"Hot" and "Cold" Punishment in Public Goods Provision*

Nanyin Yang† Sora Youn‡ Andy Cao§ Jinliang Liu¶
Phatchaya Piriyathanasak‖ Samuel Priestley** Catherine Eckel††

October 17, 2023

Abstract

Previous studies have shown that punishment opportunities can reduce free riding effectively in public goods production and that negative emotions toward free riders play an important role in precipitating punishment. By varying the timing of punishment in a public good game, we develop a novel punishment rule, the "Pre-Punishment" rule, which is designed to involve a lower level of emotional arousal compared to the classical "Post-Punishment" rule. We employed biometric measures (eye trackers and skin conductance response) in a lab experiment to capture the psychological responses, which will shed light on the mechanism mediating punishment behavior, the response to punishment, and the impact on cooperative behavior. Our results show that this new punishment rule works equally well in increasing contribution compared to the Post-Punishment rule. However, the biometric finding indicates that the effectiveness of Pre-Punishment rule does not rely on subjects' emotional arousal. This study provides useful suggestions for policymakers and managers for designing proper penalty rules to increase cooperation, and will also contribute to the public good game literature by uncovering the psychological processes underlying the effectiveness of punishment institutions.

Keywords: Public goods games, Cooperation, Punishment, Emotion, Pupil dilation
JEL Codes: C92, D87, D91, H41

*We thank seminar participants in 2023 Workshop of Recent Advances in the Economics of Philanthropy, 2022 World ESA Meeting, and the 4th workshop on Choice-Process Data for useful comments.
†Yang: Department of Economics, Texas A&M University, yangnanyin@tamu.edu
‡Youn: Department of Economics, Texas A&M University, sorayoun@tamu.edu
§Cao: Department of Economics, Texas A&M University, andycao@tamu.edu
¶Liu: Department of Economics, Texas A&M University, jl716@tamu.edu
‖Piriyathanasak: Department of Agricultural Economics, Texas A&M University, piri97@tamu.edu
**Priestley: Department of Agricultural Economics, Texas A&M University, sam.priestley@tamu.edu
††Eckel: Department of Economics, Texas A&M University, ceckel@tamu.edu

1 Introduction

The public good game is a commonly used model to study cooperation in social dilemmas. The game captures the tension between individual payoff maximization and group welfare which characterizes social dilemmas. In this game a group of four players each allocates a fixed endowment between a mutual project, the public good, and a private fund. Each player's payoff includes their private fund and a payoff from the mutual project, Equal across players, consisting of an amount for each unit contributed to the public good; this is the marginal per capita return (MPCR) of the public good and falls between 1/4 and one. In a Nash equilibrium where players maximize their individual payoffs, every player would contribute zero to the public good, since the MPCR is less than one. However, empirical studies demonstrate that the average contribution level of a group is typically higher than zero, although it usually declines over time in repeated public good games(Isaac & Walker, 1988).

When a second stage of the game is introduced allowing players to punish others at a personal cost, this is shown to increase contributions and may also improve efficiency. The presence of punishment opportunities helps sustain cooperation for longer periods of time, even though subjects incur a cost by forgoing a part of their own payoff in order to punish others (Bochet et al., 2006; Falk et al., 2005; Fehr & Gächter, 2000, 2002; Masclet et al., 2003). This observation is also at odds with the theoretical prediction that in a public good game with a costly punishment opportunity, individuals would not punish because doing so would hurt their payoff and if effective, benefit the group. Thus punishment itself is a public good.

Why do individuals punish, and why is it effective? Several studies focus on the mechanisms behind the effectiveness of costly punishment, and considers both strategic and emotional motives. From the side of a punish implementer, punishment serves as a credible threat to hurt payoff, which encourages punishment receivers to increase their contributions (Fehr & Gächter, 2000). Moreover, those who received punishment would increase their contribution to avoid a loss from being punished in the next round of interaction. On the other hand, another important explanation involves emotional mechanisms. The primary motive for assigning punishment stems from a desire to express anger and sanction free-riders who violate norms of fairness (Falk et al., 2005), rather than being strategic. Furthermore, there is also evidence showing that people use punishment as a tool for revenge, not only towards free riders but also towards group members that (may) punish themselves (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Herrmann et al., 2008; Nikiforakis, 2008). In addition, those who are punished also increase their contribution to avoid feelings of guilt or shame resulting from the punishment they received (Hopfensitz & Reuben, 2005).

While both explanations have supporting evidence, it is unclear to what extent each mechanism contributes to the observed effects. The classical public good game design where punishment opportunities are given after contribution does not allow for a clear separation of these mechanisms. In this paper, we introduce a novel punishment rule where subjects make contingent punishment decisions before punishment, suppressing the emotional mechanism. Furthermore, we equip our experiment with eye trackers to uncover subjects' psychological process to verify our manipulations. We find evidence that this new rule leverages the benefits of punishment while suppressing subjects from utilizing punishment as a tool of revenge.

Specifically, we design a public good game where we vary the timing of the punishment in order to manipulate the emotional responses of subjects. This game is based on the classical design of Fehr and Gächter (2000), and we impose two punishment rules to manipulate subjects' negative emotions in the game. Under the "Post-Punishment" treatment, subjects make punishment decisions *after* contributing to the public good, allowing them to take into account each group member's contribution level when assigning punishment points. In contrast, under the "Pre-Punishment" treatment, each subject sets and commits to their own private "cutoff" *before* making contributions, and whoever contributes below this cutoff automatically receives the punishment. The key difference between these two treatments is that unlike the Post-Punishment treatment, subjects in the Pre-Punishment treatment are making punishment decisions without knowing the contribution of others and therefore the punishment is a more general commitment without targeting specific individuals.

We hypothesize that negative emotion plays a limited role in Pre-Punishment compared to Post-Punishment treatment. Unlike Post-Punishment rule under which subjects can directly express their anger or disagreement toward free riders through punishment, under the Pre-Punishment rule, punishment is unlikely to be driven by those negative emotions toward free riders because subjects are not able to know who is the free rider. In order to verify our hypotheses, we measure each subject's pupil dilation and skin conductance response (SCR) during the experiment which serve as the proxy of their psychological arousal in the game. Both pupil dilation and SCR are used as a metric of the psychological arousal driven by emotional status (Bradley et al., 2008; Brocas et al., 2021; Joffily et al., 2014; Wang et al., 2010). Specifically, we assess the emotional arousal triggered by contribution gaps and study how it relates to the subject's punishment decisions on others. In addition, we examine whether receiving punishment causes stronger emotional arousal and whether the negative emotion predicts higher subsequent contributions.

Our findings reveal that both the Post-Punishment and Pre-Punishment rules significantly increase contributions in the public good game. However, the psychological processes underlying these rules differ. Under both rules, subjects who experience free riders in their groups exhibit heightened emotional responses. This emotional response predicts subsequent punishment decisions in the Post-Punishment rule but not in the Pre-Punishment rule, suggesting that punishment serves as a means of expressing negative emotions only in the Post-Punishment treatment.

Moreover, both rules elicit stronger emotional arousal when subjects learn that they are being punished. However, these emotional responses only predict a marginal increase in subsequent contributions under the Pre-Punishment rule, but not under the Post-Punishment rule. We therefore cautiously suggest that subjects' contributions are primarily driven by strategic considerations, such as avoiding further punishment and financial losses.

Our study makes several contributions to the public good literature. First, we vary the timing of the punishment rule and develop a new punishment rule named "Pre-Punishment rule" where subjects commit to a cutoff and whoever contributes below will be punished. This rule is designed to suppress subjects using punishment as a tool of revenge and to express negative feelings, which enables us to examine whether negative emotions play a role in the effectiveness of punishment. Our findings show that this new punishment rule is equally effective as the Post-Punishment rule, i.e., the classical punishment rule designed by Fehr and Gächter (2000).

