Submission Number: PET11-11-00130

Sustaining cooperation in social dilemma: Comparison of centralized punishment institutions

Yoshio Kamijo
Waseda Institute for Advanced Study (WIAS)

Tsuyoshi Nihonsugi
Graduate School of Economics, Waseda University

Ai Takeuchi
Faculty of Political Science and Economics, Waseda University

Yukihiko Funaki
Waseda University

Abstract

This paper investigates two centralized punishment institutions for a linear public goods game. These institutions require a certain level of contribution, and sanction those players who contributed less than the requirement. The two differ in whom, among those who do not meet the requirement, receive sanctions: In one institution, all the violators are sanctioned, and in the other, only the worst violator(s) are sanctioned. Theoretically, the public goods game with latter institution yields contributions that are equal to or greater than the former institution with the same requirement and sanction level. The results of an experiment supported this theoretical prediction.

Submitted: February 28, 2011.

Sustaining cooperation in social dilemma: Comparison of centralized punishment institutions

Yoshio Kamijo, Tsuyoshi Nihonsugi, Ai Takeuchi, Yukihiko Funaki

January, 2011

Abstract

This paper investigates two centralized punishment institutions for a linear public goods game. These institutions require a certain level of contribution, and sanction those players who contributed less than the requirement. The two differ in whom, among those who do not meet the requirement, receive sanctions: In one institution, all the violators are sanctioned, and in the other, only the worst violator(s) are sanctioned. Theoretically, the public goods game with latter institution yields contributions that are equal to or greater than the former institution with the same requirement and sanction level. The results of an experiment supported this theoretical prediction.

JEL classification: C72, C91, C92, K42.
Keywords: Linear public goods game, punishment institutions, experiment

1 Introduction

From the vast amount of research on the public goods game experiment, it became a shared wisdom that a costly personal punishment is effective in reducing free riders.[1] In earlier works of Yamagishi (1986, 1988), a Novel prized work of Ostrom, Walker and Gardner (1992), and seminal works of Fehr and Gächter (2000, 2002), a drastic increase in the average contribution to the public goods is observed with an introduction of a punishment stage, where each individual can reduce the payoff of others at his own cost, after a voluntary provision to public goods. This observation in the personal or decentralized punishment is replicated by several authors with several modifications and we are now in the place to "apply the lesson learned to 'field' setting in decentralized institution that deal with social dilemma." (Chaudhuri 2010)

Despite the possible applications including managing natural resources and labor relations, there exist many fields where the application of the decentralized punishment is not valid. One major example is a criminal offense. Another is when participants are in a competitive relationship (e.g., international relations, regulations of firm's emission of carbon dioxide). In addition, there is a finding that when a possibility for counter-punishment exists, cooperators punish less, resulting in low contribution in the provision stage (Nikiforakis 2008). This implies that even in situations where decentralized punishment is valid, individuals who are entitled not to be punished, like a police officer, plays an important role. Therefore, the importance of the centralized punishment institution should be also emphasized.

We can go back to Becker (1968) on economic analyses of crime and penalty when we consider a centralized punishment institution (for a survey, see Polinsky and Shavell 2000). In Becker's analyses, a potential offender compares the benefit from criminal act with the expected loss—the conviction

[1] For a survey, see Chaudhuri (2010). For a survey on the experiments on the public goods game before 1995, see Ledyard (1995).

probability multiplied by the disutility of punishment (fine, imprison, exclusion, etc.). If the former dominates the latter, he commits a crime. In the context of public goods game, a decision maker decides to free-ride when the gain from free-riding is greater than the expected amount of punishment. A difference between the analysis of crime by Becker (1968) and public goods game is that the latter is formulated as a strategic game so the payoff of an individual depends not only on his own decision but also on the decision of others.

This paper investigates the centralized institutions that punish the free-rider in the public goods game. Following the convention of the institutional analysis of economic regulation, the centralized institutions involve two factors: a required level of performance and a fine. An individual whose contribution does not meet the requirement pays a fine. Here, all the individuals who violate the "rule" will be punished. However, this is an ideal situation. In the real field, such ideal does not hold in many cases, due to resource constraints of enforcement agencies. Even if there are many individuals who violate the law or does not meet the requirement of regulation, the enforcement agency can detect and/or apprehend only some of them. The following two scenarios exemplify such situations. First, imagine a situation where a police officer finds several cars running in excessive speed. Then, it is readily understood that the police officer cannot arrest all of them but only one or some of them. Probably, the officer will pull in the one(s) that is worst speeding among them. Second, imagine a policeman on patrol for illegal parking. Then, due to the time limitations, the police will probably check from more congested areas to less, so only some of parking violators get fined. These examples indicate the difficulty of strictly punishing all offenders and in practice, only the distinguished offender(s) will be punished. This implies that there exists a strategic interdependence among the offenders on whether they are punished or not.

This study investigates the centralized institution that is ideally enforced and one that is limitedly enforced and thus creating the strategic environment to the potential violators. We formulate these two as an absolute and a relative institution, respectively. In an absolute punishment institution, all individuals whose contribution is less than the required threshold will pay a fine. This corresponds to the ideal enforcement of the centralized institution. On the other hand, in a relative punishment institution, among the individual whose contribution does not meet the required level, only the minimum contributor(s) is punished. The relative punishment institutions can be seen not only as inadequately enforced absolute institutions but also as institutions intentionally structured to make use of the strategic interdependence of violators like leniency for detecting cartel.

We first analyze the theoretical properties of the absolute and relative punishment institutions. We find that fixing the fine, when the requirement of a regulation is smaller than some critical level, every individual contributes the required level in a unique Nash equilibrium in both institutions. In contrast, when the requirement of regulation is larger than the critical level, every individual contributes less than the required level in both institutions. An intuition behind this result is that comparing the payoff of free-rider and the payoff of law-abider, the latter is greater than the former when the level of regulation is small, and this relation is reverse when the level of regulation is beyond some critical value. From the theoretical analysis, fixing the fine, we readily obtain the optimal requirement of regulation in both institutions. The optimal level is the same between the two institutions. It is determined by the marginal value such that if the required level is beyond this value, every individual has no incentive to abide the regulation in both institutions.

A difference between the two institutions arises when the required level of regulation is greater than the critical value. In this case, while complete free-riding is the dominant strategy for each individual in the absolute punishment institution, choosing positive amount of contribution with a positive probability is the mixed strategy Nash equilibrium in the relative punishment institution. In the relative institution, individuals contribute positive amount because the strategic interdependence of sanction creates an environment where the expected probability of sanction rises with the extent of free-riding. Similar to

marginal deterrence of Stigler (1970) where the probability of sanction increases with the amount of harm and thus "those who are not deterred from committing harmful acts have a reason to moderate the amount of harm that they cause," (Polinsky and Shavell 2000, p63) every individual who faces the high requirement in the relative institution moderates his contribution instead of complete free-riding.

This result on the relative punishment institution with high requirement may be an answer to the reason why in real life, people obey laws even when the sanction is not high enough (or the requirement is not low enough) to prevent people from free-riding. In Tyran and Feld (2006), this point is explained by the effect of referendum. They find in a laboratory experiment that if participants choose a centralized punishment institution with mild sanction through a referendum, they tend to cooperate even if the level of sanction is not enough high to prevent people from free-riding. Our theoretical analysis provides another explanation. If people regard an environment as a relative punishment institution rather than the absolute punishment institution, even the self-regarding individuals contribute to some extent rather than free-ride.

