Rise of the irrational free riding behavior under a centralized punishment authority

Ai Takeuchi[a] **Yoshio Kamijo**[b] **Yukihiko Funaki**[c]

Abstract

This study compares the effects of the amount of feedback information in a public goods game with a centralized punishment institution where each player is required to contribute a certain amount, and those who under-contributes must pay a fixed fine. We compare two different amount of feedback information—one where only the aggregate level of contribution and own payoff is provided, and the other where the individual contributions and profits of the other players are also provided—under two different strength of penalty—one where it is barely deterrent to make players contribute as required (weakly enforceable punishment), and another with stronger penalty (strongly enforceable punishment)—in two by two design. Although it is a dominant strategy to contribute as required in both punishment institutions, we find a significant decline in the contribution in the weakly enforceable punishment institution when the feedback is provided at the individual level. In the other three treatments, the average contribution is stable across the periods. These results are consistent with the theoretical analysis using the finite population evolutionary stable strategy.

Keywords: Public Goods Game, Punishment, Feedback information, Laboratory experiment
JEL classification: C72, C91, C92, H41.

[a] College of Economics, Ritsumeikan University, ai-tak@fc.ritsumei.ac.jp
[b] Kochi University of Technology, Department of Management, yoshio.kamijo@gmail.com
[c] School of Political Science and Economics, Waseda University, funaki@waseda.jp

1. Introduction

Human beings have the ability to learn, not only from their own experiences but also from others experiences through observing their behavior and the resulting outcomes. However, evidences from laboratory experiments suggest that such ability to learn from others' behavior in social interactions (henceforth, social learning) do not help to promote pro-social behavior in social dilemma situations (e.g., Fiala and Suetens, 2017). In the actual society, social dilemma situations are often not stand alone, but social norms are often enforced by additional institutions that alters peoples' incentives via rewards and punishments. Thus, we investigate the effect of social learning on the behavior in the public goods game with centralized punishment institution.

In the punishment institution we study, there is a pre-specified required level of cooperation, and those who fail to meet the requirement is punished and are imposed a fixed fine. When the penalty is weak, it is individually rational to freeride even if they are penalized. When the penalty is strong enough, it is individually rational to follow the rule and cooperate as required. This study focuses on the latter case, because when the punishment is too weak and freeriding is individually rational, we can expect a similar result as in the case of simple public goods game.

This study investigates whether and when the possibility of social learning through a detailed information on the others' decisions affects the number of free-riders. The key idea of this research is that mislearning of freeriding may take place even if it is individually rational to contribute as required, if the payoff of the free-riders is higher than that of the rule abiders. This is possible because the choice which maximizes the relative payoff among the group members and the individually rational choice does not always coincide. In this sense, it is possible that social learning may promote not only inefficient but also individually irrational behavior. Therefore, we investigate whether the ease of social learning has different effects under two different strength of penalty: one where the penalty is strong enough to make it a dominant strategy to abide by the rule, but free-rider's payoff is higher than that of the rule abiders (*weakly enforceable punishment*), and the other where the penalty is even stronger so the free-rider's payoff will be equal to that of the rule abider (*strongly enforceable punishment*).

The result of a laboratory experiment supported our hypothesis. Under the weakly enforceable punishment, the number of free-riders is significantly higher in the case where social learning is possible than in the case where it is not. On the contrary, under the strongly enforceable punishment, the possibility of social learning did not affect the number of free-riders.

2. The model and the theoretical analysis

We consider the commonly studied linear symmetric public goods game where each player $i \in \{1, \dots, n\}$ has an endowment of E from which they decide how much to contribute to the public good. Each player earns payoff from the consumption of the private goods $(E - c_i)$ and

the public goods $(\beta \sum_{j=1}^{n} c_j)$, where $1/n < \beta < 1$.

To this public goods game, we add a centralized punishment institution where there is a required level of contribution (s) and those who under-contributes must pay a fixed fine of P. The strength of penalty is a relative concept determined by the size of s and P: holding s fixed, penalty is stronger as the amount of fine P increase; and when holding P fixed, the penalty is stronger as the amount of requirement s decrease. In this game, the loss from contributing as required instead of freeriding is $s(1-\beta)$. Thus, we have the following result.

Proposition 1 (Kamijo et al., 2014): When $s < P(1-\beta)$, it is a dominant strategy to contribute as required, and when $s > P(1-\beta)$, it is a dominant strategy to contribute 0.

