanisms that may sustain cooperative social norms in societies, for the last few decades (see, e.g., [1–3] for a survey).

One consistent finding from the prior research is that some individuals have other-regarding preferences. A large body of experimental research has shown that even third parties (those who are not directly involved in the relevant dilemma interactions) frequently inflict punishment on norm violators [4–6]. Scholars have uncovered a number of features of such third-party punishment. For instance, altruistic third-party punishment is associated with the activation of specific brain regions, such as reward regions [7,8]. The intensity of third-party punishment differs according to various factors—for example, individual characteristics, such as justice sensitivity [9], population and the size of societies [10,11], and group affiliations, that is, whether norm violators belong to the punisher's group [12,13]. Such third-party punishment behavior is considered to be unique to humans [14]. The past work, however, leaves one important question unanswered: if we confront a pair of two third-party players with the same dilemma interaction and let them *jointly* inflict a *single* third-party punishment on the same target, do they punish in a similar manner to individuals alone? (The impact of having two third-party individuals that act independently, instead of a pair that makes joint punishment decisions, was studied in [6].) Instances where a pair of individuals, instead of a stand-alone individual, encounters a norm violation are readily available in our everyday lives (e.g., a couple taking a relaxing walk in a park, two university friends eating lunch together at a university cafeteria). Considering that some individuals have preferences other than self-interest, such as inequity aversion, pride, guilt, and shame, individuals may receive a greater non-material utility triggered through social comparison when acting as an enforcement pair. If this is the case, third-party pairs may punish more strongly than third-party individuals, in analog to a benefit of collective behavior in a coalition game. However, a large volume of experimental literature on team decision-making suggests that teams are more cognitively sophisticated

thanks to communication among team members and hence behave more selfishly than individuals, which undermines social welfare (see, e.g., [15–17] for surveys). In strategic environments, teams behave more strategically than individuals do in order to maximize their own material payoffs. Under this alternative hypothesis, "more cognitively able" third-party pairs would punish less relative to third-party individuals since third-party punishment acts are privately costly. Pairs' third-party punishment could then be weaker than individuals', but then, why do we not often hear that people are less punitive when they act as a pair even though enforcement coalitions are frequently formed in reality?

This study conducts two treatments with joint decision-making to explore how third-party pairs' punishment differs from third-party individuals'. In each treatment, there are two subjects in a group that play a one-shot prisoner's dilemma game with each other (PD players, hereafter). A pair of third-party players is additionally placed in each group, and jointly make punishment decisions for their respective group. The two treatments differ in how the pairs are formed. In one treatment, each pair is anonymously and randomly formed at the onset of the experiment by the computer. In the other treatment, the identity of each subject within the pair is revealed at the time of pair formation; they are then given an opportunity to introduce themselves including their names. This kind of design setup is known to reduce social distance within pairs, thereby affecting pairs' altruistic tendencies [18–20]. For example, close social distance within pairs has been shown to improve cooperation in a finitely repeated public goods game when the decision-making unit is pairs [19]. The third-party punishment behavior of pairs is compared with that of individuals in a control treatment, identically designed except that a stand-alone individual, instead of a pair, is placed in each group as a third-party punisher.

The experiment shows that the behavioral pattern of third-party punishment extends to pairs: third-party pairs impose significantly stronger punishment on a defector who interacts with a cooperator than on any other type of PD player. A closer look at the data indicates interesting team dynamics. Prior to communicating with their partners, subjects in the third-party pairs have stronger inclinations to punish than third-party individuals do. The willingness of pairs to punish, however, decreases significantly after communication, consistent with the idea that communication among team members strengthens their strategic thinking [15–17]. Nevertheless, the punishment strength of pairs does not become weaker than that of individuals. This suggests that the prior findings on third-party punishment are robust to third parties' decision-making formats. This also implies that forming a third-party coalition to enforce cooperation norms would not have negative consequences on their punitive inclinations.

The rest of the paper proceeds as follows: Section 2 describes the experimental design. Section 3 briefly discusses related literature. Section 4 reports results, and Section 5 concludes.

2. Experimental Design

The design frame is a prisoner's dilemma game with the involvement of a third-party punisher [4–6]. A between-subjects design is used in this study (Table 1). There are three treatments that vary in the third-party punishment form: whether the third party is (a) an independent individual, or (b) a pair of individuals that jointly decides a single punishment amount for their group. The three treatments are referred to as the "Individual Punishment" (I-P) treatment, the "Joint Punishment" (J-P) treatment, and the "Joint Punishment, Close Social Distance" (J-P-C) treatment. The I-P treatment is the control treatment (The same I-P treatment was also used as a control treatment in [21].). The conversion rate is as follows: five points in the experiments are equal to £1 in all three treatments. Subjects are privately paid in cash based on their accumulated points at the end of the experiment.

