Working For the Common Good

Effects of Endogenous Effort in a Public Goods Game*

Sebastian Flennerhag†, David Rydén‡

March 2015

Abstract

Public goods games have been extensively studied in homogeneous environments where punishment mechanisms are typically found to be welfare enhancing. However, the literature on public goods games in heterogeneous environments have been relatively sparse, and with considerably less robust results. We investigate the effects of a punishment mechanism when individuals control their level of endowment. We study experimentally how contribution and effort evolves over a repeated public goods game as other-regarding beliefs changes and analyse the effects of a punishment mechanism in such a setting. In particular, we compare the implications for the punishment mechanism under different normative principles of cooperation.

Keywords: Public Goods Game, Punishment, Endogenous Endowment

*Term paper, 5316 Behavioural Economics, Stockholm School of Economics

†22273

‡40784

Contents

Introduction

A pervasive feature of a society is the need for cooperation among individuals. In some of these scenarios, the interests of the common align with those of the individual. However, in a great many other settings we are not so fortunate. Ranging from small scale scenarios such as a group sharing the costs of a dinner bill, to more high-stake issues such as tax evasion, sharing infrastructure or controlling environmental emissions. (Falkinger et al, 2000; Poppe, 2005). Such scenarios constitute specific instances of the general public good dilemma, where the interests of the common are in conflict with those of the individual (Reuben and Riedl, 2012). For this reason, the public goods game is generally considered a good laboratory approximation to social dilemmas.

However, in past studies, experiments have abstracted away from the effort decision an individual face before or simultaneously with the cooperation decision. While such abstractions are immaterial to a certain class of social dilemmas, many important problems cannot be analyzed in such a framework. For instance, a tax evader must also decide how hard to work, and thus how much income to earn, in addition the actual decision of tax evasion. Perhaps even more relevant for policy making is the effects social security has on the level of effort various populations exerts. Similarly, when individuals contribute to a common infrastructure, such as when farmers share irrigation system (Reuben and Riedl, 2012), farmers simultaneously decide on how much effort to put into their farming, and how much to contribute to the maintenance of the common irrigation system. In this paper, we design an experiment to study the interrelationship between the effort decision and the contribution decision. In the absence of an enforcement mechanism, we find that contribution levels decrease, with a likely consequence that individuals compensate by increasing their private income through higher effort.

However, when one introduces an enforcement mechanism, the dynamics of the social dilemma changes materially due to the fact that a rational individual would factor in how the enforcement mechanism reacts not only to changes in contribution but also in private earnings. In this paper we study how the typical punishment mechanism introduced by Fehr and Gächter (2000) performs when individuals threat the endowment decision as an endogenous variable through a costly effort. We hypothesize[1] that while the punishment mechanism typically is welfare enhancing when effort is exogenous, it is considerably less robust in our environment, and may even be welfare destructive. This result is brought about by two findings in previous studies; i) while a majority of people conditionally cooperate, they do so imperfectly (Fischbacher and Gächter, 2010) and ii) in an environment where individuals have heterogeneous endowments, punishment appears to enforce a proportional contribution rule (Reuben and Riedl, 2012). The first result puts a downward pressure on

[1] Alas, since time restrictions for this course do not allow us to perform the actual experiment, we must make do with hypotheses.

any level of contribution, and the second result means that individuals have an incentive to mitigate punishment by reducing effort rather than increasing contributions.

A key complication for stable cooperation in an heterogeneous environment is the fact that there is no prevailing norm to which individuals automatically rally. In a homogeneous environment, studies have shown (Fehr and Gächter, 2000; see Chauduri (2009) for further references) a rather robust tendency for convergence on an equal contribution close to full capacity. However, in a heterogeneous environment, there is a multiplicity of normative equilibria, with current microeconomic theory being incapable of predicting which may prevail (Reuben and Riedl, 2012). While Reuben and Riedl show that a proportional contribution norm seem to prevail, their study keep the effort decision exogenous. The question then arises as to what enforcement norm, if any, may arise when individuals are free to choose their effort level as well. We design an experiment to disentangle the relative importance of deviations from the average level of contribution from the relative importance of deviations from the average relative contributions. We study how each type of norm enforcement affects effort and contributions decisions and what the implication are for the general level of welfare.