Next, we analyze the game using the concept of Finite population evolutionary stable strategy (FESS) by Schaffer (1988). FESS is an equilibrium concept based on evolutionary stability on a finite population. It analyzes the case where a group of finite players randomly meets and plays the game repeatedly. A strategy is a FESS if the payoff of player i who did not choose the FESS is less than or equal to the payoff of other players who did choose the FESS.[1] FESS is also a Nash equilibrium of the game transformed by the relative payoff difference between players. Thus, if a subject has a spiteful preference or cares of their relative ranking among the group members, their choice matches the FESS. Furthermore, when $s \neq P$, FESS coincides with the unique stochastically stable state of the imitation dynamics a la Vega-Redondo (1997). Thus, there are many behaviorally possible reasons why people's choices coincide with FESS especially when it is easy for the people to compare themselves with the others. Applying FESS to our game, we can show the following proposition:

Proposition 2: In the linear public good game with the centralized punishment institution, if s > P, then 0 is the unique FESS; if s = P, 0 and s are FESSs; and if s < P, s is the unique FESS.

Proposition 1 and 2 together show that, except for degenerate cases where the difference between P and $P/(1-\beta)$ is less than one, there is a range of thresholds $P < s < P/(1-\beta)$ in which the dominant strategy do not coincide with the FESS. Also, the equilibrium profit with the thresholds in this range is higher than the equilibrium profit obtained under requirements where the FESS coincides with the dominant strategy.

3. Experimental design

The experiment was conducted at Waseda University on years 2011 and 2015 using 180 undergraduate students as participants. Written informed consent was obtained from all subjects.

[1] We use a special case of the FESS. In the original definition by Schaffer (1988), it requires that the fitness of the one mutant to be lower than that of the FESS players in the situation where C players out of N players randomly meet to play the game. We use the case where out of N player population, all N players meet and play the game.

The experimental session lasted for 75 minutes. The subjects were paid ¥3 times the sum of points earned in all rounds plus ¥500 participation fee. The average payment was ¥1,784.

The main treatment variables of the experiment were the strength of the penalty (*strongly enforceable* and *weakly enforceable*) and the feedback information provision (*Individual* (IND) and *Aggregate* (AGG)). The experiment was in 2 by 2, between-subject design.

Parameters of the game were set as follows: endowment E equaled 24, the marginal per capita return β equaled 0.35 and the penalty point P equaled 12. To change the relative strength of the penalty, we varied the level of requirement s. In the strongly enforceable punishment institution, s was set to 12 and in the weakly enforceable punishment institution it was set to 18. Since $P/(1-\beta) = 18.46$, it is a dominant strategy to contribute as required in both treatments. On the contrary, since P = 12, the FESS is to contribute zero in the weakly enforceable punishment institution, and it is to contribute zero or s in the strongly enforceable punishment institution. The subjects played this game 15 times in a same group.

The difference in the amount of feedback information in each treatment were as follows. In the AGG, the subjects received information on: identification number, own contribution to the public goods, the sum of contributions of all four group members, the endowment points kept, points gained from the public goods, whether they were punished or not, and the profit gained in that period. Here, it was not possible to know the contribution made nor the profit obtained by the other group members. On the contrary, in the IND, in addition to the above feedback information, subjects received information on each group member's contribution, profit, and whether they were punished or not. Thus, it is easier to learn about the choices and the results of other group members.

4. Hypothesis

We have four hypotheses as below, one on the behavior in the strongly enforceable treatment and three on the behavior in the weakly enforceable treatment:

1. In the strongly enforceable treatment, because the dominant strategy and the FESS coincide, the information condition does not affect the subjects' behavior.
2. In the weakly enforceable treatment, subjects' behavior diverges to either 18 (the dominant strategy) or 0 (the FESS) in both information conditions.
3. In the weakly enforceable treatment, the FESS choice is more frequent in IND than in AGG.
4. In the weakly enforceable punishment treatment, if there was a free-riders in the same group in the previous round, others would lower their contribution more in the IND than in AGG.

5. Results

Here, we simply show the results using graphs, but these results can be verified with statistical tests. First, the change in the average contribution across periods clarifies that the feedback

information affects the behavior differently in the weakly enforceable and strongly enforceable treatment (see Figure 1). Under the weakly enforceable punishment institution, the average contribution shows a sharp decline in the IND but not in the AGG. On the other hand, under the strongly enforceable punishment institution, difference in the feedback information does not cause a difference in the trends. These results provide supports for Hypothesis 1.