In the later part of this paper, we examine the theoretical prediction for the two centralized punishment institutions by means of laboratory experiment. It is well known from number of experimental results that the theoretical results in the public goods game do not always coincide with the observations in the laboratory. We investigated whether the theoretical results would hold in the laboratory. For each institution, we choose three sets of parameters, called Low, Middle, and High, which differ in the value of requirement. The three sets of parameters are chosen so that (1) while in Low and Middle, contributing the requirement is an equilibrium, contributing less than the required level is an equilibrium in High, and (2) in Middle, the requirement is optimal. These three parameter sets are common for two institutions and thus we have $6\ (=2\times 3)$ treatments.

The main results of the experiment are summarized by noting the two suggestive observations. First, when the requirement was high, there were more contribution to the public goods when the punishment institution was relative than when it was absolute. This supports the theoretical prediction, and our previous argument: Contrarily to our intuitions, people may cooperate even when the sanction is non-deterrent, because the apprehension is limited and the institution is not fully enforced. Second, the efficiency gained in the middle treatment, where the requirement level is set to yield the optimal outcome were not the highest among the three parameters under both institutions. In both institutions, contributions observed in the treatments with Middle requirement were lower than the equilibrium level, and were declining with repetition. As a result, in the latter rounds, profits in treatment with Low requirement were higher than that in the Middle requirement. We discuss the possible causes of this discrepancy with theory.

The rest of the paper is organized as follows. In the next section, we present a model of voluntarily provision of public goods with centralized punishment institutions and the theoretical predictions from an equilibrium analysis. In Section 3, we explain the design of an experiment and its parameter selections. In Section 4, we provide experimental results on the absolute punishment institutions and the relative punishment institutions, and discuss how and why the results deviated from the theoretical predictions. In Section 5, we conclude the paper with policy implications obtained from our theoretical and experimental analysis to the punishment institutions.

2 Model

2.1 Basic setup

We consider a usual symmetric linear voluntary contribution game among n participants. Let $N=\{1,...,n\}$ be the set of n participants. Each participant has an e unit of resource as his endowment, which will be divided into either his private account or a contribution to a public project. If participant

$i \in N$ contributes c_i unit of resource to the project, this is multiplied by α and equally distributed to all the society members, and the remaining $e - c_i$ is kept for his private use.

Let $E = \{0, 1, ..., e\}$ be the set of contribution levels of a participant. For each i, let $c_i \in E$ be his contribution and $c = (c_1, c_2, ..., c_n) \in E^N$ be the contribution profile of all participants. Then, i's utility from the voluntary contribution game at contribution profile c is

$$\begin{aligned} v_i(c) &= e - c_i + \frac{\alpha}{n} \sum_{j \in N} c_j \\ &= e - c_i + \beta \sum_{j=1}^{n} c_j, \end{aligned}$$

where $\beta := \frac{\alpha}{n}$ is a marginal per capita return (MPCR) of the voluntary contribution. Throughout the paper, we assume $0 < \beta < 1$.

A more useful representation of i's utility is

$$v_i(c) = e - (1 - \beta)c_i + \beta \sum_{j \in N, j \neq i} c_j. \tag{1}$$

From this, it is clear that given the contribution of others, participant i loses $(1 - \beta)c_i$ unit of his utility by contributing c_i to the public project. The positive contribution lowers the utility for the participant and the marginal loss of contribution is $(1 - \beta)$ in this linear voluntary contribution game. From this insight, it is readily understood that when all participants decide their levels of contribution voluntary and independently, selecting zero contribution is the dominant strategy of this one-shot game for each player. Therefore, "all participants do not contribute any positive level of resource" is a unique Nash equilibrium of this game.

A number of researches examine this theoretical prediction by laboratory experiments. Main findings can be summarized as follows: in one-shot version of the public goods game, contributions are above the theoretical predictions; whereas when the game is repeated, contributions often starts out between 40% to 60% of the full contribution, and decrease steadily over time, approaching zero contribution (Ledyard 1995). This decline in the contributions with repetition was improved upon by the introduction of personal costly punishment. Many researches across fields investigated the effects of such decentralized punishment, and this tendency was repeatedly observed (see, for example, Yamagishi 1986, 1988, Ostrom et al. 1992, Fehr and Gächter 2000, 2002). Although the impact of personal punishment on cooperative behavior is of importance, there are many areas where application of such personal punishment is inadequate. In such cases, centralized punishment institution may play an important role in enhancing cooperation. Therefore, in this study, we investigate the characteristics of two types of centralized punishment institutions.

2.2 Centralized punishment institutions

We present the two types of centralized punishment institution, both of which entail a threshold level $s \in E$ and an amount of sanction $P > 0$.

The first centralized punishment institution is that given the pre-determined value of threshold s, any participant whose contribution is less than s is punished and receives the sanction P. We call this (P, s)-*absolute punishment institution* and denote this by a notation $G^A(P, s)$. The final payoff of a

player of $G^A(P,s)$ is given as follows: for $c \in E^N$,

$$u_i^A(c) = \begin{cases} v_i(c) - P & \text{if } c_i < s, \\ v_i(c) & \text{otherwise.} \end{cases}$$

Let $L(c;s) \subseteq N$ be the set of participants that contribute less than s. By definition, (P,s)-absolute punishment institution requires that all members in $L(c;s)$ should pay fines. This can be seen as a rigorous application of the punishment institution and at the same time is an institution in an ideal state. We know, however, that in practice, the rule of punishment is mildly applied because the complete application of the rule is essentially impossible in many cases. The second centralized punishment institution follow this view.

The second centralized punishment institution is that given the predetermined value of threshold s, the participant whose contribution is less than s and lowest in the society is punished and receives the sanction P. We call this (P,s)-*relative punishment institution* and denote this by $G^R(P,s)$. Let $B(c) = \arg\min_{i \in N} c_i$. The final payoff of a player of $G^R(P,s)$ is given as follows: for $c \in E^N$,

$$u_i^R(c) = \begin{cases} v_i(c) - P & \text{if } i \in B(c) \cap L(c;s), \\ v_i(c) & \text{otherwise.} \end{cases}$$

The relative punishment institution is an extreme form of a punishment institution in the real world where the enforcement agency faces the resource constraint and thus cannot arrest all violators. The enforcement agency generally spend more effort to punish the distinguished violators. For instance, law enforcement exerts its power to the investigation for the heavy criminal (murder, etc.) compared to lessor offense. The tax office is likely to investigate the tax avoidance of large companies, compared to small ones. These are results of maximization behavior of the enforcement agency with resource constraint and as the result, the probability of being sanctioned increases as the extent of violating the law or regulation in relative to others behavior.

Throughout the paper, we assume that $P > 2(1-\beta)$, or equivalently, $\frac{P}{1-\beta} > 2$. Thus, the amount of sanction is greater than the loss from 2-unit contribution.

2.3 Theoretical prediction

In this subsection, we analyze the two centralized punishment institutions by applying a Nash equilibrium.

Let $P > 0$ and $s \in E, s > 0$. We first consider $G^A(P,s)$. The following proposition indicates that except for some degenerate case, each player has a dominant strategy in the absolute punishment institution.

Proposition 1. *Let $i \in N$. In $G^A(P,s)$,*

(i) when $s < \frac{P}{1-\beta}$, $c_i = s$ is the dominant strategy for i,

(ii) when $s > \frac{P}{1-\beta}$, $c_i = 0$ is the dominant strategy for i, and

(iii) when $s = \frac{P}{1-\beta}$, $c_i = 0$ and $c_i = s$ are perfectly equivalent strategies and dominate any others.

