G-COE GLOPE II Working Paper Series

Sustaining cooperation in social dilemmas: Comparison of centralized punishment institutions

Yoshio Kamijo, Tsuyoshi Nihonsugi, Ai Takeuchi, and Yukihiko Funaki

Working Paper No. 45

Sustaining cooperation in social dilemmas: Comparison of centralized punishment institutions

Yoshio Kamijo[a]*, Tsuyoshi Nihonsugi[b], Ai Takeuchi[c], and Yukihiko Funaki[d]

[a] Waseda Institute for Advanced Study (WIAS). 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, JAPAN. yoshio.kamijo@gmail.com

[b] Graduate School of Economics, Waseda University. 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, JAPAN. t.nihonsugi@gmail.com

[c] Faculty of Political Science and Economics, Waseda University. 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, JAPAN. ai-tak@moegi.waseda.jp

[d] School of Political Science and Economics, Waseda University. 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, JAPAN. funaki@waseda.jp

June, 2011

Abstract

This study investigates two centralized punishment institutions for a linear public goods game. These institutions require a certain contribution level, and sanction those players who under-contribute. The two differ in whom, among those who do not meet the requirement receives 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 the 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. However, there was a discrepancy between the theory and the laboratory observation in that the institution with the theoretically optimal requirement did not yield the highest cooperation.

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

1 Introduction

From the vast amount of research on the public goods game experiment, it became a common wisdom that a costly personal punishment is effective in reducing free riders.[1] In earlier studies by Yamagishi (1986, 1988), a Nobel-prized work by Ostrom et al. (1992), and seminal works by Fehr and Gächter (2000, 2002), a marked increase in the average contribution to the public goods is observed with the introduction of a punishment stage in which each individual can reduce the payoff of others at his own cost, after a voluntary provision to public goods. These observations on the personal (or decentralized) punishment are replicated by several authors with several modifications, bringing us now to the point of "apply(ing) the lesson learned to (the) 'field' setting in decentralized institution that deal with social dilemma(s)." (Chaudhuri, 2011)

Despite the possible applications including managing natural resources and labor relations, many fields exist where the application of the decentralized punishment is not valid. One example is a criminal offense, and another is when participants are in a competitive relationship (e.g., international relations,

*Corresponding Author. TEL: +81-3-5286-2103.

[1] For a survey, see Chaudhuri (2011). For a survey on the experiments on the public goods game before 1995, see Ledyard (1995).

regulations of a firm's emission of carbon dioxide). Furthermore, Nikiforakis (2008) demonstrated that cooperators punish less in the face of counter-punishment possibilities, resulting in low contribution in the provision stage. This implies that even in situations where decentralized punishment is valid, individuals who are exempt from punishment, like police officers, play an important role. Therefore, the importance of the centralized punishment institution should be also emphasized.

We can refer to Becker (1968) for economic analyses of crime and penalty when considering centralized punishment institutions (for a survey, see Polinsky and Shavell, 2000). In Becker's analyses, a potential offender compares the benefit from a criminal act with the expected loss—the conviction 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 a public goods game, a decision maker decides to free ride when the gain from free riding is greater than the expected amount of punishment. One difference between the analysis of crime by Becker (1968) and the 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 study investigates the centralized institutions that punish free riders 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 satisfy the requirement pays a fine. Here, all individuals who violate the "rule" will be punished. However, this is an ideal situation. In the real field, such an ideal does not hold in many cases because of enforcement agencies' resource constraints: Even if many individuals violate the law or fail to meet the regulation requirement, the enforcement agency can apprehend or punish only some of them. For example, imagine a situation where a police officer finds several cars speeding. Then, it is readily understood that the police officer cannot arrest all of them but only one or a few of them. Probably, the officer will pull over only the highest speeding car(s). This example indicates the difficulty of strictly punishing all offenders, and in some cases, punishing only the worst offender(s). In such situations, punishment enforcement depends upon a strategic interdependence among the offenders.

We examine centralized institutions that are ideally enforced and limitedly enforced, formulating these two as absolute and relative punishment institutions, respectively. In an absolute punishment institution, all individuals whose contribution is less than the required threshold will pay a fine. In a relative punishment institution, among the individuals whose contribution does not meet the required level, only the minimum contributor(s) is punished. The relative punishment institutions can be viewed not only as limitedly enforced absolute institutions but also as institutions intentionally structured to make use of the strategic interdependence of violators such as leniency when detecting a cartel.