Table 1. Summary of Treatments.

Treatment Name	Third Party	Decision-Making Protocol	# of Subjects
I-P (Individual Punishment)	Individual	Individual	48
J-P (Joint Punishment)	Pair	Joint (each third-party player is not aware of who their partner is in their enforcement pair.)	72
J-P-C (Joint Punishment, Close Social Distance)	Pair	Joint (each third-party player is aware of who their partner is in their enforcement pair.)	68
Total	—	—	188

Each treatment has two stages. In stage 1, the two PD players are each endowed with 25 points, and simultaneously decide whether to send 10 points to their counterparts. If a subject sends 10 points, the sender receives 15 (= 25 − 10) points, and the counterpart receives 30 (= 10 × 3) points as an additional payoff. If the subject does not send 10 points to the other player, the subject simply receives 25 points. Each group has another individual or a pair of individuals as a third-party player. The PD players are aware that the third-party player acts as a punisher in the next stage. Third-party players make no decision in stage 1. They are just asked on their computer screens to provide guesses as to how many PD players, zero, one, or two, in the group would send 10 points in stage 1. Since the focus of this study is on third-party players' punishment decisions, this question is not incentivized to avoid subsequent effects in stage 2. The question is included merely to retain a high level of anonymity in sessions (i.e., to make the numbers of computer mouse clicks the same between PD players and third-party players).

In stage 2, third-party players decide on how to inflict punishment. Each punisher is given an endowment of 40 points. Punishment points toward each PD player must be a non-negative integer and be less than or equal to 20. Punishment is privately costly: for each punishment point imposed on a PD player, three points are deducted from the target's payoff while one point is deducted from the punisher's own payoff. A strategy method is employed to obtain as many incentive-compatible observations as possible. Third parties' punishment behaviors are known not to be affected by manipulating the use of the strategy method [22]. Specifically, each third-party punisher decides punishment points under the following four possible stage 1 outcomes:

Scenario CC: Punishment points targeted at a cooperator (a player who sent 10 points to the partner) when the partner was another cooperator in the group.
Scenario DC: Punishment points targeted at a defector (a player who did not send 10 points to the partner) when the partner was a cooperator in the group.
Scenario CD: Punishment points targeted at a cooperator when the partner was a defector in the group.
Scenario DD: Punishment points targeted at a defector when the partner was another defector in the group.

After third-party players make punishment decisions for all four potential outcomes, their choices corresponding to the realized PD players' sending decisions are applied. Once stage 2 is over, each subject learns the outcome of both stages 1 and 2. A PD player's payoff is set to zero if her payoff is negative due to punishment received. Sections 2.1 and 2.2 explain the specific design piece of each treatment.

2.1. The I-P Treatment (Control)

In the I-P treatment, subjects are randomly assigned to a group that consists of two PD players and one third-party individual. The PD players play the aforementioned prisoner's dilemma game in stage 1. The third-party players make punishment decisions as explained above in stage 2.

2.2. *The J-P and J-P-C Treatments*

In the J-P and J-P-C treatments, subjects are randomly assigned to a group of four individuals. The four subjects are randomly assigned a player number: player 1, 2, 3, or 4 so that each player has a different number. Players 3 and 4 then form a pair, and the pair acts as a third-party punisher in their group. Players 1 and 2 are assigned the role of the PD player.

2.2.1. The J-P Treatment

Stage 1 proceeds the same as the I-P treatment. Once the PD players complete their sending decisions, third-party pairs are given five minutes to freely discuss within their pairs via a free-form computer chat window, as in [19]. Subjects are neither informed at which desk their counterparts are seated in the laboratory nor given any other information that may specify the matched third-party partners' identity. Subjects are also not allowed to convey any personal information that may specify themselves (No subjects violated this rule in the experiment.). Before communicating with their partners, each person in a pair is asked to answer how many punishment points they would assign as the pair's joint punishment points under the four scenarios, in the hypothetical event that they could decide the joint punishment amounts unilaterally without communication (on the condition that the payoff consequence would be the same between the two individuals in the pair). This elicitation task is included to investigate possible social effects as a supplementary analysis (An alternative to eliciting hypothetical willingness to punish is to make the pre-communication question incentive-compatible by setting it to be realized with some probability. This approach was not employed primarily because this would make the design more complex to subjects, but also because the focus of this study is on the pairs' actual decisions to punish. The same elicitation method was included in a context of a finitely repeated two-player public goods game in [19].). If having a partner inflates punishers' non-material motives, such as guilt, shame, and pride, third parties may display an inflated willingness to punish before the communication begins [23].

