Does Punishment opportunity increase contributions to public good when envy exists?

by

Chenggang Zhou

An essay submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

August 2008

Abstract

This paper studies the free rider problem in an economy in which individuals care not only about utility from goods (private and public) but also the equality of their utility level. A two-player and two-stage Stackelberg game is introduced and players have different endowments. Envy is endogenous and varies as the ratio of players' utility; and is absent if the ratio is over an exogenous level $\lambda_0 \in [0, 1]$. Dissatisfaction incurred by envy directly leads to negative utility to a player. R, the rich individual, is the Stackelberg leader who makes contribution first at stage one. P, the poor individual, observes R's contribution and decides how much to contribute at stage two. It is shown that the private contribution to public good is not enough to be social optimal. Then the paper studies how the contributions change if P has the chance to punish R's under-provision. P carries out the punishment by destroying public goods after R's contribution rather than committing a crime directly to R. It is shown that the public good level does not necessarily increase, and there is always destruction unless envy is absent in equilibrium. R is worse off and P is not necessarily better off with the punishment power. The commitment problem is also studied and some extensions are discussed at end of the paper.

Acknowledgements

I would like to appreciate my supervisor Professor Robin Boadway for helping me define this topic and for his constructive and important guidance all the way through. I am also grateful to my wonderful classmates from the 2007/08 MA program at Queen's Economics Dept for helpful comments and suggestions.

1. Introduction

When there is a conflict between social and individual interest, all agents maximize their personal utility without regard for others. A socially sub-optimal outcome will be produced in these cases unless other mandatory contracts exist to force agents to cooperate (Samuelson, 1954; Olson, 1965). The total provision of public goods generally falls short of the level required by efficiency because people attempt to free ride on each other's contribution. In a public goods provision game, agents have to make a decision concerning how much to contribute. The higher an agent's contribution, the higher is the aggregate payoff. However, every player also has an incentive to free ride since they maximize their personal utility without regard for others. Experimental results indicate that the problem of free riding is indeed pervasive and leads to the under-provision of public goods (Ledyard, 1995).

Experimental economists have studied the solutions to the free ride problem. Ostrom et al. (1992) show that the existence of punishment opportunities that are carried out by individuals without the intervention of a central authority to free riders in a common-pool resource game increases cooperation (i.e. the aggregate contributions to public goods) between appropriators significantly. The same result is reported in Fehr and Gächter (2000, 2002) who study public good games. In the setting of their experiments, the games are finitely repeated. Punisher cannot get benefit from the punishment and the action is costly. Standard game theory implies that a utility maximizing individual will never have incentive to punish in equilibrium. In reasonable settings, the results of the experiments, however, report frequent punishments even in the last period and the fear of punishment has a strong positive effect on cooperation. Free riding generally causes very strong negative emotions among cooperators and there is a widespread willingness to punish the free riders. In Fehr and Gächter (2000, 2002), their results indicate that this holds true even if punishment is costly and does not provide any material benefits for the punisher. In their game setting, however, they did not count the negative emotion directly in the players' utility. Instead, I introduce envy as a vehicle to represent the negative emotion which affects players' action and take it as a variable in agents' utility function directly.

Foley (1967) and Varian (1974) define the concept of envy for exchange economies without uncertainty. Suppose that agent i receives a bundle x while agent j receives a bundle y. Then, agent i envies agent j if he/she prefers y over x. Envy-freeness, the absence of envy, is an appealing concept of equity of society. Combined with efficiency, it leads to a natural notion of fairness. In Feldman and David (1978), they focus on the binary relation of wealth rather than agents' goods bundle. It is equivalent to compare agents' initial wealth and goods bundle if they have the same preference and face the same competitive market prices for goods. In my paper, I assume the agents get the same utility from the same goods bundle (i.e. their utility function on commodity bundle are the same) and hence their endowments' differences are the source of envy. So I simply define envy in terms of agents' utility level: agent i envies agent j if his/her utility is less than that of j. The more the difference of utility level, the more does i envy j.

When agents have different endowment (inherited wealth, skill, ability and anything else), it is reasonable that the agents who have lower endowment will not expect to be as well off as the agents with higher endowment. Given all other things same, one with few dollars will not expect to get the same utility as one that inherits one million dollars. Chaudhuri (1986) and Diamantaras and Thomson (1990) considered the following measure of envy. An allocation is λ-equitable if no agent envies a proportion λ of the bundle of any other agent. In this paper, I extend the concept that the society is λ-equitable if no agent envies a proportion λ of the utility of any other agent. One agent will not be jealous of the other one if the ratio of their utility from goods bundle is greater or equal to λ.

In the game setting of Ostrom et al. (1992) and Fehr and Gächter (2000, 2002), the punishment processes are designed as follows. If agent i punishes agent j, j loses money value as a function of the intensity of punishment. There is a cost for i to practice the punishment. In other theoretical papers, the authors used crime as a punishment (or threaten) mechanism to incur agents to contribute more. The agents who practice the punishments (crime) get positive payoff but at a risk of penalty. The agents who suffer from crime get negative payoff certainly. In my paper, however, I do not use both of the above

methods to rule the punishment. Instead, an agent practices the punishment by destroying public goods that were contributed in the previous period. The detail is discussed in the next section.

Another mechanism for overcoming the free-rider problem is shown by Guttman (1978). In his model, agents in a first stage announce rates at which they will match the contributions of other agents. Then in the second stage, given the announced matching rates, agents choose their own contributions. Then the sub-game perfect equilibrium in such a two-stage non-cooperative game is fully efficient. As mentioned in Boadway, Song and Tremblay (2007), since the efficiency of the Guttman mechanism requires that both agents be able to commit to matching the contributions of the other in the second stage, commitment problem must be taken into consideration. In this paper, a revised version of matching mechanism is introduced. One agent announced a schedule to destroy based on the other agent's behaviour and fulfill the announcement at the last stage. Then the commitment problem is studied.

The rest of the paper proceeds as follows. Section two introduces the bench mark model. Section 3 describes the rule of the contribution game and how agents (players) make their decision. Section 4 analyzes how the aggregate contribution to public goods changes when punishment is introduced. Section 5 studies the commitment problem. Section 6 lists some extensions of the model and section 7 ends with conclusion.

2. Bench mark model

2.1 The Economy

There are two individuals, P (poor) and R (rich), with initial wealth: w^i, i=P, R and $w^P < w^R$. Both P and R have the same utility function on the commodity bundle (c^i, G), where c^i is individual i's private consumption, i= P, R, and G is the public good. Let the utility function (on the commodity bundle) be $U(c^i, G)$ with $\frac{\partial U}{\partial c^i}, \frac{\partial U}{\partial G} > 0$ *and* $\frac{\partial^2 U}{\partial {c^i}^2}, \frac{\partial^2 U}{\partial G^2} < 0, \frac{\partial^2 U}{\partial G \partial c} = \frac{\partial^2 U}{\partial c \partial G} > 0$. Utility functions are monotonic increasing and strictly quasi-concave, and both private consumption and public good are strictly normal.

The cross derivatives being greater than zero implies that the person with more private consumption goods prefers a public goods increment more than the person with less private goods. Both P and R decide how to allocate available endowments between private consumption and contribution to public goods. The aggregate public good is $G = g^P + g^R$, where g^i, i=P, R, are P and R's contribution to public good respectively. Their budget constraints are $c^i + g^i = w^i$, i=P, R. The price of the private consumption good and the cost of a unit of public good are constant and normalized to unity. There is no third-party firm to produce the public good and no government in the bench mark model. The players cannot transfer wealth to each other in this economy.