Second, we utilize the biometric measures to uncover the psychological process of a public good game, which provides direct evidence on how negative emotions play a role in the effectiveness of punishment rules. Our results indicate that the two punishment rules, although both significantly increase contributions, are operating through different mechanisms. While the punishment is utilized to express negative feelings toward free riders under the Post-Punishment rule, the punishment under the Pre-Punishment rule is not driven by negative feelings.

Overall, our study offers insights into the design of penalty rules to promote cooperation and social welfare. The Pre-Punishment rule provides an alternative approach that leverages the benefits of punishment while reducing the reliance on negative emotions. By uncovering the psychological processes behind punishment, our findings contribute to a better understanding of cooperation in social dilemmas and offer implications for policymakers aiming to design effective penalty systems.

This paper is structured as follows. Section 2 describes the experimental design, including the punishment rules in the public goods game and the use of biometric devices. Section 3 reports the experimental results. Section 4 is the conclusion.

2 Experimental Design

2.1 Round 1-10: Public Good Game without Punishment

In the first part of the experiment, subjects repeated a public goods game for 10 rounds. We emplyed the classical design of public good game in Fehr and Gächter (2000) where each subject is assigned to a group of four. At the beginning of each round, each subject is endowed with 20 tokens, and they individually decide how many tokens to contribute to a group project. The total contribution of the project will be multiplied by 1.6 and equally distributed among group members, i.e., each group member receives 0.4 times the total contribution during that round. Therefore, the subject's earning from each round consists of two parts: the return from the group project, and the remaining endowment that is not contributed. Below is the equation used to calculate a subject i's earning in a round t, with other group members denoted as j:

$$Earnings_{i,t} = 0.4 * (x_{i,t} + \sum_{j \neq i} x_{j,t}) + (20 - x_{i,t}) \tag{1}$$

Apart from the contribution decisions, we also elicited each participant's belief on the average contribution level within their group for each round. After the contribution decisions, participants received feedback on the contribution level of each group member in an anonymous manner. Finally, they received information about their contribution level and payoff from that round.

2.2 Round 11-20: Post- and Pre-Punishment Rules

After the first 10 rounds, we introduced the punishment opportunities to subjects and subjects repeated the public good game with punishment opportunities for the next 10 rounds. There are two punishment rules in our study, Post-Punishment rule and Pre-Punishment rule, with a between-subject design. The Post-Punishment rule is similar to the punishment rule in

Fehr and Gächter (2000), under which subjects make punishment decisions after receiving feedback about each group member's contribution. But we simplify the punishment into a binary decision: we present subjects with their group members' contribution levels, and they need to make binary decisions on each group member about whether to assign punishment to them or not. The punishment assigned to a group member will reduce this group member's earnings by 3 tokens at 1 token of cost from the punishing subject.

Under the Pre-Punishment rule, subjects are making punishment decisions after making contributions and eliciting beliefs by choosing a "cutoff". This cutoff represents the minimum contribution level that a subject will tolerate from each group member, and whoever contributes below this cutoff will automatically receive a reduction of 3 tokens at 1 token of cost from that subject. In addition, each subject chooses the maximum number of subjects that they would like to punish, and whenever the number of subjects contributing below the cutoff is higher than that maximum number, we will randomly choose the maximum number of subjects to receive punishment. The cutoff and the maximum number are kept private and not revealed to other group members.

Under both punishment rules, subjects will receive feedback on the amount of punishment they receive from others and they impose on others. However, the punishment is anonymous, meaning that subjects who are punished were not informed about who had punished them.

The key difference between the Post- and the Pre-Punishment rules is whether subjects identify free riders when making punishment decisions. Under the Post-Punishment rule, subjects know every group member's contribution level when assigning punishment and could use punishment to retaliate against free riders. In contrast, under the Pre-Punishment rule, subjects make punishment decisions, i.e., their cutoffs, before making contributions, therefore they are not able to identify who will be the free rider. We conjecture that this difference will change the role that negative emotions play in the effectiveness of punishment: on one hand, subjects are unlikely to use punishment to express anger or disagreement toward a certain free rider under the Pre-Punishment rule; on the other hand, subjects who receive punishment may also experience less negative emotions (shame or guilt) under the Pre-Punishment rule, because they are less likely to feel that they are targeted by someone specifically.

2.3 Biometric Devices

During this experiment, we recorded each subject's pupil dilation using the Tobii eye tracker X2-60 and Pro Spectrum, sampled at 60Hz. In addition, we record each subject's skin conductance signal by the Shimmer wristband, sampled at 128Hz. To make subjects' biological activities comparable to each other, we exposed subjects to each of the scenes of viewing punishments and viewing contributions for 10 seconds. Figure 1 describes the scene presentations for each round from round 11 to 20. The scene of viewing contributions will capture subjects' emotional arousal when learning the contribution differences between themselves and other group members. And the scene of viewing punishments will record subjects' emotional responses to the punishment that they receive.

2.4 Experimental Procedure

The experiment was performed at Human Behavior Lab, Texas A&M University. Our subjects were 108 undergraduate students recruited through the Sona System; 52 participated in the Post-Punishment treatment and 56 participated in the Pre-Punishment treatment. Subjects were assigned to groups of 4 randomly, but the group composition remained the same throughout the 20 rounds of the experiment. And we used a between-subject design for the Post- and Pre-Punishment rules, so each subject only experienced one punishment rule. The experiment on average lasted 60 minutes in the lab. 1 round out of round 1-10 and another round out of round 11-20 were randomly selected for payment. The average earning is $ 32.7, including a 10-dollar participation fee, an average earning of $ 7.3 from a pre-experimental online survey (explained below), and an average earning of $ 15.4 from the main study.

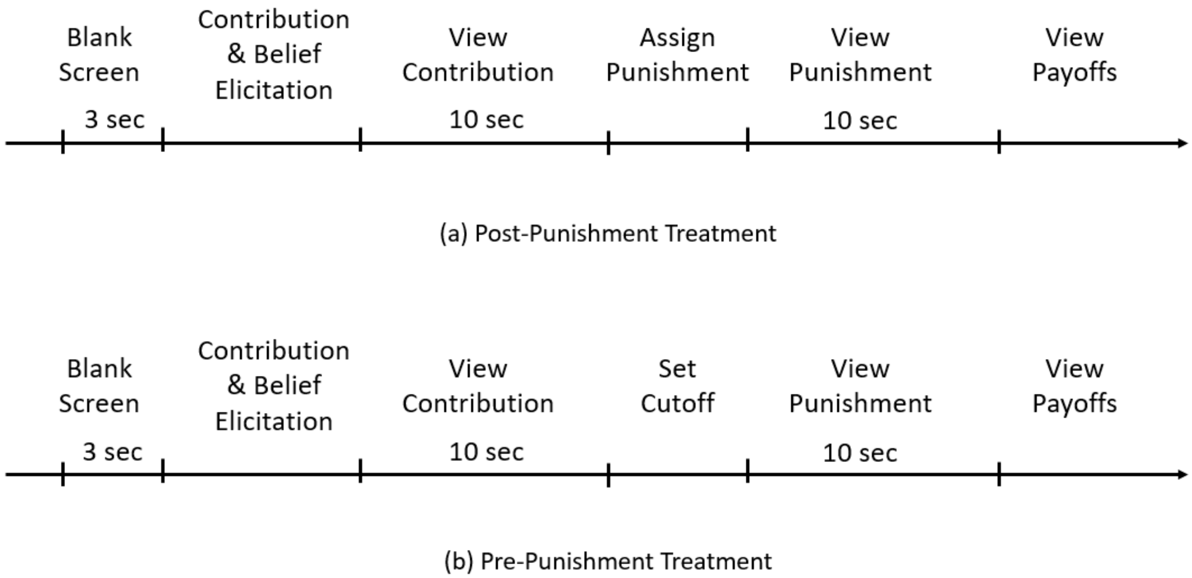


Figure 1: Scene Presentations in Round 11-20. This figure depicts the scenes that a subject observed in a round during round 11 to round 20, where punishment opportunity is available. "Blank screen" is a scene that persisted for 3 seconds during which subjects were shown a blank screen with a black cross in the middle. This scene helps re-calibrate subjects' eye fixations, and it also existed in Round 1-10. "View Contribution" is a 10-second scene where subjects were presented with the contribution level of each group member. "View Punishment" is a 10-second scene where subjects were presented with the punishment feedback, including punishment points they received and sent.