Once the five minutes of the communication stage pass, each individual in a pair submits the number of punishment points that they want to assign as a pair under the four possible scenarios. The joint decision-making rule is as follows: if two individuals in a pair submit the same punishment points for a given scenario, then the points become that pair's joint punishment points. If they submit different points, then one of the two submissions is selected randomly by the computer. Note that there are several other methods to break a tie in the case of disagreement, such as (a) a default option is then applied, (b) each teammate does not obtain any points in a given game, (c) a majority rule is applied, or (d) the average of submissions is used. This study adopted the random tie-breaking rule because not only is there no consensus regarding which method is the best, but this random tie-breaking rule also worked sufficiently in the author's earlier paper [19] in the context of a public goods game. This rule was applied for only two pairs in Scenario CC, four pairs in Scenario DC, three pairs in Scenario CD, and three pairs in Scenario DD in the J-P treatment. This rule was applied for only one pair in Scenario CC, two pairs in Scenario DC, one pair in Scenario CD, and two pairs in Scenario DD in the J-P-C treatment. There are no real decisions for PD players to make during this stage. Once all third-party punishers submit decisions, they are informed of the realized interaction outcomes. The resulting payoffs of the two individuals in a third-party pair are the same: each obtains a payoff of $40 - p_1 - p_2$, where p_1 and p_2 are the pair's joint punishment points imposed on the two PD players in their group. The same per-subject payoff consequences for individuals and pairs are usually designed in this kind of team decision-making experiment (e.g., [18–20]).

2.2.2. The J-P-C Treatment

The J-P-C treatment is identical to the J-P treatment, except that each subject in a pair is informed of the seat number of their partner and is then given two minutes to introduce themselves using a computer chat window before the five-minute discussion stage begins.

There are *no* restrictions on the contents of the communication (except for the prohibition of offensive language). Subjects are explained that the content of any communication is private, not subject to analysis, and not disclosed in any formats. Social distance within pairs is hence assumed smaller in the J-P-C than in the J-P treatment. A similar design setup was used in [19] to induce a different level of social distance within pairs in the context of a repeated public goods game. [19] demonstrated that pairs show a stronger willingness to contribute to their group when they are aware of their paired partner's identity than otherwise.

2.3. Experimental Procedure

A total of 140 subjects participated in the two joint-decision treatments at the EXEC laboratory at the University of York in December 2015 and August and September 2016 (Table 1). Another 48 subjects participated in the sessions of the I-P treatment in December 2015. The number of subjects in each of the J-P and J-P-C treatments was set to around 1.5 times more than the I-P treatment because this study wanted to analyze pairs' punishment decisions in detail (e.g., the effects of communication) as will be explained in Section 4.2. Institutional Review Board (IRB) approval has been obtained for this project from both Durham University and the University of York. All subjects were recruited through *hroot* [24]. No subjects participated in more than one session. All experiments except instructions were programmed using the z-Tree software [25]. At the onset of the experiment, subjects were provided instructions that explain all the procedures (e.g., stages 1 and 2). The instructions and any verbal explanation were neutrally framed (see supplementary material B for the instructions). Subjects had to answer several control questions before the experiments commenced.

3. Related Literature and Discussions

This paper's focus is on the third-party pair's punishment behavior. Standard theory predicts that no third parties would inflict punishment on PD players, since the former do not directly interact with the latter and the punishment acts are costly. However, some other-regarding preference models, such as inequity aversion [26], predict positive punishment behavior by third-party players in Scenarios CC and DC in the experimental environment if they are sufficiently strongly inequality-averse (see the online Appendix of [21]). Emotions, such as anger, may also account for third-party punishment [27].

Previous research where third parties are *individuals* has found that third-party punishment is particularly widespread in Scenario DC (see, e.g., [4–6,9–13,21]). However, the question of how the joint decision-making procedure affects people's third-party punishment motives remains unanswered. A rich body of experimental literature suggests that communication within pairs may sharpen strategic minds and thus induce third-party pairs to punish less. First, collective institutional choices by majority voting have been found to limit anti-social and irrational choices in public goods dilemmas (e.g., [28–30]). Because third-party punishment in any scenario would decrease the third party's payoff, the joint decision-making procedure may undermine pairs' willingness to punish. Second, and more importantly, prior experimental research indicates that a team decision-making process through communication may make people more cognitively able and thus more selfish in various settings, including ultimatum games, beauty-contest games, signaling games, and centipede games (see, e.g., [15–17] for a survey). If two-person pairs behave closer to the game-theoretic prediction than individuals, pairs would decrease their willingness to punish through communication. This paper is the first to study the impact of joint decision-making through communication in the case of third-party punishment.

It should be worth noting that having a partner in the enforcement team may enhance pairs' initial willingness to punish before within-pair communication takes place. For example, Kamei [21] showed that in a design where there is one third-party individual in a group, the punishment intensity on a norm violator is stronger when each punisher's action choice is made known to another punisher in a different group than when the punitive

actions are kept private. There is also some prior research that proposes that people have non-strategic image concerns [31,32].