Previous Research

The typical public goods game studied is linear in the returns from contributions and allocates each individual some endowment. In such settings, a robust finding is that individuals contribute a nonzero amount (We refer to surveys of the literature by Chauduri (2009) and Levitt and List (2007) for further references) and when the game is repeated contribution levels fall progressively and eventually converge on the Nash equilibrium, irrespective of whether an individual face the same participants in each rounds or a random set of "strangers".

Fehr and Gächter (2000) showed that, in contradiction to what standard game theory predicts, when individuals are able to punish each other at a personal cost cooperation can be sustained. A remarkable result is that the punishment mechanism is effective even in the "stranger" treatment, where subjects are randomly reshuffled. That is, even when subjects meet a random set of individuals in each round, punishment increases contributions over time. While punishment in a partner protocol can be thought of as a signalling strategy, punishment in a stranger treatment indicates that the punishment can be motivated by general normative principles. Also, the changes in behavior over rounds indicate that individuals update their beliefs regarding the general behavior of others.

Further studies into the dynamics of the punishment mechanism shows that the cost-efficiency ratio of the punishment mechanism is an important factor (Nikiforakis and Normann, 2008), and that the number of rounds to be played matters not only for the end

game, but for the initial behavior as well (Gächter et al, 2008). Furthermore, as the feedback dynamics of the punishment mechanism increases, its efficiency appears considerably less stable. It appears particularly sensitive to the type of information individuals receive about the behavior of their co-players. (Nikiforakis, 2008; Cinyabuguma et al., 2006; Nikiforiakis et al, 2013).

When effort is an endogenous variable, punishment can dependent not only on contribution deviations, but also on effort deviations. Heterogeneous endowments introduce considerable complexity, as it is no longer clear what normative principles may guide either contribution or punishment. Reuben and Riedl (2013) note that when individuals have social preferences, the public goods game can be thought of as an issue of coordination over multiple Nash equilibria. One way of solving the coordination problem is for the subjects to (in this case tacitly) agree on a contribution norm. However, with endowment heterogeneity there is multiplicity of potential norms and with different type of subjects, there may be no consensus as to which should prevail. They show that with heterogeneous endowments, punishment behavior seem to cluster around two types of enforcement norms; either the enforcement of full contribution, or convergence on some relative contribution norm.[2] This result indicate that the prevailing norm seems to be relative to endowment, albeit their framework only considers heterogeneous endowments. Building on this, a question of importance is why some groups achieve full efficiency and others not. Two plausible explanation are that the group distribution of player types matter and that there might be some path dependency. Unfortunately, since Reuben and Riedl use a partner protocol one cannot answer this question from their results. Furthermore, when effort is an endogenous choice, it is not clear that a relative enforcement norm will prevail, as the presence of individualistic normative principles would discourage such behavior.

When one adopts a stranger protocol as we do here, punishment loose much of strategic signalling value, yet previous studies indicate that stranger protocol generate the same typical pattern as partner protocol. Fischbacher and Gächter (2010) consider this remarkable fact and show how the gradual decline can be explained as changes in behavior induced by a participants updating their beliefs in the direction of the observed outcome in each round. While the belief updating in itself cannot cause a change, a key factor was what they call "imperfect conditional cooperators". This type of player constitute the predominant type at ≈ 55 %, and only match others contributions partly. On average, actual contributions will be lower than expected contributions and when the game is repeated this causes individuals to revise their expectations downwards, and so lowers their own contribution furtherr. If the majority engage in this type of behavior, cooperation is bound to unravel gradually.

