Collectivism Meets Individualism: The Option to Punish in Experiments of Public Goods Provision

Raine Jones*

December 16, 2022

I. Introduction

Experts in economics, politics, sociology, and a range of other disciplines have grappled with the matter of public goods experiments for years. From these studies, the dominant strategy of free-riding, meaning benefitting from a public good provision without contributing, emerges. One method that combats free-riding is punishment. Will the addition of punishment to the standard public goods game augment collective action and the size of the public good? The threat of punishment prompts players to increase their contribution, thereby reducing free-riding. However, if punishment is costly, rational players do not punish. Therefore, theoretical models indicate that the standard game and punishment version both produce the equilibrium action of not contributing. In spite of this, the majority of data from public goods games with punishment reflects that punishment maintains higher contribution levels than no-punishment. This study examines whether players will punish free-riders, despite punishment not being an optimal strategy.

This study fits into a wealth of existing research concerning public goods provision. Moreover, public goods provision situations are quite prevalent in the real world. Government services, such as national defense, public health services, public education, and access to necessities, often tax-funded, are examples of public goods provision. International relations between countries illustrate punishment and cooperative behavior in public goods provision. For instance, alliances between nations, such as NATO or the UN, provide protection and security benefits to members, who must contribute to the alliance to be a member. If a country breaks the alliance, it faces consequences. This example illustrates how sanctions can promote collective action.

This analysis investigates three versions of a public goods game – a standard version, a low-cost punishment version, and a high-cost punishment version. Section II, the literature review, summarizes and synthesizes relevant existing studies. Section III lays out the setup and particulars of this experiment.

* Thank you to my group members Cam, Collier, and Emily.

Section IV details the theoretical equilibrium solutions and Section V examines the results. Lastly, Section VI summarizes the game, presents the implications of the findings, and posits areas for improvement in further research.

II. Literature Review

The common consensus in existing research is that the theoretical equilibrium for games with and without punishment leads to the same result – all players contribute nothing and refrain from punishing when given the option. However, actual experiments show that players are willing to punish free-riders, despite punishment not being a rational action. A common finding is that when punishment occurs, the less a player contributes, the more punishment they typically receive relative to others. Some studies take this finding into consideration and redefine rational players to include a "fairness" measure. Other frequent modifications include reward and communication opportunities.

Fehr and Gächter (2000) examine a game of 24 participants split into groups of 4 where each player has an endowment of 20 tokens. The marginal payoff of the public good remains fixed at 0.4 and players can give up to 10 punishment points per round. Fehr and Gächter (2000) predict that if players are selfish, optimal behavior is the same for punishment and non-punishment games, as selfish, rational players do not punish. They find that free-riding is a dominant strategy in non-punishment games, but that punishment games can maintain higher contribution rates. The study asserts that participants are willing to punish free-riding, despite punishment not being optimal.

Fehr and Schmidt (1999) incorporate the idea that "fairness motives affect the behavior of many people."[1] This study assimilates fairness into the definition of rationality, describing fairness as "self-centered inequity aversion." Fehr and Schmidt (1999) acknowledge the theoretical equilibrium that "if selfishness and rationality are common knowledge," punishment is irrelevant because rational players will not punish.[2] Yet, their trials show that players cooperate and contribute when punishment exists. They incorporate this finding in their equilibrium solutions, showing that full cooperation is an equilibrium "if

[1] Ernst Fehr and Klaus Schmidt (1999), 817.

[2] Ernst Fehr and Klaus Schmidt (1999), 837.

there is a group of n' 'conditionally cooperative enforcers.''[3] Similarly, Rabin (1993) defines a kindness function that factors beliefs into payoffs. Rabin (1993) resolves that the lower the cost of punishment, the more people are willing to reward contributors and punish free-riders at their own expense.

Reward is a common addition to the standard game. Like Rabin (1993), Adreoni, Harbaugh, and Vesterlund (2003) conduct public goods experiments with reward incentives and find that higher contributions typically receive less punishment and more reward. The study notes that "rewards alone are relatively ineffective" in increasing contributions, but combining reward and punishment is effective.[4] Another condition that typically increases contributions is communication.[5] Dawes, McTavish, and Shaklee (1977) verify this, presenting data in which payoffs increase from 31% to 72% with relevant communication.