4. Results

Before studying pairs' punishment behaviors, this section first compares PD players' cooperation (sending) decisions in the J-P and J-P-C treatments with those in the I-P treatment. The data indicate that PD players' cooperation rates differ only slightly by treatment. The rates are 71.9% (23 out of 32 subjects), 55.6% (20 out of 36 subjects), and 70.6% (24 out of 34 subjects) in the I-P, J-P, and J-P-C treatments, respectively. The cooperation rates are not significantly different between any two treatments (Part (1), supplementary material Table S1). The distributions of third parties' expectations regarding the number of cooperators are also similar among the three treatments: their average beliefs are 1.31, 1.33, and 1.35 persons in the I-P, J-P, and J-P-C treatments, respectively. Mann–Whitney tests fail to reject the null hypothesis that the average beliefs are different between any two treatments (Part (2), Supplementary Material Table S1). This may mean that pairs' third-party punishment is widespread similar to individuals' punishment.

4.1. Punishment Decisions by Third-Party Players

Figure 1 shows the frequencies and the strength of third-party punishment. The data replicate those of past research for pairs. First, pairs' third-party punishment is common in each scenario and the frequency is especially high in Scenario DC—when the target is a defector while the counterpart cooperates (Panel (a)). Second, PD players receive much stronger punishment from third-party pairs in Scenario DC than in any other scenario (Panel (b)), as is the case for third-party individuals. Note that the punishment intensity is significantly stronger in Scenario DC than in Scenarios CC and CD for both the J-P and J-P-C treatments. This pattern is similar to that in the I-P treatment–see supplementary material Table S3. Further, the punishment intensity in Scenario DC is also significantly stronger than that in Scenario DD in the J-P treatment. These patterns of punishment strength resonate with the idea that pairs are not purely selfish, and they attempt to mitigate inequality in their groups by engaging in punishment.

(a)

Figure 1. *Cont.*

(**b**)
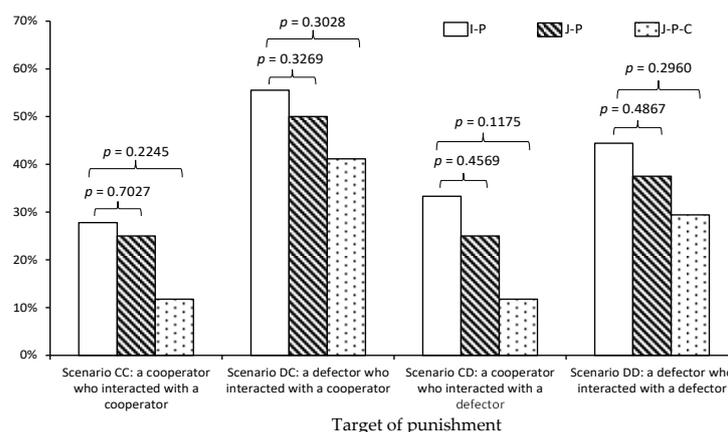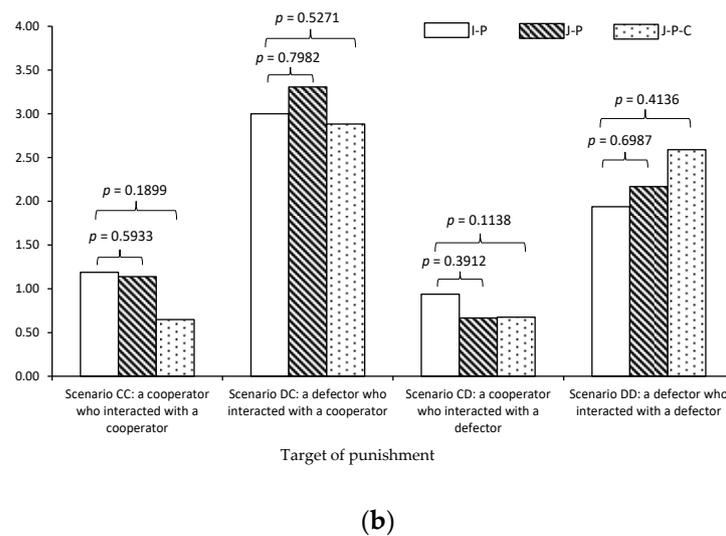
Figure 1. Frequency of Punishment, and Average Strength of Punishment: (**a**) Frequency of Punishment; (**b**) Average per PD player Punishment Points Received. *p*-values for two-sided Fisher's exact tests in panel (**a**) and two-sided Mann–Whitney tests in panel (**b**). The punishment frequency and strength are also not significantly different between the J-P and J-P-C treatments for each of the four scenarios ($p > 0.010$): *p*-values are omitted from the panels to conserve space.