In our model, the individuals are selfish. R's utility is only from consumption of the commodity bundle and he/she does not care about the equality of the society. However, P cares not only about his/her commodity bundles but also about the inequality between R and him/her, i.e. P is jealous of R if R's utility is higher than that of P and this envy makes P feels bad. The essential problem is how envy affects P's utility. The direct way is that envy is introduced to P's utility function as a variable. The affect of the inequality between P and R is represented by a negative utility to P, named "dissatisfaction", as a function of the envy. The more the dissatisfaction, the more is the negative utility, i.e. P feels more "upset" when the inequality is higher. In order to represent envy in quantitative terms, one necessary assumption is that R will never be jealous P. R will not make a contribution if his/her utility is lower than that of P. Since the public good is non-exclusive, the first assumption is necessary:

Assumption 1: *R is never jealous of P, i.e. R's utility from the commodity bundle is no less than that of P's:* $U(c^R, g^P + g^R) \geq U(c^P, g^P + g^R)$ *.This is equivalent to* $w^R - g^R \geq w^P - g^P$*, i.e.* $c^R \geq c^P$*. The equality holds only when* $w^R - g^R = w^P - g^P$ *, i.e. both players have the same commodity bundle.*

It is realistic to suppose P knows the difference of endowments between him/her and R. P will not expect the same but something less than R's utility level. I consider the λ_0-equitable envy-free concept. The proportional difference is more relevant than the absolute difference between P and R's utility since

the absolute difference cannot represent the degree of inequality well. Given the same utility difference between P and R, P with lower utility feels it is more unfair if P has higher utility. P feels it is λ_0-equitable if his/her utility from bundle $\left(c^P, g^P + g^R\right)$ is at least as much as a proportion λ_0 of R's utility from bundle $\left(c^R, g^P + g^R\right)$, $\lambda_0 \in [0,1]$. If λ_0 <1, P does not expect to be as well off as R. While λ_0=1 means P expects the same utility level as R. λ_0 is not necessarily constant. It depends on both the society attitudes and the difference of players' endowments. As mentioned in the introduction section, it is realistic that people with low ability expect less than people with high ability. Then it is reasonable to consider λ_0 as a function of players' endowments. Since endowments are exogenous, we take λ_0 as given.

Let $\lambda = \frac{U(c^P, g^P + g^R)}{U(c^R, g^P + g^R)}$. If $\lambda \geq \lambda_0$, P thinks R contributes enough to make the society λ_0-equitable and there is no envy. If $\lambda < \lambda_0$, P thinks R is too stingy to contribute enough and even the λ_0-equitable outcome cannot be reached. Then, envy exists and introduces negative utility to P. Indeed, the word "stingy" is not suitable here. As we now show, λ may become smaller even if R contributed more under some conditions. Consider the following case. For simplicity, suppose P's initial wealth is such that he/she is always a non-contributor. From $\lambda = \frac{U(w^P, g^R)}{U(w^R - g^R, g^R)} = \frac{U^P}{U^R}$, we get

$$\frac{\partial \lambda}{\partial g^R} = \frac{U_G^P * U^R - U^P * (-U_c^R + U_G^R)}{(U^R)^2} \Rightarrow \frac{\partial \lambda}{\partial g^R} \gtreqless 0 \Leftrightarrow U_G^P * U^R - U^P * \left(-U_c^R + U_G^R\right) \gtreqless 0.$$

Players' utility and marginal utility are all greater than zero. When g^R is low, $-U_c^R + U_G^R$ is positive and high. Then it is possible that $\frac{\partial \lambda}{\partial g^R}$ is negative and λ decreases as g^R increases. When g^R is big enough, for example $-U_c^R + U_G^R < 0$ (or $MRS_{c,G}^R > 1$), $\frac{\partial \lambda}{\partial g^R} > 0$ and λ increases as R contributes more.[1]

The players' utility functions are:

$$V^P = U\left(c^P, g^P + g^R\right) - E\left(\lambda_0 - \frac{U(c^P, g^P + g^R)}{U(c^R, g^P + g^R)}\right) = U^P - E\left(\lambda_0 - \frac{U^P}{U^R}\right) \text{ subject to } c^P + g^P \leq w^P$$

[1] Even if $-U_c^R + U_G^R > 0$, it is still possible that $U_G^P * U^R - U^P * \left(-U_c^R + U_G^R\right) > 0$ if the first part dominates.

$$V^R = U(c^R, g^P + g^R) = U^R \text{ subject to } c^R + g^R \leq w^R \quad \square$$

$E(\lambda_0 - \lambda)$ is the "dissatisfaction" from envy and $\lambda = \frac{U(c^P, g^R)}{U(c^R, g^R)}$, where $\frac{\partial E}{\partial(\lambda_0 - \lambda)} > 0$, $\frac{\partial^2 E}{\partial(\lambda_0 - \lambda)^2} > 0$ if $\lambda_0 - \lambda > 0$. For simplicity, I assume $E(.) = 0, \frac{\partial E}{\partial(\lambda_0 - \lambda)} = 0$ if $\lambda_0 - \lambda \leq 0$, P will not get positive utility from "satisfaction" even R's behaviour leads the ratio to be over λ_0. The more the difference between actual utility ratio and expected utility ratio, the more disappointed is P. The convexity means marginal disappointment is increasing as the difference increases.

Players move as in a Stackelberg sequential model. R is the Stackelberg leader and P is the follower. The players' decision procedure is represented as the following two stage game. At stage 1, R chooses how much to contribute: (c^R, g^R) constrained by his/her endowment. At stage two, P observes how much public good R has contributed and decides his/her contribution in order to maximize utility.

Now, suppose P has the power to punish R's action. Assume P can punish R if he/she thinks that R acts so strategically that $\lambda < \lambda_0$. A usual way of punishment is for P to commit some crime against R. I will not consider, however, using "crime" as P's way of revenging R because of the following reasons. Generally, P can get positive benefit from crime and R will lose something. It is obvious that the extra benefit and lost functions must necessarily be introduced. Also, a cost of crime is necessary to keep the model tractable (i.e. if there is no punishment for crime, everyone will go to do it). Then the model becomes more complex and hard to analyze. And, the more important reason is that generally if the cost (i.e. penalty) imposed on crime is quite high, P is deterred by the high cost, and the variability of the model becomes less.

Alternatively, P can punish R by destroying public goods. Even though the destroying of the public good makes P gets less from consumption, his/her dissatisfaction decreases at the same time under some conditions. If the latter effect dominates, P does have incentive to destroy the public goods R contributed at stage 1. And in practice, the penalty for destroying actions on public goods is slight.