1 day before taking part in the main study, all subjects completed an online survey that contained two tasks.[1] Task 1 measures subjects' other-regarding concerns using the Social Value Orientation slider measure designed by Murphy et al. (2011). This task consists of 6 decisions on money allocations between self and another anonymous subject, varying in the degree of conflict between self-benefiting and mutually beneficial options. This measure produces a continuous score that reflects subject's degree of prioritization of self-interest over other's interest, and is believed to be a valid predictor for cooperative behaviors (Balliet et al., 2009; Fiedler et al., 2013). We adopted a strategy method in this task, where all subjects made the six allocation decisions, but they would be randomly matched into pairs and only

[1] We implemented this online survey before the main lab study to avoid the spillover effect between the online tasks and the public good game, following the setup by Fiedler et al. (2013).

one subject within a pair would be randomly chosen as the actual allocator. For each chosen subject, one out of the 6 decisions would be randomly selected and implemented to determine the payoff within their pair. Task 2 is an incentivized gambles-choice task designed by Eckel and Grossman (2008) to elicit subjects' risk preferences, in which subjects choose one out of six gambles that vary in expected payoffs and variations.

After arrival, all subjects were randomly seated in the lab, and they were guided by experimenters to wear the skin conductance response device and calibrate their eye fixations using eye trackers. Then they read the instruction of the game and used computers to participate in the public good game. At the end of the experiment, all subjects filled out a questionnaire to provide basic demographic information and their opinions about the game.

3 Results

In this section, we present the results from our experiment. In section 3.1, we concentrate on cooperative behavior across treatments. We first show that both Post- and Pre-Punishment rules equally increase the average contribution. We also do a welfare analysis as a further comparison between these two punishment rules. In section 3.2, we elaborate the emotional mechanism using biometric data. The two main focuses are the emotional motives behind punishing decisions, and the impact of emotional arousals on contribution after receiving punishment. [2]

[2]Our main focus is the emotional mechanisms behind these two punishment rules. However, some readers may be interested in the dynamics between punishment decisions and contribution decisions. For the sake of completeness, we include an analysis of the interplay between punishment decisions and contribution decisions in Appendix A.

3.1 Treatment Effects of Punishment Opportunities

3.1.1 Increase of Contribution with Punishment Opportunity

We start our analyses by examining whether our treatment conditions increase contribution. While the unique Nash equilibrium predicts that a rational player would always contribute zero, the literature documents subjects contribute a positive amount. Cooperation deteriorates with repetition, but it jumps to almost a full contribution upon the introduction of punishment (Fehr & Gächter, 2000).

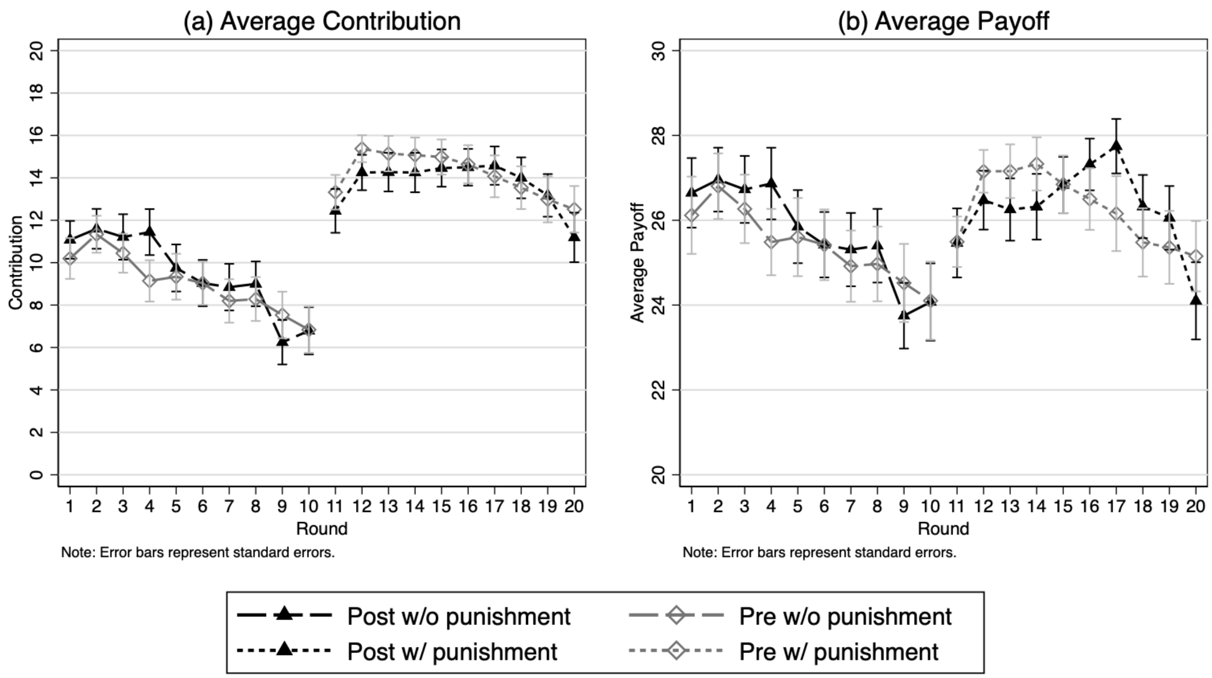


Figure 2: Treatment effects on Contributions and Individual welfare

We find consistent evidence. Figure 2(a) illustrates the average contribution of each round, conditional on treatment conditions. Recall that there is no punishment opportunity in round 1-10, and starting from round 11 punishment opportunities are introduced. The first finding from Figure 2(a) is that when there is no punishment opportunity, the average contribution is decreasing, which is true for both treatments – in round 1, the average contribution is 11.08

(sd6.43) for the Post one and 10.2 (sd 7.19) for the Pre one, and these numbers decrease to 6.79 (sd 8.01) and 6.83 (sd 8.19) in round 10. Notice that when there is no punishment, these two treatments are identical. Therefore, unsurprisingly, the averages between the two treatments in the first 10 rounds without punishment are not statistically different ($p > 0.1$ for two-sample t-test).

Furthermore, Figure 2(a) implies that when punishment opportunities are introduced to the game, subjects' contribution levels are higher under both treatments – in round 11, the average contribution jumped to 12.44 (s.d. 7.44) in Post- and 13.32 (s.d. 6.11) in Pre-. Moreover, the increase in the contribution is persistent until the very last round of the game. The averages for the last 10 rounds are 13.71 (s.d. 6.9) in Post and 14.16 (s.d. 6.84) in Pre. They do not reach to full contribution, though. Moreover, we find little difference in average contribution between the Pre-Punishment treatment condition and the Post-Punishment condition. The averages between these two treatments are not statistically different ($p > 0.1$ for tw-sample t-test). Both types of punishment rule increased contribution to the public good.

To formally estimate the effect of punishment on contribution that we find above, Table 1 reports the estimation of the following regression model. The dependent variable is the amount of contribution from subject i on round t. The effect of punishment opportunities is captured by α_1. α_2 captures the heterogeneous effect of Pre-Punishment opportunities in comparison to Post-Punishment opportunities. We control for subjects' beliefs about the average contribution of the other three group members, round effects, and group fixed effects.

$$Contribution_{it} = \alpha_0 + \alpha_1 WithPun_t + \alpha_2 WithPun_t \times PrePunishment_i + Controls_{it} + \varepsilon_{it} \quad (2)$$

In Table 1, Columns (1) and (2) report estimates from the Post-Punishment treatment and Pre-Punishment treatment separately. And estimates from the full sample are provided in Column (3). Regression results confirm the graphical evidence presented in Figure 2(a). Firstly, the implementation of a punishment rule, regardless of its type, leads to a significant increase of about 7.5 tokens in subjects' contribution level. Secondly, although punishment opportunities increase contribution 0.9 tokens more in Pre-Punishment than in Post-Punishment, this difference is not statistically significant (*p*¿0.1). Columns (4) to (6) include additional controls for individual characteristics, including SVO angle, gender, age, ethnicity, whether they are from Texas, and whether subjects are in economics-related majors. Accounting for a rich set of controls, we confirm that the positive impact of punishment opportunities on an individual's contribution remains robust. Furthermore, there are no significant disparities between the Pre-Punishment and Post-Punishment rules in terms of their effectiveness in enhancing contribution levels.