We analyze the theoretical properties of the absolute and relative punishment institutions in the next section. We fix the fine and vary the requirement as the ratio between these two determines the equilibrium of the game considered. We find that, in both institutions with requirements below a critical level, every individual contributes the required level in a unique Nash equilibrium. In contrast, with the requirement above the critical level, every individual contributes less than required. An intuition behind this result is that the free rider's payoff is less than the law-abider's payoff in the former situation but greater in the latter. From this analysis, we readily obtain the optimal requirement in both institutions. The two institutions share the same optimal requirement determined by the marginal value such that if the required contribution is beyond this value, every individual has no incentive to comply with the regulation.[2]

The difference between the two institutions arises when the required level is above the critical

[2] The optimal level equals the above-mentioned critical value if the strategy set is in real numbers. In our case, it is in integers for consistency with the experiment, so the optimal requirement equals the maximum integer less than the critical value.

value. In this case, while complete free riding is the dominant strategy in the absolute punishment institution, choosing some positive amount of contribution is the mixed strategy Nash equilibrium in the relative punishment institution. In the relative institution, the environment created by the strategic interdependence of the sanction raises the expected probability of sanction with the extent of free riding. Similar to Stigler's (1970) marginal deterrence 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, p.63) every individual who faces the high requirement in the relative institution moderates their contribution rather than completely free riding.

This result in the relative punishment institution with a high requirement may explain why in real life, people obey laws even when the sanction is not sufficiently high (or the requirement not sufficiently low) to prevent people from free riding. Tyran and Feld (2006) explain this point by the effect of referendum. They find in a laboratory experiment that if participants choose through a referendum a centralized punishment institution over a public goods game with no institutions, cooperation increase although the level of sanction is not sufficiently 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 an absolute punishment institution, even the self-regarding individuals contribute to some extent.

The rest of this paper examines the two centralized punishment institutions by means of a laboratory experiment to gain further insights into the institutions considered.[3] For each institution, we choose three sets of parameters (Low, Middle, and High), differing in the value of the contribution requirement. The three sets of parameters are chosen so that (1) contributing the requirement is an equilibrium in Low and Middle, and contributing less than the requirement is an equilibrium in High and (2) the requirement is optimal in the Middle treatment. These three parameter sets are common for the two institutions hence we have $6\ (= 2 \times 3)$ treatments. The design details are explained in Section 3.

The major results of the experiment are summarized here by noting two analytic observations. First, in High requirement treatments, we observe more contribution to the public goods in the relative punishment institution than in the the absolute punishment institution. This supports the theoretical prediction and our previous argument that people may cooperate even when the sanction is non-deterrent, because of the limited apprehension and partially enforced institutions. Second, the efficiency in the Middle treatment with the optimal requirement level, was not the highest among the three parameter sets for both institutions. In both institutions, we observe below equilibrium contributions in the Middle treatments, which decline with repetition. As a result, in the latter rounds, Low requirement treatment obtain higher profit than the Middle treatment. We show our results and discuss the possible causes of this discrepancy between the theory and the observations in Section 4.

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 e unit of resource as his endowment, and divides this into either his private account or to a public project. If participant $i \in N$ contributes c_i unit

[3] Also, experimental studies on centralized punishment institutions have increased yet remain limited, and this paper provides further empirical evidence on this issue. There are many works on absolute punishment institutions, especially those that compare the effect of changing the apprehension probabilities holding expected loss constant (see Vyrastekova and Van Soest, 2010). Kosfeld et al. (2009), Tyran and Feld (2006), Putterman et al. (2010), and Vyrastekova and Van Soest (2003, 2010) study the effects of voting on the centralized punishment institutions.

of resource to the project, all participants gain βc_i unit, where β is a marginal per capita return (MPCR) of the voluntary contribution. 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

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

Throughout the paper, we assume $1/n < \beta < 1$.

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 $(1-\beta)$. From this insight, it is readily understood that 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 studies examine this theoretical prediction through laboratory experiments. Major findings can be summarized as follows: in a one-shot version of the public goods game, contributions are above the theoretical predictions; whereas when the game is repeated, contributions often begins 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 is improved upon by the introduction of personal costly punishment. Many studies across fields investigate the effects of such decentralized punishment, and this tendency is 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 important, there are many areas where personal punishment enforcement is inadequate. In such cases, a 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 two types of centralized punishment institution, both of which entail a requirement level (henceforth, threshold level) $s \in E$ and an amount of sanction $P > 0$.

In the first centralized punishment institution, 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 the (P, s)-*absolute punishment institution* and denote it by the 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, the (P, s)-absolute punishment institution requires all members in $L(c; s)$ to pay fines. This can be viewed as a rigorous application of the punishment institution and as an institution in an ideal state. We know, however, that in practice, the rule of punishment is only moderately or inconsistently applied because the complete application of the rule is essentially impossible in many cases. The second centralized punishment institution follows this view.

In the second centralized punishment institution, given the pre-determined 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 the (P, s)-*relative punishment institution* and denote it 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 has a resource constraint and thus cannot arrest all violators. The enforcement agency generally spends more effort to punish the worst violators. For example, a law enforcement agency exerts more of its power to investigate violent crimes (e.g., murder) compared to minor offenses; and tax agencies are more likely to investigate tax evasion by large companies compared to small ones. These are results of optimization behavior by the enforcement agency with resource constraint, and consequently, the probability of being sanctioned increases as the extent of violating the law or regulation increases relative to others' violations.

