Firing the right bullets: Exploring the effectiveness of the hired-gun mechanism in the provision of public goods

LIU, Jia <http://orcid.org/0000-0001-9568-0317>, RIYANTO, Yohanes E. <http://orcid.org/0000-0001-7343-4043> and ZHANG, Ruike

Firing the Right Bullets: Exploring the Effectiveness of the Hired-Gun Mechanism in the Provision of Public Goods*

Jia Liu† Yohanes E. Riyanto‡ Ruike Zhang§

Abstract

We explore the robustness of the "hired-gun" mechanism proposed by Andreoni and Gee (2012)—a centralized punishment mechanism to promote the collective provision of public goods. In order to avoid the race to the bottom, the hired-gun mechanism relies on the use of the *unilateral* and *tie* punishment imposed on the lowest contributor(s). We examine the effectiveness of the hired-gun mechanism under varying sizes of the unilateral and tie punishment by theoretically deriving and experimentally testing a range of punishment parameters that would lead to full contribution. We show that, to some extent, the effectiveness of the mechanism depends on the size of both types of punishments. In particular, the lack of unilateral punishment renders the mechanism less effective.

Keywords: Public goods provision; Hired-gun mechanism; Unilateral punishment; Tie punishment; Lab experiment

JEL Classification: C63, C72, C92, D7, D83, H41

---

* We thank James Andreoni, Jasmina Arifovic, Tim Cason, Gary Charness, and the seminar participants at Nanyang Technological University, the World Meeting of the Economic Science Association in San Diego, and the ESA European Meeting in Vienna. Riyanto would also like to acknowledge the AcRF MOE Tier 1 Research grants awarded by Nanyang Technological University (NTU). The funding source had no involvement in the preparation of this article. We wish to confirm that no known conflicts of interest are associated with this article.

†(Corresponding author) The Business School, Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, United Kingdom, tel: +44(0)1142254450, email: jia.liu@shu.ac.uk .

‡ Division of Economics, Nanyang Technological University, 14 Nanyang Drive HSS-04-83, Singapore 637332, tel: +65-65921578, email: yeriyanto@ntu.edu.sg.

§ Division of Economics, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332, email: RZHANG005@e.ntu.edu.sg.

1 Introduction

Law enforcement agencies often devote resources to prosecuting the most heinous crimes (e.g., murder) instead of misdemeanors. For instance, not every car on the road exceeding the speed limit gets a speeding ticket. Instead, more commonly, a ticket is issued to the fastest car (i.e., the largest deviator) going over the speed limit. The second largest deviator, therefore, dodges the sanction. Intuitively, the competition of not being the largest deviator should eventually drive out any deviations and lead to a socially optimal outcome.

Following this line of reasoning, Andreoni and Gee (2012) proposed the “hired-gun” mechanism to overcome the free-rider problem in a voluntary contribution setting. The hired-gun mechanism is a centralized and automatic mechanism that punishes the lowest contributor to the extent that the contributor would rather have been the second lowest. The hired-gun mechanism essentially works through imposing a *unilateral punishment* and a *tie punishment*, which would transform the public goods game from the one that has a unique free-riding equilibrium into one that has a unique full-contribution equilibrium. Andreoni and Gee (2012, 2015) show that the hired-gun mechanism is effective in promoting contribution.

The original hired-gun mechanism in Andreoni and Gee’s (2012) paper features a specific set of punishment parameters (heavy punishment) in a setting with a relatively small range for the decision space (6 possible contribution levels in total). We aim to test the robustness of the mechanism by substantially expanding the decision space (to 21 possible contribution levels) and by varying the size of the unilateral punishment and tie punishment. We formalize the idea that the effectiveness of the mechanism is sensitive to the size of both punishment components. Our generalized model shows that the magnitude of unilateral and tie punishment can be softened to a certain degree while the full-contribution equilibrium remains intact. The generalized mechanisms were also tested experimentally. Our experimental results show that the effect of the mechanism is, to some extent, sensitive to the size of both punishment components. Results from treatments with positive unilateral punishment are consistent with the theory. However, removing the unilateral punishment renders the mechanism ineffective regardless of the size of the tie punishment, which contradicts the theory. A decaying trend of contributions exists in such treatments. Possible explanations for the discrepancy are explored.

We contribute to the literature on centralized punishment mechanisms in the following ways: First, to the best of our knowledge, our study is the first to test the robustness of the hired-gun mechanism—an important centralized punishment mechanism in the literature. Its desirability stems from having relatively low-cost implementation and empirically demonstrated effectiveness. We explore whether such effectiveness is robust with varying sizes of punishment components.

Second, varying sizes of punishment components allow us to examine what the appropriate size of punishment should be to achieve socially optimal outcomes—a question of paramount importance in the implementation stage but one understudied in the literature. Studying the appropriate size of punishment broadly echoes a long-established recognition in criminal law and policy that enforcing the right amount of punishment is crucial for encouraging obedience to law and minimizing the social cost associated with the administration of such punishments (Becker, 1968). Empirical evidence indicates that more severe punishments (i.e., longer prison sentences), which require more societal resources to implement, often fail to reduce crime rates as desired (Darley, 2005). As a result, how to design a sanction policy that effectively deters offenders and minimizes implementation costs bears significance from the social welfare perspective. Durlauf and Nagin (2011) have championed such policies: "Sanction policies that reduce both crime and punishment have the desirable feature of avoiding not only the costs of crime but also the costs of administering punishment" (p. 13). These simultaneous effects are also recognized by Kleiman (2009) and Kennedy (2009). Our study takes a step toward the same direction by exploring the right size of punishment in the centralized punishment mechanism.

Third, we go beyond specifying a limited number of punishment conditions (e.g., the no, mild, and severe sanction conditions in Tyran & Feld, 2006). We generalize the hired-gun mechanism and theoretically derive a class of punishment parameters that would lead to the full-contribution equilibrium. This approach highlights the possibility of using less punishment to achieve the same full-contribution outcome.

The free-rider problem in public goods provision has received the lion's share of attention in the literature. The public goods provision is a classic social dilemma where collective interests are at odds with individual interests, leading to a non-socially optimal outcome. There is a burgeoning literature on punishment mechanisms for improving cooperation in public goods provision (see Chaudhuri, 2011 for an excellent survey). Depending on the administration body of punishment, there are informal and formal punishments. Informal punishment, which is also commonly referred to as peer-to-peer punishment, is usually administrated by fellow group members. Though it is costly to exercise such punishments, people are willing to punish free-riders, which results in higher contribution levels (e.g., Fehr & Gächter, 2000, 2002; Cason & Gangadharan, 2015). Despite its effectiveness, the informal punishment mechanism suffers from some drawbacks. The punishment exerted by peers may trigger revenge or anti-social punishment (Denant-Boemont, Masclet, & Noussair, 2007; Herrmann, Thöni, & Gächter, 2008; Nikiforakis, 2008). This outcome is commonplace in reality as well. For example, there are laws in place to protect whistleblowers from retaliation (e.g., Whistleblower Protection Act of 1989 in the United States).[1] In addition to revengeful punishment, there could be second-order free-riding problems (Panchanathan & Boyd, 2004; Fowler, 2005; Gross, Méder, Okamoto-Barth, & Riedl, 2016). As punishment is only costly to the punisher and benefits the whole group, individuals could free ride on others who exercise punishments. Moreover, people with certain non-standard preferences may enjoy punishing others, and such joy could lead to over-punishment (Casari & Luini, 2009). Over-punishment exacerbates as the number of bystanders increases (Kamei, 2018). Further, since the presence of punishment does not necessarily improve group payoff (Fehr & Gächter, 2000), over-punishment could lead to substantial efficiency loss (Kamei, 2018). Despite the drawbacks, however, the merits of peer punishment are obvious. It is autonomous and has a relatively low implementation cost.

Formal punishment or centralized punishment mechanisms provide an alternative to solving free-rider problems in social dilemmas. The punishment is often implemented by central authorities according to pre-set rules, which prevents the anti-social punishment, over-punishment, and second-order free-rider problems that exist in informal punishment systems. Moreover, the concentration of power in a centralized body is a hallmark of civilization (Mann, 1986). However, setting up a centralized punishment system (i.e., court system, police force) is usually substantially costly. This cost could be one of the reasons that centralized punishment systems have received much less attention than informal punishment systems in the literature. The hired-gun mechanism falls into the centralized punishment category and is relatively low cost. Within this category, the autonomous nature of how the punishment amount is determined mitigates the potential efficiency loss caused by over-punishment when punishment is administrated by peers. In addition, the hired-

[1] See https://www.congress.gov/bill/101st-congress/senate-bill/20/text.

gun mechanism is not as costly to implement as other centralized punishment mechanisms. The mechanism does not require the enforcer to document all norm violators and instead only the largest and the second-largest violator. Thus, the hired-gun mechanism reduces the cost of implementation.[2]

Experimental evidence on centralized punishment systems is abundant. Putterman, Tyran, and Kamei (2011) ask subjects to vote on formal punishment systems and find that people mostly learn to choose the system that helps to resolve the free-rider problem. Markussen, Putterman, and Tyran (2014) and Kamei, Putterman, and Tyran (2015) explore the choice between informal and formal sanction systems, finding that people prefer informal sanctions if implementing formal sanctions is costly. The abovementioned formal sanctions are meted out to all deviators and thus are absolute punishment systems (see also Falkinger, Fehr, Gachter, & Winter-Ebmer, 2000). The relative punishment system instead only punishes the largest deviator. Yamagishi (1986) and Andreoni and Gee (2012) propose formal punishment systems that only target the lowest contributor (see also Xiao & Houser, 2011; DeAngelo & Gee, 2017). They show that these mechanisms are successful in mitigating free-rider problems. Kamijo et al. (2014) theoretically and experimentally compare the use of absolute and relative punishment systems. They find that the relative punishment system results in equal or higher contributions than those in absolute punishment systems. The typical centralized mechanism, including the hired-gun mechanism, usually has an exogenously assigned central authority to carry out the sanctions. Notably, research has shown that the legitimacy of authority affects cooperation; people are more cooperative if the authority is endogenously elected (Tyran & Feld, 2006; Baldassarri & Grossman, 2011; Grossman & Baldassarri, 2012;).

A key question in relation to the formal punishment mechanism is how to define the right size of punishment, which had not yet been studied for the hired-gun mechanism. Previous studies show that a specific set of punishment parameters works. We test its robustness by varying the sizes of both punishment components. This approach adds to the strand of literature exploring the appropriate size of punishment. If the exogenously imposed punishment is too lenient, it fails to

[2] In the speeding ticket example, if there are a number of cars exceeding the speed limit, the police could focus on the most reckless drivers instead of documenting the speed of every car.

improve cooperation (Tyran & Feld, 2006).[3] McFatter (1982) indicates that the medium-size punishment is optimal in achieving the goal of punishment whether it is retributive or rehabilitative and that too-lenient or too-harsh punishments are suboptimal in this regard. In the literature, various approaches have been adopted in deciding the size of punishment. The punishment in Yamagishi (1986) depends on the punishment fund raised by the group. In Andreoni and Gee (2012), the punishment meted out on the lowest contributor depends on the difference between the lowest and the second-lowest contributions. Kamijo et al. (2014) use a fixed penalty in centralized punishment systems. In addition to punishment size considerations, a good centralized sanction system should be low cost to implement as a low-cost and deterrent sanction system is preferred by voters (Markussen et al., 2014; Kamei et al., 2015). The hired-gun mechanism is a relatively low-cost centralized mechanism. We generalize the hired-gun mechanism and find that it is possible to soften punishment but still achieve the full-contribution equilibrium. Softening punishments without, in principle, compromising deterrence is welfare-enhancing from the social welfare perspective.

The rest of this paper proceeds as follows: Section 2 details our generalized model for the hired-gun mechanism, which is followed by the experimental design and procedures in section 3. Section 4 discusses the results, and section 5 concludes the paper.