Examples include the following: i. dumping garbage in a park to destroy the environment; ii. driving slowly on roads to slow down the traffic. These kinds of "destroying" actions, named “soft” crime, incur a slight cost for P so it is more possible for P to undertake them. This leads to the second assumption:

Assumption 2: *Since the punishment for the “soft” destroying actions to public good is slight, I assume there is no cost incurred by the action in the benchmark model. This will be relaxed in the extension section.*

In summary, it is a reasonable punishing method that P has the chance to destroy public goods. Let D denotes the amount of public good that P destroys. Then the actual public good is the contribution of R minus the amount destroyed by P. One thing needs to be mentioned is that R’s contribution to public good has done before P’s “punishing” action. While the variable λ is variable after R’s contribution since $\lambda = \frac{U(w^P, g^R - D)}{U(c^R, g^R - D)}$ depends on D. When D is negative, P is a contributor and constrained by his budget. Finally, the utility functions that the players actually receive are:

$$V^P = U(c^P, g^R - D) - E\left(\lambda_0 - \frac{U(c^P, g^R - D)}{U(c^R, g^R - D)}\right) = U^P - E\left(\lambda_0 - \frac{U^P}{U^R}\right) \text{ subject to } c^P + \max\{0, -D\} \leq w^P \quad (3)$$

$$V^R = U(c^R, g^R - D) = U^R \text{ subject to } c^R + g^R \leq w^R \quad (4)$$

We want to study how much R contributes when P has the chance to punish. The destruction is based on the fact that P knows how much R contributed. Then R makes his/her decision based on the fact that he/she knows what P will do. So the perfect information assumption is necessary:

Assumption 3: *Both players have perfect information. Each player knows: (i) their own information and P has perfect recall, i.e. P can observe how much R contributes; (ii) the rival’s endowment; (iii) the rival’s action functions, for example, R knows how much P will destroy and hence both players know their rival’s response function.*

Before discussing the strategies the players used to contribute, it is desirable to consider the problem of commitment. The commitment problem is about whether P actually destroys the public good (i.e., if destroying action makes him/her better off) at stage two. This is essential for R to make a decision at stage one since if R believes that P cannot commit, he/she can ignore the "threaten" of punishment from P and make decision as the case without punishment. In our model with the above settings, however, the commitment problem does not exist. P does not promise he/she will destroy or not in state one and his/her destroying action is a function of R's action in stage one. Given the observable variable from stage one, P makes a decision in order to maximize his/her utility. So whether P destroys or not can be perfectly anticipated by R and P will act as R's anticipation. Thus the requirement for commitment problem to exist, i.e. inconsistence between expectation and actuality, is not valid. If we change the game's rule, the commitment problem exists. Suppose P pronounces a destruction level at period 0. R maximizes utility based on P's pronouncement. Then P destroys the level pronounced before at stage 2. If P can commit, we need to find the best destruction level pronounced to give P the highest payoff. If R believes that P will not commit, he/she will make decision as the basic sequential game. The commitment problem will be discussed later.

3 How do the players make their decisions?

The following analysis uses the backward induction method with perfect information to find the Subgame Perfect Nash Equilibrium of this Stackelberg game.

At stage two, only P has chance to "move". He/she decides how much to destroy (contribute). P's decision is based on the following optimal solution for (3). As mentioned above, P observes how much R contributes and takes g^R as given; P can only choose D to maximize V^P. If D<0, P will contribute public good. From the objective function, we can see that the variable D has two effects: (i) when D increases, the amount of public good decrease and hence the utility from consuming commodity bundle decreases. This is a negative effect on V^P; (ii) the ratio λ changes at the same time and hence the dissatisfaction E. $U(c^R, g^R - D)$ also decreases as D increases. To consider how λ changes when D increases, let's denote

the players' utilities as U^P, U^R and the marginal utilities of G given private consumption level as $U_G^P|_{w^P}$, $U_G^R|_{c^R}$ when D is greater than zero. If λ decreases as D increases, there is incentive for P to reduce destruction (or increase contribution) since more destruction will introduce both utility loss and more dissatisfaction. So it is interesting to study the conditions under which λ changes in the same direction as D. A necessary condition required for λ to increase when D increases is $\frac{U^P - U_G^P|_{w^P}}{U^R - U_G^R|_{c^R}} > \frac{U^P}{U^R} \Rightarrow \frac{U^P}{U^R} > \frac{U_G^P|_{w^P}}{U_G^R|_{c^R}} \Rightarrow \lambda = \frac{U(w^P, g^R - D)}{U(c^R, g^R - D)}$ increases and $\lambda_0 - \lambda$ decreases, hence the dissatisfaction E decreases. Since both private goods and public good are normal goods and $U(c^R, g^R - D) \geq U(w^P, g^R - D)$ (from assumption 1) when D>0, the inequality $\frac{U^P}{U^R} > \frac{U_G^P|_{w^P}}{U_G^R|_{c^R}}$ is valid only when the RHS is less than 1. This requires that $\frac{\partial^2 U}{\partial G \partial c} = \frac{\partial^2 U}{\partial c \partial G} > 0$, which means the person with more private consumption goods prefers a public goods increment more than the person with less private goods. In our case, R holds more private goods than that of P. It is obvious that separable utility functions are not valid since their marginal utility of public good does not depend on private consumption goods. And Cobb-Douglas utility does also not work since λ is constant. It is also simple to see that $\frac{U_G^P|_{w^P}}{U_G^R|_{c^R}} < \lambda_0$ is required, otherwise P is satisfied and has no incentive to destroy. Rearrange the above inequality to $\frac{U_G^R|_{c^R}}{U^R} > \frac{U_G^P|_{w^P}}{U^P}$. Since λ is the ratio of P's utility to the R's, it will increases when D increases if the percentage change of P's utility is less than that of R's (i.e. the percentage loss of R is higher than that of P). An example is shown in Appendix A.1. P chooses D to solve the utility maximizing problem. The solution is $D^*(g^R; w^P, w^R, \lambda_0)$. The existence of a positive solution of D depends on special settings and can be proved for the example A.1.The proof is shown in Appendix A.2. In general, there are three possible D^*: (i) $D^* < 0$; optimal solution is that P is a contributor; (ii) $0 \leq D^* < g^R$; P will destroy some of the public good from the first stage; (iii) $D^* \geq g^R$, even P destroys all public good, he/she cannot get the optimal point. This is the end of analysis of stage two. Now go back to the first stage.

At the stage one, R makes a contribution decision. With perfect information, R takes w^P, w^R , λ_0 as given and he/she also knows what D^* will be at stage two based on his/her contribution. Then R chooses g^R and $c^R = w^R - g^R$ to solve the optimal solution of (4). The solution is R's optimal contribution to public good: $g^{R*}(w^P, w^R, \lambda_0)$. By substituting this into the P's destroy function, we can get $D^*(w^P, w^R, \lambda_0)$.

4. Does R contribute more when P has the chance to punish R?

When it is social planner who allocates private consumptions and provides public good, the aggregate marginal rate substitution is equal to one and the outcome is envy-free (proved in Appendix A.3).

Let us first consider the Stackelberg model when P cannot punish R, referred to as the "basic Stackelberg model". R takes P's contribution as a function of g^R rather than taking it as given. If P is non-contributor at the equilibrium, R's optimal solution must be the contribution which set $MRS^R_{G,c} = 1$, denote the solution as g^R_0. Suppose R's contribution at equilibrium is not equal to g^R_0 and P is a non-contributor. Then R's utility at g^R_0 must be higher than his utility at equilibrium whether P is a contributor or not at g^R_0 since $g^P \geq 0$. If P is a contributor at equilibrium, we can get $MRS^R_{G,c} = 1/(1 + \frac{\partial g^{P*}(g^R)}{\partial g^R})$ from first-order condition of (2). Appendix A.4 proves that it is possible that P is contributor and does not have the same utility as R (i.e. envy may be present) at equilibrium in the basic Stackelberg game. This is different from the case where P and R make their decisions simultaneously, in which case envy is absent and players get the same utility following the Warr Theorem (Warr 1983). Whether P is a contributor or not, the aggregate MRS is greater than one and not enough public goods are provided.

Proposition 1*: When P is contributor in the basic Stackelberg model, the equilibrium is the same even if P can punish.*

Proof:

Denote R's equilibrium contribution as g^{R*} in the basic model. Whether P can punish R or not, P always contributes at g^{R*}. So R still contributes the same as before since g^{R*} give him/her the highest payoff.