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 certain degenerate cases, 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. In contrast, $c_i = 0$ 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 because all free riders are detected and 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 $s \leqq \frac{P}{1-\beta}$, all players contribute s unit of resource in a Nash equilibrium.

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) The proof can be divided into the following two steps.

Step 1. For any integer $s>0$, 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 improves.

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' contributions fixed. Then, the payoff difference is

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

Thus, i's payoff improves.

Step 2. $(s,s,...,s)$ is a Nash equilibrium.

If all players contribute s, then 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

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

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

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

Because $a \leqq s$, 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 sufficient to show that for $m < s$, $(m,m,...,m)$ is not a Nash equilibrium. This is easily verified by considering a participant i's changes in contribution to $m+1$. Then, i's utility increases by $P - (1-\beta) > 0$. □

Propositions 1 and 2 imply that when the required level in a regulation or law is reasonably low with respect to the degree of punishment, the equilibrium behaviors for the two punishment institutions are the same and everyone complies with the regulation. We sometimes refer to these institutions with $s < P/(1-\beta)$ as institutions having effective sanction. In contrast, when the threshold level is high, the regulations are not complied with in either institution, and therefore we sometimes refer to these institutions as having ineffective sanction. More importantly, when the threshold level is high, the equilibrium behaviors in the two institutions are distinct from each other. The following proposition shows that if the value of the 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 Step 1 in the proof of Proposition 2, we know that for $m \neq s$, $(m, m, ..., m)$ is not a Nash equilibrium. Thus, it is sufficient 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, i's payoff improves.

Next, we show that there exists no pure strategy Nash equilibrium. We prove this by contradiction. Assume that $c = (c_1, c_2, ..., c_n)$ is a pure strategy Nash equilibrium such that $c_1 = c_2 = ... = c_n$ does not 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 the need to consider a mixed strategy Nash equilibrium to obtain some prediction for the (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 natural number $m \leqq \frac{P}{1-\beta}$, 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. Furthermore, there is no other symmetric mixed strategy Nash equilibrium.

Proof. Let $(q, q, ..., q)$ be a symmetric mixed strategy Nash equilibrium. The proof can be divided into the following four 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 withholding the contribution, $k(1-\beta)$, is greater than the loss from punishment, P.

Step 2. There exist no 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 $j = \arg\min\{h \in E : q(h) > 0, h \geqq k+1\}$. By the assumption and Step 1, $j \geqq k + 1$ and $j \leqq \frac{P}{1-\beta}$. Because all players follow the mixed strategy q, there is a 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 withholding contribution, $(j - k)(1 - \beta)$, compared to choosing j contribution, whereas 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 $k = m + 1, m + 2, ..., e$. Since $(0, ..., 0)$ is not a pure strategy Nash equilibrium, m must be greater than 0.

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 assigned a 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 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 when he contributes $m + 1$ and others follow $\hat{q}$. We know that the expected payoff of a player at $(\hat{q}, \hat{q}, ..., \hat{q})$ is $Eu(0) = e + (n - 1)\beta A - P$ and the expected payoff in 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 on 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 the aforementioned difference between the relative and absolute punishment institutions when $s > \frac{P}{1-\beta}$. 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 probability of sanction for each contribution level explains this non free riding in the relative punishment institution. When all 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 indicates that the expected probability of receiving a sanction rises with the extent of free riding. Similar to Stigler's (1970) marginal deterrence 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, p. 63), every individual confronting 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 results in the previous section. We focus on the level of optimal threshold s for two centralized punishment institutions, given the sanction amount P. Clearly, it is the ratio of the threshold level and the sanction that affects the results, hence they are two sides of the same coin. Two reasons support focusing on the threshold rather than on the sanction. One is that in the real field, the amount of sanction is determined from many perspectives, and choosing P from the viewpoint of economic performance alone is often controversial. The other is that if any amount of sanction is allowed, any level of threshold is easily attained by imposing a high fine, 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 proofs of (i), (ii) and (iii) of this proposition are readily obtained from Propositions 1, 2 and 4. Thus we omit the proofs. □

In the rest of the paper, we investigate the two centralized punishment institutions by means of a laboratory experiment. We test whether these theoretical results hold in the experiment, particularly 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 the Computer Laboratory of Waseda University in Japan.