III. Description of the Game

15 participants split into 3 groups of 5 players each and an experimenter monitors each group. Participants log their contributions via Google Sheets. While participants know who is in their group, participants cannot match the player on the spreadsheet to the person, as each participant has a separate tab in the spreadsheet. This permits a degree of anonymity. Participants play the standard game first and are not yet aware there is a second game. Players begin with an endowment of 100 theoretical dollars. Each round, players decide how much of their endowment to allocate to the public good pot. Contributions can range from none to the entirety of their endowment. Players simultaneously input their contributions. The total of all contributions is multiplied by a factor of 1.5 and then evenly distributed to all players, no matter whether a player did or did not contribute. A player's payoff equals their remaining endowment plus their slice of the public good. That payoff becomes that player's endowment for the next round. Players can see the total contribution amount and the public good benefit in their spreadsheet tab. The game's length is 4 rounds.

[3] Fehr and Schmidt (1999), 841.
[4] Adreoni, Harbaugh and Vesterlund (2003), 901.
[5] Ledyard (1995), 141.

Once the game ends, the experimenter announces that another game with punishment opportunities will ensue. The setup begins the same as the standard game. However, once all players input their contributions, everyone can see each player's contribution. After players have time to process this information, they can assign dollar punishment amounts to each player. For rounds 1-2, giving $1 of punishment costs a player $0.50. A player's total punishment dollars given cannot exceed their endowment. A player's payoff equals their remaining endowment, minus their contribution, plus their slice of the public good, minus their total punishment received, and minus their total punishment dollars given times 0.5. In rounds 3-4, there is a one-for-one cost to punish, meaning $1 of punishment costs a player $1. This study analyzes data from one of the three aforementioned groups.[6]

IV. *Theoretical Equilibria*

Consider the first game, the standard version with no punishment:

Players: $n \geq 2$ (n = 5 in this study)
Actions: Contribute x_i amount of dollars, $x_i \in [0, 100]$
Preferences: Maximize individual payoff, $U_i = e_i - x_i + \frac{1}{n}(1.5)\sum_{i=1}^{n} x_i$
e_i - initial individual endowment
x_i - individual contribution
n - number of players in the group

If nobody contributes, all players' payoffs are $100 each after round 1. If everyone contributes their entire endowment, all players' payoff are $150 each after round 1. All players have a higher payoff when everyone contributes their entire endowment than when nobody contributes. However, a player can increase their individual payoff by contributing less. For example, if the other players each contribute $100, player 5's payoff is $150 if she contributes $100. If she contributes nothing, her payoff is $220. The work below shows the full free-riding equilibrium:

Either contribute $0, case (2), or contribute a > $0, case (1).

(1) $U_i = e_i - a + \frac{1}{n}(1.5)(\sum_{i=1}^{n} x_i + a)$

(2) $U_i = e_i + \frac{1}{n}(1.5)\sum_{i=1}^{n} x_i$

[6] I analyze data for the group for which I was the experimenter.

Set (2) ≥ (1)

$$e_i + \frac{1}{n}(1.5 * \sum_{i=1}^{n} x_i) \geq e_i - a + \frac{1}{n}(1.5)(\sum_{i=1}^{n} x_i + a)$$

Subtract $e_i + \frac{1}{n}(1.5) \sum_{i=1}^{n} x_i$ from both sides to obtain the following inequality:

$$0 \geq \frac{1}{n}(1.5)(a) - a$$

Since n ≥ 2, $\frac{1}{n}(1.5)(a) - a$ is always negative, contributing $0 always yields a higher payoff than contributing a positive amount. Player$_i$ is better off not contributing than contributing; not contributing is a Nash equilibrium.

Consider the second game, the version with low-cost punishment in rounds 1-2 and high-cost punishment in rounds 3-4:

Rounds 1-2, low-cost to punish:
Players: n ≥ 2 (n = 5 in this study)
Actions: Contribute x_i amount of dollars, $x_i \in$ [0, 100]
Assign punishment to each player, $p_i \in [0,\ 2e_i]$ (p_i is the total amount of punishment dollars given)
Preferences: Maximize individual payoff where $U_i = e_i - x_i + \frac{1}{n}(1.5) \sum_{i=1}^{n} x_i - 0.5p_i - r_i$

e_i - initial individual endowment
x_i - individual contribution
n - number of players in the group
p_i - an individual's total punishment given
r_i - an individual's total punishment received

Punishing always results in a lower payoff; therefore, it is not part of a Nash equilibrium. It is common knowledge that no rational player will punish and that no rational player will contribute. Following similar logic to the first game, the work below depicts the Nash equilibrium of not contributing and not punishing:

Either contribute $0, case (2), or contribute a > $0, case (1). Do not punish, p_i = $0.