2 Theoretical background

We start by introducing the hired-gun mechanism proposed by Andreoni and Gee (2012), followed by our generalization of the model. The punishment in the hired-gun mechanism consists of a unilateral punishment, which aims to discourage people from being the lowest contributor, and a tie punishment, which aims to prevent people from coordinating a tie at a below-full-contribution level. The two punishment components, in principle, have different roles. We are interested in whether and to what extent these two punishment components are connected and whether it is possible to achieve full contribution with less punishment. Should the answer be yes,

[3] Tyran and Feld (2006) also find that mild sanctions substantially improve cooperation if they are endogenously chosen. This aligns with findings in Baldassarri and Grossman (2011) and Grossman and Baldassarri (2012). As suggested by evidence in the literature, the endogeneity nature of the mechanism has significant implications on compliance. As a result, the optimal size of punishment might be hugely different from what is required in the exogenously imposed environment. Though exploring the optimal size of punishment in an endogeously chosen environment is interesting, it is beyond the scope of our paper. It would be a promising direction for future research. In this paper, we only focous on the exogenously imposed mechanism.

there might exist a class of such mechanisms that are effective in promoting contribution and wherein punishment is minimized to reduce potential welfare loss. We generalize the hired-gun mechanism and demonstrate theoretically that such a possibility exists. As is the case in criminal justice legislation, imposing lower punishments while maintaining the deterrence effect is a desirable sanction policy.

2.1 The model

In the classic linear public goods game, players form a group of n, and each is endowed with w. All group members decide independently and simultaneously how much of w to contribute to the public goods. Each unit contributed to the public goods generates a payoff of α ($\frac{1}{n} < \alpha < 1$) for each group member regardless of each group member's contributed amount. α is referred to as marginal per capita return (MPCR). If player i contributes g_i $(0 \leq g_i \leq w)$, the payoff π_i can be expressed as follows:

$$\pi_i = w - g_i + \alpha \sum_{j=1}^{n} g_j \tag{1}$$

Given that $\partial\pi_i / \partial g_i = -1 + \alpha < 0$, the dominant strategy would be not to contribute at all, which leads to the zero-contribution inefficient equilibrium. Everyone would be better off if all group members contribute to the whole endowment. This result is referred to as the socially optimal outcome, which is usually hard to achieve.

The idea of the hired-gun mechanism is to punish the lowest contributor so that the person would rather have been the second-lowest contributor. Define the set of contributors as S and $L(g) \subseteq S$ as a set of contributors with the lowest contributions. The size of $L(g)$ could range from 1 to n. For example, if $g_z \leq g_j$ holds for all j, then $z \in L(g)$. Also, let g_y be the second lowest contribution. That is, $y \notin L(g)$, and if $g_y \leq g_j$ holds for all, then $j \notin L(g)$. For a player who contributes g_i when the choice vector is g of all players, the hired-gun mechanism administers the punishment $P(g_i, g)$ as follows:

$$P(g_i, g) = \begin{cases} 1, & if\ L(g) = S\ and\ \ g_i < w \\ 0, & if\ L(g) = S\ and\ \ g_i = w \\ g_y - g_i + 1, & if\ L(g) \subset S\ and\ \ i \in L(g) \\ 0, & if\ L(g) \subset S\ and\ \ i \notin L(g) \end{cases} \tag{2}$$

Equation (2) depicts the following punishment rules: 1) if all contributors are tied at a below-full-contribution level, all are punished by 1 unit; 2) if all contribute the full endowment, no one is punished; 3) if player *i* belongs to the set of the lowest contributors, then player *i*, together with the other lowest contributors, will be punished, and the punishment amount for player *i* is equal to the difference between the second-lowest contribution and the lowest contribution plus 1 unit. It is straightforward to infer from the repeated elimination of dominated strategies that full contribution is the unique equilibrium.

The punishment can be dissected into two components: the unilateral punishment and the tie punishment. The former is to discourage people from being the lowest contributor; that is, for $P(g_i, g) = g_y - g_i + 1\ if\ L(g) \subset S$ and $i \in L(g)$, the 1-unit punishment on top of the difference between the second lowest and lowest contribution is defined as the *unilateral* punishment.

The tie punishment, on the other hand, is to discourage people from settling at a below-full-contribution tie; that is, for $P(g_i, g) = 1$, if $L(g) = S$ and $g_i < w$, the 1-unit punishment is defined as the tie punishment. The original hired-gun mechanism assumes both the unilateral and the tie punishment are equal to the value of the smallest unit of the private good (i.e., 1 unit). We relax this assumption by allowing the unilateral and the tie punishment to be different and by removing the 1-unit restriction. Let u $(u \geq 0)$ be the unilateral punishment and t $(t \geq 0)$ be the tie punishment. Equation (2) could then be rewritten as follows:

$$P(g_i, g) = \begin{cases} t, & if\ L(g) = S\ and\ \ g_i < w \\ 0, & if\ L(g) = S\ and\ \ g_i = w \\ g_y - g_i + u, & if\ L(g) \subset S\ and\ \ i \in L(g) \\ 0, & if\ L(g) \subset S\ and\ \ i \notin L(g) \end{cases} \tag{3}$$

2.2 Equilibrium analysis

For this analysis, we assume that $\alpha \geq 0.5$, which includes the cases analyzed by Andreoni and Gee (2012).[4] The relative magnitude of the tie punishment (t) and the unilateral punishment (u), together with the value of MPCR (α), will determine the equilibrium outcome.

Proposition 1. *If both the tie punishment (t) and the unilateral punishment (u) are sufficiently harsh, such that* $t > 1 - \alpha$ *and* $u \geq 0$*, then the unique equilibrium of the game is characterized by everyone contributing the full endowment.*

Proof. Following Andreoni and Gee (2012), we also adopt the repeated elimination of dominated strategies to find the equilibrium. Let Δ_{-1} be the payoff change from decreasing the contribution by 1 unit and Δ_{+1} be the payoff change from increasing the contribution by 1 unit. It is informative and tractable, and without loss of generality, to start with 2 players. Since the player's payoff varies depending on the distribution of contributions in the group, we discuss the player's move separately in the different cases.

Case 1. The group ties at a below-full-contribution amount.

This is the case where $L(g) = S$ and $g_i < w$ in (3). Both players receive a punishment $P = t$. Increasing the contribution by 1 unit incurs a loss of $1 - \alpha$ based on (1) but avoids the punishment t. Decreasing the contribution by 1 unit increases the payoff by $1 - \alpha$ and avoids the punishment t, but a punishment $(1 + u)$ kicks in, as shown below:

$$\Delta_{+1} = t - (1 - \alpha) \tag{4}$$

$$\Delta_{-1} = t - (\alpha + u) \tag{5}$$

To encourage people only to move the contribution upward, it is necessary to have the following:

$$\Delta_{+1} > 0 \Rightarrow t > 1 - \alpha \tag{6}$$

$$\Delta_{+1} > \Delta_{-1} \Rightarrow u > 1 - 2\alpha \tag{7}$$

If (6) and (7) are both satisfied, players would have incentive to increase the contribution to break the tie.

[4] See Appendix A.1 for the equilibrium analysis for the case of $\alpha < 0.5$.

Case 2. The group has heterogeneous contributions (no tie).

If the difference in contribution is more than 1 unit, it is straightforward that the higher contributor will decrease the contribution and that the lower contributor will increase the contribution as doing so would improve the payoff for both before the relative position of contributions changes. Therefore, we discuss the case where the difference in contribution has been shortened to 1 unit. As $t > 1 - \alpha$ according to (6), the higher contributor does not have incentive to decrease the contribution by 1 unit to reach a tie. For the lower contributor, the change in payoff from the 1-unit-contribution increase is $\Delta_{+1\_lower} = \alpha + u - t$. If $\Delta_{+1\_lower} = \alpha + u - t < 0$, then the lower contributor is better off by staying put rather than increasing the contribution by 1 unit to reach a tie. If $\Delta_{+1\_lower} = \alpha + u - t > 0$, then the lower contributor should increase the contribution by 1 unit. Besides the option of increasing the contribution by 1 unit to reach a tie, the lower contributor could also increase the contribution by 2 units to avoid any kind of punishment. The resulting payoff change $\Delta_{+2\_lower} = 2\alpha + u - 1 > 0$ always holds according to (7). Thus, the lower contributor is better off by increasing the contribution by 2 units rather than maintaining the status quo. When the contribution is close to the full-contribution level, which prevents the lower contributor from increasing the contribution by 2 units, the lower contributor is better off by increasing the contribution by only 1 unit. As tie punishment no longer applies if the two players tie at the full-contribution level, the payoff change from a 1-unit contribution increase is $\Delta_{+1\_lower} = \alpha + u > 0$. Overall, the lower contributor always has incentive to increase the contribution until both the lower and higher contributors reach the full-contribution level. ■

To summarize, in the 2-player case, as long as $t > 1 - \alpha$ and $u > 1 - 2\alpha$, the game has the unique full-contribution equilibrium. In other words, the tie punishment and the unilateral punishment need to be harsh enough to reach the full-contribution equilibrium. As we assume $\alpha \geq 0.5, 1 - 2\alpha \leq 0$ always holds. As a result, we can rewrite $u > 1 - 2\alpha$ as $u \geq 0$. The same reasoning applies to the case with more than 2 players. It is straightforward that increasing the group size does not change the theoretical prediction.[5]

[5] See Appendix A.2 for the equilibrium analysis with group size = 4.

Proposition 2. *If the tie punishment (t) is lenient and the unilateral punishment (u) is sufficiently harsh or* $t < 1 - \alpha$ *and* $u \geq 0$*, the game becomes a coordination game.*[6]

Proof. We start with the tie case. If $t < 1 - \alpha$ and $u > 1 - 2\alpha$, then $\Delta_{-1} < \Delta_{+1} < 0$, which means deviating from the tie in either direction tends to decrease one's payoff. As we assume $\alpha \geq 0.5, 1 - 2\alpha \leq 0$ always holds. As a result, we can rewrite $u > 1 - 2\alpha$ as $u \geq 0$. Next, we discuss the no-tie situation. In a no-tie situation $\Delta_{+1\_lower} = \alpha + u - t > 1 - \alpha - t > 0$, which indicates that the lower contributor always has incentive to increase the contribution to a tie. Alternatively, the change in the relative ranking of contributions could come from the higher contributor's actions. If the higher contributor decreases the contribution by 1 unit to reach a tie, the resulting change in payoff would be $\Delta_{-1\_higher} = 1 - \alpha - t > 0$. Thus, it is in the higher contributor's best interest to decrease the contribution to reach a tie. If the decrease goes beyond what is needed to reach a tie (e.g., 2 units), the higher contributor then becomes the lower contributor, and the payoff change would be $\Delta_{-2\_higher} = 1 - 2\alpha - u < 0$. In other words, the higher contributor is better off by decreasing the contribution just enough to reach a tie. From both players' perspectives, the tied situation is beneficial, and no one has incentive to break from the tie. Players could be tied at any contribution level, which indicates there are multiple equilibria in this game. Therefore, it ends up with a coordination problem. ■

Note that the theory does not give a clear prediction on which contribution level people would coordinate on in equilibrium. That level depends on the starting contribution level; people would coordinate on a contribution level that is no lower than the starting point. The tie could be reached by either the lower contributor increasing or the higher contributor decreasing their contributions.

Figure 1 illustrates all possible cases described in the equilibrium analysis for various parameters of the unilateral punishment (u) and the tie punishment (t), with $\alpha = 2/3$. The figure also positions the 7 treatments in the game outcome map based on the size of t and u, which we will explain in detail in the next section. The dark-shaded area on the right represents the required condition for an effective hired-gun mechanism where the unique full-contribution Nash

[6] Notably, lenient unilateral punishment ($u < 1 - 2\alpha$) is impossible under the assumption of $\alpha \geq 0.5$. Since $\alpha \geq 0.5$ leads to $1 - 2\alpha \leq 0$, it is impossible to have negative unilateral punishment ($u < 1 - 2\alpha \leq 0$). However, it is possible to have lenient unilateral punishment ($u < 1 - 2\alpha$) for $\alpha < 0.5$. This possibility will generate two additional propositions: lenient unilateral punishment with lenient tie punishment, and lenient unilateral punishment with harsh tie punishment. See details in Appendix A.1.

equilibrium is achieved. The light-shaded area on the left indicates the case where the game degenerates into a coordination problem.

[Insert Figure 1 here]

3 Experimental design and procedures

3.1 Design and predictions

In Andreoni and Gee (2012), the contribution level is close to the full-contribution equilibrium where the hired-gun mechanism was implemented. They had an endowment of 5 with contributions being integers and $u = 1,\ t = 1$. The relatively small choice set of contributions in their study might have impacted the outcome. For instance, it might be easier for people to play the equilibrium strategy without realizing it. It would be more difficult to guess it correctly with a relatively large choice set. Evidence of equilibrium outcomes with a large choice set would further substantiate the effectiveness of the hired-gun mechanism. To this end, we run our treatments with an endowment of 20 to expand the choice set.