QED

So only when P is a non-contributor at equilibrium in the basic model is the outcome different if we give punishment power to P. As mentioned above, R's optimal solution is at g_0^R, which set $MRS_{G,c}^R = 1$ in the basic model. For simplicity, let us assume:

Assumption 4: *In the basic (without punishment) model, P is non-contributor no matter how much R contributes.*

This means that P's reaction function given R's contribution, $g^{P*}(g^R) = 0$ for any $g^R \in [0,\ w^R - w^P]$. Given this assumption, P's marginal utility from private consumption is always higher than the marginal utility from contribution. Then $g^P = 0,\ G = g^R$and players' utilities received are:

$$V^P = U(w^P, g^R - D) - E\left(\lambda_0 - \frac{U(w^P, g^R - D)}{U(c^R, g^R - D)}\right) = U^P - E\left(\lambda_0 - \frac{U^P}{U^R}\right) \quad (5)$$

$$V^R = U(c^R, g^R - D) = U^R \text{ subject to } c^R + g^R \leq w^R \quad (6)$$

The fact that P is a non-contributor has two implications. On the one hand, P is better off by consuming all endowment on private consumption at every $g^R \in [0, w^R - w^P]$. In detail, P's contribution has three effects: (i) his/her private consumption decreases; (ii) more public goods are supplied; and (iii) λ changes. Effect (i) is negative and (ii) is positive. Effect (iii) depends on the relative change of utility from commodity bundle between P and R. As shown in Appendix A.1, more contribution from P leads λ to decrease even without considering the loss due to less private consumption by P. Thus P's dissatisfaction increases as he/she contributes more. Hence for any $g^R \in [0, w^R - w^P]$, the aggregate effect is negative (or zero) at $g^P = 0$ if P is non-contributor and for any $g^P > 0$, the aggregate effect is negative. The latter statement is true because the more g^P, the less does P gain from contribution and the

more does P loss due to effect (i) and (iii). In a word, P's utility decreases as g^P increases and hence the lowest value of $g^P = 0$ is the optimal contribution when P does not have chance to destroy. On the other hand, R knows P is non-contributor and contributes g^R to set $MU_c^R = MU_g^R$ if $g^R > 0$ at SPNE, denoted as g_0^R in the basic model.

Given the assumption that P is a non-contributor, suppose the marginal effect of a contribution by P when $g^P = 0$ is negative for any $g^R \in [0, w^R - w^P]$ in the basic Stackelberg model. P is better off by decreasing g^P to negative. If P's budget constraint still holds (i.e. the negative contribution is transferred to private consumption), P's utility curve can reach a maximum value with some negative g^P. The negative contribution can be considered as destruction of the public good. But private consumption cannot be greater than P's endowment. So the actual part of P's utility curve is flatter on the left side of $g^P = 0$. P will always have incentive to destroy since the slope of P's utility curve is negative at $g^P = 0$. The function E, however, is equal to zero when $\lambda_0 - \lambda \leq 0$. Let $g_{\lambda_0}^R$ be R's contribution that sets $\lambda_0 - \lambda = 0$ and λ is increasing for any $g^R > g_{\lambda_0}^R$. Then for any $g^R > g_{\lambda_0}^R$, P's utility decreases when D increases from zero since P cannot gain from less dissatisfaction anymore. So when P can punish and assumption 4 is true, D>0 for $g^R \in [0, g_{\lambda_0}^R)$ and D=0 for $g^R \in [g_{\lambda_0}^R, w^R - w^P)$

R knows P is a non-contributor and destroyer for some g^R. Based on P's reaction function, R chooses g^R to maximize his/her utility. R's utility is less than the case where P does not have a chance to punish if D>0 and equal if D=0 for any $g^R \in [0, w^R - w^P)$. Recall R's optimal contribution is g_0^R in the basic model. Let us consider the optimal solution g^{R*}:

(i) $g_{\lambda_0}^R < g_0^R$

Since D=0 for $g^R \in [g_{\lambda_0}^R, w^R - w^P)$, R still contributes to set $MU_c^R = MU_g^R$ and $g^{R*} = g_0^R$ and there is no destruction.

(ii) $g^R_{\lambda_0} > g^R_0$ and $V^R\left(w^R - g^R_{\lambda_0}, g^R_{\lambda_0}\right) > V^R\left(w^R - g^R, g^R\right)$ for any $g^R \in [0,\ g^R_{\lambda_0})$

Since D=0 for $g^R \in [g^R_{\lambda_0}, w^R - w^P)$, R's utility curve is the same as in the case where there is no punishment. And, for any $g^R > g^R_0$, $\frac{\partial U^R}{\partial g^R} < 0$, then $g^{R*} = g^R_{\lambda_0} > g^R_0$ and there is no destruction.

(iii) $g^R_{\lambda_0} > g^R_0$ and $V^R\left(w^R - g^R_{\lambda_0}, g^R_{\lambda_0}\right) < V^R\left(w^R - g^{R*}, g^{R*}\right)$ for some $g^{R*} \in [0,\ g^R_{\lambda_0})$.

Then $g^{R*} \in [0,\ g^R_{\lambda_0})$ and there has to be destruction. If $g^{R*} \in [0,\ g^R_0]$, the final public goods level is definitely less than g^R_0. If $g^{R*} \in (g^R_0, g^R_{\lambda_0})$, it is ambiguous whether the final public good is greater than g^R_0 since there is destruction. And, the final public good is not zero (i.e. P does not destroy all R's contribution) for any $g^{R*} \in (0,\ g^R_{\lambda_0})$; otherwise R is better off by contributing nothing. This is the interior solution that satisfies: (i) R's marginal utility from one more contribution to public good is equal to his/her marginal utility from one more private consumption; (ii) P's marginal utility from destroying public good is equal to zero.

For the case (iii), an interior solution exists. Let us obtain R's marginal utility from contribution by substituting $c^R = w^R - g^R$.

$$\frac{\partial U}{\partial g^R} = -U^R_c + U^R_G * (1 - \frac{\partial D^*}{\partial g^R}) \quad (7)$$

R's utility increment from one more unit of contribution has to consider how much P will destroy for this unit of public good. When $\frac{\partial D^*}{\partial g^R} = 1$, P will destroy the whole unit of incremental contribution and R can be better off by consuming this unit of wealth as a private good (given that the private good is strictly normal). $\frac{\partial D^*}{\partial g^R}$ cannot be greater than one by similar reasoning. Hence $\frac{\partial D^*}{\partial g^R} < 1$ is required for R to be a contributor at equilibrium. R's utility curve is decided by the value of $\frac{\partial U}{\partial g^R}$. If an interior solution exists, we need $\frac{\partial U}{\partial g^R} = 0$. Let us rearrange (7) to obtain

$$\frac{\partial U}{\partial g^R} = U_G^R * \left(-\frac{U_c^R}{U_G^R} + 1 - \frac{\partial D^*}{\partial g^R}\right) = U_G^R * \left(1 - MRS_{c,G}^R - \frac{\partial D^*}{\partial g^R}\right) \quad (8)$$

Since $U_G^R > 0$, R's utility curve is upward sloped when $1 - MRS_{c,G}^R - \frac{\partial D^*}{\partial g^R} > 0$, and vice versa. For optimal solution (iii), (8) =0 at g^{R*}. We know that $1 - MRS_{c,G}^R$ is greater than zero for $g^R < g_0^R$ and less than zero for $g^R > g_0^R$. So, $\frac{\partial D^*}{\partial g^R}$ is positive and less than 1 for $g^{R*} \in [0, g_0^R)$ and negative for $g^{R*} \in (g_0^R, g_{\lambda_0}^R)$.