(1) $U_i = e_i - a + \frac{1}{n}(1.5)(\sum_{i=1}^{n} x_i + a) - r_i$

(2) $U_i = e_i + \frac{1}{n}(1.5) \sum_{i=1}^{n} x_i - r_i$

Set (2) ≥ (1)

$$e_i + \frac{1}{n}(1.5) \sum_{i=1}^{n} x_i - r_i \geq e_i - a + \frac{1}{n}(1.5)(\sum_{i=1}^{n} x_i + a) - r_i$$

Subtract $e_i + \frac{1}{n}(1.5)\sum_{i=1}^{n} x_i - r_i$ from both sides to obtain the following inequality:

$0 \geq \frac{1}{n}(1.5)(a) - a$

Since n ≥ 2, $\frac{1}{n}(1.5)(a) - a$ is always negative, contributing $0 yields a higher payoff than contributing a positive amount and not punishing gives a higher payoff than punishing. Player$_i$ is better off not contributing and not punishing than contributing and punishing, and this set of actions forms a Nash equilibrium.

Rounds 3-4, high-cost to punish:

Players:	n ≥ 2 (n = 5 in this study)
Actions:	Contribute x_i amount of dollars, $x_i \in [0, 100]$ Assign punishment to each player, $p_i \in [0, e_i]$ (p_i is the total amount of punishment dollars given)

Preferences: Maximize individual payoff where $U_i = e_i - x_i + \frac{1}{n}(1.5)\sum_{i=1}^{n} x_i - p_i - r_i$

e_i - initial individual endowment

x_i - individual contribution

n - number of players in the group

p_i - an individual’s total punishment given

r_i - an individual’s total punishment received

Similar reasoning finds that high and low-cost cases produce the same equilibrium.

V. Results and Analyses

Table 1: No Punishment Game (n = 5 players)

Round	Public Good Benefit	Average Contribution	Average Endowment
1	$30.30	$20.20	$100.00
2	$18.30	$12.20	$110.10
3	$3.00	$2.00	$116.20
4	$0.30	$0.20	$117.20

Table 2: Punishment Game (n = 5 players)

Round	Public Good Benefit	Average Contribution	Average Endowment	Average Punishment Given/Received
Low Cost to Punish				
1	$87.00	$58.00	$100.00	$15.00
2	$64.95	$43.30	$106.50	$10.00
High Cost to Punish				
3	$37.50	$25.00	$113.15	$2.00
4	$12.00	$8.00	$121.65	$8.00