3.1 Subjects

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

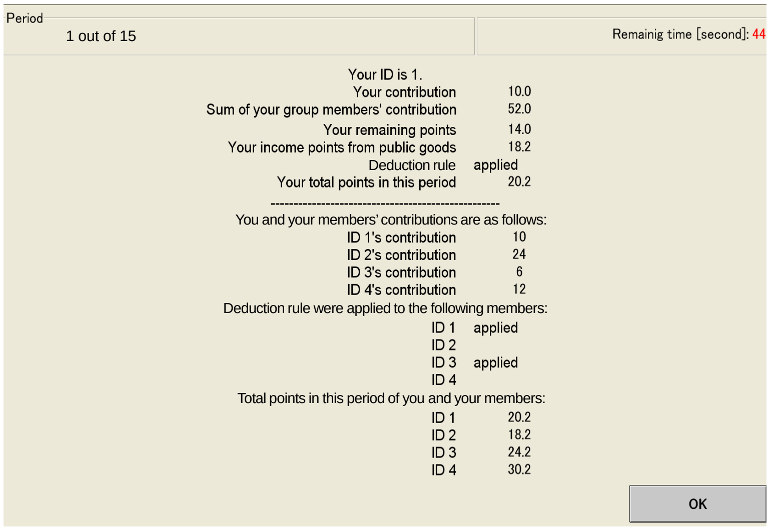
3.2 Tasks and procedures

We conducted the six experimental treatments explained below. Twenty-eight or thirty-two different subjects participated in each treatment. In all treatments, at the beginning of the 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 instructions explaining the game, payoffs, sanction rule, and procedures, and read it on their computer screen at their own pace. The instructions used neutral wording, as is common practice in experimental economics.[4] After reading the instructions, the 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 the test, the subjects were randomly and anonymously allocated to groups of size $n = 4$, and these groups played the linear public goods game with a centralized punishment institution for 15 periods. Group composition remained the same throughout 15 periods ("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 to eliminate the effect of reputation. Then, each group member had to determine how many of 24 points to keep and how many points to contribute to a public good on the 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 following results of the period appeared on their computer screen (Figure 1): each member's contribution points, sum of the group members' contribution points, each member's outcome, and who received the punishment (i.e., $P = 12$ points deduction). After each experiment, all subjects returned the questionnaire.

We used z-Tree (Fischbacher, 2007) to conduct the experiments. Each session averaged approximately 1.25 hours to complete. The mean payoff per subject was \$19.86 (\$1 = 85yen). The maximum

[4] Instructions are available at: `http://aitakeuchi.web.fc2.com/materials/absolute.html`



```
Period
        1 out of 15                                   Remainig time [second]: 44

                              Your ID is 1.
                          Your contribution          10.0
         Sum of your group members' contribution     52.0
                     Your remaining points           14.0
           Your income points from public goods      18.2
                              Deduction rule     applied
                  Your total points in this period   20.2
                  ------------------------------------------------
             You and your members' contributions are as follows:
                          ID 1's contribution        10
                          ID 2's contribution        24
                          ID 3's contribution        6
                          ID 4's contribution        12
           Deduction rule were applied to the following members:
                                          ID 1   applied
                                          ID 2
                                          ID 3   applied
                                          ID 4
             Total points in this period of you and your members:
                                          ID 1       20.2
                                          ID 2       18.2
                                          ID 3       24.2
                                          ID 4       30.2
```

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

payoff was \$23.76, and the minimum payoff was \$12.00. Average per-hour earnings exceeded the average hourly wage of a typical student job in the location of Waseda University.

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 institutions, "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, which we called Low (L), Middle (M), and High (H), respectively. By doing so, we could investigate the effects of each 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 in Section 2 with our parameters for the experiment, the critical value which determines whether the sanctions are deterrent is $P/(1-\beta) \approx 18.46$. Therefore, it is obvious that the equilibrium contribution in L and M threshold treatments are 12 and 18, respectively, for both institutions. When the threshold is H, the prediction differs between the two institutions. For ABS-H, the theoretical prediction of expected contribution is 0, and for REL-H, it is approximately 13.38, calculated by 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 expected contribution	12	18	0	12	18	$\approx$ 13.38
Theoretical prediction of expected profit	28.8	31.2	12	28.8	31.2	$\approx$ 26.35

4 Results

Table 3: Summary statistics per treatment (mean and standard deviation in parenthesis)

	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

This section aims to analyze experimental observations and investigate the two centralized punishment institutions. We first analyze the effects of thresholds on behavior holding institutions fixed. Then, we compare the two punishment institutions holding the threshold fixed. 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 discussing 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.[5] 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.[6]

[5] A comparison of Table 2 and 3 shows some difference between the theory and experimental observations, in both contribution and profit, although these differences are not statistically significant. For each treatment, we used the Wilcoxon signed-rank test to determine whether the median of the group-average contributions was the same as the theoretical prediction. For all treatments, the results were insignificant at the 10% level, as were those for profits. This may partly result from the limited number of samples: we used group averages over all periods as units of observation (recall that the experiment used the partner matching protocol). For most of the statistical tests we conducted, the number of observations equals the number of groups in each treatment. Also, we used two-tailed tests throughout the analysis.

[6] Another possible measure of efficiency is to calculate the average gain per subject without subtracting the sanction, since the collected sanctions are usually redistributed back to the society. This type of efficiency can be calculated by simply transforming the contribution linearly.