Now let us consider players' payoff change. When P does not have chance to destroy, R's utility is $U(w^R - g_0^R, g_0^R)$ and P's utility is $V^P(g_0^R)|_{D=0}$. Let us consider players' payoffs under the cases discussed above when P has the power to punish. If R's optimal contribution is case (i), there is no change of utility for both players. Under case (ii), it is obvious that R is worse off and P is better off. For case (iii), R is definitely worse off. But P's payoff change is not necessary positive. Let us use P's utility-maximizing problem to analyze P's utility change as g^R increases (optimal D is function of g^R and let W^P be P's indirectly utility function given g^R).

$$\frac{\partial W^P}{\partial g^R} = U_G^P * \left(1 - \frac{\partial D^*}{\partial g^R}\right) + E' * \frac{U_G^P * \left(1 - \frac{\partial D^*}{\partial g^R}\right) * U^R - U^P * \left(-U_c^R + U_G^R * \left(1 - \frac{\partial D^*}{\partial g^R}\right)\right)}{(U^R)^2}$$

Since P's marginal utility from destroying the public good is equal to zero for case (iii),

$$\frac{\partial V^P}{\partial D} = -U_G^P + E' * \frac{-U_G^P * U^R + U^P * U_G^R}{(U^R)^2} = 0$$

$$\Longrightarrow \frac{\partial W^P}{\partial g^R} = \left(1 - \frac{\partial D^*}{\partial g^R}\right) * \left(U_G^P * + E' * \frac{U_G^P * U^R - U^P * U_G^R}{(U^R)^2}\right) + E' * \frac{U^P * U_c^R}{(U^R)^2} = E' * \frac{U^P * U_c^R}{(U^R)^2} > 0$$

Hence P's utility is strictly increasing for $g^R \in [0, g_{\lambda_0}^R)$ when $0<D<g^R$. And we also know that $V^P(g_0^R)|_{D=0} \leq W^P(g_0^R)$. Then in case (iii), P's utility is definitely higher than his/her utility when P

does not have the power to destroy if $g^{R*} \in (g_0^R, g_{\lambda_0}^R)$. But $W^P(g^{R*}) < W^P(g_0^R)$ when $g^{R*} \in (0, g_{\lambda_0}^R)$, we cannot say the relationship between $W^P(g^{R*})$ and $V^P(g_0^R)|_{D=0}$. Hence P is not necessarily better off with the possibility of destroying. In conclusion, we can summarize the results as follows:

Proposition 2*: Under assumptions 1-4, when P has the chance to punish,*

(i) R contributes more if $g^{R*} > g_0^R$. One necessary condition is that λ_0-equality is not reached at g_0^R, i.e. we need λ_0 to be big enough.

(ii) Another necessary condition for R to contribute more is that P has an incentive to decrease destruction if R contributes more at g^{R*}, i.e. $\frac{\partial D^*}{\partial g^R} < 0$.

(iii) When λ_0-equality is not reached, the destruction of the public good always happens;

(iv) R is worse off. P's payoff, however, does not necessarily increase with the power to punish R.

(v) The final public good level is higher if $g^{R*} = g_{\lambda_0}^R > g_0^R$, lower if $g^{R*} < g_0^R$ and ambiguous if $g^{R*} \in (g_0^R, g_{\lambda_0}^R)$.

5. Commitment problem

In this section, commitment problems are discussed. As mentioned before, for the benchmark model, there is no commitment problem, so in order to explore commitment possibilities we need to change the game's rules. Suppose that P first announces a level of destruction D^a at stage 0 no matter how much R contributes. R then makes a decision based on P's announcement at stage 1. Then P fulfils his/her announcement at stage 2.

First of all, we need to analyze the equilibrium assuming P can commit. Given P's announcement of destruction level D^a, R solves his/her utility maximization problem $\max_{g^R} V^R = U(w^R - g^R, g^R -$

$D^a) = U^R$. Given D^a, P will destroy all contributions from R if $g^R < D^a$. Hence R is better off to be a non-contributor and receive utility $U(w^R, 0)$ when $g^R < D^a$. Once R contributes more than D^a, it is possible to get an interior solution g^{R*} that satisfies $\frac{\partial V^R}{\partial g^R} = F(g^R; D^a) = -U_c^R + U_G^R = 0$. And, the second-order condition is $\frac{\partial^2 V^R}{\partial {g^R}^2} = \frac{\partial F}{\partial g^R} = -(-U_{cc}^R + U_{cG}^R) + (-U_{Gc}^R + U_{GG}^R) < 0$. So g^{R*} is a local maximum solution.

Suppose $\text{Arg}\max_{g^R} V^R = g^{R*} > D^a$. Let us study how R's utility and contribution change as D^a changes. From R's first-order condition $F(g^R; D^a) = -U_c^R + U_G^R = 0$, we get

$$\frac{\partial F}{\partial D^a} = U_{cG}^R - U_{GG}^R \Longrightarrow \frac{\partial g^{R*}}{\partial D^a} = -\frac{\partial F/\partial D^a}{\partial F/\partial g^R} = \frac{(U_{cG}^R - U_{GG}^R)}{[(U_{cG}^R - U_{GG}^R) + (U_{cG}^R - U_{cc}^R)]} \in (0,1) > 0$$

Let $H^R = U(w^R - g^{R*}(D^a), g^{R*}(D^a) - D^a) = U^R$ be R's indirect utility function given D^a. Then,

$$\frac{\partial H^R}{\partial D^a} = -U_c^R * \frac{\partial g^{R*}}{\partial D} + U_G^R * \left(\frac{\partial g^{R*}}{\partial D^a} - 1\right) < 0 \text{ since } \frac{\partial g^{R*}}{\partial D^a} \in (0,1)$$

The more P announced to destroy, the more is R's contribution and the less is R's utility if R's optimal contribution is greater than zero and P can commit. When D^a=0, R contributes $g_{D^a=0}^R$ and gets the highest utility level as $U(w^R - g_{D^a=0}^R, g_{D^a=0}^R)$. As D^a increases, R contributes more, and his/her utility level $U(w^R - g^{R*}, g^{R*})$ decreases. Once $U(w^R - g^{R*}, g^{R*}) < U(w^R, 0)$, R becomes a non-contributor. So R's utility level $\in [U(w^R, 0), U(w^R - g_{D^a=0}^R, g_{D^a=0}^R)]$ and $g^{R*} \in (\text{Max}\{D^a, g_{D^a=0}^R\}, w^R - w^P)$ or $g^{R*} = 0$. We can summarize these results in the following proposition.

Proposition 3: *When P can commit to a certain level of destruction no matter how much R contributes, the optimal solution of R's contribution given D^a is: $Max\ \{U(w^R, 0), U(w^R - g^{R*}, g^{R*} - D^a)\}$ and $g^{R*} \in (D^a, w^R - w^P]$ or $g^{R*} = 0$. And R's contribution increases as D^a increases if an interior solution exists.*

Now go back to P's decision in stage 0. P chooses the announced destruction level to maximize his/her utility.

$$\max_{D} V^P = U(w^P, g^{R*}(D^a) - D^a) - E(\lambda_0 - \frac{U(w^P, g^{R*}(D^a) - D^a)}{U(w^R - g^{R*}(D^a), g^{R*}(D^a) - D^a)} = U^P - E(\lambda_0 - \frac{U^P}{U^R})$$

If D^a is so high that $\text{Arg}\max_{g^R} V^R = U(w^R - g^R, g^R - D^a) = 0$, P's utility is $U(w^P, 0) - E(\lambda_0 - \frac{U(w^P,0)}{U(w^R,0)}$. P has an incentive to decrease D^a to lead R to be contributor. Let $D^{a\prime}$ be the destruction level at which R is indifferent between a non-contributor and a contributor, i.e. $U(w^R - g^{R*}(D^{a\prime}), g^{R*}(D^{a\prime})) = U(w^R, 0)$. If P decreases $D^{a\prime}$ a little bit, R becomes a contributor and P is more likely to be better off (g^{R*} is high enough that λ increases as g^R increased). So P will not announce D^a so high as to deter R from being a non-contributor. Of cause, P also will not announce D^a too low since R contributes more and P can be better off with higher D^a. From the point view of P, the first best action is that resolving his/her utility maximizing problem at stage 2 (given R's contribution at stage 1).