Figure I
Round 1: Punishment Received vs. Contribution Amount for Each Player

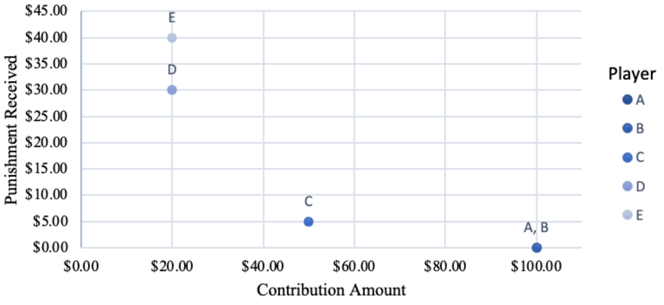


Figure II
Round 2: Punishment Received vs. Contribution Amount for Each Player

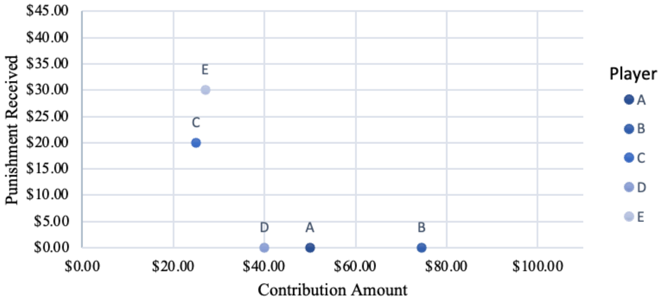


Figure III
Round 3: Punishment Received vs. Contribution Amount for Each Player

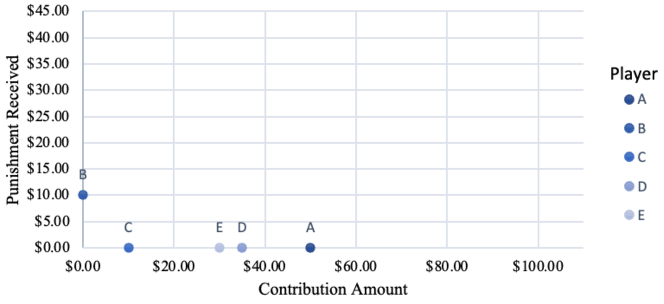


Figure IV
Round 4: Punishment Received vs. Contribution Amount for Each Player

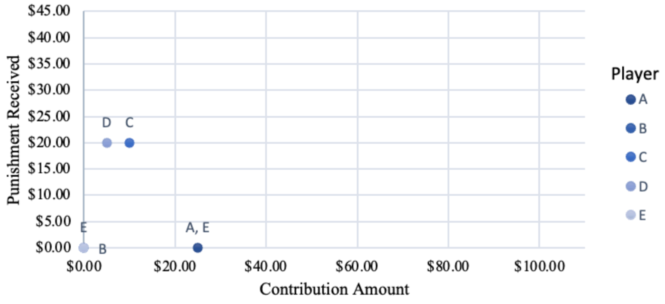


Figure V
No Punishment Game: Total Contributions Across Rounds

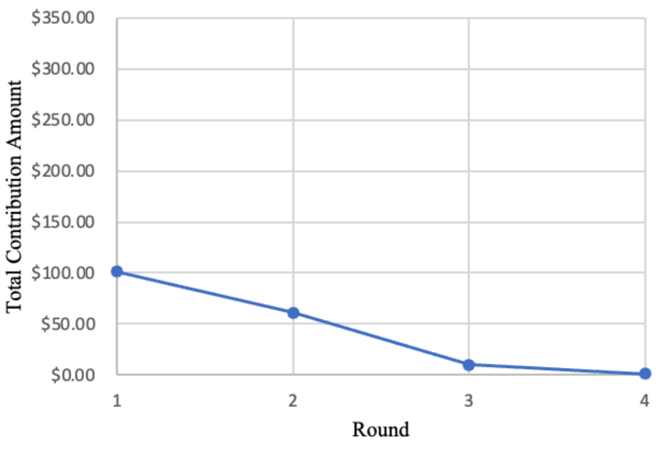
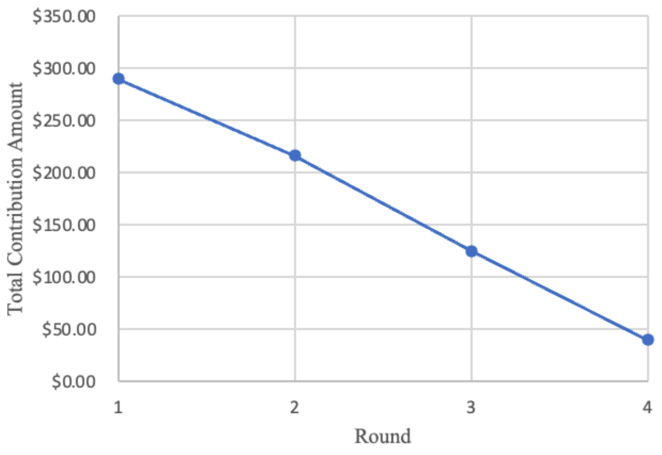


Figure VI
Punishment Game: Total Contributions Across Rounds



*note that punishment conditions change between round 2 and 3

As the theoretical equilibria and existing studies predict, in the no-punishment game, contributions converge to a low level. Figure VI displays the downward trend in total contribution across rounds, and Table 1 shows the fall in average contribution from $20.20 to $0.20. In the punishment version, the threshold of average and total contributions is higher relative to the no-punishment case. The scaling of the axes in Figures V and VI is the same to accentuate this difference. Contributions converge to a low level in the punishment game, though not as low as no-punishment levels. These results contrast the 50% to 90% levels of endowment contributions in Fehr and Gächter (2000) and the high level in Andreoni, Harbaugh, and Vesterlund (2003). This contribution data bears a closer resemblance to the theoretical equilibrium prediction.