Table 1: Treatment effects on contribution

	Dependent Variable: Contribution					
	(1) Post	(2) Pre	(3) Pooled	(4) Post	(5) Pre	(6) Pooled
w/ punishment	7.687***	8.345***	7.553***	7.650***	8.434***	7.556***
	(1.141)	(1.185)	(0.796)	(1.225)	(1.239)	(0.842)
Pre w/ punishment			0.917			0.962
			(0.872)			(0.907)
Constant	4.556***	11.555***	11.623***	22.974**	-3.437	3.798
	(0.985)	(0.865)	(0.803)	(10.337)	(10.167)	(6.518)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Total Observations	1040	1120	2160	940	1060	2000
Number of Individuals	52	56	108	47	53	100

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

Result 1. *The Post- and the Pre-Punishment rules both significantly increase the average group contribution, and there is no significant difference in the increase of average contribution between these two punishment rules.*

3.1.2 Efficiency of Punishment Opportunities

Although we have shown that both types of punishment rules are equally effective in increasing contributions to the public good, since punishment is costly in our setting, it is possible that the tokens lost from punishment outweigh the increase in contributions. Therefore, in this subsection, we compare the efficiency of punishment between the two punishment rules, with subjects' payoffs as the outcome of interest.

Figure 2(b) demonstrates the average individual payoff taking into account costs related to punishment. At first glance, we see introducing punishment opportunities seems to increase the average payoff, for both the Pre- and the Post-Punishment rules.

To further investigate this finding above, Table 2 reports the estimate the following regression analysis.

$$Payoff_{it} = \alpha_0 + \alpha_1 WithPun_t + \alpha_2 WithPun_t \times PrePunishment_i + Controls_{it} + \varepsilon_{it} \quad (3)$$

The effect of punishment opportunities on individual payoff is captured by α_1. α_2 captures the heterogeneous effect of Pre-Punishment opportunities in comparison to Post-Punishment opportunities. In Table 2, we control for rounds and group fixed effects for all columns of regressions. Columns (4) through (6) add individual controls to the model. On average, despite the cost of punishment, subjects earn 2.19 and 3.03 more tokens when introducing the Post- and the Pre-Punishment opportunities respectively (Column (4) and (5)). And we find no difference between the effect of Post- and Pre-Punishment rules, as the interaction of Pre-Punishment indicator and the indicator of punishment rounds is not significant in columns (3) and (6). This finding indicates that the two punishment rules increase individual payoff equally well.

Result 2. *Both the Post- and the Pre-Punishment rules significantly increase the welfare, and there is no significant difference between these two punishment rules in increasing welfare.*

Table 2: Individual Payoff

	Dependent Variable: Individual Payoff					
	(1) Post	(2) Pre	(3) Pooled	(4) Post	(5) Pre	(6) Pooled
w/ punishment	2.632***	2.973***	2.677***	2.185**	3.035***	2.469***
	(0.922)	(0.804)	(0.623)	(0.927)	(0.835)	(0.631)
Pre w/ punishment			0.253			0.314
			(0.659)			(0.685)
Constant	22.780***	27.911***	27.888***	32.814***	41.943***	36.097***
	(0.696)	(0.834)	(0.785)	(7.104)	(9.058)	(5.696)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Total Observations	1040	1120	2160	940	1060	2000
Number of Individuals	52	56	108	47	53	100

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

3.2 Emotional Mechanisms of the Effectiveness of Punishment Rules

To delve into by what extent emotional mechanisms can explain contribution and punishment behavior observed in our experiment, we link the choice data with the biometric data. In the paper, we primarily use pupil dilation as the proxy measuring subjects' emotional arousal[3]. To address the concern that subjects' absolute pupil diameters are highly variant, we follow Sirois and Brisson (2014) and pre-process the pupil dilation data in the following way: for each certain stage of interest (e.g. ten seconds when a subject is viewing contribution feedback in a certain round), we subtract each subject's pupil dilation by a pre-trial baseline value, which is the average pupil diameter within 0.5 seconds before entering this certain stage; then we take the average of the relative-change measures for left and right eyes, and calculate the average across the whole stage of interest, for example within ten seconds of viewing contribution feedback. In this section, when we use the term "pupil dilation" or "pupil diameter", we are indicating the relative change of pupil diameters from the pre-trial baseline value. In addition,

[3] We used Tobii eye tracker X2-60 and Pro Spectrum to record subjects' gaze data. Following Sirois and Brisson (2014), we use linear interpolation for samples where data are missing for one or both eyes.

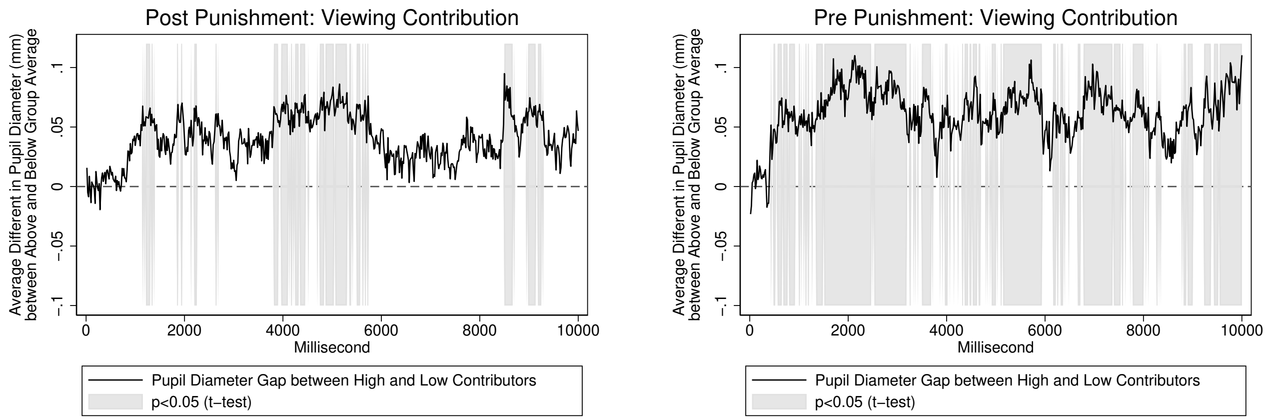
we also provide analyses using skin conductance responses in the Appendix.[4].

3.2.1 Emotional Motives behind Punishment Decisions

In this sub-section, we investigate subjects' psychological process when learning group members' contribution levels, and whether emotional arousal predicts their punishment decisions.

We start by analyzing subjects' pupil dilation when they are viewing contribution feedback, as presented in Figure 3. Based on the experimental design, after every group member made the contribution decision, subjects had to view a screen showing the contribution levels of their group members for 10 seconds. We define subjects in a round who contribute strictly higher than the average of their group members as high contributors, and subjects who contribute equal to or lower than their group members' average as low contributors. Based on this classification, high contributors in a round are free-ridden by others and they receive payoffs lower than the initial endowment in that round. Therefore, we expect that high contributors, when viewing the contribution feedback screen, should experience stronger negative emotions (for example anger or disappointment) than low contributors. Figure 3(a) and 3(b) present the average differences in pupil diameters between high and low contributors during the 10-second viewing period, with round 11 to 20 pooled together, conditional on Post- and Pre-Punishment rules separately. In these two panels, a positive difference indicates larger pupil dilation among high contributors compared to low contributors. Additionally, we conducted two-sample t-tests at each time point within the 10 seconds, shading in grey those points where the p-value was below 0.05, indicating a significant difference in pupil dilation.