Proposition 4: *No matter whether an interior solution exists (i.e. $D^{a*} \in (0, D^{a\prime})$ or not, P is better off if he/she does not commit the announcement.*

The proof is in Appendix A.5.

Let us now discuss the two players' payoffs. Suppose whether P can commit is decided exogenously. For simplicity, let us consider only the case where $g^{R*} \in \left(\text{Max}\{D^a, g^R_{D^a=0}\}, w^R - w^P\right)$

Case 1: R does not believes P can commit

Since P always has incentive to deviate from his/her announcement, R is more likely to believe P cannot commit. Instead, R believes P will maximize utility given R's contribution. Then the outcome is the same as the Stackelberg model we discussed before. Let us use g^{R*}_S to denote the optimal solution in the Stackelberg game.

Case 2: R believes that P can commit but P will not

P will maximize utility given R's contribution. R is worse off for any $g^{R*} \neq g_S^{R*}$. Recall the Stackelberg game, P's utility increases as R's contribution increased. So for P to be better off, P has to announce the destruction level which leads the highest g^{R*}. Since $\frac{\partial g^{R*}}{\partial D^a} > 0$, P will announce $D = D^{a'}$ (recall that $D^{a\prime}$ is the announced destruction level at which R is indifferent between non-contributor and contributor. Suppose $g^{R*}(D^{a'}) > g_S^{R*}$, P is better off.

Case 3: R believes that P can commit and P commits his/her announcement

P is not necessarily better off than his/her payoff in the Stackelberg game. Even if R contributes more, P's destruction (commit D^{a*}) is not optimal given $g^{R*}(D^{a*})$. R is definitely worse off compared with in the Stackelberg game if $g_S^{R*} \neq 0$. (Proof in Appendix A.6)

An alternative method is the QCM (Quantity-Contingent Mechanism), where P's commitment to destroy is contingent on R's choice. P announces that he will not destroy if R contributes over or equal to some threshold amount of public good (say g^{Ra}), or P will destroy a certain amount (as before D^a). P commits to his/her announcement for sure. The above is a special case that g^{Ra} is infinite (or more specific, $G^a \geq w^R - w^P$ since R will not contribute more than $w^R - w^P$). P's optimal strategy is (i) choosing the threshold g^{Ra*} such that $U(w^R - g^{Ra*}, g^{Ra*}) = U(w^R, 0)$, if $g^{Ra*} < w^R - w^P$. If $g^{Ra*} \geq w^R - w^P$, P choose $g^{Ra*} = w^R - w^P$ since R will not contribute more than $w^R - w^P$ from assumption 1. (ii) choosing D^a such that $U(w^R, 0) > U(w^R - g^R, g^R - D)$ for any $g^R \in (0, g^{Ra})$. R contributes at g^{Ra*} if he/she believes that P can commit. Otherwise the outcome is the same as the bench mark model. A more general version of QCM would be for P to promise to contributes if R contributes more than g^{Ra}. This will encourage R to contribute more.

6. Extensions

The model can be more realistic and complex by introducing more agents and more rules. The following cases are discussed separately. Of cause, some of them can be combined together.

6.1. Introduce Government

A third party, the government named G, is the authority whose goal is to maximize players' aggregate utility. G can achieve its goal by the following methods.

First, G has the power to post regulations for P's destruction. If P does some destruction, P has to suffer from fine prescribed by G's regulation. Suppose it is costless for G to post the regulation and there is no way for P to escape from the fine. It is equivalent to posting a cost for P's destruction. Hence, everything else is the same as before, but P's utility function changes to $V^P = U(w^P - F(D), g^R - D) - E(\lambda_0 - \lambda)$; $F(D) = 0$ if $D \leq 0$; $\frac{\partial F}{\partial D} > 0$ and $\frac{\partial^2 F}{\partial D^2} > 0$.

P has to consider the cost he/she must pay for the action of destruction. For any g^R such that $D(g^R) \in (0, g^R)$ in the basic Stackelberg model, P will destroy less when the regulation is activated. It is not clear that R will contribute less or more at equilibrium with the regulation since it depends on the specific parameters and utility functions. But if R contributed less, P is worse off and R is better off. The former is because P's utility is increasing in g^R. Even at the initial optimal contribution level, R is better off (or at least the same as before) since P destroys less (or the same). So R's lower contribution implies R is better off than before. The change of aggregate utility is ambiguous. If aggregate utility increases, G should introduce the regulation. Otherwise, G should not. In addition, it is generally not free for G to post and practise the regulation since management and supervision are costly. The cost is fixed once G adopts the regulation. G has to balance its budget. The cost can be shared between P and R or paid by R only. Players' utility functions adjust correspondingly. Then G can compare the aggregate utility with and without the regulation and decide whether introduce the regulation.

In the above paragraph, G collects revenue from players to cover its cost for regulation. Now, let us consider the case when G can transfer wealth from R to P. The public good, however, is still provided by players R and P in the Stackelberg game. Hence G's problem is $\mathrm{Max}_{w^P, w^R}\, V^P + V^R$ Subject to: $w^P + w^R = W$. For any pair (w^P, w^R), by solving P and R's utility maximization problems as before, we can get aggregate utility as $V^{P*}\left(w^P, w^R\right) + V^{R*}\left(w^P, w^R\right)$. An optimal solution exists since $0 \leq w^P, w^R = W - w^P \leq W$. It is not necessary for the optimal solution that G redistributes W such that the game is λ_0-equitable. It is possible that aggregate utility is higher in the case that envy is present (hence there is destruction) than in the case that envy is absent. If the parameter λ_0 is a function of w^P and w^R, it must be taken into consideration when we solve G's problem.

In addition, if G provides public good rather than players, G acts as the social planner with the power of transferring wealth. P and R get the same utility level and envy is absent. The Samuelson condition is satisfied.

6.2. Introduce R's self-protection mechanism

Instead of introducing G and regulation, P's destruction is not free if R can do something to protect himself. The simplest case is that R can spend money to increase the cost of P's destruction. R's utility function becomes $V^R = U\left(w^R - g^R - p, g^R - D\right)$, where p is the amount of wealth R spent to provide self-protection. P's utility function is now $V^P = U\left(w^P - C(p, D), g^R - D\right) - E(\lambda_0 - \lambda)$; $C(p, D) = 0$ if $D \leq 0$; $\frac{\partial C}{\partial D} > 0, \frac{\partial C}{\partial p} > 0$ *and* $\frac{\partial^2 C}{\partial D^2} > 0, \frac{\partial^2 C}{\partial p^2} < 0, \frac{\partial^2 C}{\partial D\, \partial p} > 0$. The cost of P's destruction is increasing in D and p. The higher p is, the higher is the cost P has to pay for the same amount of D.

6.3. Some other considerations

All the above analysis is based on perfect information. The outcome may be different if uncertainty is introduced. Suppose there are two types of P. "Normal" P follows the setting above. "Mild"

type P will never destroy. The "mild" can be understood in the following two ways: (i) the mild P does not care about inequality; or (ii) the mild P's cost to destroy is infinite high, thus he/she is deterred from destroying. The types of P follow some natural probability distribution. Then Bayesian equilibrium is expected.

When there are more than one individual in group P and group R, the analysis becomes more complex. One thing that has to be considered is whether group members act cooperatively or not. For example, if agents in group P make their decision separately, the aggregate destruction is higher than the case they act cooperatively, all other things being the same.