The punishment data resembles the theoretically optimal strategy of not punishing. Figures I-IV plot the punishment each player receives against the amount they contribute for each round. Figures I-IV have axes with the same scale, highlighting that, in general, punishments decrease over rounds. Unlike players in Rabin (1993) and Fehr and Schmidt (1999) whose behaviors emphasize fairness, players in this experiment trend more towards rational, selfish behavior. This data does not demonstrate punishment fully eliminating free-riding, but it displays a link between the threat of punishment and increased contributions. When punishment occurs, Figures I-IV indicate that low contributions typically incur high punishments.[7] This concurs with the finding from Fehr and Gächter (2000) that, "The more an individual negatively deviates from the contributions of the other group members, the heavier the punishment."[8]

One detail to note is the cost of punishing changes between rounds 2 and 3, but the game does not restart. The cost of punishment impacts a player's decision about not only how much to punish others, but also how much to contribute in advance of punishment. If a player knows punishment is costly, they may predict that this will dissuade others from dispensing punishment dollars. Therefore, a player may focus less on the threat of punishment in their choice of contribution amount at the start of each round. As punishing cost increases after round 2, players take into account the heightened cost in their punishing

[7] Figure IV is an outlier oof this tendency, which this paper addresses in section VI.
[8] Fehr and Gächter (2000), 993.

and contributing decisions. This may partly explain why the downward-sloping trend of total contributions steepens after round 2 in Figure VI. Figures I-IV find that, along with contributions, punishments typically decrease across rounds. The one-for-one punishment cost may partly explain this finding.

VI. Conclusion

While low contributions typically incur higher punishments, Figure IV, which analyzes the final round, departs from this tendency, as two players who contributed received substantial punishment. This aberration questions whether the number of rounds being finite and common knowledge impacts players' actions. Ledyard (1995) touches on this phenomenon, noting, "Towards the last iteration, the rational players will not contribute."[9] Ledyard debates if strategy or learning causes this and calls for further research.

Once this classroom experiment ended, student participants talked casually, indicating that, in the non-punishment game, they recognized their contribution behavior in the final round would not have future implications. They brought similar reasoning to the punishment game, alluding that they did not see much point to punishing in round 4, as there would be no future behavior for that punishment to impact. However, not all players had this mindset. The full spreadsheet shows that player A used a substantial amount of their endowment to punish players C and D $20 each in round 4. This is curious because, unlike other players, players C and D contributed in round 4. Player 4 indicated, after the game concluded, that this punishment was random and just to spite other players. Other participants remarked that they may have treated their money differently had their endowment been actual and not theoretical dollars.

Classic economic models of selfishness and rationality predict noncooperative behavior, i.e., nobody contributes, in punishment and no-punishment cases. This experiment finds that in the no-punishment scenario, contributions fall toward $0. With punishment, contributions converge toward a low level, though not as low as the no-punishment level. Punishments decline across rounds, but never

[9] Ledyard (1995), 148.

fully vanish, suggesting that players will punish despite this not being a payoff-maximizing action. One suggestion for improvement is to increase the number of rounds of each version of the game. Four rounds may not be enough for a stable trend to arise. Additionally, restarting the game before the punishment cost increases may allow a clearer comparison between low and high-cost punishments. This data also suggests that researchers could consider how to deal with anomalies in the final round of public goods games where the number of finite rounds is common knowledge. Lastly, studies could examine societal and environmental motives behind players' decision to deviate from selfish behavior and frequently choose to punish in public goods experiments.

References:

Andreoni, James, William Harbaugh, and Lise Vesterlund. 2003. “The Carrot or the Stick: Rewards, Punishments, and Cooperation.” *American Economic Review* 93(3): 893–902.

Dawes, Robyn, Jeanne McTavish, Harriet Shaklee. 1977. “Behavior, Communication, and Assumptions about Other People’s Behavior in a Commons Dilemma Situation.” *Journal of Personality and Social Psychology* 35(1): 1-11.

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *The Quarterly Journal of Economics* 114(3): 817-868.

Fehr, Ernst and Simon Gächter. 2000. “Cooperation and Punishment in Public Goods Experiments.” *The American Economic Review* 90(4): 980-994.

Marwell, Gerald and Ruth E. Ames. 1979. "Experiments on the Provision of Public Goods I: Resources, Interests, Group Size, and the Free Rider Problem." *American Journal of Sociology* 84(4): 926-937.

Rabin, Matthew. 1993. “Incorporating Fairness into Game Theory and Economics.” *American Economic Review* 83(5): 281–302.