Result 1. *Pairs' third-party punishment is common. PD players receive much stronger punishment from third-party pairs in Scenario DC than in any other scenario.*

A comparison between individuals and pairs finds that not only the punishment frequency but also the punishment intensity in each scenario does not differ between the decision-making formats (see Figure 1). This result is also independent of how pairs are formed. Here, individual-level data (the I-P treatment) and pair-average data (the joint-decision treatments) are used for comparing punishing intensity between the treatments. Results are similar also when we compare punishment behaviors using a tobit regression with clustered standard errors (the results are omitted to conserve space). These findings, along with Result 1, suggest that the main findings of previous research into individuals' third-party punishment ([4–6,9–13,21]) are robust to the third parties' format (individuals versus pairs). It should be noted, however, that care needs to be exercised and more data should be collected before the result can be generalized, since this result may be due to a lack of statistical power. Even a small difference could be detected if a sufficiently large number of observations were collected.

4.2. The Impact of Communication

Each third-party punisher in the J-P and J-P-C treatments has a partner. How does having a partner itself affect subjects' inclinations to punish? The impact of having a partner can approximately be measured by using the third-party players' willingness to punish elicited before the communication stage. Figure 2 shows both the average pre-communication willingness to punish and the average actual punishment points. Two clear phenomena were found. First, the average pre-communication willingness to punish under Scenario DC is significantly higher than the punishment strength observed in the I-P treatment (see supplementary material Table S4). The former is almost at the same level as the punishment imposed in that scenario in the Visibility treatment of [21]—see Section 3 for [21]. The former (6.19 and 5.62 points in the J-P and J-P-C treatments, respectively) is not significantly different from the latter (6.63 points in the Visibility treatment) according to a Mann–Whitney test (see again supplementary material Table S4). This implies that a social effect of having a pair mate can be at work in the J-P and J-P-C treatments, as is the case for the Visibility treatment in [21].

(a)

(b)
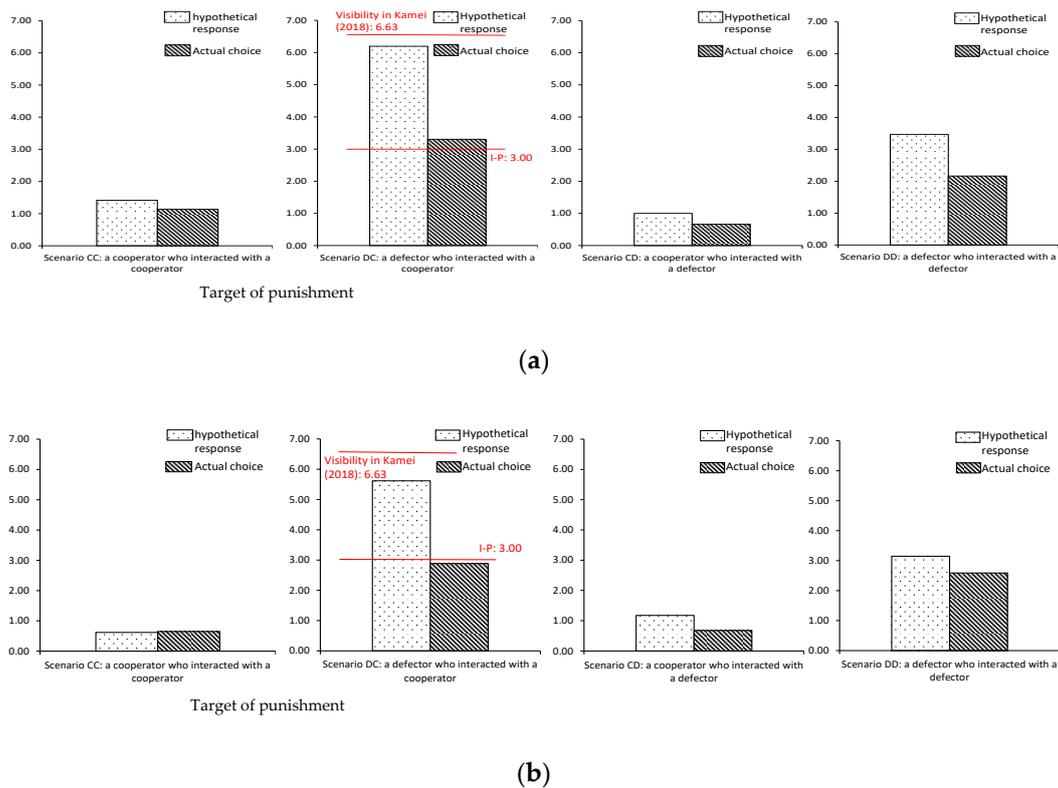
Figure 2. Pairs' Pre-communication Willingness to Punish versus After-Communication Willingness to Punish: (**a**) The J-P treatment; (**b**) The J-P-C treatment. The vertical axes indicate average punishment points per PD player. Each figure shows pre-communication and after-communication punishment points. See supplementary material Table S5 for comparisons between the hypothetical responses and actual punishment in each scenario.