7 Conclusions

When envy is introduced to a player's utility function directly, that player has to consider not only consumption bundle but also envy induced by inequality. The ability of punishment to improve public goods contributions is dubious. P punishes R by destroying public goods rather than doing crime on R directly. In our two-player, λ_0-equality seeking and Stackelberg sequential game, one necessary condition for punishment opportunity to have effect is that P should not be a contributor at equilibrium in the basic model (i.e. the model in which P cannot punish R). With the assumption that P is never a contributor, the public good level is not necessarily greater when P has the power to destroy. R is worse off but P is not necessarily better off. At equilibrium, if λ_0-equality is not reached, P will always destroy some of R's contribution. Aggregate utility is not necessarily greater when P can punish. When P promises a certain amount of destruction, R contributes more for a higher announced destruction level, and P will always have incentive to deviate from the announcement. The public good level depends on whether (i) P can commit; (ii) R believes P can commit and (iii) specific functions and parameters.

Appendix

A.1 A utility function satisfies the requirement that λ increase when D>0

$U(c, G) = c^{\frac{1}{2}}G^{\frac{1}{2}} + A$, where $A > 0$, P is non-contributor and $G = g^R$.

Proof:

$$\lambda^1 = \frac{U^P}{U^R} = \frac{w_P^{\frac{1}{2}} g^{R\frac{1}{2}} + A}{c_R^{\frac{1}{2}} g^{R\frac{1}{2}} + A} = \frac{a + A}{b + A} = \frac{1}{T}$$

With destroying D,

$$\Delta U^P = w_P^{\frac{1}{2}} g^{R\frac{1}{2}} + A - \left(w_P^{\frac{1}{2}}\left(g^R - D\right)^{\frac{1}{2}} + A\right) = w_P^{\frac{1}{2}}(g^{R\frac{1}{2}} - \left(g^R - D\right)^{\frac{1}{2}})$$

$$\Delta U^R = c_R^{\frac{1}{2}} g^{R\frac{1}{2}} + A - \left(c_R^{\frac{1}{2}}\left(g^R - D\right)^{\frac{1}{2}} + A\right) = c_R^{\frac{1}{2}}(g^{R\frac{1}{2}} - \left(g^R - D\right)^{\frac{1}{2}})$$

$$\Rightarrow \frac{\Delta U^P}{\Delta U^R} = \left(\frac{w_P}{c_R}\right)^{\frac{1}{2}} = \left(\frac{w_P g^R}{c_R g^R}\right)^{\frac{1}{2}} = \frac{a}{b} = \frac{1}{t}$$

Since $w_P^{\frac{1}{2}} g^{R\frac{1}{2}} < c_R^{\frac{1}{2}} g^{R\frac{1}{2}}$ with assumption that $U\left(c^R, g^R\right) > U\left(w^P, g^R\right)$

$$\Rightarrow \frac{1}{T} = \frac{a + A}{b + A} > \frac{a}{b} = \frac{1}{t}$$

$\Rightarrow T < t, given\ t, T\ \in [1, \infty)$

Then

$$\lambda^2 = \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A} = \frac{U^P - \Delta U^P}{U^R - \Delta U^R} = \frac{U^P - \Delta U^P}{TU^P - t\Delta U^P} > \frac{U^P - \Delta U^P}{TU^P - T\Delta U^P} = \frac{1}{T} = \lambda^1$$

It is shown that λ is increasing as destroying action is undertaken. Hence the destroying action reduces P's disappointment.

QED

A.2 Positive solution of D exists for example A.1

Proof: Assume the disappointment function $E = \frac{1}{2}B * (\lambda_0 - \lambda)^2$, where $B > 0$ and $\lambda_0 - \lambda > 0$

$$V^P = w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A - \frac{1}{2}B * \left(\lambda_0 - \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}\right)^2 \text{ and } V^R = c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}}$$

At stage two, P chooses D to maximize V^P, given g^R:

$$\underset{D}{Max}\, V^P = w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A - \frac{1}{2}B * \left(\lambda_0 - \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}\right)^2$$

$$\frac{\partial V^P}{\partial D} = 0 = -\frac{1}{2}w_P^{\frac{1}{2}}(g^R - D)^{-\frac{1}{2}} - B * \left(\lambda_0 - \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}\right) * (-1)$$

$$* \frac{-\frac{1}{2}w_P^{\frac{1}{2}}(g^R - D)^{-\frac{1}{2}} * \left(c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A\right) + \frac{1}{2}c_R^{\frac{1}{2}}(g^R - D)^{-\frac{1}{2}} * (w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)}{(c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)^2}$$

D=g^R or

$$w_P^{\frac{1}{2}} + B * \left(\lambda_0 - \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}\right) * \frac{w_P^{\frac{1}{2}} * \left(c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A\right) - c_R^{\frac{1}{2}} * (w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)}{(c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)^2} = 0$$

$$\Rightarrow w_P^{\frac{1}{2}} + B * \left(\lambda_0 - \frac{w_P^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}{c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A}\right) * \frac{w_P^{\frac{1}{2}} * A - c_R^{\frac{1}{2}} * A}{(c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)^2} = 0$$

For simplicity, suppose $\lambda_0 = 1$

$$\Rightarrow (c_R^{\frac{1}{2}}(g^R - D)^{\frac{1}{2}} + A)^3 = w_P^{-\frac{1}{2}} * AB * (c_R^{\frac{1}{2}} - w_P^{\frac{1}{2}})^2 * (g^R - D)^{\frac{1}{2}}$$

⇒ Given appropriated value of A and B, there is non-negative solution for g^R-D.

QED

A.3 The solution of social planner's problem

Social planner has the power to allocate resources to private consumptions and provide public good.

$$\underset{c^P,c^R,G}{\text{Max}}\ V = U(c^P, G) + U(c^R, G) - E(\lambda_0 - \frac{U(c^P, G)}{U(c^R, G)}) = U^P + U^R - E(\lambda_0 - \frac{U^P}{U^R})$$

Subject to: $c^P + c^R + G \leq w^P + w^R$

First order conditions with respect to c^P, c^R, G:

$$\frac{\partial V}{\partial c^P} = U_c^P + E' * \left(\frac{U_c^P}{U^R}\right) = \mu$$

$$\frac{\partial V}{\partial c^R} = U_c^R - E' * \left(\frac{U^P * U_c^R}{(U^R)^2}\right) = \mu$$

$$\frac{\partial V}{\partial G} = U_G^P + U_G^R + E' * \left(\frac{U_G^P * U^R - U^P * U_G^R}{(U^R)^2}\right) = \mu$$

We know that the shadow price is greater than zero and $U_c^P > U_c^R$ with assumption 1 and $\frac{\partial^2 U}{\partial c^2} < 0$. Then from the first two equations, we get $E' * \left(\frac{U_c^P}{U^R}\right) < -E' * \left(\frac{U^P * U_c^R}{(U^R)^2}\right) \Rightarrow E' = 0$ since all elements except E' are greater than zero and $E' \geq 0$. With the definition of function E, we know that E=0 when$E' = 0$. The above equations can be simplified to

$$\frac{\partial V}{\partial c^P} = U_c^P = u = U_c^R = \frac{\partial V}{\partial c^R}$$

$$\frac{\partial V}{\partial G} = U_G^P + U_G^R = \mu$$

$$\Rightarrow MRS_{G,c}^P + MRS_{G,c}^R = 1$$

In conclusion, for the social optimal, the aggregate marginal rate substitution is equal to one and the envy is free.