While their study was within a homogeneous framework. The results of Cherry et al

[2]Enforcing a relative contribution norm stipulates that individuals are punished from deviating from the average relative contribution.

(2005) indicate that this effect might be even stronger in an heterogeneous environment. In line with Fischbacher and Gächter (2010), they find what they call an "anticipatory reciprocity effect" by which subjects with high endowment will contribute less since others cannot reciprocate a higher contribution on their part. .

To the best of our knowledge, the effect of effort on behavior in public goods game has only employed a design where subject perform a one-shot competitive task prior to playing the public goods game. This type of design cannot endogenize the effort mechanism fully, since individuals are not allow to change their effort level in response to updated beliefs. Nevertheless, these studies can shed light into a potential "endowment effect", and so far the results indicate an immaterial effect, although results are not robust (Clark, 2002; Harrison, 2007; Cherry et al, 2005; Antinyah et al, 2014). Most relevant for our purposes are the findings of Muehlbacher and Kirchler (2009), who show that the propensity to contribute is negatively correlated with the level of effort subjects exert.

A Simple Model of Endogenous Effort

When effort is endogenous, the decision dynamics are different from the typical public goods game. Since effort comes at a cost outside the game itself, one must consider how the choice of effort relates to individuals preferences outside the game itself. To facilitate such an analysis, we develop a simple model of the decision dilemma facing an individual and infer likely outcomes of the experiment.

The effort level e_i an individual i ultimately decides on depends on two factors, i) the payoff π_i from the public goods game and ii) the resulting utility U_i an individual derives from such an outcome. As for i), this is the payoff from the public goods game, either with or without punishment,

$$\pi_i^p = \alpha e_i + (\beta - 1)c_i + \beta C - P_j - P_i \tag{1}$$

$$\pi_i = \alpha e_i + (\beta - 1)c_i + \beta C \tag{2}$$

Where α is an earnings multiplier, c_i is individual i's contribution, β denotes MPCR[3], C denotes the sum of contributions from the $(n-1)$ other participants, P_j the punishment received from other participants and P_i punishment allocated to other participants. As for ii) we follow the literature on experimental models and assume a rational utility maximizing agent with a utility function that is additively separable in material and moral preferences[4]; $U_i(\cdot) = W_i(\cdot) - K_i(\cdot) + M_i(\cdot)$, where K_i denotes some effort cost function.

We will impose that these functions are continuous and everywhere differentiable. The

[3]For the social dilemma to occur we must have that $\beta \in (0,1)$, $n\beta > 1$

[4]See Levitt and List (2007) for a discussion of this approach

wealth function W_i, being a standard self-regarding utility function, is concave and non-decreasing in payoff π_i, while the effort cost function K_i is convex and strictly increasing in effort (e_i). Assume that M_i depends on deviations in the contribution level from other participant's average, $\bar{c}$.[5] Also, for every $\bar{c}$, M_i is strictly concave in own contributions c_i and has a unique maximum. Through this characterization, M_i accommodates moral effects from both negative and positive deviations. In the case without punishment, the resulting utility function is

$$\begin{aligned} U_i(e_i, c_i) &= W_i(\alpha e_i + (\beta - 1)c_i + \beta C) - K_i(e_i) + M_i(c_i - \bar{c}) \\ &= W_i(\pi_i) - K_i(e_i) + M_i(c_i) \end{aligned} \tag{3}$$

Implications in the Absence of Punishment

From above, one implication is that if an agent is morally indifferent, such a person adhere to standard preferences. Also, since effort is undertaken before knowing what actual contributions are, an individual must form some expectation of others contribution. And so for a given belief about the contributions of others, maximization implies

$$\frac{\partial W_i}{\partial e_i} = \frac{\partial K_i}{\partial e_i} \qquad , \qquad -\frac{\partial W_i}{\partial c_i} = \frac{\partial M_i}{\partial c_i} \tag{4}$$

Thus, i decides on a contribution level up to the point where the marginal moral utility is off-set by the marginal material disutility. Simultaneously, i decides on an effort level where the marginal utility of gaining more wealth is off-set by the cost of exerting more effort. Note that since W_i is concave and K_i is convex, greater returns from the public good results in lower effort, ceteris paribus.