Result 2. *The pre-communication willingness to punish in Scenario DC in the J-P and J-P-C treatments is much higher than the punishment strength in the I-P treatment. The former is not significantly different from punishment strength in Scenario DC observed in the Visibility treatment of [21].*

Second, subjects' willingness to punish under Scenario DC decreases substantially after communicating with their pair partner (Figure 2). The decreases are significant ($p < 0.05$ for each treatment; see supplementary material Table S5). This suggests that consistent with prior research on team decision-making, communication within pairs strengthens selfish tendencies. It should, nevertheless, be emphasized that pairs still inflict sizeable punishment even after communication and that their inclinations to punish do not become significantly lower than individuals' inclinations to punish (Figure 1).

Does the impact of communication differ by pairs' punitive disposition? To address this question, this paper follows the approach of [18] by classifying a pair as "self-regarding" ("other-regarding") if the pair's average pre-communication willingness to punish is below (above) the session average—see also [19]. As shown in Table 2, nine (eight) pairs are classified as self-regarding and eight (nine) pairs are classified as other-regarding in the J-P (J-P-C) treatment. The data indicate that the other-regarding pairs' willingness to punish in Scenario DC decreases significantly through communication in both the J-P and J-P-C treatments. Further, a significant decrease in the self-regarding pairs' punishment intensity is also observed through communication in the J-P treatment (the punishment intensity was already at a modest level before communication for the self-regarding pairs in the J-P-C treatment). This suggests that the impact of communication is not affected by pairs' intrinsic inclinations to punish.

Table 2. Shifts in the Third-Party Pairs' Willingness to Punish through Communication.

	J-P Treatment		J-P-C Treatment	
	Self-Regarding (1)	**Other-Regarding (2)**	**Self-Regarding (3)**	**Self-Regarding (4)**
The number of pairs	9 [#1]	8 [#1]	8 [#2]	9 [#2]
(i) Average pre-communication punishment points in Scenario DC (C_i^{pre})	3.61	9.44	2.25	8.61
(ii) Average after-communication punishment points in Scenario DC (C_i)	1.39	5.63	3.13	2.67
Direction of Shifts:				
(a) Upward shift ($C_i > C_i^{pre}$)	2	2	2	1
(b) No shift $\left(C_i = C_i^{pre} \right)$	1	0	2	0
(c) Downward shift $\left(C_i < C_i^{pre} \right)$	6	6	4	8
Wilcoxon signed ranks test for the null H_0: (i) = (ii)				
p-value (two-sided)	0.0363 **	0.0793 *	0.6180	0.0105 **

[#1] The total number of pairs in the J-P treatment is 18. Because one pair's pre-communication average punishment points in Scenario DC were the same as the session average, this pair was not assigned either self-regarding or other-regarding pair. [#2] The total number of pairs in the J-P-C treatment is 17. * and ** indicate significance at the 0.10 level and at the 0.05 level, respectively.

Result 3. *Communication within pairs decreases their inclinations to punish in Scenario DC, regardless of whether pairs are classified as other-regarding or self-regarding.*

5. Discussion and Conclusions

This paper experimentally explored whether third-party punishment is frequently observed for the case of pairs, as is the case for individuals. The experiment data reveal that pairs' altruistic punishment is common and similar in size to that of third-party individuals. This finding is useful considering that third-party punishment can be taken not only by individuals but also by teams in the real world.

Lastly, it should be acknowledged that there are many areas of further research possibilities. For example, while this study used a simple one-shot prisoner's dilemma game to eliminate any reputation concerns, one may wonder how the difference in decision-making format, pairs versus individuals, may affect the evolution of cooperation among PD players if the game is repeated. The experiment data, unfortunately, imply a possible negative answer to this question: irrespective of the format, cooperation may not evolve if third parties exhibit similar levels of punitive inclinations consistently over time. While third-party pairs' punishment was frequently observed, its strength was in fact not strong enough to induce a PD player to choose cooperation only from material motives in this study. A calculation indicates that the incentive structure that the PD player faces is still one of a prisoner's dilemma game even when punishment amounts (Figure 1) are subtracted from the payoffs; the expected payoff of a PD player who selects cooperation (defection) is 30.86 (35.11) and 34.21 (37.79) points in the J-P and J-P-C treatments, respectively. The detailed calculation can be found in the working paper version of this paper [33]. However, this result shares similarities with prior research based on third-party individuals [4–6,21]. Having said this, prior research on team decision-making in repeated games suggests that teams are less myopic loss averse and invest more in building materially beneficial cooperative relationships than individuals under certain conditions ([19,34,35]). If third parties also have some repeated dilemma interactions in their communities like the PD players, third-party pairs may realize indirect benefits from costly punishment, and hence use stronger punishment strategically than individuals do. This possibility could not be

verified unless more data are collected. Further experiments will be required regarding how repeated interactions may change third parties' punishment attitudes.