Proof. For any $i \in N$, we have

$$u_i^A(c) = \begin{cases} e - (1-\beta)c_i - P + \beta \sum_{j \in N, j \neq i} c_j & \text{if } c_i < s, \\ e - (1-\beta)c_i + \beta \sum_{j \in N, j \neq i} c_j & \text{otherwise.} \end{cases}$$

Thus, irrespective of others contributions, $c_i = s$ dominates any c_i greater than s. On the other hand, $c_i = 0$ dominates any c_i in $\{1, 2, ..., s-1\}$. The payoff difference between $c_i = s$ and $c_i = 0$ is

$$-(1-\beta)s + P$$

Thus, $c_i = s$ is the dominant strategy when $-(1-\beta)s + P > 0$; $c_i = 0$ is the dominant strategy when $-(1-\beta)s + P < 0$; and $c_i = 0$ and $c_i = s$ are equivalent and dominate any others when $-(1-\beta)s + P = 0$. □

An intuition of this result is as follows. The benefit of free-riding, compared with contributing the threshold, is $s(1-\beta)$ and the expected loss from free-riding is P since free-riders are absolutely detected and thus punished. If the former is larger (smaller) than the latter, free-riding (law-abiding) is the dominant strategy.

Next, we consider $G^R(P, s)$. The next proposition shows that if $s \leqq \frac{P}{1-\beta}$, in a Nash equilibrium, all players contribute just s unit of resource.

Proposition 2. *In $G^R(P, s)$,*

(i) when $s < \frac{P}{1-\beta}$, $(s, s, ..., s)$ is a unique Nash equilibrium, and

(ii) when $s = \frac{P}{1-\beta}$, $(s, s, ..., s)$ is a unique symmetric Nash equilibrium.

Proof. (i) We first show that if $c = (c_1, c_2, ..., c_n) \neq (s, s, ..., s)$, c is not a Nash equilibrium.
Case 1: Suppose that there exists $i \in N$ such that $c_i = m > s$. The payoff of i is $u_i^R(c) = e - m(1-\beta) + \beta \sum_{j \in N, j \neq i} c_j$. If i changes his contribution to $m-1$, the difference in his payoff is

$$(1-\beta) > 0.$$

Thus, i's payoff is improved.
Case 2: Suppose $L(c; s) \neq \emptyset$. Then, there exists some $i \in L(c; s) \cap B(c)$. Let $c_i = m < s$. Consider that i changes his contribution from m to s, keeping others contribution fixed. Then, the payoff difference is

$$-(s-m)(1-\beta) + P > 0.$$

Thus, i's payoff is improved.

Next, we will show that $(s, s, ..., s)$ is a Nash equilibrium. If all players contribute s, the i's payoff is $u_i^R(s, ..., s) = e - s(1-\beta) + \beta(n-1)s$. Consider that i changes his contribution to $s+a$, $0 < a \leqq e-s$. Then, the difference in his payoff is

$$-a(1-\beta) < 0.$$

On the other hand, if i changes his contribution to $s-a$, $0 < a \leqq s$, the difference in his payoff is

$$a(1-\beta) - P.$$

Since $a \leqq s$, we have

$$a(1-\beta) - P \leqq s(1-\beta) - P \leqq 0.$$

Therefore, i cannot improve his payoff by changing his contribution from s.

(ii) From the proof of (i), it is enough to show that for $m < s$, $(m, m, ..., m)$ is not a Nash equilibrium. This is easily verified as follows. For $i \in N$, consider that i changes his contribution to $m+1$. Then, his utility increases by $P - (1-\beta) > 0$. □

From Propositions 1 and 2, we know that when the required level of regulation or law is not so high, the equilibrium behaviors for the two punishment institutions are the same, everyone contributing the required level. However, when the level of threshold is high, the equilibrium behaviors in the two institutions are quite differing. The following proposition shows that if the value of threshold is higher than $P/(1-\beta)$, there is no pure strategy Nash equilibrium in $G^R(P,s)$.

Proposition 3. *When $s > \frac{P}{1-\beta}$, there exists no pure strategy Nash equilibrium in $G^R(P,s)$.*

Proof. We first show that there exists no pure strategy *symmetric* Nash equilibrium when $s > \frac{P}{1-\beta}$. From the proof of Proposition 2, we know that for $m \neq s$, $(m, m, ..., m)$ is not a Nash equilibrium. Thus, it is enough to show that $(s, s, ..., s)$ is not a Nash equilibrium.

Suppose that all players contribute s. The payoff difference of i when he changes his contribution to 0 is

$$s(1-\beta) - P.$$

By the condition of this proposition, this is positive. Thus, the i's payoff is improved.

Next, we show that there exists no pure strategy Nash equilibrium. We prove this by the way of contradiction. Assume that $c = (c_1, c_2, ..., c_n)$ is a pure strategy Nash equilibrium such that $c_1 = c_2 = ... = c_n$ doesn't hold.

By the same proof of Case 1 in Proposition 2, c_i is less than or equal to s for all $i \in N$. Let $i \in \arg\min_{j \in N} c_j$. Then, $c_i < s$ because c is not a symmetric strategy profile. Then, c_i must be 0 because otherwise, i can improve his payoff by changing his contribution from c_i to 0. Since $c_i = 0$, the best response of $j, j \neq i$ is to choose $c_j = 1$. However, if any player other than i contributes 1, i can improve his payoff by changing his contribution from 0 to 2 because $P > 2(1-\beta)$. This means that there exists no pure strategy Nash equilibrium. □

Proposition 3 indicates that we have to consider a mixed strategy Nash equilibrium to obtain some prediction for (P,s)-relative punishment institution when $s > \frac{P}{1-\beta}$. The mixed strategy of $i \in N$, denoted by $q_i \in [0,1]^E$, assigns to each pure strategy $k \in E$ the probability of k being chosen. For $k \in E$, let $q_i(k)$ denote the probability assigned to pure strategy k by the mixed strategy q_i. Thus, $\sum_{k \in E} q_i(k) = 1$ and $q_i(k) \geqq 0$ for any $k \in E$ must hold. Let $(q_1, q_2, ..., q_n)$ be the profile of the mixed strategies of all players.

Proposition 4. *Assume that $s > \frac{P}{1-\beta}$. For some integer $m \leqq \frac{P}{1-\beta}$, define $\hat{q}(k)$ for any $k \in E$ as follows.*

- *for all $k = 0, ..., m-1$,*

$$\hat{q}(k) = \left(1 - \frac{k(1-\beta)}{P}\right)^{\frac{1}{n-1}} - \left(1 - \frac{(k+1)(1-\beta)}{P}\right)^{\frac{1}{n-1}}.$$

- $\hat{q}(m) = 1 - \sum_{h=0}^{m-1} \hat{q}(h)$
- $\hat{q}(k) = 0$ *for all $k = m+1, ..., e$.*

If m is the integer satisfying $\frac{P}{1-\beta} - 1 \leqq m \leqq \frac{P}{1-\beta}$, then $(\hat{q}, \hat{q}, ..., \hat{q})$ is a mixed strategy Nash equilibrium. Moreover, there is no other symmetric mixed strategy Nash equilibria.

Proof. Let $(q, q, ..., q)$ be a symmetric mixed strategy Nash equilibrium. The proof can be divided into the following steps.

Step 1. For any $k > \frac{P}{1-\beta}$, $q(k) = 0$.

When $k > \frac{P}{1-\beta}$, $c_i = 0$ dominates $c_i = k$ because gain from refraining the contribution, $k(1-\beta)$, is greater than the loss from punishment, P.