We replicate the hired-gun treatment in Andreoni and Gee (2012) as our *Control* treatment, where the endowment is 5, $u = 1,$ and $t = 1$. Likewise, we set α to $2/3$. We have an additional 6 treatments varying the unilateral punishment (u) and the tie punishment (t). Those 6 treatments all have an endowment of 20. The 6 treatments are defined as follows: 1) treatment *Rescale*: we only rescale the endowment to 20, but keep $u = 1$ and $t = 1$ (u and t are normalized based on the size of the endowment; the same logic applies hereafter); 2) Treatment *Coordination*: $u = 1$ and $t = 0$; 3) Treatment *LoTNoU* (the name is short for "low t no u"): $u = 0$ and $t = 0.5$; 4) Treatment *HiTNoU* (short for "high t no u"): $u = 0$ and $t = 1$; 5) Treatment *LoTLoU* (short for "low t low u"): $u = 0.5$ and $t = 0.5$; and 6) Treatment *LoTHiU* (short for "low t high u"): $u = 1$ and $t = 0.5$.

The treatments Control, Rescale, LoTNoU, HiTNoU, LoTLoU, and LoTHiU follow Proposition 1 and locate in the dark-shaded area in Figure 1, where the conditions for an effective hired-gun mechanism are satisfied. The predicted outcome would be full contribution in these 6 treatments. The treatment Coordination follows Proposition 2 and locates in the light-shaded area, where the game becomes a coordination game. Table 1 outlines the treatment parameters and theoretical predictions.

[Insert Table 1 here]

3.2 Procedures

The experiment was conducted at Nanyang Technological University, Singapore in January 2017 and August 2019.[7] Subjects were recruited through mass university emails and were from various majors. The experiment was programmed in Z-tree (Fischbacher, 2007). There were 4 sessions for each treatment. Each session had a size of 12 except for one had 8 participants. In total, there were 332 participants in 28 sessions. The between-subject design was implemented; that is, no subjects participated in more than 1 session.

We closely followed the procedure in Andreoni and Gee (2012), including the instructions. All instructions used are available in Appendix D. Instructions were read aloud at the beginning of the session, and questions were answered privately. Subjects had to correctly answer some control questions before proceeding to the real experiment. Then, subjects played the public goods game for 20 periods in total. The first 10 periods were the public goods game without the hired-gun mechanism. Starting from period 11, subjects played the game with the introduction of the hired-gun mechanism for another 10 periods. Instructions for the last 10 periods were given after subjects had finished the first 10 periods.

Subjects played in groups of 4 and were randomly re-matched from period to period within the session. The random rematch was designed to mitigate repeated game effects. As a result, one session is considered 1 independent observation, which results in 4 independent observations for each treatment. Subjects were informed of everyone's contribution and earnings in the group at the end of every period. To follow the exact procedure of Andreoni and Gee (2012), which did not elicit belief during the experiment, we only elicited belief[8] at the end of the experiment to avoid any potential effects on contribution. The ex-post belief elicitation aims to provide a glimpse of motivations behind contributions and is not the focus of our analysis. One out of 20 periods was

[7] We conducted 4 treatments (Control, Rescale, LoTNoU, and Coordination) in 2017. Following the results in the experiment that contributions do not converge to the full-contribution equilibrium in the LoTNoU treatment, we conducted an additional 3 treatments in 2019 to further understand the role of u and t in the hired-gun mechanism. Based on the parameters in the LoTNoU treatment, we vary the parameters by increasing t only (HiTNoU), adding a mild u only (LoTLoU), or adding a large u only (LoTHiU). We want to capture the separate effects of u and t through those additional treatments.

[8] Subjects were asked how much they expected other group members to contribute on average for the first 10 periods and the last 10 periods, respectively.

randomly selected for payment. After 20 periods, subjects filled in a post-experiment questionnaire before collecting payments. The average payment for this experiment was around 15 Singapore dollars (equivalent to around 11 U.S. dollars).

4 Results

4.1 Experimental results

We start with descriptive results, followed by econometric analysis. Table 2 summarizes the distribution of contributions. LPG represents the linear public goods game in the first 10 periods. HG represents the public goods game with the hired-gun mechanism in the last 10 periods. The mean contribution substantially increases with the introduction of the hired-gun mechanism in all treatments. The contributions in the Control, Rescale, LoTLoU, and LoTHiU treatments are close to full contribution (with mean over 90%). The distribution of ties indicates that the majority (87.5%) are tied at full contribution in the Control treatment, which aligns with the theoretical prediction. Though this percentage is lower (53.3%) in the Rescale treatment, the average contribution remains close to full contribution. The larger endowment size in the Rescale treatment might make a full-contribution tie more difficult to reach within the time horizon in the experiments, but it does not change the tendency to reach the full-contribution equilibrium. When the unilateral punishment is removed, full-contribution ties are surprisingly rare regardless of the size of the tie punishment (4.2% and 12.5% in the LoTNoU and HiTNoU, respectively), and the mean contribution seems a bit lower than in other treatments with the same full-contribution equilibrium prediction. As for the treatment Coordination where the tie punishment is removed, the theory predicts that people will coordinate on contribution levels that are no lower than the starting point. Except for those who coordinate on the full contribution (20.9%), successful coordination on below-full contribution is rare (0.9%).

[Insert Table 2 here]

Figure 2 depicts the mean contribution over time across treatments. Note that subjects played the same standard linear public goods game in the first 10 periods in all treatments except that the endowment in the Control treatment is 5 tokens, instead of the 20 tokens in the other 6 treatments. Though the first 10 periods are not the focus of this study, we present the results for completeness. The decaying trend is consistent with the classic findings in the literature (e.g., Houser & Kurzban, 2002; Fischbacher & Gaechter, 2010). Contributions across treatments are remarkably similar, and most of the pairwise comparisons (17 out of 20) are not significant (two-sided Mann-Whitney tests[9]), which suggests that there is no substantial difference in individual idiosyncrasy across treatments. Obviously, a difference in the amount of endowment does not affect the contribution.

[Insert Figure 2 here]

Result 1: Expanding the choice set while maintaining the severity of both punishment components does not hinder the effectiveness of the hired-gun mechanism. The contribution level remains high, but full-contribution ties are less frequent.

The introduction of the hired-gun mechanism in period 11 substantially boosts contribution regardless of the size of the unilateral and tie punishments. Such a boost in contribution squares with the findings in Andreoni and Gee (2012). Their results are successfully replicated in our Control treatment with the exact same parameters ($Endowment = 5, u = 1, t = 1$). In the Rescale treatment with a 20-token endowment but the same unilateral and tie punishment ($u = 1, t = 1$), the results remain comparable. The difference between the Control and Rescale treatments is not statistically significant (two-sided Mann-Whitney tests, p = 0.248). The results in these two treatments align with theoretical predictions. This evidence further substantiates the effectiveness of the hired-gun mechanism.

Result 2: The absence of the unilateral punishment regardless of the size of the tie punishment leads to decaying contributions, which contradicts theoretical predictions. Removing the unilateral punishment renders the hired-gun mechanism ineffective.

The theory prediction does not find support in the LoTNoU treatment. The predicted outcome is the contribution converging to the full-contribution equilibrium. In contrast, the contribution shows a decaying trend over time, especially in the last few periods. The contribution level in the LoTNoU treatment is significantly lower than that in the Rescale treatment (two-sided Mann-Whitney tests, p = 0.043). This decaying trend also happens in the HiTNoU treatment. The low percentage of full-contribution ties in the LoTNoU and HiTNoU treatments shown in Table 2

[9] Independent sessions are used as the unit of observation. The rest of the non-parametric tests in this article use the same unit of observation unless noted otherwise. The pairwise comparisons between LoTNoU and HiTNoU treatments (p = 0.043), LoTNoU and LoTLoU treatments (p = 0.083), and Coordination and HiTNoU treatments (p = 0.083) are significant.

further demonstrates that the mechanism is much less effective in driving the contribution to the full-contribution equilibrium. Both treatments have the unilateral punishment removed. Thus, positive unilateral punishment is seemingly a necessary driver for a full-contribution equilibrium. Such necessity is not captured by the theory.

A possible reason behind the discrepancy between the theory and experimental results in the LoTNoU and HiTNoU treatments is that the theory assumes people are fully rational. People are perceived to be able to play the dominant strategy even if it is difficult to empirically identify such a strategy. This condition does not occur in the actual play of the game. In fact, if the best strategy is too difficult to identify, people evolve to play the second-best strategy. In our experimental setting, if the best strategy is to increase the contribution by some amount that falls into a narrow range, the difficulty in identifying such a range may instead make people decrease their contributions. In a broader sense, the same reasoning could be applied to many other contexts in explaining the discrepancy between theories and empirical results. Following this line of reasoning, we aim to explain this discrepancy using individual evolutionary learning (IEL). IEL has been used to explain contribution patterns in public goods games (Arifovic & Ledyard, 2012) and to explain other economic behaviors (Arifovic, 1994; Arifovic & Ledyard, 2007). Details about the implementation of IEL are in Appendix B. The results from IEL simulations without sample size and time horizon constraints square with the experimental results. A similar decaying trend happens in the LoTNoU and HiTNoU treatments in simulations.

Result 3: It is possible to soften one or both punishment components while maintaining the effectiveness of the hired-gun mechanism.

Compared to the Rescale treatment ($u = 1, t = 1$), the tie punishment is softened in the LoTHiU ($u = 1, t = 0.5$) treatment, and both the unilateral and tie punishments are softened in the LoTLoU ($u = 0.5, t = 0.5$) treatment. Nevertheless, the contribution level remains high in these two treatments (mean over 90%) and is not significantly different from the Rescale treatment (two-sided Mann-Whitney tests, $p > 0.5$). As shown in Figure 2, the contribution pattern over time is also remarkably similar in both treatments.

Though reducing the punishment without compromising the effectiveness of the mechanism is possible, such a reduction must occur in a certain range. As discussed in result 2, if the unilateral punishment is reduced to 0, the mechanism becomes much less effective. Therefore, unilateral punishment needs to be positive. As for the tie punishment, the theory suggests that, if the tie punishment falls below $1 - \alpha$, the game becomes a coordination game, and the full contribution is no longer the unique equilibrium. The theory fails to predict the specific contribution level that people will coordinate on. To a large degree, that level depends on the starting contribution level. Our experiments show that the contribution level in the Coordination treatment is consistently high over time in the last 10 periods. The contribution level is not significantly different from the Rescale treatment (two-sided Mann-Whitney tests, p = 0.564). The starting point in period 11 is high, which might have contributed to overall high contributions over time. The results also indicate that successful coordination on below-full-contribution levels, though possible in theory, is unlikely to happen empirically (only occurred 0.9% of the time). Successful coordination mostly happens at the full-contribution level in our experiments. Individual contribution plots are available in Appendix C.

Table 3 reports the regression analysis of contribution by treatment in the last 10 periods where the hired-gun mechanism is introduced. The table shows the estimations for the 6 treatments where the theoretical prediction is the full-contribution equilibrium. The overall model is not significant for the Coordination treatment and thus is excluded in Table 3. The game played in the Coordination treatment is essentially different from the game played in the rest of the treatments, which might explain why the same estimation does not work universally. Additional analyses of the Coordination treatment are provided in Appendix C.

Multilevel mixed-effects linear estimation is adopted. Observations are clustered by subject.[10] The dependent variable is defined as the percentage of endowment contributed. Explanatory variables include 1) period (*Period*); 2) a dummy variable taking a value of 1 if the payoff in the previous period was the group's minimum and the group did not tie at full contribution, and 0 otherwise (*Last min profit*); 3) a dummy variable taking the value of 1 if the payoff in the previous period was the group's maximum and the group did not tie at full contribution, and 0 otherwise (*Last max profit*)[11]; 4) a dummy variable indicating if there was a tie in the previous period (*Tie_t-*

[10] In addition, we have tried the multilevel mixed-effects Tobit regression. The Tobit model turns out to be a poor fit for the Control treatment as the log likelihood does not converge. We have also tried the individual-level fixed-effects linear regression. The model appears to be a poor fit for the LoTHiU treatment as the model overall is insignificant.