Third parties' punitive inclinations may differ by their demographic attribute or co-operative type in the prisoner's dilemma. For example, certain pairing, such as gender composition, parochial ties, or ages, may affect third parties' punitive inclinations. As discussed, while joint decision-making was conducted anonymously in the J-P treatment, each third-party player in the J-P-C treatment was aware of the identity of their partner, whose aspect makes additional analyses on attribute/type possible. However, the number of third-party pairs in the J-P-C treatment was 17 in this study, rendering empirical analyses by splitting data into subcategories difficult. Studying the role of third-party pair composition is an exciting further research direction. For another example, while this study did not let third parties play as a PD player, for instance [36] showed, in an adversarial, criminal setup, that not only ordinary citizens, but also some criminal defectors punish other defectors by reporting their misconduct, and the defectors' pro-social reporting plays a meaningful role in sustaining cooperation norms in communities. An investigation on the relationship between players' cooperative dispositions in the community and their third-party punishment behaviors also remains an interesting avenue for future research.

Supplementary Materials: The followings are available online at https://www.mdpi.com/2073-4336/12/1/22/s1, Figure S1: Percentages of Third Party Players who Engage in Punishment in the J-P and J-P-C treatments, Table S1: PD Players' Sending Rates and Third Party Punishers' Beliefs on the Number of Cooperators, Table S2: The Differences in the Frequency of Third party Punishment between Scenarios in the Two Joint-Decision Treatments, Table S3: The Differences in Average Punishment Points between Scenarios in the Two Joint-Decision Treatments, Table S4: Pre-Communication Willingness to Punish in the J-P and J-P-C treatments versus Punishment Intensity in the I-P treatment and in the Visibility treatment of Kamei (2018) [21] in Scenario DC, Table S5: Decreases in the Willingness to Punish through Communication in the J-P and J-P-C treatments, Appendix B: Instructions used in the Experiment.

Institutional Review Board Statement: Institutional Review Board approvals was obtained from both Durham University and the University of York.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Each subject was given and signed on a consent form before the sessions started.

Data Availability Statement: Data available upon request.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Exp. Econ.* **2011**, *14*, 47–83. [CrossRef]
2. Fehr, E.; Fischbacher, U. Social norms and human cooperation. *Trends Cogn. Sci.* **2004**, *8*, 185–190. [CrossRef] [PubMed]
3. Fehr, E.; Schmidt, K. The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories. In *Handbook of the Economics of Giving, Altruism and Reciprocity*; Kolm, S.-G., Ythier, J.M., Eds.; Elsevier: North Holland, The Netherlands, 2006; pp. 615–691, ISBN 9780444506979.
4. Fehr, E.; Fischbacher, U. Third-party Punishment and Social Norms. *Evol. Hum. Behav.* **2004**, *25*, 63–87. [CrossRef]
5. Lergetporer, P.; Angerer, S.; Glätzle-Rützler, D.; Sutter, M. Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6916–6921. [CrossRef] [PubMed]

6.   Kamei, K. Group Size Effect and Over-Punishment in the Case of Third Party Enforcement of Social Norms. *J. Econ. Behav. Organ.* **2020**, *175*, 395–412. [CrossRef]

7.   Buckholtz, J.; Asplund, C.; Dux, P.; Zald, D.; Gore, J.; Jones, O.; Marois, R. The Neural Correlates of Third-Party Punishment. *Neuron* **2008**, *60*, 930–940. [CrossRef]

8.   Strobel, A.; Zimmermann, J.; Schmitz, A.; Reuter, R.; Lis, S.; Windmann, S.; Kirsch, P. Beyond revenge: Neural and genetic bases of altruistic punishment. *NeuroImage* **2011**, *54*, 671–680. [CrossRef]

9.   Lotz, S.; Baumert, A.; Schlösser, T.; Gresser, F.; Fetchenhauer, D. Individual Differences in Third-Party Interventions: How Justice Sensitivity Shapes Altruistic Punishment. *Negot. Confl. Manag. Res.* **2011**, *4*, 297–313. [CrossRef]

10.   Henrich, J.; McElreath, R.; Barr, A.; Ensminger, J.; Barrett, C.; Bolyanatz, A.; Cardenas, J.; Gurven, M.; Gwako, E.; Henrich, N.; et al. Costly punishment across human societies. *Science* **2006**, *312*, 1767–1770. [CrossRef] [PubMed]

11.   Marlowe, F.; Berbesque, C.; Barrett, C.; Bolyanatz, A.; Gurven, M.; Tracer, D. The 'spiteful' origins of human cooperation. *Proc. R. Soc. Lond. B Biol. Sci.* **2010**, *278*, 2159–2164. [CrossRef]