To analyse how effort interacts with the belief updating process, suppose for simplicity that each player in this game is "imperfectly conditionally cooperative"[6]. Now, for any given initial belief C_1, the outcome will be less since players contribute imperfectly[7]. When the game is repeated, participants may update their beliefs, and if we suppose this process to follow that of Fischbacher and Gächter (2010), where beliefs in the next period is a weighted average of beliefs in the prior period and the outcome in that period, beliefs are revised downwards which induces lower contributions. Simultaneously, with lower expected

[5] This abstracts away from the fact that people may be guided not only by what people contribute, but also by how much they could have contributed, i.e. their endowments. However, since our main purpose is to study the incremental effect from punishment, M_i is only of secondary importance and with endowments being earned, the moral implications are ambiguous

[6] While a more realistic distribution of types in accordance with e.g. those found by Fischbacher and Gächter (2010), the fact that other types tend to be more self-regarding would not change our conclusions.

[7] A possible way to understand this finding is to note that if M_i is symmetric around the optimum, this would imply perfect conditional cooperation. However, increasing cooperation yields lower material utility which shifts the optimal contribution towards slightly less contribution. Thus, material incentives distort moral incentives.

returns from the public good, the marginal utility of effort increases. And so as the game progresses and the typical convergence towards free-riding materialize, individuals would compensate by earning more privately.

Prediction I: Without punishment, contributions will deteriorate over rounds. As contributions deteriorate, effort will increase.

Introducing Punishment

Now, consider the effects of a punishment mechanism. As the previous section illustrated, it is not entirely clear whether punishment is driven by deviations from the level contributions or contributions relative to earnings. While Reuben and Riedl (2012) found evidence for a relative enforcement norm, a key difference in with this experiment is the endowment is earned and not based on random sorting. How individuals factor in this fact is an ambiguous matter. In the one extreme, one normative principle suggest that money earned above and beyond what others contribute is for the individual to dispose. This suggest an enforcement norm that is dependent on deviations in level contribution. In the other extreme, a type of normative principle suggest that individuals should contribute according to their capacity. This imply a relative contribution norm. Most likely, people use a little bit of both. In this paper, we study how each of these norms affect behavior, and which norm seems to be the strongest guiding principle when both are available.

We define a pure "level enforcement norm" as $\theta^l[c_i - \bar{c}]$, and a pure "proportional enforcement norm" as $\theta^r[c_i/e_i - \bar{c}/\bar{e}]$, where θ is some function determining the resulting punishment. To simplify matters, we restrict punishment only to negative deviations (which excludes "anti-social" punishment) and impose that i herself will not punish others. The payoff from the public goods game is then

$$\pi_i^l = \alpha e_i + (\beta - 1)c_i + \beta C - \theta^l(c_i - \bar{c})$$
$$\pi_i^r = \alpha e_i + (\beta - 1)c_i + \beta C - \theta^r\Big(\frac{c_i}{e_i} - \frac{\bar{c}}{\bar{e}}\Big)$$

To introduce punishment in the utility function, assume that i deviates from the norm, expects to be punished and derives disutility $\Phi = \phi(\theta)$ from being punished. Then

$$U_i(e_i^k, c_i^k) = W_i(\pi^k) - K_i(e_i^k) + M_i(c_i^k) - \Phi_i^k(\cdot) \quad , \quad k = \{l, r\} \tag{5}$$

For ease of exposition, we suppress the subscripts in the following treatment.