Step 2. There don't exist two integers k and k' in E such that $k < k'$, $q(k) = 0$ and $q(k') > 0$.

Assume in negation that there exist such k and k'. Let $j = \arg\min\{h : q(k) > 0, h = k+1, 1, , ..., e\}$. By assumption and Step 1, $j \geqq k+1$ and $j \leqq \frac{P}{1-\beta}$. Since all players follow the mixed strategy q, there is positive probability, say $\eta > 0$, of being punished when a player contributes j unit of resource. On the other hand, if a player chooses k contribution, he obtains the gain from refraining contribution, $(j-k)(1-\beta)$, compared to choosing j contribution, while the probability of being punished is not changed. Therefore, the player becomes better off by modifying his mixed strategy.

From Steps 1 and 2, we know that there exists some $m \leqq \frac{P}{1-\beta}$ such that $q(k) > 0$ for any $k = 0, 1, ..., m$ and $q(k) = 0$ for any $k = m+1, m+2, ..., e$.

Step 3. A symmetric mixed strategy Nash equilibrium, $(q, q, ..., q)$, must be $(\hat{q}, \hat{q}, ..., \hat{q})$ defined in this proposition for $m \leqq \frac{P}{1-\beta}$.

Let a be an expected contribution of a player who follows mixed strategy q. Let $Q(k)$ and $Eu(k)$ be the probability of a player being punished and the expected payoff of a player, respectively, when he contributes k and other players follow mixed strategy q. By simple calculation, we have

$$a = \sum_{h=0}^{m} hq(h),$$

$$Q(0) = 1 \text{ and } Q(k) = (1 - \sum_{h=0}^{k-1} q(h))^{n-1} \text{ for any } k = 1, 2, ..., m, \text{ and}$$

$$Eu(k) = e - (1-\beta)k + \beta(n-1)a - Q(k)P \text{ for any } k = 0, 1, ..., m.$$

In a mixed strategy Nash equilibrium, the pure strategies that are assigned positive probability by the mixed strategy must give the same expected payoff. Therefore, we have

$$Eu(0) = Eu(1) = ... = Eu(m).$$

From these equations, we are forced to have $q = \hat{q}$.

Step 4. $(\hat{q}, \hat{q}, ..., \hat{q})$ is a mixed strategy Nash equilibrium if and only if m is the integer satisfying $\frac{P}{1-\beta} - 1 \leqq m \leqq \frac{P}{1-\beta}$.

To show that $(\hat{q}, \hat{q}, ..., \hat{q})$ is a mixed strategy Nash equilibrium, it suffices to compare the expected payoff of a player at $(\hat{q}, \hat{q}, ..., \hat{q})$ with the payoff of that player wherein he chooses $m+1$ contribution and others follow $\hat{q}$. We know that the expected payoff of the player at $(\hat{q}, \hat{q}, ..., \hat{q})$ is $Eu(0) = e + (n-1)\beta a - P$ and the expected payoff at the latter case is

$$e - (1-\beta)(m+1) + (n-1)\beta a.$$

Thus, the former payoff is greater than or equal to the latter payoff if and only if

$$m + 1 \geqq \frac{P}{1-\beta}.$$

Since m satisfies $m \leqq \frac{P}{1-\beta}$, we obtain the desired result.

From Steps 3 and 4, this proposition is proved. □

A remark of this proposition is that if $\frac{P}{1-\beta}$ is not an integer, m in this proposition is the maximal

integer less than $\frac{P}{1-\beta}$, and thus, $(\hat{q}, \hat{q}, ..., \hat{q})$ is the unique symmetric mixed strategy Nash equilibrium. If $\frac{P}{1-\beta}$ is an integer, there exist two symmetric mixed strategy Nash equilibria (one for $m = \frac{P}{1-\beta}$ and the other for $m = \frac{P}{1-\beta} - 1$).

This proposition indicates that when $s > \frac{P}{1-\beta}$, players' behavior in the relative punishment institution is completely different from the one in the absolute punishment institution. While perfect free-riding is the dominant strategy in the latter institution, players choose the contribution levels from zero to $\frac{P}{1-\beta}$ with a positive probability in the former. The reason for the non free-riding in the relative punishment institution would be explained by looking at the probability of sanction for each contribution level. When every players follow $\hat{q}$ described in Proposition 4, the probability of receiving a sanction when a player chooses contribution level $k, 0 \leqq k \leqq m$, is

$$\hat{Q}(k) = (1 - \sum_{h=0}^{k-1} \hat{q}(h))^{n-1} = 1 - \left(\frac{1-\beta}{P}\right) k.$$

This means that the expected probability of receiving a sanction rises with the extent of free-riding. Similar to marginal deterrence by Stigler (1970) where the probability of sanction increases with the amount of harm and thus "those who are not deterred from committing harmful acts have a reason to moderate the amount of harm that they cause," (Polinsky and Shavell, 2000, p63) every individual who faces the high requirement in the relative institution moderates his contribution instead of complete free-riding.

An expected level of contributions of i who follows the mixed strategy $\hat{q}$ is

$$\hat{a} = \sum_{k=0}^{m} k\hat{q}(k) = \sum_{k=1}^{m} \left(1 - \frac{k(1-\beta)}{P}\right)^{\frac{1}{n-1}}. \tag{2}$$

Clearly, this is positive and smaller than $m \leqq \frac{P}{1-\beta}$.

2.4 Optimal threshold

In this subsection, we consider optimal mechanisms based on the theoretical results in the previous section. We focus on the level of optimal threshold s for two centralized punishment institutions, given the amount of sanction P. There are two reasons why we focus only on the threshold. One is that in the real field, the amount of sanction is determined from many perspectives other than economic performance. Choosing P from the viewpoint of an economic performance is often controversial. Another is that if any amount of sanction is allowed, by imposing a tremendous amount of fine, any level of requirement is easily attained and thus any level of performance is possible. This is unrealistic.

The following proposition shows that the optimal levels of the threshold are uniquely determined for the two punishment institutions. The optimal thresholds for the two institutions are the same value, and the two institutions with optimal threshold demonstrate the same performance.

Proposition 5. *Assume that $\frac{P}{1-\beta}$ is not an integer. Let m^* be the maximal integer less than $\frac{P}{1-\beta}$.*

(i) $G^A(P, m^*)$ gives the highest equilibrium payoff of a player among the absolute punishment institutions with sanction P and any threshold s.

(ii) $G^R(P, m^*)$ gives the highest equilibrium payoff of a player among the relative punishment institutions with sanction P and any threshold s.

(iii) Equilibrium payoffs of a player in $G^A(P, m^*)$ and $G^R(P, m^*)$ are the same.

Proof. The proof of (i), (ii) and (iii) of this proposition is readily obtained from Propositions 1, 2 and 4. Thus we omit the proof. □

In the later part of the paper, we investigate the two centralized punishment institutions by means of laboratory experiment. We test whether these theoretical results would hold in the experiment, especially, we check whether the optimal mechanism is really optimal in a laboratory, and how the subjects in the two institutions choose their level of contribution, with respect to the threshold level.

3 Experimental design

In this section, we explain our experimental design. We conducted the experiment in October 2010 at Computer Laboratory of Waseda University in Japan.

3.1 Subjects

Our subjects were 184 undergraduate students (76 females; mean age of 20.4 years) from various disciplines. All subjects were recruited from Waseda University by via the internet. Written informed consent was obtained from all subjects.