[11] Since the punishment is exerted in a specific way so that the relative payoff position may or may not be altered (i.e., the lowest contributor might no longer have the highest payoff because of the punishment), the resulting payoff

1); and 5) a dummy variable taking the value of 1 if one was punished in the previous period and 0 otherwise (*Last punished*).[12]

[Insert Table 3 here]

Period does not have significant effects on the contribution in the Control, Rescale, LoTLoU, and LoTHiU treatments, which is consistent with the findings shown in Figure 2. The negative coefficient speaks to the decaying trend in the LoTNoU and HiTNoU treatments, which again squares with the results depicted in Figure 2. Earning the group minimum payoff or the group maximum payoff in the previous period has significantly positive effects on the contribution in all treatments with an endowment of 20. The highest contributor increasing the contribution might occur due to expectations that others will increase the contribution. Since the lowest contributor is punished with a payoff that is lower than the second-lowest contributor in the Rescale, LoTLoU, and LoTHiU treatments, the second-lowest contributor has the maximum payoff in the group. Regressions (2), (5), and (6) suggest that the second-lowest contributor tends to increase the contribution, possibly expecting that the lowest will increase the contribution as well.[13] Those who have the group maximum payoff in the LoTNoU and HiTNoU treatments are the lowest and the second-lowest contributors as the unilateral punishment is zero. Those people, on average, tend to increase their contributions in the next period. This phenomenon broadly echoes the essence of the hired-gun mechanism—the competition to avoid being the lowest contributor drives up the contribution.

The positive effect of being tied in the previous period is universal in all treatments with an endowment of 20. The presence of the tie punishment creates a pull to the full-contribution equilibrium, just as the theory predicts. The significant and negative sign of *Last punished* suggests that the lowest contributor tends to decrease the contribution in the next period, which might have contributed to the declining trend in the LoTNoU and HiTNoU treatments.

---

ranking should be more relevant for the decisions on future contributions. We want to capture how people respond to the relative payoff position. In light of this, we use the minimum and maximum profit dummies as independent variables.

[12] We have run the regression using the punishment amount instead of the dummy as the independent variable; the results remain qualitatively similar.

[13] This effect, together with effects captured by other explanatory variables, is not observed in the Control treatment. This outcome might occur because subjects achieved the full-contribution equilibrium shortly after the introduction of the hired-gun mechanism and then mostly stayed in the equilibrium.

5 Conclusion

In this paper, we explore the robustness of a centralized punishment mechanism—the hired-gun mechanism proposed by Andreoni and Gee (2012). We are interested in the hired-gun mechanism because it effectively promotes cooperation and has relatively low-cost implementation. The hired-gun mechanism punishes the lowest contributor to the extent that the person would rather have been the second-lowest contributor. The suggested punishment involves two components: a unilateral punishment and a tie punishment. The former is imposed to discourage people from wanting to be the lowest contributor, and the latter is added to prevent people from coordinating on a tie at a below-full-contribution level. By theoretically generalizing the hired-gun mechanism, we find that there is essentially a range of values for the relative magnitude of these two components that would sustain a full-contribution equilibrium. We are interested in whether the effectiveness of the mechanism is sensitive to the sizes of both punishment components.

Broadly speaking, we aim to shed light on how severe the unilateral and tie punishments should be to achieve the full-contribution equilibrium. In other words, we are interested in the size of the "bullets" that the "hired-gun" should carry. We vary the magnitude of the unilateral and tie punishments in such a way that the full-contribution equilibrium sustains. We theoretically derive a class of punishment mechanisms, which are also tested by experiments, that would lead to full-contribution equilibrium. We also run an experimental treatment wherein we only eliminate the tie punishment but keep the unilateral punishment intact so that the voluntary contribution game is transformed into a coordination game. Our results show that the effect of the mechanism is, to some extent, sensitive to the size of both punishment components. It is possible to soften both punishment components to a certain degree while maintaining the full-contribution equilibrium. However, reducing the unilateral punishment to zero renders the mechanism ineffective and leads to a decaying trend of contributions, which is at odds with theoretical predictions. This discrepancy is successfully explained by individual evolutionary learning.

We conclude by suggesting some promising directions for future research. One possible direction for future research may be more extensive empirical investigations of hired-gun mechanisms that lead to the full-contribution equilibrium. Our theory clearly specifies the size of the unilateral and tie punishments required to achieve such an outcome. There is a well-defined boundary specifying the size of the punishments between the effective and ineffective (i.e., the full-contribution equilibrium is not achieved) mechanisms. Our results indicate that this boundary might be different empirically given that the two mechanisms deemed theoretically effective do not necessarily lead to the full-contribution equilibrium as expected. One could empirically explore the boundary separating the effective and ineffective mechanisms. Our experiment results indicate that the zero unilateral punishment condition leads to the decaying trend of contributions. However, it is unclear whether the effect is due to the size of the unilateral punishment or the zero unilateral punishment itself. As people sometimes overreact to zero price/cost (Shampanier, Mazar, & Ariely, 2007; Kamei, 2017), one could have a treatment with very small unilateral punishment to explore this angle. In addition, it would be interesting to explore whether the generalized hired-gun mechanism applies to other types of public goods games, such as those with provision points, with asymmetric payoffs, sequential public goods games, and so on. Andreoni and Gee (2015) pursued this line and studied a hired-gun mechanism with one specific size of the unilateral and tie punishments in threshold public goods games. One could extend the study by adding more versions of hired-gun mechanisms with different unilateral and tie punishments. Another direction may be studying endogenously chosen hired-gun mechanisms. Subjects decide endogenously on the size of the unilateral and tie punishments. It would be interesting to see what type of hired-gun mechanism is preferred and whether the endogenously chosen mechanism has different effects compared to the exogenously imposed mechanism. This line of research would add to the literature on endogenous sanctions in public goods games (Tyran & Feld, 2006; Baldassarri & Grossman, 2011; Putterman et al., 2011; Grossman & Baldassarri, 2012; Markussen et al., 2014; Kamei et al., 2015; Fehr & Williams, 2018). We leave all of these interesting extensions for future research.

References

Andreoni, J., & Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, *96*(11–12), 1036–1046. https://doi.org/10.1016/j.jpubeco.2012.08.003

Andreoni, J., & Gee, L. K. (2015). Gunning for efficiency with third party enforcement in threshold public goods. *Experimental Economics*, *18*(1), 154–171. https://doi.org/10.1007/s10683-014-9392-1

Arifovic, J. (1994). Genetic algorithm learning and the cobweb model. *Journal of Economic Dynamics and Control*, *18*(1), 3–28. https://doi.org/10.1016/0165-1889(94)90067-1

Arifovic, J., & Ledyard, J. (2007). Call market book information and efficiency. *Journal of Economic Dynamics and Control*, *31*(6), 1971–2000. https://doi.org/10.1016/j.jedc.2007.01.006

Arifovic, J., & Ledyard, J. (2011). A behavioral model for mechanism design: Individual evolutionary learning. *Journal of Economic Behavior & Organization*, *78*(3), 374–395. https://doi.org/10.1016/j.jebo.2011.01.021

Arifovic, J., & Ledyard, J. (2012). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, *96*(9–10), 808–823. https://doi.org/10.1016/j.jpubeco.2012.05.013

Arifovic, J., & Maschek, M. K. (2006). Revisiting individual evolutionary learning in the cobweb model – An illustration of the virtual spite-effect. *Computational Economics*, *28*(4), 333–354. https://doi.org/10.1007/s10614-006-9053-3

Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11023–11027. https://doi.org/10.1073/pnas.1105456108

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 139–146.

Casari, M., & Luini, L. (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior and Organization*, *71*(2), 273–282. https://doi.org/10.1016/j.jebo.2009.03.022

Cason, T. N., & Gangadharan, L. (2015). Promoting cooperation in nonlinear social dilemmas through peer punishment. *Experimental Economics*, *18*(1), 66–88. https://doi.org/10.1007/s10683-014-9393-0

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. https://doi.org/10.1007/s10683-010-9257-1

Darley, J. M. (2005). On the unlikely prospect of reducing crime rates by increasing the severity of prison sentences. *Journal of Law and Policy*, *13*(1), 189–208.

DeAngelo, G., & Gee, L. K. (2017). Peers or police? Detection and sanctions in the provision of public goods. *Working paper*.

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, *33*(1), 145–167. https://doi.org/10.1007/s00199-007-0212-0

Devetag, G., & Ortmann, A. (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics*, *10*(3), 331–344. https://doi.org/10.1007/s10683-007-9178-9

Durlauf, S. N., & Nagin, D. S. (2011). Imprisonment and crime. *Criminology & Public Policy*, *10*(1), 13–54. https://doi.org/10.1111/j.1745-9133.2010.00680.x

Falkinger, B. J., Fehr, E., Gachter, S., & Winter-Ebmer, R. (2000). A simple mechanism for the efficient provision of public goods: Experimental evidence. *American Economic Review*, *90*(1), 247–264.

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

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

Fehr, E., & Williams, T. (2018). Social norms, endogenous sorting and the culture of cooperation. *University of Zurich, Department of Economics, Working Paper No 267, Revised Version. Available at https://Ssrn.com/abstract=3062528 or http://dx.doi.org/10.2139/ssrn.3062528*

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fischbacher, U., & Gaechter, S. (2010). Heterogeneous social preferences and the dynamics of free riding in public goods. *American Economic Review*, *100*(1), 541–556.

Fowler, J. H. (2005). Human cooperation: Second-order free-riding problem solved? *Nature*, *437*(7058), E8–E8. https://doi.org/10.1038/nature04201

Gross, J., Méder, Z. Z., Okamoto-Barth, S., & Riedl, A. (2016). Building the Leviathan – Voluntary centralisation of punishment power sustains cooperation in humans. *Scientific Reports*, *6*(June 2015), 20767. https://doi.org/10.1038/srep20767

Grossman, G., & Baldassarri, D. (2012). The impact of elections on cooperation: Evidence from a lab-in-the-field experiment in Uganda. *American Journal of Political Science*, *56*(4), 964–985. https://doi.org/10.1002/9781118762707.ch08

Herrmann, B., Christian, T., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367.

Houser, D., & Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *American Economic Review*, *92*(4), 1062–1069.

Kamei, K. (2017). Endogenous reputation formation under the shadow of the future. *Journal of Economic Behavior and Organization*, *142*, 189–204. https://doi.org/10.1016/j.jebo.2017.07.012

Kamei, K. (2018). Group size effect and over-punishment in the case of third party enforcement of social norms. *Journal of Economic Behavior and Organization*, 1–18. https://doi.org/10.1016/j.jebo.2018.04.002

Kamei, K., Putterman, L., & Tyran, J.-R. (2015). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics*, *18*(1), 38–65. https://doi.org/10.1007/s10683-014-9405-0

Kamijo, Y., Nihonsugi, T., Takeuchi, A., & Funaki, Y. (2014). Sustaining cooperation in social dilemmas: Comparison of centralized punishment institutions. *Games and Economic Behavior*, *84*, 180–195. https://doi.org/10.1016/j.geb.2014.01.002

Kennedy, D. M. (2009). *Deterrence and crime prevention: Reconsidering the prospect of sanction*. London: Routledge.

Kleiman, M. A. R. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton, New Jersey: Princeton University Press.

Mann, M. (1986). *The sources of social power, vol. I: A history of power from the beginning to 1760 AD*. Cambridge: Cambridge University Press.

Markussen, T., Putterman, L., & Tyran, J. R. (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies*, *81*(1), 301–324. https://doi.org/10.1093/restud/rdt022

McFatter, R. M. (1982). Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness. *Journal of Applied Psychology*, *67*(3), 255–267.

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. https://doi.org/10.1016/j.jpubeco.2007.04.008

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, *432*(7016), 499–502. https://doi.org/10.1038/nature02978

Putterman, L., Tyran, J. R., & Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics*, *95*(9–10), 1213–1222. https://doi.org/10.1016/j.jpubeco.2011.05.001

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science*, *26*(6), 742–757. https://doi.org/10.1287/mksc.l060.0254

Tyran, J. R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent.