[4]The skin conductance signal was continuously recorded by the Shimmer wristband. We pre-process the skin conductance signal data by using the R Notebook by iMotions.Inc, whose algorithm is constructed based on standardized methods (Benedek, n.d.; Fowles et al., 1981; Greco et al., 2016). The phasic data is extracted from the original skin conductance signal using a median filter over an 8000 ms window. The onsets/offsets are identified when the phasic signal crosses above/below the onset/offset threshold of 0.01 μS. The SCR amplitude threshold is 0.02 μS. The SCR amplitude is the difference between the signal amplitude at the peak and the onset. Following Joffily et al. (2014), the SCR magnitude analysis of a certain stage of interest takes the whole stage into account and assigns zero's to those subjects without a measurable response within the stage.



((a)) Post-Punishment

((b)) Pre-Punishment

Figure 3: Pupil Dilations When Viewing Contributions

Note: These two figures depict the differences of pupil dilation when viewing contribution feedback between high contributors (with contribution strictly higher than the other three group members' average contribution) and low contributors (with contribution lower or equal to other group members' average contribution), with positive values indicating higher pupil diameters among high contributors than low contributors. The pupil diameter is recorded at every 60 milliseconds. We pool round 11 20 together and calculate the average pupil diameter differences for each time point. For each time point, we conduct a pairwise t-test between subjects above and below average. We shade in grey those time points where the p-value is below 0.05, indicating statistically significant differences.

Figure 3(a) and 3(b) show that under both punishment rules, high contributors' pupil dilations are significantly higher than low contributors frequently during the 10-second period of viewing contribution feedback. Regression analysis from Table B1 in the Appendix confirms this finding, and this table further demonstrates that there is no difference in pupil dilation among high contributors between these two treatments. These findings confirm our hypothesis that subjects experience heightened arousal when they learn about being free-ridden by others, indicating the presence of negative emotions in response to free-riding behavior.

Next, we investigate whether higher emotional arousal from being free-ridden by others predicts more punishment towards others. We define "High Contributor" as a binary indicator of whether a subject contributed strictly above the group average in a certain round, i.e., whether the subject is a high contributor. And for each subject in each round, we take an average of pupil dilation across the 10 seconds when viewing the contribution feedback as the measure of emotional arousal. To analyze the relationship between emotional arousal and punishment decisions under the Post-Punishment rule, we regress the punishment points assigned to others in round t on the pupil dilation and the indicator of being above group average:

$$\begin{aligned} PunishmentAssigned_{it} =& \alpha_0 + \alpha_1 PupilDilationit + \alpha_2 HighContributor_{it} \\ & + \alpha_3 HighContibutor_{it} \times PupilDilation_{it} + Controls_{it} + \varepsilon_{it} \end{aligned} \tag{4}$$

As for the Pre-Punishment rule, because the punishment decision (i.e. cutoff choices) occurs before subjects viewing contribution feedback, we use subjects' pupil dilation in round $t-1$ and subjects' status of being a high contributor in round $t-1$ to predict subjects' punishment points assigned to others in round t, with the model specification as follows:

$$\begin{aligned} PunishmentAssigned_{it} =& \alpha_0 + \alpha_1 PupilDilationit - 1 + \alpha_2 HighContributor_{it-1} \\ & + \alpha_3 HighContibutor_{it-1} \times PupilDilation_{it-1} + Controls_{it-1} + \varepsilon_{it} \end{aligned} \tag{5}$$

Table 3 reports the regression results from the two specifications above. For the Post-Punishment rule (Columns (1) and (2)), high contributors are found to assign more punishment points to others. More importantly, high contributors with larger pupil diameters during the 10 seconds of viewing contribution feedback exhibit a greater tendency to impose punishments. This aligns with our predictions and previous literature, suggesting that under the Post-Punishment rule, subjects express their anger or disagreement by targeting specific individuals after learning about group members' contributions. In contrast, the regression results for the Pre-Punishment rule (Columns (3) and (4) in Table 3) reveal distinct patterns. Firstly, being a high contributor in the previous round does not predict increased punishment points in the current round. Moreover, the interaction between the high contributor indicator and pupil dilation has a negative and insignificant coefficient, indicating that as a high contributor, being more aroused by free riders does not increase punishment points assigned to others. Findings from Table 3 are consistent with our predictions: under the Pre-Punishment rule, the punishment decision is made before learning about the group's contributions in the current round, reducing the likelihood of using the cutoff to express anger towards specific group members; additionally, emotional arousal from the previous round does not spill over to influence punishment decisions in the current round.

Table 3: Decision to Punish

	DV: Punishment Points Assigned to Others			
	Post		Pre	
	(1)	(2)	(3)	(4)
Pupil Dilation	-0.335*	-0.238		
	(0.202)	(0.204)		
High Contributor	0.391***	0.334***		
	(0.094)	(0.089)		
High Contributor=1 × Pupil Dilation	0.645**	0.621**		
	(0.288)	(0.289)		
L.Pupil Dilation			0.366	0.308
			(0.296)	(0.301)
L.High Contributor			0.130	0.096
			(0.114)	(0.119)
1L.High Contributor × L.Pupil Dilation			-0.127	-0.132
			(0.348)	(0.366)
Constant	0.391	1.774*	0.257	2.216***
	(0.292)	(0.955)	(0.293)	(0.518)
Group FE	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	No	Yes
Total Observations	298	279	291	281
Number of Individuals	37	34	35	34

Note: Standard errors are in parentheses. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity. Pupil dilation is measured within the scene when subjects were viewing the contribution of each group member.

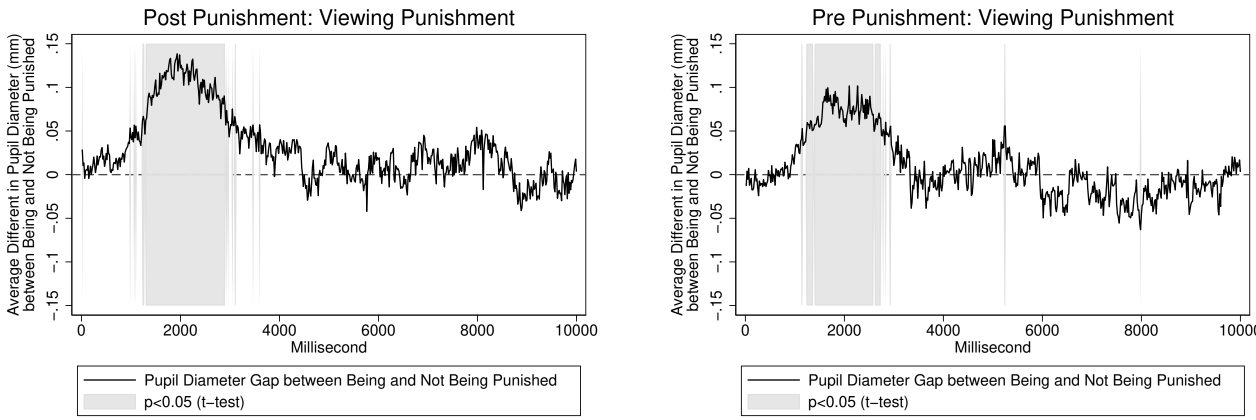
Result 3. *(1) Under both the Pre- and the Post-Punishment rules, compared with those who contribute less or equal to the other group members' average, subjects who contribute strictly above the other group members' average (namely, high contributors) experience stronger emotional arousal when learning the contributions of other group members; (2) These high contributors' higher emotional arousal predicts more punishment only under the Post-Punishment rule, but not under the Pre-Punishment rule.*

3.2.2 Emotion Trigger by Receiving Punishment and the Impact on Contribution Decisions

In this subsection, we are examining the psychological process for subjects who received punishment. While in Section 3.1, we demonstrate that both punishment rules increase contributions equally well, in this section, we will examine the psychological process of subjects who are punished and whether their emotional arousal predicts a subsequent increase in contribution.