3.2 Tasks and Procedures

We conducted six experimental treatments explained below. Twenty-eight or thirty-two different subjects participated in each treatment. In all treatments, at the beginning of experiment, subjects were randomly assigned to their booths in the laboratory. The booths separated the subjects visually and ensured that every individual made his or her decision anonymously and independently. Subjects were provided with written instruction explained the game, payoffs, sanction rule, and procedures, and read it on their computer screen at their own pace. Instruction used neutral wording, as is common practice in experimental economics. After the instruction, subjects were tested to confirm that they understood the rules and how to calculate their payoffs. We did not start the experiment until all participants had answered all questions correctly. Therefore, all subjects completely understood the rules of this game and were able to readily calculate their payoffs.

After test, the subjects were randomly and anonymously allocated to groups of size $n = 4$ and these groups played linear public goods game with centralized punishment institution for 15 periods. Group composition remained the same throughout 15 periods (so called "partners design"). Each group member was endowed with $e = 24$ points in each period. Also, each group member was assigned a new identification number (1, 2, 3 or 4) in each period in order to eliminate the effect of reputation. Then, each group member had to decide on how many of 24 points to keep and how many points to contribute to a public good on their computer screen. All members simultaneously made this decision. Each subject's income from the public good was the sum of contributions by all group members to the public good, multiplied by $\beta = 0.35$. Every subject had the same payoff function and every subject knew this fact. After their decisions, the results of the period: each member's contribution points, sum of group member's contribution points, each member's outcome, and who received the punishment $P = 12$ points, appeared on their computer screen (Figure 1). After experiment, all subjects returned the questionnaire.

We used z-Tree (Fischbacher 2007) to conduct the experiments. Each session required approximately 1.25 hours on average to complete. The mean payoff per subject was \$19.86 (\$1 = 85yen). The maximum payoff was \$23.76, and the minimum payoff was \$12.00. Average earnings exceeded the average hourly wage of a typical student job in the location of Waseda University.

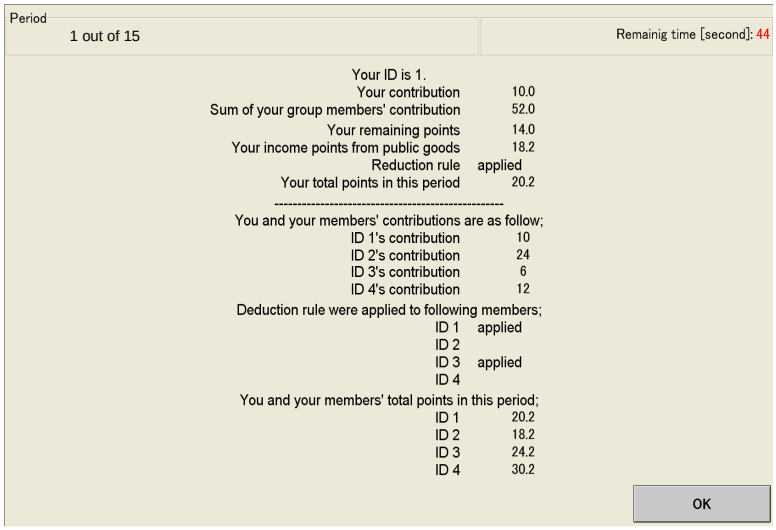

Figure 1: An example of feedback screen (from ABS-L)

3.3 Treatments and theoretical predictions

Our experiment consisted of six treatments. We modulated two conditions: punishment institution and threshold level. There were two punishment institution; "Absolute punishment institution (ABS)" and "Relative punishment institution (REL)" as described in the previous section. Also, there were three threshold levels $s = 12, 18$ and 24 points, we called these threshold levels "Low (L)", "Middle (M)" and "High (H)", respectively. By doing so, we could investigate the effects of centralized punishment institution by threshold level. We called our treatments ABS-L, ABS-M, ABS-H, REL-L, REL-M, and REL-H (Table 1).

Table 1: Experimental treatments

		Condition 2: threshold level		
		Low (L) $s = 12$	Middle (M) $s = 18$	High (H) $s = 24$
Condition 1: punishment institution	Absolute (ABS)	ABS-L	ABS-M	ABS-H
	Relative (REL)	REL-L	REL-M	REL-H

Regarding our theoretical predictions from Section 2, with our parameters of the experiment, the critical value which determines whether the sanctions are or are not deterrent is $P/(1-\beta) \approx 18.46$. Therefore, it is obvious that the contribution in L and M threshold treatments are 12 and 18 respectively for both institutions. When the threshold is H, the prediction differs in the two institutions. For ABS-H, it is 0; and for REL-H, it is approximately 13.38, which is calculated from Eq. (2). The parameters and theoretical predictions in each treatment are summarized in Table 2.

Table 2: Overview of parameters and predictions in each treatment

	ABS-L	ABS-M	ABS-H	REL-L	REL-M	REL-H
Number of subjects	32	32	28	32	32	28
Group size (n)	4	4	4	4	4	4
Endowment (e)	24	24	24	24	24	24
MPCR (β)	0.35	0.35	0.35	0.35	0.35	0.35
Punishment institution	ABS	ABS	ABS	REL	REL	REL
Amount of punishment (P)	12	12	12	12	12	12
Threshold level (s)	12	18	24	12	18	24
Theoretical prediction of mean contribution	12	18	0	12	18	≈ 13.38

Table 3: Summary statistics per treatment

	ABS-L	ABS-M	ABS-H	REL-L	REL-M	REL-H
Contribution	12.54	15.94	6.55	12.31	12.43	14.55
	(4.39)	(7.9)	(9.83)	(4.54)	(6.36)	(8.37)
Profit	28.21	27.98	17.19	28.05	26.55	26.9
	(3.35)	(6.21)	(6.72)	(3.72)	(4.68)	(5.69)
Total # of Sanctions Imposed	32	96	330	35	97	102

4 Results

The main purpose of this section is to analyze experimental observations of the two centralized punishment institutions and to investigate institutions more capable of sustaining cooperation. We first analyze the effects of thresholds on behavior holding institutions fixed. Then, we compare the two punishment institutions. These analyses of the data demonstrate that the theoretically optimal institution was not optimal. We therefore complete the section by discussing the possible reasons for the discrepancies between the theory and observations.

Before going into each comparison, we provide a table with the summary statistics of all treatments. Table 3 lists the averages and standard deviations of contributions, profits, and total number of imposed sanctions for each treatment.[2] We regard average profits as a measure of efficiency: In this set-up, the maximum and minimum total profit is the same for all treatments, making average profit suitable for measuring efficiency. The maximum attainable average profit is 33.6, obtained when all fully contributes. and the minimum is 12, obtained when all free ride and receive sanction.

[2]A comparison of this table and Table 2 already shows some difference between the theory and experimental observations, both in contribution and profit. However, these differences are not statistically significant. For each treatment, we used Wilcoxon signed-rank test to test whether the median of the group-average contributions is the same as the theoretical prediction. For all treatments, the results were insignificant at the 10% level. The same hold for profits. This may partly due to the limited number of samples: we used group averages over all periods as units of observation for independence of samples (recall that the experiment used partner matching protocol). For most statistical test we conduct, the number of observations equals the number of groups in each treatment. Also, we use two-tailed tests throughout the paper.