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 the dominant strategy is to free ride, and when $s < P/(1-\beta)$, the sanction is effective and the dominant strategy is to contribute s. Theoretically the optimal threshold is thus 18. Therefore, we predict the following: 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 prediction but the not the latter two predictions.

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 in 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. Furthermore, the two treatments varied in the dynamics of contribution choices, as it decreased with repetition in ABS-M but had leveled-off at s in ABS-L. In the last rounds, the profits were higher in ABS-L than ABS-M.

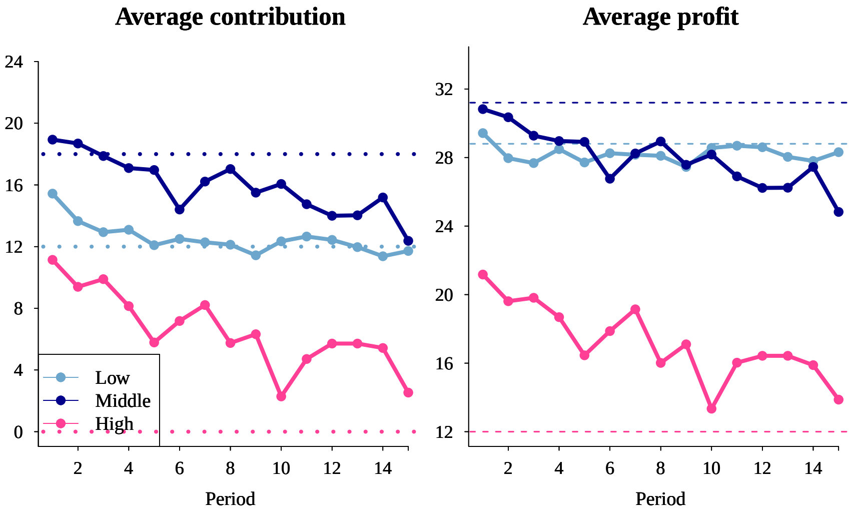


Figure 2: Comparison of contributions and profits across thresholds in absolute punishment institutions (dotted lines are theoretical predictions)

Evidence for Observation 1 is 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 the 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 the Wilcoxon rank-sum test for each pair of treatments with Holm's p-adjustment method. There were significant differences between ABS-H and the other two treatments. For contribution, p-values were 0.008 for the low and high comparison, and 0.007 for the middle and high, while for profit, they were less than 0.001 and exactly 0.002, respectively.

Further evidence is presented in Figure 2 which plots, separately for each treatment, the per-period average contribution and profit in the left and right panels, respectively. The dynamics of the contribution and profit in ABS-H were similar to the common observations in public goods game experiments:

They started above the theoretical prediction and declined over time (see Ledyard, 1995). The Spearman rank-order correlation between the per-period average contributions and periods was negative and significant ($\rho = -0.878, p < 0.001$), as it was 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 this evidence, we state that ineffective punishment institutions cannot sustain cooperation.

Next, we found several unexpected differences between the two effective punishment institutions. In accordance with the theoretical prediction, the mean contributions in ABS-M were higher than in ABS-L at the aggregate level (see Table 3). The profits in ABS-M were slightly lower than those in ABS-L, but these differences were not statistically significant. The result of the abovementioned Wilcoxon rank-sum test with p-adjustment was insignificant at the 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 2 reveals the following. The mean profit in ABS-L leveled off at the theoretical prediction, whereas in ABS-M, they decline over time, moving away from the theoretical prediction (Spearman rank-order correlation; ABS-L: $\rho = 0.07, p = 0.81$; ABS-M: $\rho = -0.875, p < 0.001$).[7] In the last five rounds, there is, in fact, a reversal in the profits of ABS-L and ABS-M. This difference remains, even in group level analysis: In ABS-L, no large group difference exists; whereas it does in the ABS-M group, with four out of eight groups contributing to the threshold level and others contributing much less. Thus, ABS-L was more effective in sustaining cooperation. We consider the possible causes of these differences in Section 4.4.

4.2 The effect of thresholds in relative punishment institutions

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

Observation 2. In the relative punishment institution, the threshold level does not significantly affect the subjects' behavior. Differences between the different thresholds in the contributions and profits were statistically insignificant.

Support for Observation 2 is illustrated in Table 3 and Figure 3. First, the comparison of average contributions and profits in Table 3 reveals that the differences among the three treatments in relative punishment institutions were less than those in absolute institutions. Comparing the distribution of per-group average contributions across threshold levels, the null hypothesis that all the distribution was 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 difference between the mean observations in the experiment and the theoretical predictions was low in REL-L and REL-H.[8] However, contributions observed in REL-M were much lower than the

[7] In ABS-L, the results of the Spearman rank-order correlation test shows difference in contribution. The decline of the first five periods drives the overall ρ to be negative and significant ($\rho = -0.746, p = 0.002$). However, the subset of data from period five shows that the results are as the profit: ρ is positive and not significantly different.