Under both treatments, in every round with punishment opportunities, subjects had to view a screen of punishment feedback for 10 seconds, during which they learned whether they were punished by others and the number of punishment points they received. For each round, we classify subjects to those who are punished and who are not, and calculate the average difference in pupil dilations between these two groups. Figure 4 shows this average difference during the 10 seconds of viewing punishment, pooling round 11-20 together. In addition, we color in grey for those time points when the p-value is lower than 0.05 from the t-test on pupil dilation between subjects being punished and not being punished. Figure 4(a) focuses on the Post-Punishment rule and 4(b) concentrates on the Pre-Punishment rule. These two sub-figures both present a significant difference in pupil dilation between those who receive and do not receive punishment, mostly happening between 1000 and 3000 milliseconds.

Regression analysis in Table B2 in the Appendix further confirms the finding that being punished leads to stronger arousal. Furthermore, the coefficient associated with the interaction



((a)) Post-Punishment ((b)) Pre-Punishment

Figure 4: Pupil Dilations When Viewing Punishment Results

Note: These two figures depict the differences of pupil dilation when viewing punishment results between those who are punished and those who are not punished, with positive values indicating larger pupil diameters of those who are punished than those who are not. The pupil diameter is recorded at every 60 milliseconds. We pool round 11-20 together and calculate the average pupil diameter differences for each time point. For each time point, we conduct a t-test between subjects receiving and not receiving punishment. We shade in grey those time points where the p-value is below 0.05, indicating statistically significant differences.

between the Pre-Punishment indicator and the number of punishment points received is significant and negative, which indicates that the arousal from receiving punishment is stronger under the Post-Punishment rule than under the Pre-Punishment rule. This finding is in line with our prediction that under the Post-Punishment rule, subjects who receive punishment will experience stronger negative emotions (e.g. shame or guilt) because they know that they are targeted by some group members, while under the Pre-Punishment rule, the punishment is a general commitment of a threshold, leading to less negative emotions.

Since arousal does arise due to being punished by others, the next question of interest is whether this arousal predicts higher contribution in the next round. We take the average of pupil dilation across the 10-second scene of viewing punishment feedback for each subject in each round to construct the emotional arousal measure. Table 4 regresses the contribution change from round $t-1$ to t to an indicator of subject receiving punishment in round $t-1$, subject' pupil dilation when viewing punishment in round $t-1$, and the interaction between these two variables. Columns (1) and (2) analyze the Post-Punishment rule, while Columns (3) and (4) analyze the Pre-Punishment rule. This table yields consistent findings with Table A2 in the Appendix: receiving punishment is positively associated with an increase in contribution in the subsequent round under both the Post- and Pre-Punishment rules. Notably, the coefficient of the interaction between being-punished indicator and pupil dilation in Columns (1) and (2) is negative and statistically insignificant. This coefficient suggests that for Post-Punishment rule, higher arousal resulting from knowledge of being punished does not further enhance subjects' contribution in the subsequent round. Therefore, while receiving punishment significantly arouses subjects, our results do not support our prediction that subjects will increase their contribution to assuage their negative feelings triggered by punishment from others. This result suggests that under the Post-Punishment rule, the motivation behind an increase in contribution after receiving punishment is more likely to be a strategic avoidance of further punishment.

However, surprisingly, in Column (3) of the Pre-Punishment rule, the coefficient of the interaction between being-punished indicator and pupil dilation is positive. Furthermore, in Column (4) after adding a rich set of individual controls, this coefficient becomes marginally significant and maintains a similar magnitude to that in Column (3). Although findings from Table B2 suggests that subjects are less aroused from being punished under the Pre-Punishment rule than under the Post-Punishment rule, Columns (3) and (4) in Table 4 suggests that the arousal in Pre-Punishment rule predicts higher contribution. Therefore, we find evidence that under the Pre-Punishment rule, subjects' negative feelings (e.g. disappointed; shamed) play a role in the effectiveness of punishment. This again deviates from our predictions.

Result 4. *Under both Post- and Pre-Punishment rules, receiving punishment causes higher emotional arousal, and this arousal is stronger under the Post-Punishment rule. However, this arousal does not predict an increase in subsequent contribution under the Post-Punishment rule, but predict some increase of contribution under the Pre-Punishment rule.*

Table 4: Changes in Contribution

	DV: Δ Contribution			
	Post		Pre	
	(1)	(2)	(3)	(4)
1L.Being punished	3.508***	3.341***	6.197***	6.884***
	(0.843)	(0.910)	(1.254)	(1.442)
L.Pupil Dilation	2.740	3.180	-0.819	-0.541
	(1.966)	(2.245)	(2.587)	(2.731)
1L.Being punished × L.Pupil Dilation	-3.537	-3.421	6.135	6.878*
	(5.357)	(5.777)	(3.884)	(4.036)
Constant	4.274***	1.265	5.922***	7.708**
	(1.269)	(3.989)	(1.714)	(3.099)
Group FE	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	No	Yes
Total Observations	284	265	264	255
Number of Individuals	37	34	36	35

Note: Standard errors are in parentheses. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity. Pupil dilation is measured within the scene when subjects were viewing the punishment feedback scene, where they learned the punishment points that they received in that round.

4 Conclusion

Punishment after the group contribution is a commonly used modification in a public good game to reduce free-riding and increase contribution. In this study, we varied the timing of the punishment decision in a public good game and design a new punishment rule, the "Pre-Punishment" rule, under which subjects make punishment decisions before contribution decisions. We adopted a between-subject design to compare these two punishment rules, Post-Punishment rule and Pre-Punishment rule. This novel design allows us to manipulate the negative emotion triggered in the game, which will shed light on how negative emotions play a role in the efficacy of these two distinct punishment rules. Moreover, we incorporated eye trackers in the experiment to measure subjects' pupil dilation which served as a proxy of their psychological arousal driven by emotional status.

Our results show that the Post- and the Pre-Punishment rule increase group contribution equally well and there is no difference in terms of the efficiency of the costly punishment. Moreover, under both punishment rules, free riders are more likely to be punished, and receiving punishment significantly increases their subsequent contribution. However, the pupil dilation data indicates that the psychological processes between these two punishment rules are not identical. When realizing being free-ridden by others, subjects in both treatments were aroused (e.g. angry or disappointed), but this arousal predicted more punishment only under the Post-Punishment rule. In addition, when realizing being punished by others, subjects under the Post-Punishment rule are more aroused (e.g. guilt or shame) than those under the Pre-Punishment rule. However, surprisingly, the arousal triggered by punishment does not play a role in increasing the subsequent contribution under the Post-Punishment rule but only under the Pre-Punishment rule.

Our study makes a significant contribution to the public good literature by revealing the psychological mechanisms behind the effectiveness of punishment and introducing a novel punishment rule, the Pre-Punishment rule, which is compared to the commonly used Post-

Punishment rule in a public good game. We found that while both punishment rules increase group contribution equally well, the negative emotions triggered by these two punishment rules are distinct, and punishment's efficacy relies less on negative emotions under the Pre-Punishment rule. This novel punishment rule could be implemented in organizations or teams seeking to improve cooperation without relying on negative emotions among group members.

References

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547.

Benedek, M. (n.d.). Kaernbach c.(2010b). *Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology*, *47*(4), 647–658.

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *60*(1), 11–26.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Brocas, I., Carrillo, J. D., & Montgomery, M. (2021). Shaming as an incentive mechanism against stealing: Behavioral and physiological evidence. *Journal of Public Economics*, *194*, 104351.

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, *9*, 265–279.

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic theory*, 145–167.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results*, *1*, 1061–1073.

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, *73*(6), 2017–2030.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational behavior and human decision processes*, *120*(2), 272–284.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology*, *18*(3), 232–239.

Greco, A., Valenza, G., & Scilingo, E. P. (2016). *Advances in electrodermal activity processing with applications for mental health*. Springer.

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367.

Hopfensitz, A., & Reuben, E. (2005). The importance of emotions for the effectiveness of social punishment.

Isaac, R. M., & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, *103*(1), 179–199.