4.1 The effect of thresholds in absolute punishment institutions

In absolute punishment institutions, when $s > P/(1-\beta) \approx 18.46$, the sanction is ineffective and it is dominant strategy to free ride and when $s < P/(1-\beta)$, the sanction is effective and it is dominant strategy to contribute s. Theoretically optimal threshold is thus 18, the largest integer less than $P/(1-\beta)$. Thus, we hypothesize the lowest contributions and profits in ABS-H; a similar distribution of contribution, simply shifted with respect to s, in ABS-L and ABS-M; and the largest contributions and profits in ABS-M. The experimental results supported the first but the not the latter two hypothesis.

Observation 1. The difference between the effective and ineffective punishment institutions were as predicted by theory: In ABS-H, contributions and profits were lower than the other two threshold treatments. The differences between the two effective punishment institutions were, however, not as predicted. No significant difference in average contributions and profits existed between the two treatments. Moreover, the two treatments varied in the dynamics of contribution choices. In the last rounds, the profits are higher in ABS-L than ABS-M.

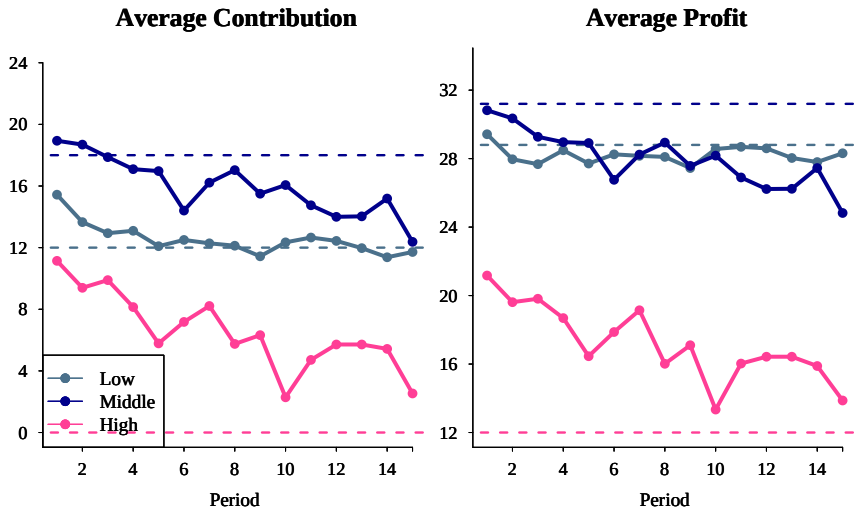

Figure 2: Comparison of contributions and profits across thresholds in absolute punishment institutions (dotted lines are theoretical predictions)

Evidences for Observation 1 are as follows. First, the contributions and profits of ABS-H were lowest among the three treatments. This is clear from Table 3. To test this statistically, we used Kruskal-Wallis test, and tested the null hypothesis that the distribution of groups' average contributions is the same among all treatments. The null hypothesis was rejected with p-value less than 0.001. The same was true for profits. To analyze which of the three treatments were different, we conducted Wilcoxon rank sum test for each pair of treatments with p-adjustment method of Holms. There were significant differences between ABS-H and the other two treatments. For contribution, p-values were 0.008 for low and high comparison, and 0.007 for middle and high. For profit, they were less than 0.001 and equaled to 0.002 respectively.

Further evidence is presented in Figure 2 which plots, separately for each treatment, per period average contribution and profit in the left and right panel, respectively. The dynamics of the contribution and profit in ABS-H is similar to the common observations in public goods game experiments: They start above theoretical prediction and decline over time (see Ledyard, 1995). The Spearman rank or-

der correlation between the per-period average contributions and periods was negative and significant ($\rho = -0.878, p < 0.001$); and similarly for profit ($\rho = -0.845, p < 0.001$). Also, the average contribution and profit in each period of ABS-H were never above those of ABS-L and ABS-M. From these evidences, we state that ineffective punishment institutions cannot sustain cooperation.

Next, there were some unexpected differences between the two effective punishment institutions. In accordance with the theoretical prediction, the mean contributions and profits in ABS-M were higher than ABS-L at the aggregate level (see Table 3). These differences were, however, not statistically significant. The result of abovementioned Wilcoxon rank sum test with p-adjustment was insignificant at 10% level. Although the central tendencies were similar, observed behaviors in the two treatments were distinct in ways different from the theoretical prediction. The comparison of time-trends in Figure **??** reveals this. The mean profit in ABS-L leveled off at the theoretical prediction, whereas in ABS-M, they decay over time, moving away from the theoretical prediction (Spearman rank order correlation; ABS-L: $\rho = 0.07, p = 0.81$;ABS-M: $\rho = -0.875, p < 0.001$).[3] In the last 5 rounds, there is even a reversal in the profit of ABS-L and ABS-M. This difference remains, even if analyzed at the group level: In ABS-L, no large group difference exists; whereas in ABS-M group difference exists, with 4 out of 8 groups contributing the threshold level and others contributing much less. Thus, ABS-L was more effective in sustaining cooperation. We discuss about the possible causes of these difference in section 4.4.

4.2 The effect of thresholds in relative punishment institutions

Compared to the theoretical predictions of absolute, predictions in relative punishment institutions differ only in one way. Even when the sanctions are ineffective, we expect positive contributions in mixed strategy Nash equilibrium. With the parameters in experiment, the expected value of contributions in equilibrium is 12, 18, and 13.38 in REL-L, REL-M, and REL-H respectively. However, the threshold level did not affect the behavior in the relative punishment institution.

Observation 2. In the relative punishment institution, the threshold level does not largely affect the behavior. Difference in the contributions and profits were insignificant between the different thresholds.

Support for Observation 2 comes from Table 3 and Figure 3. First, the comparison of average contributions and profits in Table 3 reveals that the difference between the three treatments in relative punishment institutions are less than those of absolute. Comparing the distribution of per-group average contributions across threshold levels, the null hypothesis that all the distribution is the same cannot be rejected (Kruskal-Wallis test, p-value $= 0.85$). The dynamics of the contributions and profits are depicted in Figure 3. The graphs are intertwined both for contributions and profits. Also, these three treatments share the tendency to decline over periods. The correlation between periods and per-period average contributions was negative and significant for all treatments (Spearman rank order correlation; REL-L: $\rho = -0.86, p < 0.001$; REL-M: $\rho = -0.88, p < 0.001$; REL-H:$\rho = -0.56, p = 0.03$).

The insignificant difference observed between REL-L and REL-H is reasonable: The difference in the theoretical prediction between the two treatments is small.[4] However, contributions observed in REL-M are much lower than the theoretical prediction. Consequently, average contributions were

[3] In ABS-L, the results of Spearman rank order correlation test looks different with contribution. The decline of the first 5 periods drives the overall ρ to be negative and significant ($\rho = -0.746, p = 0.002$). However, when we take the subset of data from period 5 on, the results are back to that of profit: ρ is positive and not significantly different.

[4] Since our theoretical prediction for REL-H is based on mixed strategy Nash equilibrium, we compared the theoretical prediction with the empirical distribution of the contribution in REL-H. With our parameters, the support of the mixed strategy Nash equilibrium $\hat{q}(k)$ is $0, 1, ..., 18$. Subjects in our experiment were contributing more than predicted. There were many contributions above the support of the mixed strategy Nash equilibrium. Especially, there were many threshold contributions.