[8] Since our theoretical prediction for REL-H is based on the 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 equilibrium mixed strategy is $\{0, 1, ..., 18\}$. Subjects in our experiment contributed more than predicted. There were many contributions

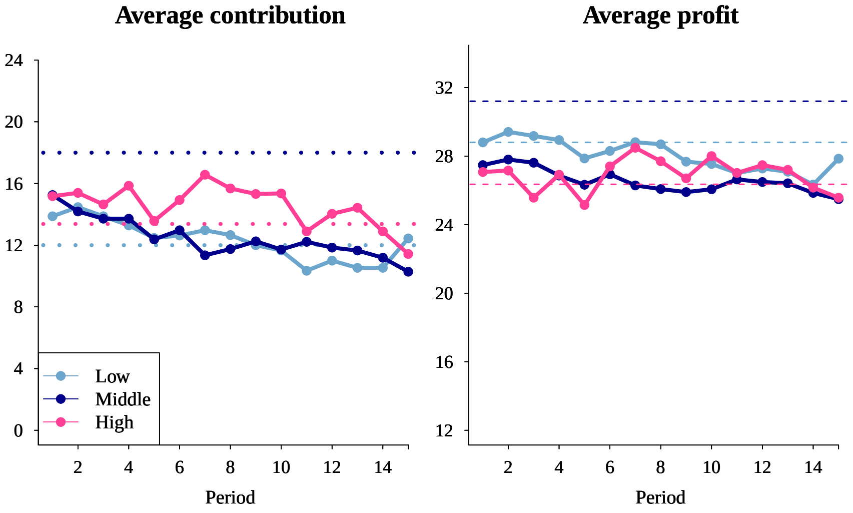


Figure 3: Comparison of contributions and profits across thresholds in relative punishment institutions (dotted lines are theoretical predictions)

theoretical prediction. Consequently, average contributions were highest not in REL-M but in REL-H for most periods, and the average profits in REL-L were consistently higher than that in 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 institutions with the same sanctions and threshold levels differ only in designating violators for punishment. One may intuitively think that cooperation is better sustained by arresting all violators, but our theoretical model disproves this intuition. When the institution is effective, there is no difference between the two, but when it is ineffective relative punishment institutions yield higher contributions than absolute institutions. The experimental results in general supported 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 were higher than in the absolute institution.

Details of the observations are as follows. Let us start with the effective case. Referring to Table 3, controlling for the threshold level, the average contributions, profits, and even the number of punished individuals were similar across the ABS and REL. Contributions were higher in the ABS-M than in the REL-M, but these differences were 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 in ABS-H. Thus, the relationship between the relative and the absolute punishment institutions were as predicted by theory.[9]

above the support of the mixed strategy Nash equilibrium. Notably, there were many threshold contributions.

[9] Some may argue against relative institutions because the perceived fairness may be low. In the post-experiment questionnaire we used Sondak and Tyler's (2007) procedural fairness questions to obtain subjects' evaluation of the punishment rule. We had seven questions: two questions each on procedural desirability, procedural justice, and outcome valence, and one on

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

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

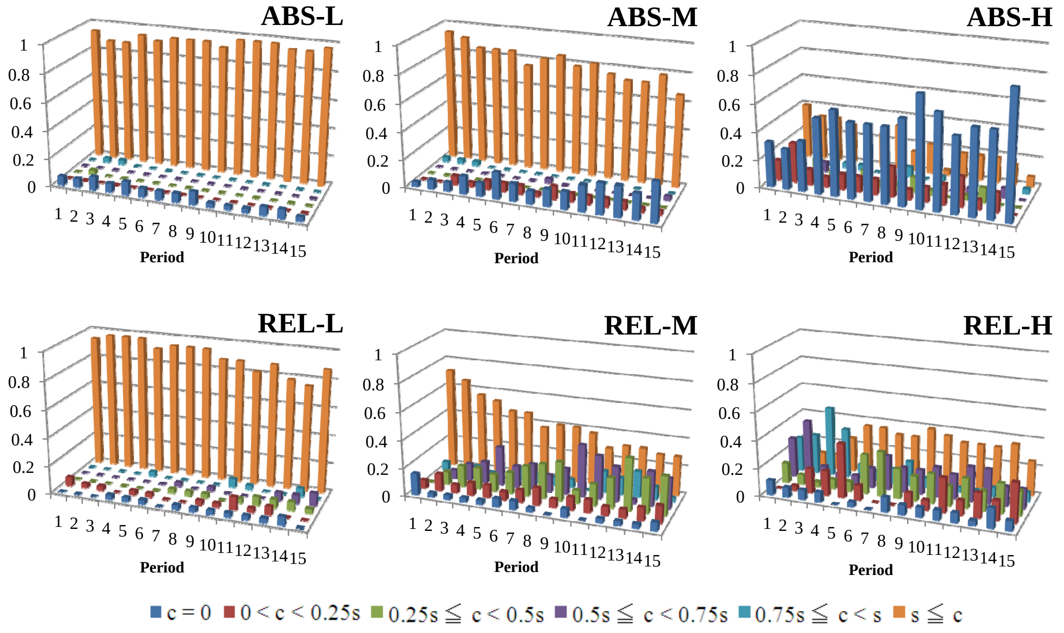


Figure 4: Comparison of the distributions of contributions with respect to threshold