Scandinavian Journal of Economics, *108*(1), 135–156. https://doi.org/10.1111/j.1467-9442.2006.00444.x

Xiao, E., & Houser, D. (2011). Punish in public. *Journal of Public Economics*, *95*(7–8), 1006–1017. https://doi.org/10.1016/j.jpubeco.2010.11.021

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*(1), 110–116. https://doi.org/10.1037/0022-3514.51.1.110

Tables and Figures

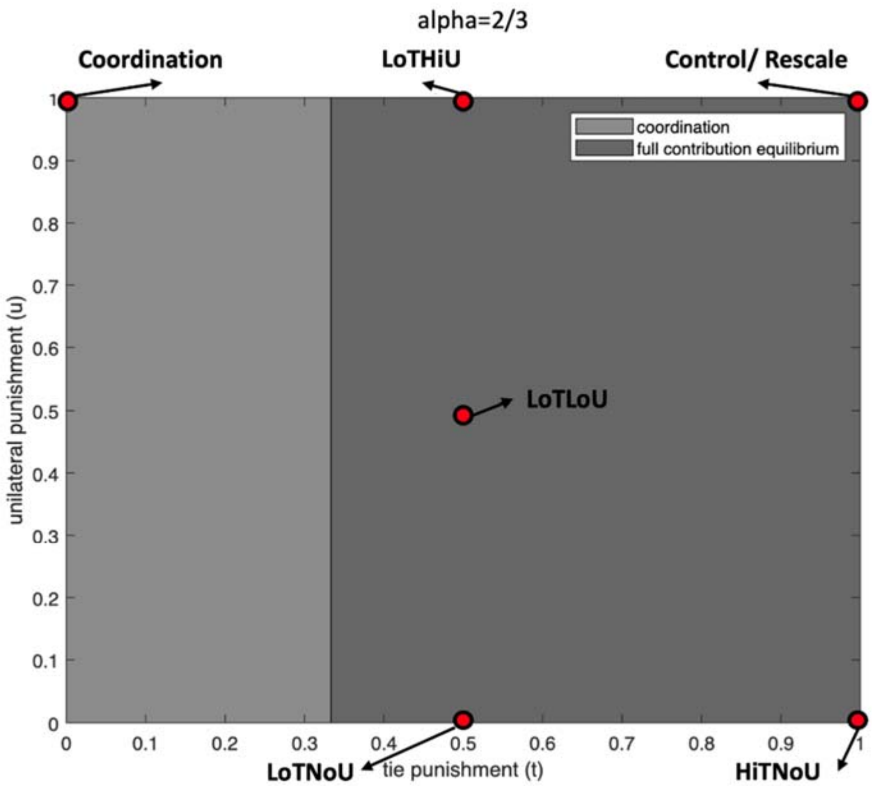


Figure 1. Game outcomes conditional on t, u, and the treatments

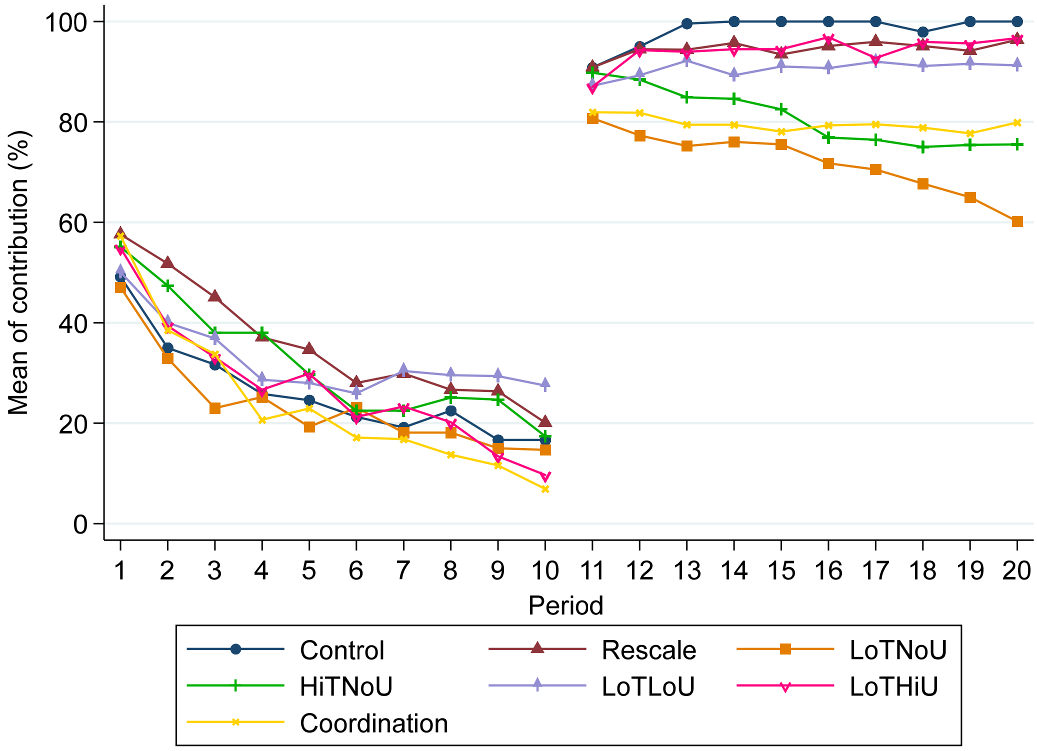


Figure 2. Mean contribution over time

Table 1. Treatment parameters and predictions

Treatment	Endowment	t	u	Relationship of t, u, α	Predicted Outcome
Control	5	1	1		
Rescale	20	1	1		
LoTNoU	20	0.5	0	$t > 1 - \alpha,$	Full contribution
HiTNoU	20	1	0	$u > 1 - 2\alpha$	
LoTLoU	20	0.5	0.5		
LoTHiU	20	0.5	1		
Coordination	20	0	1	$t < 1 - \alpha,$ $u > 1 - 2\alpha$	Coordination

Table 2. Contribution and ties across treatment

Treatment	Male	Game	Contribution (%)		Ties		
			Mean	Std Dev.	Below Full	At Full	Total
Control	39.6%	LPG*	26.3%	0.264	4.2%	0	4.2%
		HG*	98.3%	0.029	0	87.5%	87.5%
Rescale	50.0%	LPG	35.7%	0.247	5.0%	0	5.0%
		HG	94.6%	0.079	0	53.3%	53.3%
LoTNoU	41.7%	LPG	23.7%	0.254	5.0%	0	5.0%
		HG	72.0%	0.195	0	4.2%	4.2%
HiTNoU	47.9%	LPG	32.0%	0.329	16.7%	0	16.7%
		HG	80.9%	0.291	0	12.5%	12.5%
LoTLoU	50%	LPG	32.6%	0.341	0.8%	0	0.8%
		HG	90.6%	0.187	0	36.7%	36.7%
LoTHiU	50%	LPG	27.2%	0.313	4.2%	0	4.2%
		HG	94.2%	0.151	0	40%	40%
Coordination	38.6%	LPG	23.9%	0.241	10.9%	0	10.9%
		HG	79.6%	0.118	0.9%	20.9%	21.8%

*LPG: linear public goods game; HG: public goods game with the hired-gun mechanism

Table 3. Determinants of contribution: Multilevel mixed-effects linear estimation

Dependent variable: Contribution percentage

	Control	Rescale	LoTNoU	HiTNoU	LoTLoU	LoTHiU
	(1)	(2)	(3)	(4)	(5)	(6)
Period	0.000290	-0.000312	-0.0157***	-0.0156***	0.000937	0.00117
	(0.00143)	(0.00210)	(0.00324)	(0.00323)	(0.00208)	(0.00222)
Last min profit	-0.0860	0.150***	0.137***	0.0862***	0.0907***	0.0707***
	(0.0701)	(0.0295)	(0.0223)	(0.0228)	(0.0210)	(0.0269)
Last max profit	-0.00676	0.179***	0.0949***	0.0547**	0.0641***	0.0829***
	(0.0761)	(0.0335)	(0.0231)	(0.0235)	(0.0239)	(0.0303)
Tie_t-1	0.0101	0.187***	0.232***	0.186***	0.108***	0.107***
	(0.0757)	(0.0351)	(0.0517)	(0.0419)	(0.0267)	(0.0311)
Last punished	0.0141	0.0239	-0.0615***	-0.0412*	-0.0176	-0.0399*
	(0.0330)	(0.0242)	(0.0229)	(0.0227)	(0.0221)	(0.0226)
Constant	0.982***	0.783***	0.843***	0.959***	0.822***	0.855***
	(0.0770)	(0.0458)	(0.0694)	(0.0700)	(0.0449)	(0.0451)
Observations	432	432	432	432	432	432
Log. Likelihood	550.5	305.9	83.46	102.7	298.5	297.4
Chi-squared	33.23	37.05	94.82	59.27	38.86	32.54

Notes: Standard errors are in parentheses. Observations are clustered by subject.
Data includes periods12–20 where the hired-gun mechanism is in place.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A: Additional results on equilibrium analysis

A1. Equilibrium analysis with $\alpha < 0.5$

In this appendix, we provide equilibrium analysis with $\alpha < 0.5$. When $\alpha < 0.5$, $u < 1 - 2\alpha$ and $u \geq 0$ could be satisfied simultaneously; therefore, lenient unilateral punishment is a possibility. We next discuss cases where both types of punishment are harsh enough, one type is lenient and one type is harsh enough, and both types are lenient.

Proposition A1.1. *If both the tie punishment and unilateral punishment are harsh enough, or* $t > 1 - \alpha$ and $u > 1 - 2\alpha$, *the game's unique equilibrium occurs when everyone contributes the full endowment.*

Proof. It follows the proof of Proposition 1.■

Proposition A1.2. *If the tie punishment is lenient and the unilateral punishment is harsh enough, or* $t < 1 - \alpha$ and $u > t - \alpha$, *the game becomes a coordination game.*

Proof. Part (i). $t < 1 - \alpha$ and $u > 1 - 2\alpha$; it is the same as the proof of Proposition 2.

Proof. Part (ii). $t < 1 - \alpha$ and $t - \alpha < u < 1 - 2\alpha$; it leads to $\Delta_{+1} < 0$ and $\Delta_{-1} < 0$, which suggests that players are better off by staying tied rather than breaking from the tie. In the no-tie case, $\Delta_{+1\_lower} = \alpha + u - t > 0$, and the lowest contributor has incentive to increase the contribution until a tie is reached. In other words, the game becomes a coordination game.■

Proposition A1.3. *If both the tie punishment and the unilateral punishment are lenient, or* $t < 1 - \alpha$ and $u < t - \alpha$, *the game's unique equilibrium is that everyone makes zero contribution.*

Proof. When $\alpha < 0.5$, it is possible to satisfy $0 < u < 1 - 2\alpha$. This outcome leads to $\Delta_{+1} < 0$ and $\Delta_{-1} > \Delta_{+1}$. $u < t - \alpha$ leads to $\Delta_{-1} = t - \alpha - u > 0$. In the tie case, players are better off by decreasing the contribution by 1 unit to break from the tie. In the no-tie case (the difference in the contribution has been shortened to 1 unit), $\Delta_{+1\_lower} = \alpha + u - t < 0$, and $\Delta_{+2\_lower} = 2\alpha + u - 1 < 0$, which means the lower contributor is better off by staying at the status quo rather than increasing the contribution to a tie or to be the higher contributor. Besides the possible change in the relative payoff positions caused by actions of the lower contributor, the higher contributor's actions could also cause such changes. If the higher contributor decreases the contribution by 1 unit and the resulting payoff change is $\Delta_{-1\_higher} = 1 - \alpha - t > 0$, then the higher contributor has incentive to decrease the contribution to a tie. The game then ends up with the zero-contribution equilibrium following the elimination of dominated strategies.■

Proposition A1.4. *If the tie punishment is harsh enough and the unilateral punishment is lenient, or* $t > 1 - \alpha$ and $u < 1 - 2\alpha$, *the game outcome is dependent on the starting point.*

Proof. As discussed earlier, it is possible to satisfy $u < 1 - 2\alpha$ only if $\alpha < 0.5$. These two conditions of t and u lead to $0 < \Delta_{+1} < \Delta_{-1}$ (i.e., players are better off by deviating 1 unit negatively from the tie). In the no-tie case, $\Delta_{+1\_lower} = \alpha + u - t < 0$, which means the lower contributor is better off staying as the lower rather than increasing the contribution to a tie. It is straightforward to see that the higher contributor is better off by staying holding the contribution steady rather than decreasing it to a tie ($\Delta_{-1\_higher} = 1 - \alpha - t < 0$). In this case, there is no single dominant strategy; thus, the logic of the elimination of dominated strategies does not apply. The game outcome is dependent on the starting contributions in the group. Specifically, if both players start with full contributions, neither of them would have incentive to deviate. If they start with different individual contributions, they would stay with the initial contribution as no one has incentive to deviate from the starting point. If they start with a below-full-contribution tie, they would deviate negatively until a difference in contributions is restored. ■