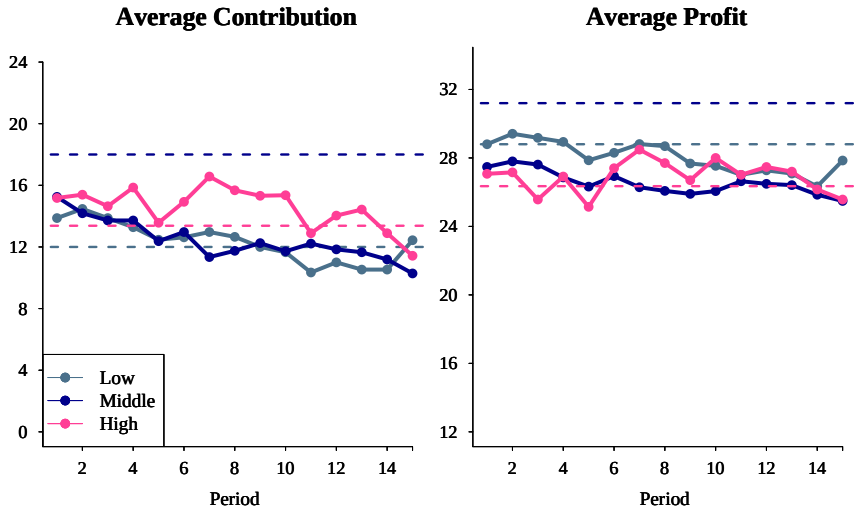

Figure 3: Comparison of contributions and profits across thresholds in relative punishment institutions (dotted lines are theoretical predictions)

highest not in REL-M but in REL-H for most periods, even the average profits in REL-L were constantly higher than that of REL-M in all periods. Observed behaviors in the relative punishment institution with theoretically optimal threshold level were not optimal.

4.3 Comparison of absolute and relative punishment institutions

The absolute and the relative punishment institution with same sanction and threshold level only differ in which violators are punished. One may intuitively think that cooperation is better sustained by arresting all violators. Our theoretical model denies this intuition. When the institution is effective there is no difference between the two, and when it is ineffective relative punishment institutions yield higher contributions than absolute. The experimental results in general support this prediction.

Observation 3. When the sanction is effective, there were no significant differences in the average contributions and profits in the two punishment institutions. When the sanction is ineffective, contributions in the relative punishment institution are higher than the absolute.

Details of the observations are as follows. Let us start with the effective case. Going back to Table 3, controlling for the threshold level, the average contributions, profits, and even the number of punished individuals are similar across the ABS and REL. Contributions are higher in the ABS-M than REL-M, but these differences are not statistically significant using the Wilcoxon rank-sum test. Table 3 also depicts differences for the effective case. Contributions and profits in REL-H were higher than those of ABS-H. Thus, the relationship between the relative and the absolute punishment institutions were as predicted by theory.[5]

When analyzed in further detail, there were some differences between the two institutions.

[5]Some may argue against the relative institutions because the perceived fairness may be low. In the post experiment questionnaire we asked for subjects' evaluation of the punishment rule, using procedural fairness questions of Sondak and Tyler (2007). We had two questions on procedural desirability, procedural justice, procedural desirability, and outcome valence, and one on the effectiveness. We compared the distribution of subjects' answers across punishment institutions, controlling for the threshold level. When the sanction is effective, there were no differences in the distribution of subjects' answers in the absolute and the relative, but when ineffective, there were some differences—the relative punishment institution

Observation 4. In the absolute institutions, most subjects either contributed above threshold or became complete free-rider; whereas in the relative institutions, contributions below threshold were distributed more uniformly between zero and the threshold.

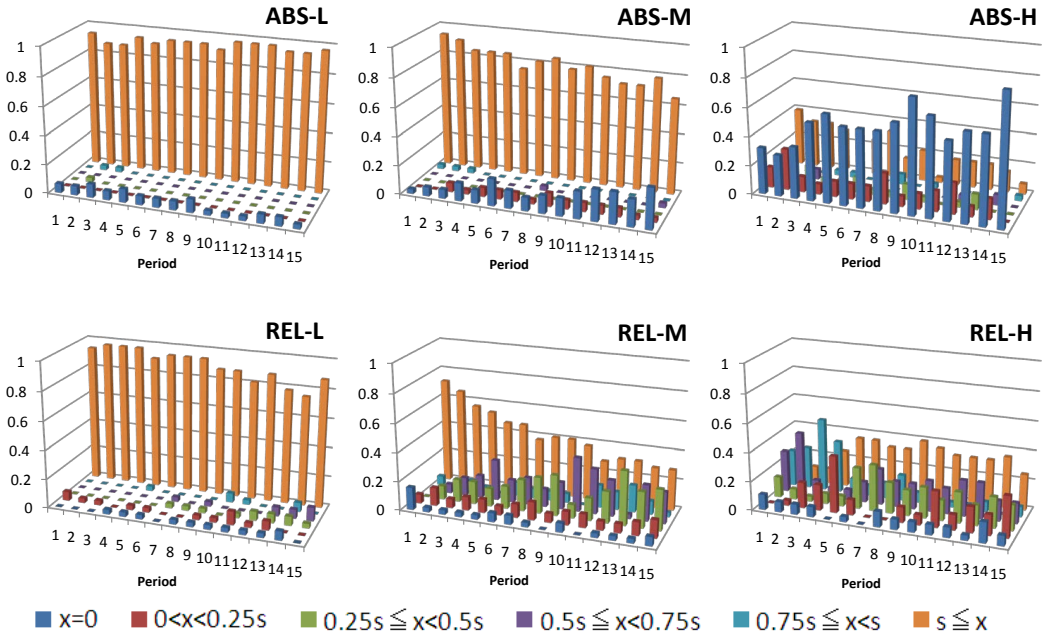

Figure 4: Comparison of the distributions of contributions with respect to threshold

The support of this observation is provided in Figure 4, which plots the frequency of contributions, separated with respect to the threshold. The bars in the furthest back are the frequencies of the above threshold contributions, and the bars in the front are the frequencies of the zero contribution. Most contributions in absolute institutions were either above the threshold or zero, whereas more diverse in relative institutions. The standard deviations of below threshold contributions were about 1.5 times larger in the relative than in the absolute treatment.[6] Another point to note is that when the sanction is effective, the number of above threshold contributions is higher in absolute than the relative punishment institutions. Therefore, there is a tradeoff between absolute and relative institution when the sanction is effective. On one hand, above threshold contributions are observed more frequently in absolute, but violations are mostly complete free-riding. On the other, more contributes less than threshold in relative, but the violations are milder.

4.4 Discussion on the discrepancy between the theory and experimental observations

Experimental results by and large supported theoretical predictions in treatments with low and high thresholds. The discrepancies between the theory and the observations lie in the middle threshold treatment. Theoretically, this is the optimal threshold level for both institutions; however, the average

was perceived as fairer rule compared to the absolute. The difference was statistically significant for 4 out of 7 questions at 5% level. Although the relative punishment institution might seem unfair from the objective point of view, subjects in the institution did not evaluate it as being unfair.

[6]The standard deviations were 2.46 and 3.56 for ABS-L and REL-L; 3.15 and 4.56 for ABS-M and REL-M; and 4.12 and 6.87 for ABS-H and REL-H.

contributions are lower than expected. As depicted in Figure 2 and 3, contribution declines, moving away from the theoretical prediction. This decline raises another key issue, the reversal of the profit in low and middle threshold treatments. The mean profit obtained in REL-L is constantly higher than that in REL-M, and for ABS-L, it is higher than that in ABS-M after the 10th period. Thus theoretically optimal threshold may not be the optimal threshold, and we need to take other behavioral factors into account.