The support of this observation is provided in Figure 4, which plots the frequency of contributions, separated with respect to the threshold. The furthest back bars are the frequencies of above-threshold contributions, and the bars in the front are the frequencies of zero contribution. Most contributions in absolute institutions were either above the threshold or zero, but 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.[10] Another noteworthy point is that when the sanction is effective, the number of above-threshold contributions was higher in absolute than in the relative punishment institutions. Therefore, there is a tradeoff between absolute and relative institution when the sanction is effective. On the one hand, above threshold contributions were observed more frequently in absolute institutions, but violations were mostly complete free riding. On the other hand, more subjects contributed below threshold in relative institutions, but the violations were smaller.

This observation well illustrates the difference between the two institutions. In the relative insti-

effectiveness. We compared the distribution of subjects' answers across punishment institutions, controlling for the threshold level. When the sanction was effective, there were no differences in the distribution of subjects' answers between the absolute and the relative institutions, but when ineffective, subjects perceived the relative punishment institution as having a fairer rule than they did for the absolute. The difference was statistically significant at 5% for four out of seven questions. Although the relative punishment institution might seem unfair from the objective viewpoint, subjects in the institution did not evaluate it as being unfair.

[10] 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.

tutions, there is a chance to contribute below the threshold without punishment. More importantly, the possibility of getting punished increases with decreased contribution. Thus determining the below-threshold contribution amount could greatly affect profits. In the absolute institutions, every subject who contributes below the threshold receives sanction regardless of the level of below-threshold contribution. Thus, the choice is simply between completely free riding or contributing at the threshold level.

4.4 Discrepancies between the theory and experimental observations

Overall, the experimental results supported the theoretical predictions for 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 contributions were lower than expected. As depicted in Figure 2 and 3, contribution declined, departing from the theoretical prediction. This decline raises another key issue, the reversal of profits in the low and middle threshold treatments. The mean profit obtained in REL-L is consistently higher than that in REL-M, and for ABS-L, it is higher than that in ABS-M after the tenth period. Thus the theoretically optimal threshold may not be the actual optimal threshold, and we need to consider other behavioral factors.

Why did the subjects in the low threshold treatment continue contributing as predicted while the subjects in the middle threshold treatment reduce their contribution? This is especially puzzling in absolute punishment institutions where it is a monetary-payoff-maximizing strategy to contribute to the threshold level no matter what the other players contribute (i.e., it is the dominant strategy). Therefore, let us concentrate our discussion on this institution.[11]

Clearly, models with monetary profit maximizing players will not explain the difference between the two thresholds. Also, conditional cooperation and spiteful preference cannot explain the behavioral difference as they both predict a decline in cooperation when there are free riders.[12] There is one dissimilarity between the low and middle threshold treatments that, we suspect, caused this behavioral difference. In the low threshold treatment, even if a player deviates from the equilibrium and become a free rider, there is no difference between his and the contributors' profit. Both will be 24.9. In the middle threshold, however, this is not the case. The free rider's profit of 30.9 is higher than the contributors' profit of 24.6.[13] Therefore, contributors may envy free rider in the middle threshold, but not in the low threshold. We suspect that an equilibrium where the payoff earned at the deviation from the equilibrium

[11] Once there is a free rider in the group and the subjects are myopic, there is an incentive to lower contributions in the relative punishment institution. If a contributor changes their contribution to a level slightly higher than that of 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 a monetary-profit-maximizing action, this will yield a higher payoff than contributing the threshold level. This may be one reason why, as depicted in Figure 4, the below-threshold contributions were much more frequent in REL-M than in ABS-M. Once below-threshold contribution is observed in the relative institution, it is unlikely to return to the equilibrium level. There were two groups that maintained above-threshold contributions throughout the experiment, but in both groups, no subject contributed below the threshold. For the other six groups, once they observed one or two below-threshold contributions, they could not return to the equilibrium threshold contribution.

[12] 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 behaviorally based and does not include discussion about people's profits when cooperating or defecting. Therefore, in both the low and middle treatments, when there are free riders in the group, conditional cooperators should also free ride. We observed free riding in the early rounds of both the low and middle threshold treatments, thus, if we rigorously apply the conditional cooperation argument, we should see declines 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, because it reduces their own and the others' payoffs. However, it is also spiteful strategy to free ride in ABS-L, so the spiteful strategy would also not explain the difference between the two treatments.