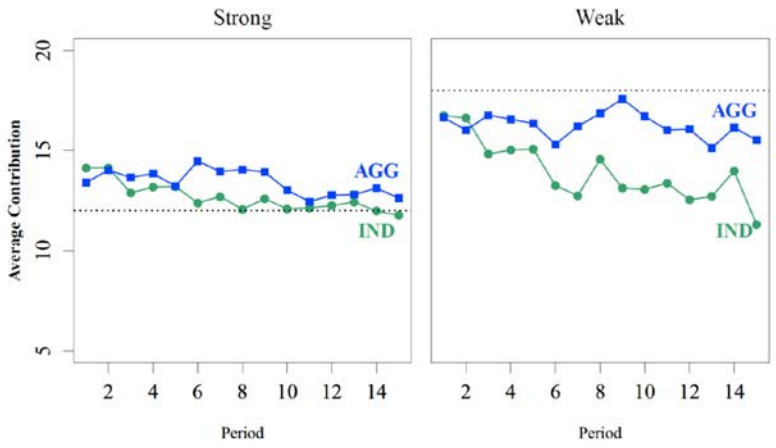

Figure 1

Next, we provide supports for Hypothesis 2 and 3 using data of weakly enforceable punishment. In both information treatments, more than 90% of contributions were either 0 or 18 or more. Among those who contributed 18 or more, only 12% in IND and 20% in AGG contributed strictly more. Hence, we can support Hypothesis 2 and say that most choices diverged to either 0 or 18. As is stated in Hypothesis 3, the frequency of FESS choice is more frequent in IND than in AGG. We also observed that the number of free-riders increase in IND but not in AGG.[2] In IND, the frequency of zero contribution gradually increased from 5%. After about period 5, it levels off at about 20 to 30%. On the contrary, in the AGG, it is constantly about 10%.

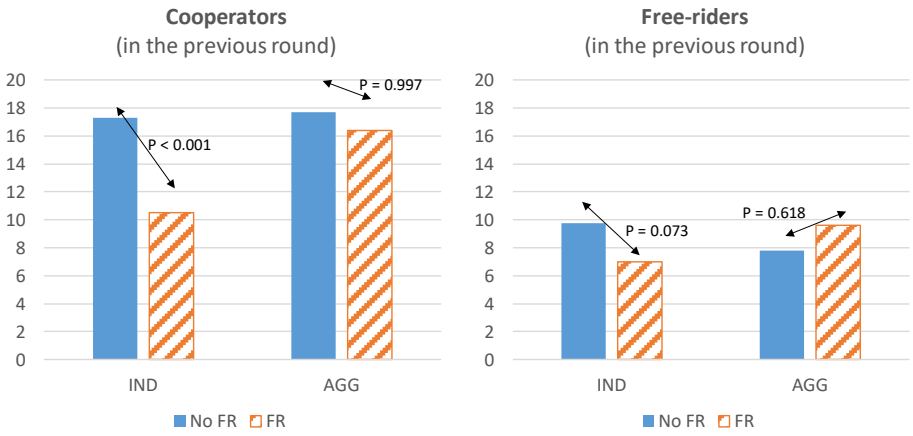

Figure 2

[2] Correlation between periods and frequency of freeriders is significant at 1% level in IND.

Finally, we provide supports for Hypothesis 4 by analyzing people's behavior conditional on the existence of free-rider in the previous round. Figure 2 shows the average contribution of those who cooperated in the previous round (in the left) and of those who freeride (in the right), conditional on the amount of feedback information and the existence of free-rider among the other group member in the previous round. As the figure shows, on the one hand, in the AGG, the average contribution does not significantly differ with the existence of free-rider for both the cooperators and the free-riders. On the other hand, in the IND, the existence of free-rider among the other group members significantly lowers the average contribution in the next round for both the cooperators and the free-riders. These findings are consistent with Hypothesis 4.

6. Conclusion

We conclude by extrapolating our findings to the field of law and economics and to institution design in general. First, this research adds new insights on the optimal punishment system to make people abide by the rules, which is studied in depth in the field of economics of institution and law and economics. In assessing whether the punishment can make people abide by the rules, commonly used assumption in the field is that people maximize their own payoffs. The results of the theoretical and experimental analysis suggest that the strength of punishment calculated based on such assumption is not enough to make people follow the rules. In environments where social learning is possible, not only should it be optimal for the people to abide by the rules, but penalty should be strong enough so that the violators are worse off than the rule abiders. Second, this finding may be extrapolated into designing of any incentive schemes, given the increase in the ease of social learning via internet. When the visibility of others behavior and outcomes are high, designing an incentive scheme with the assumption of simple profit maximization at the individual level may not be able to achieve the intended results. In addition to the property that profit maximizing individuals would act as intended, incentive schemes should also satisfy the property that makes the grass to not look greener on the others side.

Reference

Fiala, L. and S. Suetens, 2017. Transparency and cooperation in repeated dilemma games: a meta study. Experimental Economics 20, 755–771.

Kamijo, Y. et al., 2014. Sustaining Cooperation in Social Dilemmas: Comparison of Centralized Punishment Institutions. Games and Economic Behavior 84, 180–195.

Schaffer, M. 1988. Evolutionarily Stable Strategies for a Finite Population and a Variable Contest Size. Journal of Theoretical Biology 132, 469–478.

Vega-Redondo, F. 1997. The evolution of Walrasian behavior. Econometrica 65, 375–384.