Why did the subjects in low threshold treatment kept on contributing as predicted while the subjects in middle threshold treatment reduced their contribution? This is especially puzzling in absolute punishment institutions where it is a monetary payoff maximizing strategy to contribute the threshold level no matter what the other players contribute (i.e., it is a dominant strategy). Therefore, let us concentrate our discussion on this institution.[7]

Clearly, models with monetary profit maximizing players will not explain the difference in the two thresholds. Also, conditional cooperation and spiteful preference cannot explain the behavioral difference: they both predict a decline in cooperation in face of free-rider.[8] There is one dissimilarity between the low and middle thresholds treatment that, we suspect, caused this behavioral difference. In low, even if a player deviates from equilibrium and become a free-rider, there is no difference in his and the contributors' profit. Both will be 24.9. In middle, however, this is not the case. The free-rider's profit will be 30.9, which is higher than the contributors' profit of 24.6.[9] Therefore, contributors may envy free-riders in the middle threshold, but not in the low. We suspect that an equilibrium where the payoff earned at the deviation from equilibrium is envied from others is unstable in the long run. The dominant strategy equilibrium of ABS-M is of this type, but ABS-L is not. Our rational are as follows: When a contributor observes a deviation in the middle threshold treatment, there are many possible reasons—for example, imitation learning (c.f., Apesteguia et al., 2007) and inequality aversion (c.f., Fehr and Schmidt, 1999)—for the contributors to become a free-rider in the following round, creating a chain of decline in contributions.

Going back to Figure 4, one can see the gradual increase in the number of complete free-rider in ABS-M, which is in stark contrast to ABS-L where the below threshold contribution remains low throughout the experiment. Also, the difference in the average contribution of subjects when there was and was no zero contribution in the previous period is much larger in ABS-M. In ABS-M, the average contribution when there was no zero contribution by other group members was 18.08, but when there was zero contribution by other group members, it drops to 9.07. This difference in ABS-M was

[7]Once there is a free-rider in group and subjects are myopic, there is an incentive to lower contributions in the relative punishment institution. If a contributor changes contribution to a level slightly higher than the free-rider, say ε, they will earn a payoff of $24 - \varepsilon + 0.35(2 \times 18 + \varepsilon) = 36.6 - 0.65\varepsilon$. Unlike in ABS-M where changing the contribution to zero is not monetary profit maximizing action, this will yield higher payoff than contributing the threshold level. This may be one reason why, as depicted in Figure 4, the below threshold contribution are much more frequent in REL-M than ABS-M. Once contribution below threshold is observed in the relative institution, it is unlikely to return to the equilibrium level. There were 2 groups that retained above threshold level contributions throughout the experiment, but in both groups, no subject contributed below the threshold. For the other 6 groups, once they observe one or two contribution below threshold, they could not go back to the equilibrium threshold contribution.

[8]Conditional cooperation is "people's propensity to cooperate (in the lab and field environments) provided others cooperate as well (Fischbacher and Gächter, 2010)." Notice that this definition is purely behavioral based and does not include discussion about people's profits when cooperating or defecting. Therefore, in both low and middle treatment, when there are free-riders in group, conditional cooperators should change their behavior to free-riding. In both low and middle threshold treatments, there are free-riders in early periods, thus if we take the conditional cooperation argument rigorously, we should see a decline in both treatments. Cason et al. (2004) defines a spiteful strategy as a strategy reducing both their own payoff and the other subjects' payoffs in comparison to the payoffs earned when choosing a monetary payoff maximizing strategy. It is a spiteful strategy to free ride in ABS-M, since it reduces their own and the others' payoffs. However, it is also spiteful strategy to free ride in ABS-L, so this would also not explain the difference between the two treatments.

[9]The free-rider would have achieved 31.2 if he had contributed the threshold level, thus it is dominant strategy to contribute.

statistically significant (Wilcoxon rank sum test; $p < 0.001$). In ABS-L, they were 12.56 and 11.23 respectively, and the difference was not statistically significant (Wilcoxon rank sum test; $p = 0.122$).

In sum, these results suggest the importance of designing an institution where the free-riders' grass does not look green from the contributors.

5 Conclusion

We have two main findings: One is derived from a between-institutions comparison, and the other is from within-institution comparisons.

First, when we compare between institutions, both theoretical and experimental results suggest that when the sanction is ineffective, relative punishment institution yields higher performance in the associated public goods game than the absolute punishment institution. This result has two implications. First, it is unnecessary to fully enforce the punishment institution. If the resources are limited, it is not necessary to arrest every violation, but more effort should be devoted into arresting the worst one. Second implication is as additional empirical evidence on the possible explanation for the success of punishment institutions with non-deterrent sanctions. Tyran and Feld (2006) state that there is a "lack of empirical evidence on whether and why a law backed by non-deterrent sanctions ... induces people to abide by the law (p. 136)." Their experimental results suggest that endogenous selection of the institution is one possible reason behind compliance to mild sanctions. Because the ABS-H is a centralized punishment institution with mild sanctions, our model and experimental comparison of ABS-H and REL-H raise another possibility: although the institution is built as being absolute, due to limited enforcement, the actual game being played may be that of relative. Then, as our model and the experiment show, contributions will be higher than that expected in the absolute institution.

Second, by comparing across threshold levels within-institution, we found some discrepancy between the theoretical and the experimental results in the optimal level of threshold. In the treatment with theoretically optimal level of threshold, the number of free-riders increased with repetition in both institutions. This increase in the number of free-riders was not observed in the treatment with lower threshold level, and as a result, the profit in the lower threshold level surpassed that of optimal level in the last few rounds. From this difference in the results, we implied and discussed the importance of a property that "the outcome of the most likely deviation from the equilibrium to be envy-free." In a repeated setting, if this property is not satisfied and a deviation from equilibrium occurs, there are many behavioral reasons for the experimental observations to depart from the equilibrium, such as inequality aversion and imitation learning. Therefore, this property may be important when designing an institution, and is worth further investigations.

We conjecture that the last result may be vulnerable to differences in the information provided. In our experiment, we provided feedback information about the contribution and profit of each individual in the same group. In the experimental literature on the public goods game, observed behavior differs with information provided (see e.g., Bigoni and Suetens, 2010). Our institution is different in the sense that, unlike the public goods games studied in the above mentioned literature it is an equilibrium strategy to contribute. Still, if we only provided the information about the aggregate level of contribution, the result could have differed, especially in the treatment with optimal level of threshold. To sustain cooperation in institution that is supported as equilibrium, hiding the existence of free-rider in group may be one effective strategy. However, to elaborate on this point, further investigation is in need, and is a topic for future research.

References

Becker, G. (1968): "Crime and punishment: an economic approach," *Journal of Political Economy*, 76, 169–217.

Chaudhuri, A. (2010): "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," forthcoming in *Experimental Economics*.

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

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

Ledyard, O. (1995): "Public goods: some experimental results," in *Handbook of experimental economics*, ed. by J. Kagel, and A. Roth. Princeton University Press (Chap 2).

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

Ostrom, E. Walker, J. and Gardner, R. (1992): "Covenants with and without a sword: self-governance is possible," *American Political Science Review*, 38, 45–76.

Polinsky, A.M. and Shavell, S. (2000): "The economic theory of public enforcement of law," *Journal of Economic Literature*, 38, 45–76.

Stigler, G. (1970): "The optimum enforcement of laws," *Journal of Political Economy*, 78, 526–536.

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

Yamagishi, T. (1986): "The provision of a sanctioning system as a public good," *Journal of Personality and Social Psychology*, 51, 110–116.

Yamagishi, T. (1988): "The provision of a sanctioning system in the United States and Japan," *Social Psychology Quarterly*, 51, 265–271.