Joffily, M., Masclet, D., Noussair, C. N., & Villeval, M. C. (2014). Emotions, sanctions, and cooperation. *Southern Economic Journal*, *80*(4), 1002–1027.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, *93*(1), 366–380.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. (2011). Measuring social value orientation. *Judgment and Decision making*, *6*(8), 771–781.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1-2), 91–112.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive*

Science, *5*(6), 679–692.

Wang, J. T.-y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American economic review*, *100*(3), 984–1007.

Appendix

A Dynamics of Behaviors with Punishment Opportunities

In this section, we investigate the dynamics related to punishment opportunities. First, we are interested in the determinants of receiving punishment. Next, we study how receiving punishment changes subjects' subsequent contributions. Finally, we also examine whether receiving punishment also encourages subjects to punish others, as found in previous studies of punishments being used as retaliation (Nikiforakis, 2008).

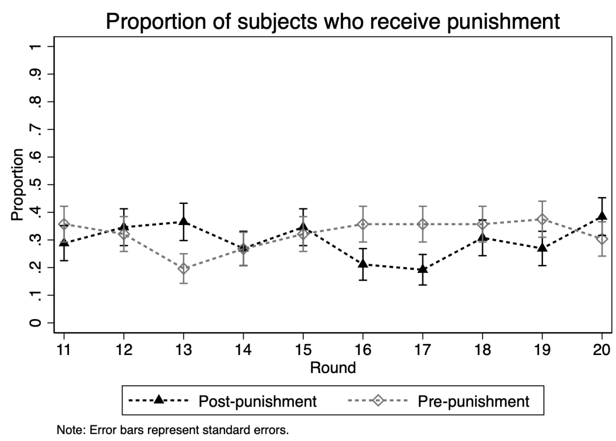


Figure 5: Proportion of subjects who receive punishment

Figure 5 shows that the proportions of subjects being punished are comparable between Pre- and Post-Punishment conditions. Pooling all rounds with punishment opportunities, the proportion of subject s who received punishment is not statistically different between the two punishment rules (p=0.407). To further investigate the effect of individual contribution on receiving punishment, we run regressions using the following model specification, with the dependent variable $BeingPunished_{it}$ as an indicator variable that equals 1 if subject i received punishment in round t. :

$$\begin{aligned} BeingPunished_{it} =& \alpha_0 + \alpha_1 Contribution_{it} + \alpha_2 PrePunishment_i \\ & + \alpha_3 Contribution_{it} \times PrePunishment_i + Controls_{it} + \varepsilon_{it} \end{aligned} \tag{6}$$

Table A1 reports the regression results. First, we find that in both treatment conditions, the probability of being punished significantly decreases by about 5 percentage points for every unit of increase in a subject's contribution to the public good. This suggests that high contributors are less likely to receive punishment compared with low contributors. And when we interact the contribution level with the Pre-Punishment-rule indicator, the coefficient estimated in columns (3) and (6) suggest that there is not significant difference in how the contribution level affects the probability of being punished between the Post- and the Pre-Punishment rules. [5]

Result 5. *Both the Post- and Pre-Punishment rules follow the same pattern, where subjects who contribute less have a higher probability of being punished, and the decrease in punishment probability per token increase in contribution is equal between the two rules. However, overall, subjects in the Pre-Punishment condition are more likely to receive punishment, regardless of their contribution level.*

Given that subjects who contribute less are more likely to receive punishment, the next question of interest is whether receiving punishment increases subjects' contributions in the next round, and whether there exists a difference between the two punishment rules. To address this question, we regress the change in contributions in round t compared to round $t-1$, i.e.,

[5] Interestingly, we also find that the Pre-Punishment condition is associated with a significantly positive coefficient in column (3), indicating that given the same contribution level, subjects under the Pre-Punishment rule are naturally more likely to receive punishment. This effect is robust when adding individual controls in column (6). One explanation for this finding is the difference in the timing of punishment. Under the Post-Punishment condition, subjects are making the decisions of whether to punish others, while under the Pre-Punishment condition, punishments are sent automatically. Therefore, the costs from punishment may be more salient under the Post-Punishment rule, making subjects more hesitant to impose punishment. In comparison, such hesitancy is absent in the Pre-Punishment condition because subjects are not manually sending out punishment.

$\Delta Contribution_{it} = Contribution_{it} - Contribution_{it-1}$, on punishment points subject i received in round $t-1$. In this equation, $Others'AverageContribution_{it-1}$ is the average contribution of the three group members of subject i in round $t-1$.

$$\begin{aligned}\Delta Contribution_{it} =& \alpha_0 + \alpha_1 Others'AverageContribution_{it-1} + \alpha_2 PunishmentReceived_{it-1} \\ &+ \alpha_3 PunishmentReceived_{it-1} \times PrePunishment_i + Controls_{it} + \varepsilon_{it}\end{aligned} \tag{7}$$

We report our estimates from the regression analysis in Table A2. Column (1) reports estimates from the full sample. On average, receiving one more point of punishment in the previous round increases subject's contribution in the current round by around 0.7 tokens. We find the interaction of Pre-Punishment indicator and punishment point received is marginal (0.137) and not statistically significant (p >0.1), which implies that there is lack of a treatment differences between the Pre- and the Post-Punishment rules in increasing contribution through receiving punishment. In Column (4) after adding individual controls, the findings are still robust.

In columns (2) and column (3), we divide the full sample into those whose contribution is strictly lower than the average of the other group members (low contributors) and is strictly above the average (high contributors) and report estimates from each sub-sample respectively. From Column (2) we find that low contributors significantly increase their contribution in the current round by 0.46 tokens if they received punishment in the previous round. In contrast, receiving punishment significantly decreases high contributor's subsequent contribution by 0.58 tokens. This decrease indicates that high contributors are discouraged from further cooperation when receiving punishment. However, we also find that in Column (3), the interaction between the Pre-Punishment rule and the punishment points received is significantly positive (1.056), which indicates that the detrimental effect of punishment on high contributors is lower and even eliminated under the Pre-Punishment rule. One explanation is the fact that under the Pre-Punishment rule, subjects are less inclined to use punishment as a means of retaliating against high contributors. As a result, high contributors do not view the punishment as adversarial, and are not dissuaded from cooperating.

Result 6. *(1) For subjects who contribute equals to or below their group average, receiving punishment significantly increases their subsequent contribution, and there is no significant difference between these two punishment rules in this positive impact. (2) For subjects who contribute above their group average, receiving punishment decreases their subsequent contributions under the Post-Punishment rule, but not under the Pre-Punishment rule.*

Finally, we are interested in the impact of receiving punishment on subjects' subsequent punishment decisions. The underlying hypothesis is that the previous experience of receiving punishment may raise subjects' hatred and encourage them to seek revenge through punishment. In order to verify this conjecture, we run the regressions using the following equation for the Post-Punishment condition:

$$\begin{aligned} PunishmentSent_{it} = &\alpha_0 + \alpha_1 Others' AverageContribution_{it} + \alpha_2 BeingPunished_{it-1} \\ &+ \alpha_3 AbsoluteNegativeDeviation_{it} \\ &+ \alpha_4 BeingPunished_{it-1} \times AbsoluteNegativeDeviation_{it} \\ &+ \alpha_5 PositiveDeviation_{it} + \alpha_6 BeingPunished_{it-1} \times PositiveDeviation_{it} \\ &+ Controls_{it} + \varepsilon_{it} \end{aligned} \tag{8}$$

where $AbsoluteNegativeDeviation_{it}$ is the absolute value of the deviation of subject's contribution from other three group members' average contribution conditional on the subject's contribution below this average, and $PositiveDeviation_{it}$ is the absolute deviation from group members' average conditional on the subject contributing higher than the average.