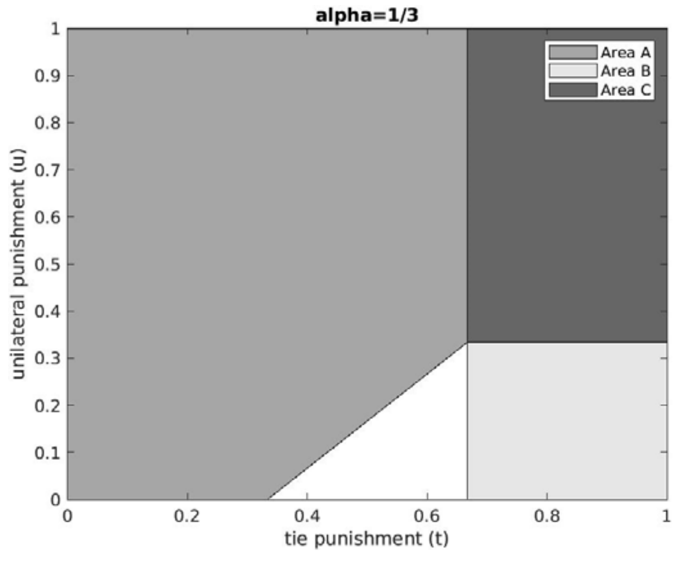


Figure A1.1. Game outcomes conditioned on t and u ($\alpha = 1/3$)

Figure A1.1 summarizes and illustrates all possible cases described above for various parameter values for the unilateral punishment (u) and the tie punishment (t), with $\alpha = 1/3$. Area C represents the required condition for an effective hired-gun mechanism where the unique full-contribution Nash equilibrium is achieved. Area A indicates the case where the game degenerates into a coordination problem. Area B represents the situation where the game outcome is dependent on the starting point. The blank area represents the situations where the game results in the zero-contribution equilibrium. Table A1.1 summarizes the game outcomes and parameter conditions.

Table A1.1. Game outcomes conditioned on t and u

Area	t	u	Outcome
A	$t < 1 - \alpha$	$u > t - \alpha$	Coordination
B	$t > 1 - \alpha$	$u < 1 - 2\alpha$	Dependent on the starting point
C	$t > 1 - \alpha$	$u > 1 - 2\alpha$	Full contribution
Blank	$t < 1 - \alpha$	$u < t - \alpha$	Zero contribution

A2. Equilibrium analysis with group size = 4

Case 1. *The 4 players in the group tie at a below-full-contribution amount.*

The proof is the same as the 2-player case. All 4 players receive a punishment $P = t$. Increasing the contribution by 1 unit incurs a loss of $1 - \alpha$ but avoids the punishment t. Decreasing the contribution by 1 unit increases the payoff by $1 - \alpha$ and avoids the punishment t, but a punishment $(1 + u)$ kicks in, as shown below:

$$\Delta_{+1} = t - (1 - \alpha) \quad \text{(A2.1)}$$

$$\Delta_{-1} = t - (\alpha + u) \quad \text{(A2.2)}$$

To encourage people only to move the contribution upward, it is necessary to have the following:

$$\Delta_{+1} > 0 \Rightarrow t > 1 - \alpha \quad \text{(A2.3)}$$

$$\Delta_{+1} > \Delta_{-1} \Rightarrow u > 1 - 2\alpha \quad \text{(A2.4)}$$

If (A2.3) and (A2.4) are both satisfied, players would have incentive to increase the contribution to break the tie.

Case 2. *The group has heterogeneous contributions (i.e., at least one player's contribution is different from others).*

There are 7 possible distributions of contributions. Suppose the 4 players are a, b, c, and d, and their contributions are ranked from highest (top row) to lowest (bottom row) in Table A2.1, followed by an equilibrium analysis of each distribution.

Table A2.1. Possible distributions of contributions

<table>
<tr><th>(1)</th><th>(2)</th><th>(3)</th><th>(4)</th><th>(5)</th><th>(6)</th><th>(7)</th></tr>
<tr><td>a</td><td rowspan="2">a, b</td><td rowspan="2">a</td><td rowspan="2">a</td><td rowspan="3">a, b, c</td><td rowspan="3">a</td><td rowspan="3">a, b</td></tr>
<tr><td>b</td></tr>
<tr><td>c</td><td>c</td><td>b, c</td><td>b</td></tr>
<tr><td>d</td><td>d</td><td>d</td><td>c, d</td><td>d</td><td>b, c, d</td><td>c, d</td></tr>
</table>

If the difference between the second-lowest and the lowest contribution level is more than 1 unit, it is straightforward that the higher contributor will decrease the contribution, and the lower contributor will increase the contribution as doing so would improve the payoff for both before the relative position of contributions changes. Therefore, we discuss the case where the difference between the second-lowest and the lowest contribution has been shortened to 1 unit. The equilibrium analyses of the 7 possible distributions of contributions are as follows:

(1) All four contributions are different.

For the top two contributors—players a and b—decreasing the contribution by 1 unit generates a payoff gain of $1 - \alpha > 0$ as long as neither player becomes the lowest contributor. In other words, players a and b will keep decreasing their contributions until they reach the contribution level of player c. The contributions of players a, b, and c are identical and are 1 unit above the lowest contribution from player d. The initial four contribution levels are reduced to two contribution levels. The situation is therefore equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(2) Top two contributors (a and b) are tied.

The same logic in (1) above applies. Players a and b will keep decreasing their contributions until their contributions, together with player c's contribution, are 1 unit above player d's contribution. The initial three contribution levels are reduced to two contribution levels. The situation is therefore equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(3) Two middle contributors (b and c) are tied.

The same logic in (1) applies. Player a will keep decreasing the contribution until it, together with the contributions of players b and c, is 1 unit above player d's contribution. The initial three contribution levels are reduced to two contribution levels. The situation is therefore equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(4) Bottom two contributors (c and d) are tied.

The same logic in (1) applies. Player a will keep decreasing the contribution until it, together with player b's contribution, is 1 unit above the contributions of players c and d. The initial three contribution levels are reduced to two contribution levels. The situation is therefore equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(5) Top three contributors (a, b, and c) are tied.

There are two contribution levels to start with. The situation is therefore equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(6) Bottom three contributors (b, c, and d) are tied.

The same logic in (5) applies. The situation is equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2.

(7) Top two contributors (a and b) are tied. Bottom two contributors (c and d) are tied.

The same logic in (5) applies. The situation is equivalent to the 2-player situation. The rest of the proof is the same as the proof for Proposition 1 under Case 2. ■

Appendix B: Additional results on simulations

Individual evolutionary learning (IEL)

Following the discussion on the theoretical and empirical game outcome, we explore the evolution path under the hired-gun mechanism with different parameters. The evolution path is interesting for several reasons. First, the contribution in the LoTNoU and HiTNoU treatments has a decaying trend instead of converging to the full-contribution equilibrium as the theory predicts. The individual evolutionary learning (IEL) model is an attempt to explain this deviation from the theory. Second, the theory cannot predict the speed of convergence to the equilibrium, so it would be intriguing to learn about the convergence path and speed in general.[14] Third, the experiments shed light on how people actually behave in the hired-gun context, but the experiments are bounded by the number of subjects and periods one could use. Thus, simulation provides a robustness check using a longer horizon and a larger sample.

We adopt the IEL approach (see more details in Arifovic & Maschek 2006; Arifovic & Ledyard 2012, 2011). IEL has been shown to be successful in simultaneously explaining all five stylized facts in public goods games (Arifovic & Ledyard 2012). The idea of IEL is that subjects carry a finite set of remembered strategies, which are evaluated and replicated in a certain way, and the better strategies have better chances of being carried forward to future periods. As a result, subjects evolve to play the optimal strategies over time.

Following the definition of IEL in Arifovic and Ledyard (2012), we let X^i be subject i's action space, $I^i(x_t)$ be the information revealed to i at the end of period t, A_t^i be i's remembered set of strategies (i.e., contribution levels) in period t, and ψ_t^i be the probability measure on A_t^i. Next, A_t^i has a dimension of J, which represents the subject's memory capacity. In period t, each subject randomly chooses a strategy (i.e., a contribution level) from A_t^i based on the probability measure ψ_t^i and ends up with the action $x_t^i = a_{j,t}^i$ $(j \in \{1, \ldots J\})$. A^i and ψ^i are updated from period to period. At the end of period t, subjects use a process to compute A_{t+1}^i, ψ_{t+1}^i based on A_t^i, ψ_t^i, $I^i(x_t)$. Experimentation, replication, and selection are the three crucial steps in this process. We

[14] The experimental results do not provide much information on convergence as the starting contribution level is close to the full-contribution equilibrium. Though the experiment is not designed to study different convergence speeds across treatments, simulations could provide valuable insights in this regard.

next describe this process to explain how the decision is made at $t + 1$ based on what happened at t.

The process starts with experimentation. Experimentation introduces a dash of randomness into the remembered strategy set, which contributes to some level of diversity in the process. For each strategy in A_t^i ($a_{j,t}^i, j = 1, \ldots J$), a new strategy (contribution) from X^i is randomly chosen with probability ρ to replace the existing strategy $a_{j,t}^i$. The new strategy has a normal distribution $N\left(a_{j,t}^i, \sigma\right)$. In other words, the mean of the normal distribution where the new contribution is drawn from is equal to the value of the existing strategy $a_{j,t}^i$ to be replaced. This normal distribution has a standard deviation σ, and that σ is normalized as the proportion of endowment. Note that ρ and σ are free parameters that can be set to various numbers in the simulations.

Replication follows experimentation. Further, replication reinforces strategies that would have been relatively high paying in the past. It is also an opportunity for potentially better strategies to replace lower-paying ones. The key is to identify potentially good strategies. A strategy $a_{j,t}^i$ is evaluated based on the corresponding payoff if $a_{j,t}^i$ had been played at t regardless of the strategy actually played. Let $p^i(a_{j,t}^i | I^i(x_t))$ be subject i's payoff at t if the subject had played strategy $a_{j,t}^i$ given the information $\mathrm{I}^i(x_t)$. Then, $p^i(a_{j,t}^i | I^i(x_t))$ is used to screen good strategies. For each strategy in A_t^i ($a_{j,t}^i, j = 1, \ldots J$), $a_{j,t}^i$ is replicated in the following way. Two strategies in A_t^i are randomly selected with replacements, and the selection uses uniform probability. Let those two selected strategies be $a_{k,t}^i$ and $a_{l,t}^i$. Then, the strategy set for period $t + 1$ is updated as follows:

$$a_{j,t+1}^i = \begin{cases} a_{k,t}^i, & if\ p^i(a_{k,t}^i | I^i(x_t)) \geq p^i(a_{l,t}^i | I^i(x_t)) \\ a_{l,t}^i, & if\ p^i\left(a_{k,t}^i \middle| I^i(x_t)\right) < p^i(a_{l,t}^i | I^i(x_t)) \end{cases} \quad (9)$$

The replication repeats for $j = 1, \ldots J$. It is straightforward that strategies with more replicates at t, and those that would have resulted in higher payoffs had they been used at t, are more likely to be carried forward into $t + 1$. If the information $I^i(x_t)$ (i.e., other group members' contributions at t) makes strategy $a_{j,t}^i$ a high-paying strategy, this strategy will tend to have more replicates in the strategy set. In other words, this strategy is favorably remembered and thus has a higher chance of being actually played. As a result, the replication process averages over the past periods. In the long term, the remembered strategy set consists of best-response strategies.

Selection happens after replication and is the last step in the process. Each strategy $a^i_{k,t+1}$ in A^i_{t+1} is selected with a probability $\psi^i_{k,t+1}$. The probability is decided by each strategy's relative fitness in the overall strategy set, which is measured by the proportional payoff:

$$\psi^i_{k,t+1} = \frac{p^i(\, a^i_{k,t+1} | I^i(x_t))}{\sum_{j=1}^{J} p^i(\, a^i_{j,t+1} | I^i(x_t))} \tag{10}$$

Those three steps constitute the process describing how an IEL subject proceeds from t to $t+1$. The last piece missing is the initial values in period 1, A^i_1 and ψ^i_1. Following the practice in Arifovic and Maschek (2006), Arifovic and Ledyard (2011), and Arifovic and Ledyard (2012), we also use the same simple initialization in period 1. Every strategy forming A^i_1 is drawn randomly with uniform probability from the action space X^i. Each strategy in A^i_1 stands an equal chance of being selected in period 1. That is, $\psi^i_{k,1} = \frac{1}{J}\ for\ all\ k$.