12.   Bernhard, H.; Fehr, E.; Fischbacher, U. Group Affiliation and Altruistic Norm Enforcement. *Am. Econ. Rev.* **2006**, *96*, 217–221. [CrossRef]

13.   Lieberman, D.; Linke, L. The effect of social category on third party punishment. *Evol. Psychol.* **2007**, *5*, 289–305. [CrossRef]

14.   Riedl, K.; Jensena, K.; Calla, J.; Tomaselloa, M. No third-party punishment in chimpanzees. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14824–14829. [CrossRef]

15.   Charness, G.; Sutter, M. Groups Make Better Self-Interested Decisions. *J. Econ. Perspect.* **2012**, *26*, 157–176. [CrossRef]

16.   Kugler, T.; Kausel, E.; Kocher, M. Are groups more rational than individuals? A review of interactive decision making in groups. *WIREs Cogn. Sci.* **2012**, *3*, 471–482. [CrossRef] [PubMed]

17.   Kerr, N.; Tindale, R.S. Group Performance and Decision Making. *Annu. Rev. Psychol.* **2004**, *55*, 623–655. [CrossRef]

18.   Cason, T.; Mui, V.-L. A Laboratory Study of Group Polarisation in the Team Dictator Game. *Econ. J.* **1997**, *107*, 1465–1483. [CrossRef]

19.   Kamei, K. Power of Joint Decision-Making in a Finitely-Repeated Dilemma. *Oxf. Econ. Pap.* **2019**, *71*, 600–622. [CrossRef]

20.   Kocher, M.; Sutter, M. Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica* **2007**, *34*, 63–88. [CrossRef]

21.   Kamei, K. The Role of Visibility on Third Party Punishment Actions for the Enforcement of Social Norms. *Econ. Lett.* **2018**, *171*, 193–197. [CrossRef]

22.   Jordan, J.; McAuliffe, K.; Rand, D. The effects of endowment size and strategy method on third party punishment. *Exp. Econ.* **2016**, *19*, 741–763. [CrossRef]

23.   Bowles, S.; Gintis, H. *Prosocial emotions, In The Economy as a Complex. Evolving System III: Essays in Honor of Kenneth Arrow*; Blume, L., Durlauf, S., Eds.; Oxford University Press: Oxford, UK, 2005; pp. 337–367, ISBN 9780195162592.

24.   Bock, O.; Nicklisch, A.; Baetge, I. hroot: Hamburg Registration and Organization Online Tool. *Eur. Econ. Rev.* **2014**, *71*, 117–120. [CrossRef]

25.   Fischbacher, U. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Exp. Econ.* **2007**, *10*, 171–178. [CrossRef]

26.   Fehr, E.; Schmidt, K. A Theory of Fairness, Competition, and Cooperation. *Q. J. Econ.* **1999**, *114*, 817–868. [CrossRef]

27.   Nelissen, R.; Zeelenberg, M. Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgm. Decis. Mak.* **2009**, *4*, 543–553.

28.   Ertan, A.; Page, T.; Putterman, L. Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free-Rider Problem. *Eur. Econ. Rev.* **2009**, *53*, 495–511. [CrossRef]

29.   Kamei, K.; Putterman, L.; Tyran, J.-R. State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Exp. Econ.* **2015**, *18*, 38–65. [CrossRef]

30.   Putterman, L.; Tyran, J.-R.; Kamei, K. Public goods and voting on formal sanction schemes. *J. Public Econ.* **2011**, *95*, 1213–1222. [CrossRef]

31.   Ariely, D.; Bracha, A.; Meier, S. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *Am. Econ. Rev.* **2009**, *99*, 544–555. [CrossRef]

32.   Bénabou, R.; Tirole, J. Incentives and Prosocial Behavior. *Am. Econ. Rev.* **2006**, *96*, 1652–1678. [CrossRef]

33.   Kamei, K. Altruistic Norm Enforcement and Decision-Making Format in a Dilemma: Experimental Evidence. Unpublished Working Paper. 2017. Available online: https://papers.ssrn.com/abstract=2912543 (accessed on 3 March 2021).

34.   Feri, F.; Irlenbusch, B.; Sutter, M. Efficiency Gains from Team-based Coordination—Large-scale Experimental Evidence. *Am. Econ. Rev.* **2010**, *100*, 1892–1912. [CrossRef]

35.   Gillet, J.; Schram, A.; Sonnemans, J. The Tragedy of the Commons Revisited: The Importance of Group Decision-Making. *J. Public Econ.* **2009**, *93*, 785–797. [CrossRef]

36.   D'Orsogna, M.; Kendall, R.; McBride, M.; Short, M. Criminal Defectors Lead to the Emergence of Cooperation in an Experimental, Adversarial Game. *PLoS ONE* **2013**, *8*, e61458.