A.4 Players do not need to get the same utility when P is also contributor at equilibrium in the sequential game

Let's suppose P is contributor at the equilibrium. From the first order conditions of (1), we get

$$\frac{\partial V^P}{\partial c^P} = U_c^P + E' * \left(\frac{U_c^P}{U^R}\right) = U_G^P + E' * \left(\frac{U_G^P * U^R - U^P * U_G^R}{(U^R)^2}\right) = \frac{\partial V^P}{\partial g^P}$$

$$\Rightarrow MRS_{G,c}^P = 1 + \frac{E' * U^P * U_G^R}{U_c^P * ((U^R)^2 + E' * U^R)} \quad (3)$$

Suppose $\lambda_0 - \lambda > 0$, then $E' > 0$. With assumption 1, we need $MRS_{G,c}^R > MRS_{G,c}^P$ and hence P and R do not have the same utility.

$$\Rightarrow \frac{1}{1 + \frac{\partial g^{P*}}{\partial g^R}} > 1 + \frac{E' * U^P * U_G^R}{U_c^P * ((U^R)^2 + E' * U^R)} > 1$$

$$\Rightarrow \frac{\partial g^{P*}}{\partial g^R} \in (-1, \frac{1}{1 + \frac{E' * U^P * U_G^R}{U_c^P * ((U^R)^2 + E' * U^R)}} - 1)$$

So it is possible that P is contributor at equilibrium if $\frac{\partial g^{P*}}{\partial g^R}$ locates in the above domain and players have different utility level at end of the game.

A.5 (Proof of Proposition 4)

Proof:

Case (i), $g^{R*}(D^{a*}) \in (g_{\lambda_0}^R, w^R - w^P)$

By Proposition 2, P definitely has incentive to not commit since any destruction introduces loss of utility.

Case (ii), $g^{R*}(D^{a*}) = 0$

P cannot commit since there is no public good to destroy.

Case (iii), $g^{R*}(D^{a*}) \in (D^{a*}, g^{R}_{\lambda_0})$

The λ_0-equality is not reach and we have to consider the effect of dissatisfaction.

Let's consider P's marginal utility from D^a:

$$\frac{\partial V^P}{\partial D^a} = \left(\frac{\partial g^{R*}}{\partial D^a} - 1\right)\left(U^P_G + E' * \frac{U^P_G * U^R - U^P * U^R_G}{(U^R)^2}\right) + E' * \frac{U^P * U^R_c * \frac{\partial g^{R*}}{\partial D^a}}{(U^R)^2} \quad (9)$$

(a) If $D^{a*} \in (0, D^{a'})$, (9) is equal to zero. The second part of RHS is greater than zero since $\frac{\partial g^{R*}}{\partial D^a} \in (0, 1)$.

$$\Longrightarrow U^P_G + E' * \frac{U^P_G * U^R - U^P * U^R_G}{(U^R)^2} > 0$$

$$\Longrightarrow -U^P_G + E' * \frac{-U^P_G * U^R + U^P * U^R_G}{(U^R)^2} < 0 \text{ at } D^{a*}$$

At stage 2, g^{R*} is taken as given since it is decided at stage 1. Then P chooses destruction D^* to maximize utility.

$$\frac{\partial V^P}{\partial D} = -U^P_G + E' * \frac{-U^P_G * U^R + U^P * U^R_G}{(U^R)^2} = 0 \text{ Given } g^{R*}(D^a)$$

While from above analysis, we know $-U^P_G + E' * \frac{-U^P_G * U^R + U^P * U^R_G}{(U^R)^2} < 0$ when P commits at D^{a*}. So P will not destroy less than D^{a*} and had incentive to not commit.

(ii) If $D^{a*} = D^{a'}$, (9) is greater than zero (if equal to zero, it is the same as above analysis). Since the second part of RHS is greater than zero, we get:

$$U^P_G + E' * \frac{U^P_G * U^R - U^P * U^R_G}{(U^R)^2} \gtrless 0$$

P has incentive to destroy more if ">" holds or destroy less if "<" hold than $D^{a*} = D^{a'}$. But P will not commit D^{a*}.

(iii) If $D^{a*} = 0$, $g^{R*} = g_0^R$. P has incentive to destroy with the assumption that P is non-contributor.

In conclusion, no matter what the optimal destruction level P announced at stage 0, P always has incentive to deviate from the announcement.

QED

A.6 R is definitely worse off compared with in the Stackelberg game if $g_S^{R*} \neq 0$ when P can commit and R believes P can commit

Proof:

R is contributor in Stackelberg game imply $U(w^R, 0)$ is less than R's utility at g_S^{R*}. As discussed above, D^{a*} is only less than P's optimal destruction given R contributes $g^{R*}(D^{a*})$ when $D^{a*} = D^{a'}$. But at $D^{a'}$, R's utility is equal to $U(w^R, 0)$, which is less than R's utility in Stackelberg game when $g_S^{R*} \neq 0$.

If $D^{a*} < D^{a'}$, for $g^{R*}(D^{a*})$ R contributes, R is worse off since D^{a*} is higher than P's optimal destruction in the Stackelberg game and hence R's utility with D^{a*} is lower than before.

It is obvious that R's utility is higher in case 3 than in case 2 if $D^{a*} < D^{a'}$ and lower if $D^{a*} = D^{a'}$

QED

Reference

Bergstrom, T.C., Goodman, R.P., 1973. *Private demands for public goods*. American Economic Review 63, 280–296.

Boadway, R., Song, Z., Tremblay, J., 2007. *Commitment and matching contributions to public goods.* Journal of Public Economics, Volume 91, Issue 9, page 1664-1683

Chaudhuri, A., 1986. *Some implications of an intensity measure of envy*. Social Choice and Welfare 3, 255–270.

Diamantaras, D., Thomson, W., 1990. *A refinement and extension of the no-envy concept*. Economics Letters 33, 217–222.

Epple, D., Romano, R.E., 1996. *Public provision of private goods*. Journal of Political Economy 104, 57–84.

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

Fehr, E., Gächter, S., 2002. *Altruistic punishment in humans*. Nature 415, 137–140.

Feldman, Allan and Weiman, David., 1978. *Envy, wealth, and class hierarchies*. Journal of Public Economics, (1979) 81-91.

Foley, D., 1967. *Resource Allocation and the Public Sector*. Yale Economic Essays 7.

Guttman, J. (1978), *'Understanding Collective Action: Matching Behavior'*, American Economic Review 68, 251—55.

Helsley, R.W., Strange, W.C., 1998. *Private government*. Journal of Public Economics 69 (2), 281– 304.

Helsley, R.W., Strange, W.C., 2000a. *Social interactions and the institutions of local government.* American Economic Review 90 (5), 1477–1490.

Ledyard, J., 1995. *Public goods: a survey of experimental research*. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics. Princeton University Press.

Nikiforakis, N., 2007. *Punishment and counter-punishment in public good games: Can we really govern ourselves?* Journal of Public Economics, 92 (2008) 91–112

Nishimura, Y., 2001. *Optimal non-linear income taxation for reduction of envy.* Journal of Public Economics, 87 (2003) 363–386

Olson, M., 1965. *The Logic of Collective Action.* Harvard University Press, Cambridge, Massachusetts.

Ostrom, E., Walker, J., Gardner, R., 1992. *Covenants with and without a sword: self governance is possible*. American Political Science Review 86, 404–417.

Samuelson, P., 1954. *The pure theory of public expenditure*. The Review of Economics and Statistics 36 (4), 387–389.

Varian, H., 1974. *Equity, Envy and Efficiency*. Journal of Economic Theory 9, 63-91.

Warr P. (1983). *'The private provision of a public good is independent of the distribution of income'*. Economic Letters, vol. 13, pp. 207—211.