We now have a complete picture of the IEL model from the very beginning. We set the free parameters $\rho = 1$ and $\sigma = 0.2$ in our simulations. Figure B1 shows the simulation results for all treatments with 1,000 repetitions and $t = 80$. As we use the same simple initialization across treatments, the contribution always starts at 50% of endowment. Using the actual starting point in our experiment generates qualitatively similar results to the one using 50% of endowment as the starting point (see Figure B2). There is a decaying trend in the linear public goods game, and contributions converge to zero in the long run. This phenomenon is robust regardless of the size of the endowment (J).

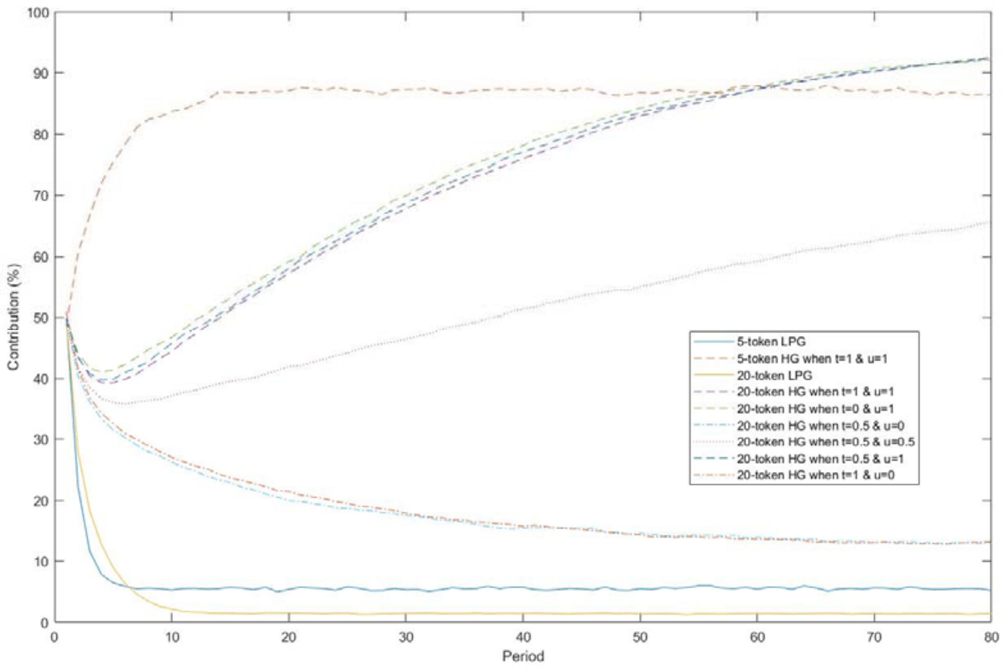


Figure B1. Simulation results with 1,000 repetitions and 80 periods

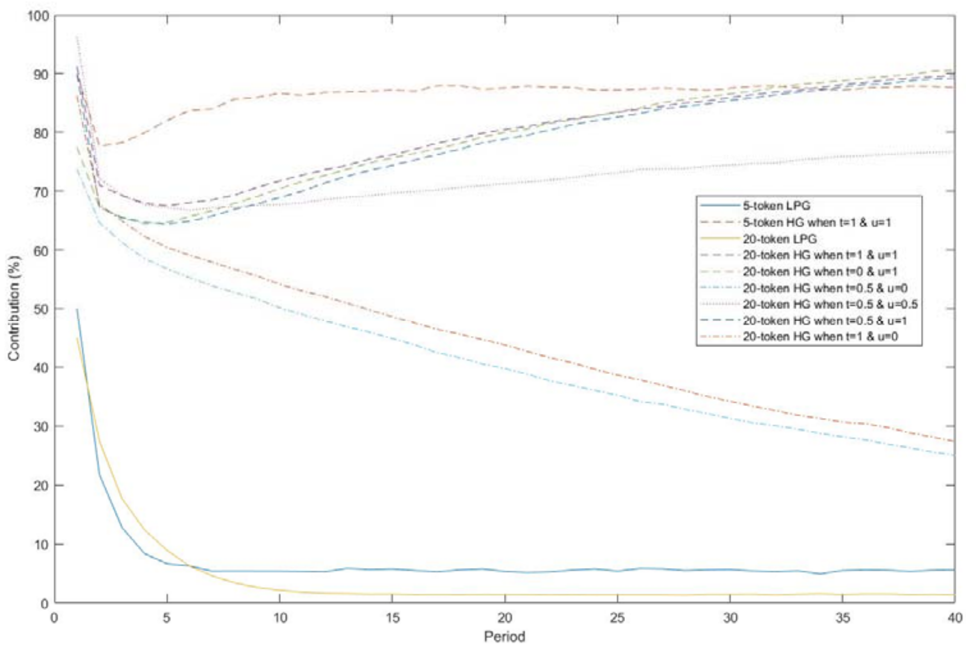


Figure B2. Simulation results with 1,000 repetitions and 40 periods (actual contribution starting point)

Endowment size plays a role in the evolution trend when the hired-gun mechanism is introduced. Contributions in the Control treatment (5-token HG when $t = 1$ and $u = 1$ in Figure B1) converge much more quickly to the full contribution than those in the Rescale treatment (20-token HG when $t = 1$ and $u = 1$ in Figure B1). This finding indicates that, if subjects start at the same medium contribution level (50%), smaller endowment size results in faster convergence to the equilibrium in the hired-gun mechanism. This effect is intuitive as the smaller set of available choices associated with the smaller endowment makes it easier to locate the dominant strategy and thus leads to a quicker convergence to the equilibrium. The difference in convergence speed is not as obvious in experiments as it is in simulations, possibly because the starting contribution level is already very high (91%) for both treatments in experiments, which limits the presence of different convergence speeds. That said, we do observe a substantial difference in the percentage of full-contribution ties (87.5% in Control vs. 53.3% in Rescale), as shown in Table 2. We have also run simulations using the actual starting contribution level (i.e., contribution in period 11) in the experiment instead of 50% as the starting point, and the results remain qualitatively unchanged.

Little difference exists in the evolution trend between the Rescale (20-token HG when $t = 1$ and $u = 1$) and the Coordination treatments (20-token HG when $t = 0$ and $u = 1$) in simulations. The contribution difference is larger in experiments than in simulations. The difference between these two treatments is the removal of tie punishment in the Coordination treatment, which theoretically would make people coordinate on a certain contribution level depending on the starting point, and the game therefore ends up with a tie. However, this is not the case in simulations. The difference might occur because successfully coordinating on a below-full-contribution level is difficult, which is a solid finding in the literature (Devetag & Ortmann, 2007). A tie also rarely exists in our simulations and experiments (only 1 out of 470 groups under the hired-gun mechanism).

Intriguingly, there is a decaying trend in the LoTNoU (20-token HG when $t = 0.5$ and $u = 0$) and HiTNoU (20-token HG when $t = 1$ and $u = 0$) treatments. This trend aligns with the experiment results but contradicts theoretical predictions. Importantly, the game is predicted to end up with a full-contribution equilibrium. This divergence from theory might result from subjects' difficulty in locating the dominant strategy. Subjects are assumed to be rational and smart in the sense that they are always able to find the dominant strategy and that the decision environment is irrelevant. For instance, suppose 1 out of 5 strategies is the dominant strategy in situation A, and 1 out of 20 strategies is the dominant strategy in situation B; theoretically, there would not be any difference in the outcome as the dominant strategy will be played in both situations. However, this result might not be the case empirically. In situation B, subjects might not be able to pick up the dominant strategy over the choice set. In the same spirit, if it is not easy for subjects to locate the "right" choice, they would ultimately play the less optimal strategy. This reality is indeed the case in the LoTNoU and the HiTNoU treatments.

We use an example where the contributions in the group are not tied to illustrate what might have happened in the LoTNoU and HiTNoU treatments. Let the contribution difference between the lowest and the second lowest be j_1, the contribution difference between the second lowest and the third lowest be j_2, the change in the lowest contributor's contribution be k, and the change in the lowest contributor's payoff resulting from the contribution change k be Δ. It is straightforward that one is never better off by decreasing the contribution when one is already the lowest contributor. Therefore, we assume $k > 0$. There are several possibilities for k: 1) the lowest contributor can increase the contribution but remains the lowest contributor, i.e., $k < j_1, \Delta = \alpha k > 0$; 2) the lowest contributor can increase the contribution to match the second-lowest contributor's contribution, i.e., $k = j_1, \Delta = \alpha j_1 - j_2$; or 3) the lowest contributor can increase the contribution to become the second-lowest contributor, i.e., $k > j_1, \Delta = j_1 + u - k(1 - \alpha)$. In the second case, if $j_2 < \alpha j_1$, increasing the contribution by exactly j_1 would make the contributor better off rather than staying put. In the third case, if $k > (j_1 + u)/(1 - \alpha)$, increasing the contribution by k would make the contributor worse off. Therefore, any one condition from (11) to (13) would be a sufficient condition to ensure that increasing k is a dominant strategy:

$$k < j_1 \tag{11}$$

$$k = j_1 \text{ and } j_2 < \alpha j_1 \tag{12}$$

$$j_1 < k < \frac{j_1 + u}{1 - \alpha} \tag{13}$$

When u is small, the range of k in (13) becomes small. If this is the case, the lowest contributor could easily increase the contribution too much, which in turn decreases the contributor's payoff. Figure B3 is a pseudocolor (checkerboard) plot of contributions in the 20th period with $\sigma = 0.2$, $w = 5$, and 100 repetitions. In the pseudocolor plot, both the tie punishment t (*x-axis*) and the unilateral punishment u (y-*axis*) occur in increments of 0.02. The color represents the contribution percentage, as indicated in the color bar on the right. Thus, the proper size of both u and t are seemingly necessary for a high contribution outcome. It might be the case that, for a small choice set associated with the 5-token endowment, it is relatively easy to find the dominant strategy. Therefore, harsh enough tie and unilateral punishments are both needed to discourage people from settling at below-full-contribution ties and being the lowest contributor. When u is sufficiently small, increasing k runs the risk of k being too much and therefore backfiring. The blue band at the bottom of Figure B3 supports this conjecture.

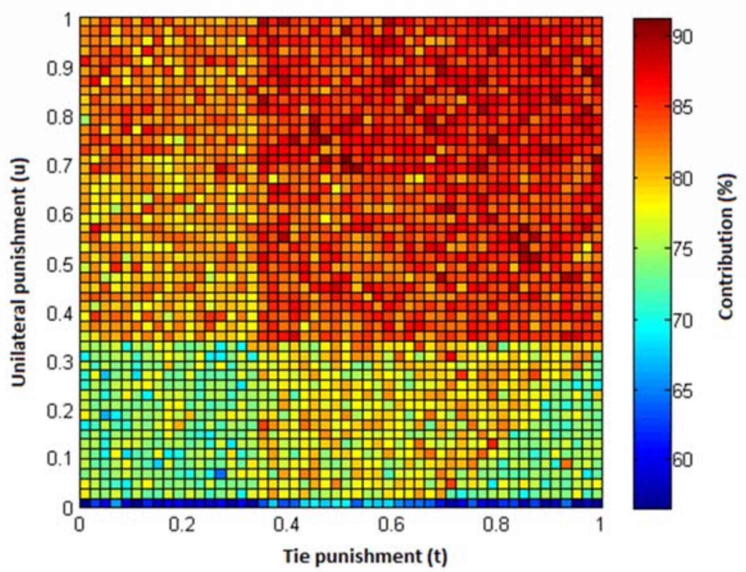


Figure B3. Contribution with 5-token endowment in the 20th period (100 repetitions)

Figure B4 plots the contributions in the 40th period with $\sigma = 0.2$, $w = 20$, and 100 repetitions.[15] When endowment increases to 20, the tie punishment no longer matters that much, and the contribution level is substantially affected by the unilateral punishment. This outcome could occur because the large choice set associated with the 20-token endowment makes ties rare, and therefore, empirically, subjects do not have many chances to learn to avoid the tie punishment. As a result, tie punishment loses its function. Similar to the 5-token case, the lowest band colored with darker blue at the bottom Figure B4 indicates lower contributions when u is very small. There are also

[15] We use the same standard deviation parameters as those in the IEL simulations. Since it takes longer to converge to the equilibrium for the 20-token endowment situations, we prolong the time frame to 40 periods.

clear cut-offs at $u = 1 - \alpha$ and $u = 1 - 2\alpha$. Consistent with (13), the range of k is sensitive to the size of u, which affects the contribution levels.