Table A3 reports the regression results, focusing on the Post-Punishment condition. The immediate finding is that when others' average contribution is lower, subjects are more likely to send out punishment, indicating a tendency to punish lower contributors. Additionally, a significant and positive coefficient for the variable $PositiveDevation$ suggests that subjects who contribute more than the group average tend to assign more punishment points, indicating a tendency to punish free riders. Furthermore, the indicator of being punished in the previous round is significant and positive, implying that being punished in the previous round is also an important driver of the current punishment decision. However, no significant impact is observed from the interaction between this indicator and any of the absolute deviations. Thus, regardless of being above or below the group average in the current round, subjects who were punished in previous rounds are more likely to administer punishment. These findings suggest that under the Post-Punishment rule, two factors drive punishment decisions: punishing free riders and retaliating against prior experiences of being punished.

Next, we examine the determinants of punishment under the Pre-Punishment rule. Due to the difference in the timing of punishment, we specify a different model specification as follows, where all variables of interest are from the previous round $t-1$:

$$\begin{aligned} PunishmentSent_{it} = &\alpha_0 + \alpha_1 Others' AverageContribution_{it-1} + \alpha_2 BeingPunished_{it-1} \\ &+ \alpha_3 AbsoluteNegativeDeviation_{it-1} \\ &+ \alpha_4 BeingPunished_{it-1} \times AbsoluteNegativeDeviation_{it-1} \\ &+ \alpha_5 PositiveDeviation_{it-1} + \alpha_6 BeingPunished_{it-1} \times PositiveDeviation_{it-1} \\ &+ Controls_{it} + \varepsilon_{it} \end{aligned} \tag{9}$$

Table A4 presents the regression results from the Pre-Punishment rule. Similar to the findings under the Post-Punishment rule, subjects who contribute higher than the average among other group members in the previous round tend to send out more punishment. However, we do not observe any significant impact of the previous group member average on the punishment assigned in the current round. Additionally, a similar pattern emerges, where subjects who are punished in the previous round assign more punishment in the current round ($\beta = 0.269$ for $BeingPunished_{t-1}$). Hence, under the Pre-Punishment rule, the subjects' punishment is also driven by the lower contributions of group members and the previous experience of being punished.

Result 7. *In both the Post-Punishment and Pre-Punishment conditions, subjects are more likely to punish when confronted with free riders and when they have previously experienced punishment.*

Table A1: Probability of being punished

	Dependent Variable: Indicator of Being Punished					
	(1)	(2)	(3)	(4)	(5)	(6)
	Post	Pre	Pooled	Post	Pre	Pooled
Contribution	-0.051***	-0.050***	-0.051***	-0.049***	-0.048***	-0.050***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Pre w/ punishment			0.226**			0.244**
			(0.100)			(0.117)
Contribution × Pre w/ punishment			0.001			0.000
			(0.004)			(0.004)
Constant	0.668***	1.089***	0.894***	0.248	1.556***	0.975***
	(0.100)	(0.110)	(0.090)	(0.381)	(0.370)	(0.256)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Total Observations	520	560	1080	470	530	1000
Number of Individuals	52	56	108	47	53	100

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

Table A2: Spillover effect

	Dependent Variable: Contribution_t - Contribution_t-1					
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low Contributions	High Contributors	All	Low Contributions	High Contributors
L.Other's Average Contribution	0.067	-0.401***	0.304***	0.075	-0.392***	0.312***
	(0.079)	(0.126)	(0.104)	(0.081)	(0.136)	(0.105)
L.Punishment Points Received	0.779***	0.462***	-0.739**	0.796***	0.456***	-0.740**
	(0.161)	(0.115)	(0.352)	(0.168)	(0.131)	(0.372)
Pre × L.Punishment Points Received	0.137	-0.016	1.095**	0.187	0.019	1.062**
	(0.213)	(0.190)	(0.509)	(0.224)	(0.215)	(0.533)
Constant	2.808***	5.776***	4.835***	1.611	0.309	-5.731
	(0.928)	(1.779)	(1.223)	(3.347)	(5.208)	(6.277)
Group FE	Yes	Yes	Yes	Yes	Yes	Yes
Round	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Total Observations	972	327	423	900	304	394
Number of Individuals	108	83	95	100	77	87

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

Table A3: Post-Punishment Condition: Determinants of Sending Punishments

	Dependent Variable: Punishment Points Assigned to Others	
	(1)	(2)
Other's average contribution	-0.060***	-0.062***
	(0.011)	(0.012)
1L.Being punished	0.188**	0.189**
	(0.092)	(0.096)
Absolute negative deviation	0.008	0.008
	(0.010)	(0.010)
1L.Being punished × Absolute negative deviation	-0.001	0.002
	(0.015)	(0.016)
Positive deviation	0.063***	0.058***
	(0.013)	(0.014)
1L.Being punished × Positive deviation	-0.012	-0.010
	(0.026)	(0.027)
Constant	0.979***	1.601
	(0.255)	(1.060)
Group FE	Yes	Yes
Round RE	Yes	Yes
Individual Controls	No	Yes
Total Observations	468	423
Number of Individuals	52	47

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

Table A4: Pre-Punishment Condition: Determinants of Sending Punishments

	Dependent Variable: Punishment Points Assigned to Others	
	(1)	(2)
L.Other's average contribution	-0.018	-0.019
	(0.012)	(0.013)
1L.Being punished	0.269**	0.221*
	(0.109)	(0.116)
L.Absolute negative deviation	-0.040	-0.021
	(0.037)	(0.038)
1L.Being punished × L.Absolute negative deviation	0.015	-0.004
	(0.038)	(0.039)
L.Positive deviation	0.033**	0.032**
	(0.014)	(0.014)
1L.Being punished × L.Positive deviation	0.069	0.080
	(0.053)	(0.054)
Constant	0.232	0.878
	(0.345)	(0.782)
Group FE	Yes	Yes
Round RE	Yes	Yes
Individual Controls	No	Yes
Total Observations	504	477
Number of Individuals	56	53

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include gender, age, ethnicity, whether subjects are from Texas, whether subjects are in economic-related majors, and SVO angle. Results are corrected for heteroskedasticity.

B Supplementary Tables

Table B1: Physiological Arousals When Viewing Group Contribution Information

	DV: Pupil Dilation	
	(1)	(2)
Pre-Punishment	0.017	0.041
	(0.034)	(0.043)
Above Group Average	0.059*	0.065**
	(0.030)	(0.031)
Pre-Punishment × Above Group Average	-0.051	-0.063
	(0.040)	(0.045)
Contribution Gap	-0.006	-0.005
	(0.006)	(0.006)
Pre-Punishment × Contribution Gap	-0.002	-0.003
	(0.008)	(0.008)
Above Group Average × Contribution Gap	0.010	0.007
	(0.008)	(0.009)
Pre-Punishment × Above Group Average × Contribution Gap	0.011	0.012
	(0.011)	(0.012)
Round		0.001
		(0.003)
SVO Angle		0.002
		(0.001)
Constant	-0.129***	-0.113
	(0.029)	(0.303)
Individual Controls	No	Yes
Total Observations	578	549
Individuals	72	68

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include group dummies, gender, age, ethnicity, whether subjects are from Texas, and whether subjects are in economic-related majors. Results are corrected for heteroskedasticity.

Notes: Pupil Dilation is measured within the scene when subjects were viewing the contribution of each group member.

Table B2: Physiological Arousals when Receiving Punishment

	DV: Pupil Dilation	
	(1)	(2)
Punishment Points Received	0.006*	0.010***
	(0.003)	(0.003)
Pre-Punishment	0.076***	0.099***
	(0.019)	(0.033)
Pre-Punishment × Punishment Points Received	-0.006	-0.009**
	(0.004)	(0.005)
SVO Angle		-0.000
		(0.000)
Constant	-0.062***	0.692***
	(0.018)	(0.185)
Individual Controls	No	Yes
Total Observations	600	569
Individuals	73	69

Notes: Standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Individual controls include group dummies, gender, age, ethnicity, whether subjects are from Texas, and whether subjects are in economic-related majors. Results are corrected for heteroskedasticity.

Notes: Pupil Dilation is measured within the scene when subjects were viewing the punishment feedback scene, where they learned the punishment point that they sent and they received in that round.