[13] The free rider will achieve 31.2 if he contributes to the threshold level, thus, it is a dominant strategy to contribute.

is envied by others is unstable in the long run. The dominant strategy equilibrium of ABS-M is of this type, but of ABS-L is not. Our rationale is 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 free riders in the following rounds, creating a chain of decline in contributions.

Referring to Figure 4, one can see the gradual increase in the number of complete free riders in ABS-M, in stark contrast to ABS-L where below-threshold contribution remains low throughout the experiment. Also, the difference in the average contribution of subjects in group with and without 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 dropped to 9.07. This difference in ABS-M was 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 greener than that of the contributors.

5 Conclusion

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

First, when we perform a comparison between institutions, both theoretical and experimental results suggest that when the sanction is ineffective, the relative punishment institution yields higher performance in the associated public goods game than does the absolute punishment institution. This result has two implications. First, it is unnecessary to fully enforce the punishment institution. If the institutions' resources are limited, it is not necessary to arrest every violation, but more effort should be devoted to arresting the worst one.[14] The second implication provides 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 moderate sanctions. Because the ABS-H is a centralized punishment institution with moderate sanctions, our model and experimental comparison of ABS-H and REL-H raise another possibility: although an institution is built as being absolute, due to limited enforcement, the actual game being played may be that of relative institution. Then, as our model and the experiment reveal, contributions will be higher than those expected in the absolute institution.

Second, from a within-institution comparison across threshold levels, we found a discrepancy between the theoretical and the experimental results in optimal threshold level. In the treatment with the theoretically optimal threshold level, 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 the lower threshold level, and as a result, the profit in the lower threshold level surpassed that of the optimal level in the last few rounds. From this difference in the results, we implied and discussed the importance of the property that "the outcome of the most likely deviation from the equilibrium is envy-free." In a repeated setting, if this property is not satisfied and a deviation from an equilibrium occurs, there are many behavioral reasons for the experimental observations to depart from the equilibrium, such as

[14] Notice that, as discussed in the last paragraph of Section 4.3, the important property of the relative institution is that the probability of being punished increases with the degree of free riding. Thus the relative institution does not encourage institutions that randomly punish any player among those that do not satisfy the requirement.

inequality aversion and imitation learning. Therefore, this property may be important when designing an institution, and is worth further investigation.

We conjecture that the last result may be vulnerable to differences in the information provided. In our experiment, we provided feedback 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 the information provided (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, contributing is an equilibrium strategy. Still, if we provided only the information about the aggregate level of contribution, the result could have differed, especially in the treatment with an optimal threshold level. To sustain cooperation in an institution that is supported by an equilibrium, hiding the existence of free rider in the group may be one effective strategy. This point merits further investigation as a topic for future research.

Acknowledgements

We thank Charles Noussaire, Robert Vestzeg, Jana Vyrastekova, participants of 2011 Japanese Economic Association Spring Meeting and PET11 for their helpful comments. This work was supported by the Grant-in-Aid for the Global COE "Political Economy of Institutional Construction" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. Ai Takeuchi was supported by JSPS Grant-in-Aid for Young Scientists (B) 23730195.

References

Apesteguia, J., Huck, S. and Oechssler, J., (2007): "Imitation—theory and experimental evidence," *Journal of Economic Theory*, 136 (1), 217–235.

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

Bigoni, M. and Suetens, S., (2010). "Ignorance is not always bliss: Feedback and dynamic in public good experiments," Tilburg University, Discussion Paper No. 2010-64.

Cason, T., Saijo, T. Yamato, T. and Yokotani, K. (2004): "Non-excludable public good experiments," *Games and Economic Behavior*, 49 (1), 81–102.

Chaudhuri, A. (2011): "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, 14 (1), 47–83.

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.

Fehr, E., and Schmidt, K. M. (1999). "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*. 114(3), 817–868.

Fischbacher, U. (2007): "z-Tree: zurich toolbox for ready-made economic experiments," Experimental Economics, 10(2), 171-178.

Fischbacher, U. and Gächter, S. (2010): "Social preferences, beliefs, and the dynamics of free riding in public goods experiments," *American Economic Review*, 100 (1), 541–556.

Vyrastekova, J., and van Soest, D. (2003). “Centralized common-pool management and local community participation,” *Land Economics*, 79(4), 500–514.

Vyrastekova, J., and van Soest, D., (2010). “Higher fines, lower conviction probabilities, and the support for government regulation,” mimeo.

Kosfeld, M., Okada, A., and Riedl, A., (2009): “Institution formation in public goods games,” *American Economic Review*, 99(4): 1335–55.

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.

Putterman L., Tyran, J.-P., and Kamei, K., (2010): “Public goods and voting on formal sanction schemes: An experiment,” *University of Copenhaven Dept. of Economics Discussion Paper*, No. 10-02.

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

Sondak and Tyler. (2007): “How does procedural justice shape the desirability of markets,” *Journal of Economic Psychology*, 28 (1), 79–92.

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.