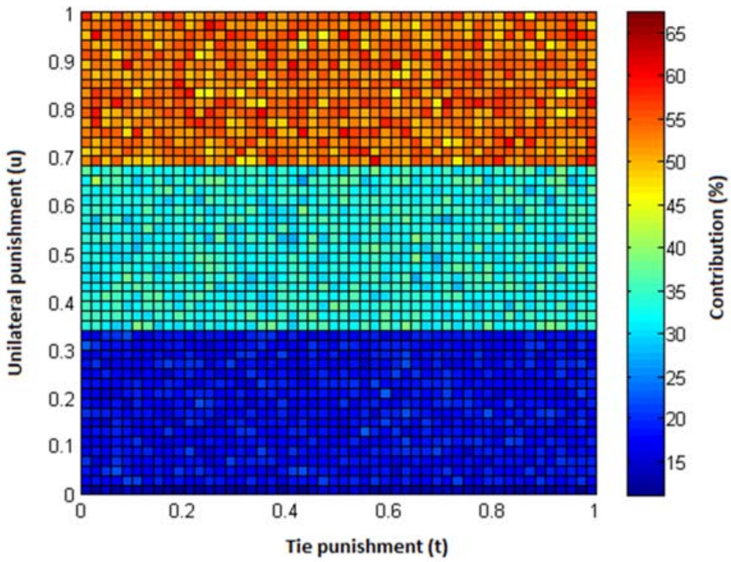
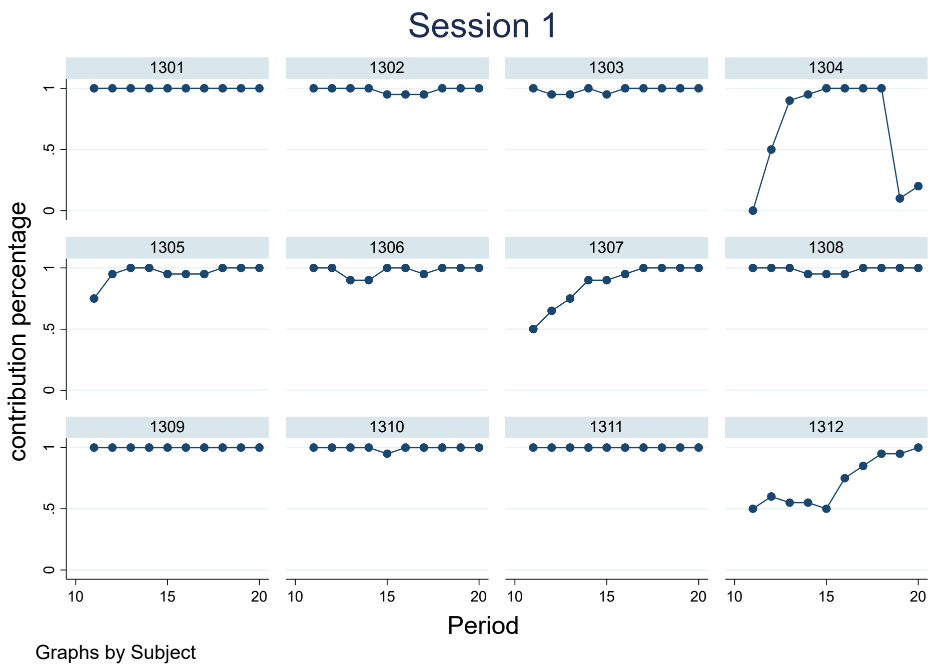
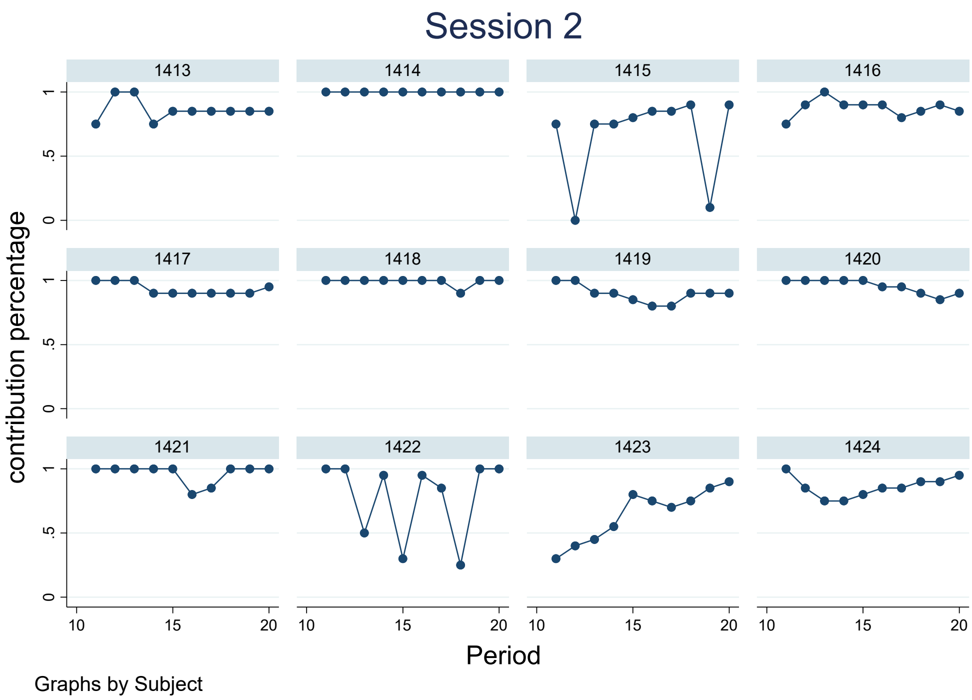


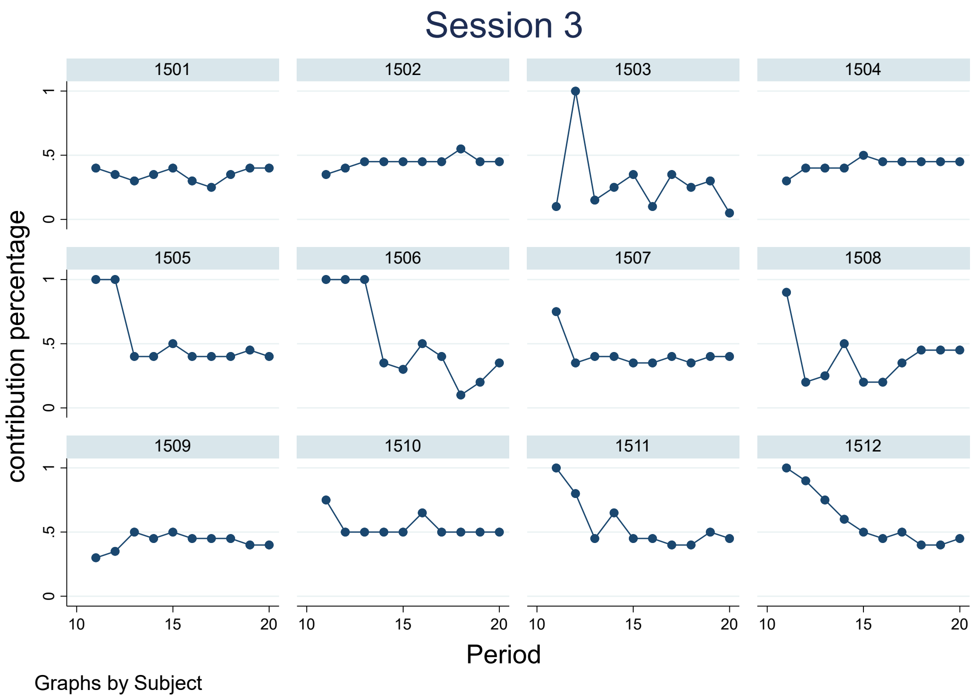
Figure B4. Contribution with 20-token endowment in the 40th period (100 repetitions)

Appendix C: Additional results on experiments

Analyses of contributions in the Coordination treatment







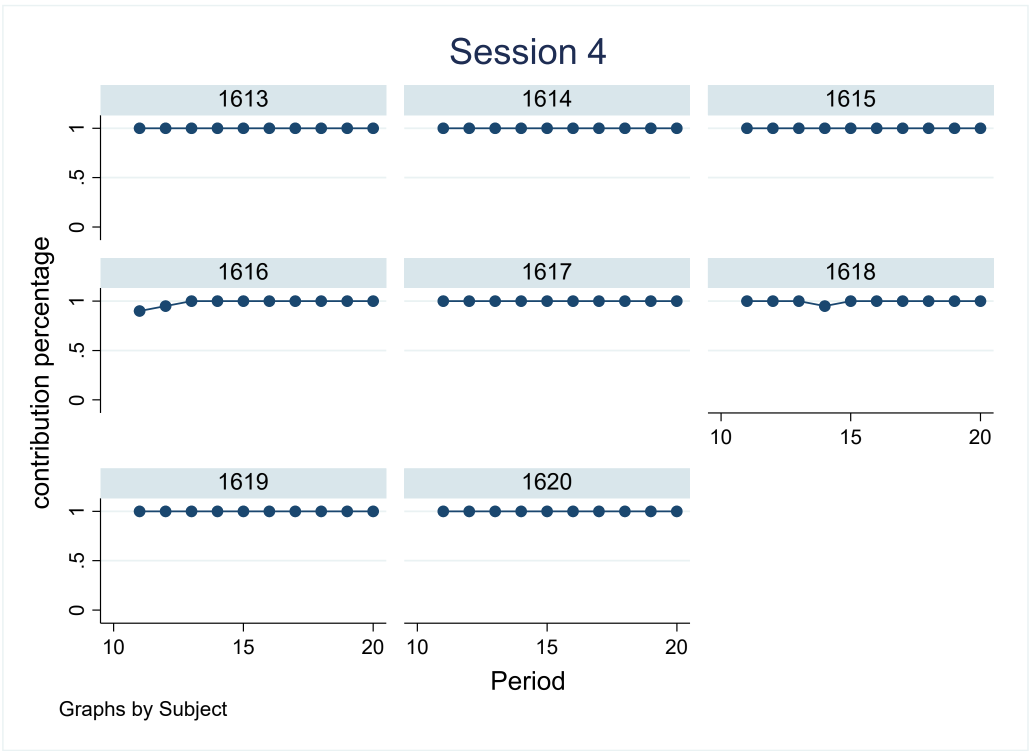


Figure C1. Individual contribution plots by session in the Coordination treatment

Table C1. Determinants of contribution (Coordination treatment): Multilevel mixed-effects linear estimation

Dependent variable: Contribution percentage				
	(1)	(2)	(3)	(4)
Period	-0.00275	-0.00330	-0.00283	-0.00318
	(0.00257)	(0.00258)	(0.00263)	(0.00262)
Punishment_t-1	-0.0267***	-0.0387***	-0.0271***	-0.0384***
	(0.00958)	(0.0124)	(0.0101)	(0.0124)
If the lowest contributor_t-1		0.0351		0.0368
		(0.0229)		(0.0238)
If the highest contributor_t-1			-0.00263	0.00482
			(0.0177)	(0.0183)
Constant	0.847***	0.851***	0.849***	0.846***
	(0.0546)	(0.0548)	(0.0570)	(0.0571)
Observations	396	396	396	396
Log. Likelihood	167.7	168.9	167.7	168.9
Chi-squared	8.407	10.78	8.429	10.85

Notes: Standard errors are in parentheses. Observations are clustered by subject.

Data includes periods 12–20 where the hired-gun mechanism is in place.

Variables: *Punishment_t-1*: the punishment amount experienced in the last period; *If the lowest contributor_t-1:* a dummy variable taking value 1 if the contribution in the previous period was the group's minimum and the group did not tie at full contribution, and 0 otherwise; and *If the highest contributor_t-1:* a dummy variable taking value 1 if the contribution in the previous period was the group's maximum and the group did not tie at full contribution, and 0 otherwise.

* p < 0.10, ** p < 0.05, *** p < 0.01