A Level Enforcement Norm

Consider the contribution decision facing i,

$$-\frac{\partial W_i}{\partial \pi_i^l}\frac{\partial \pi_i^l}{\partial c_i} = \frac{\partial M_i}{\partial c_i} - \frac{\partial \Phi^l}{\partial c_i} \tag{6}$$

That is, the marginal effect on wealth from changing contribution may now be positive for certain changes, as the reduction in punishment may be greater. Therefore, there is now a material incentive to increase contributions. Similarly, moral preferences now include an additional term describing the reduced disutility from lower punishment, and being positive in increases in contribution, the moral optimum shifts towards higher contribution. Since there are now both material and moral incentives to increase contribution, we predict the following;

Prediction II: With a level enforcement norm, contributions rise over rounds. Therefore, compared to a no punishment environment, a level enforcement norm has a positive impact on average contributions.

Now, consider the effort decision

$$\frac{\partial W_i}{\partial \pi_i^l}\frac{\partial \pi_i^l}{\partial e_i} = \frac{\partial K_i}{\partial e_i} \tag{7}$$

The effect on effort depends on i) the net value of punishment and own contribution, which is weakly less than 0, and ii) the change in C stemming from the altered contribution behavior of others. Without imposing further restrictions on the parameters of all individuals, the effect on effort is ambiguous.

Nevertheless, some predictions can be made. In the case where C would remain constant, the net return from the public goods game is negative, which reduces wealth for all levels of effort. The consequence is that marginal utility for all levels of effort increase, and so i's effort level would increase. Such a scenario is likely if i is free-riding. If the other participants are adhering relatively well to the norm, punishment may only induce minor alterations in contribution from one round to the next, while the free rider would increase her contribution by much more. Thus, if punishment is still not avoided and the increase in contribution surpasses the additional increase in the public good, the net effect would be negative and therefore lead to increased effort.

Prediction III: With a level enforcement norm, a subset of participants will increase both contributions and effort. Thus the proportion of such behaviors increase with punishment.

An important implication of this is that a level enforcement norm does not seem to induce free-riding through reduced effort. Nevertheless, a successful punishment mechanism would on average reduce effort. To see this, note that as the punishment mechanism induce higher average contributions, the returns from the public good increases which reduce the marginal return of effort.

Prediction IV: If the level enforcement norm is successful in increasing contributions over rounds, this would induce lower effort in general. Consequently, effort will on average be lower with punishment compared to without.

Note that as contribution approaches the entire endowment, individuals reach a minimum effort equilibrium. A question of interest is whether this state is welfare enhancing. If we for a moment disregard the intermediate rounds and compare only the end result of no punishment with punishment one can see that three factors determine whether punishment is welfare enhancing. i) The change in the sum total of contributions, ii) the reduction in effort and iii) the level of punishment. From previous research, a fairly typical finding is that in a socially effective end state the amount of punishment is minimal (Chauduri, 2009; Antinyah et al, 2014). In this case, it is likely that that punishment is welfare enhancing.

A Proportional Enforcement Norm

When one considers a pure proportional enforcement norm a complication from the above analysis is that effort is no longer only a matter of material utility. Under a proportional enforcement norm the choice of effort has a moral component, where increasing the effort level causes greater punishment and thus increased moral disutility. One can clearly see that i has palpable incentives to *reduce* effort,

$$\frac{\partial W_i}{\partial \pi_i^r}\frac{\partial \pi_i^r}{\partial e_i} = \frac{\partial K_i}{\partial e_i} + \frac{\partial \Phi^r}{\partial e_i} \tag{8}$$

Note that the marginal effect on payoff from increasing effort is lower than under level enforcement, since increasing effort increases punishment. Similarly, the marginal cost beyond material considerations has increased, since apart from the usual cost function, increasing effort now induced moral disutility. This leads to the prediction that effort should decrease under a proportional enforcement norm.

Prediction V: With a proportional enforcement norm, effort levels will in general decrease. This imply that effort will be lower with a proportional enforcement norm than with a level enforcement norm.

However, another way to avoid punishment is to increase contributions. Consider the decision problem facing the individual,

$$-\frac{\partial W_i}{\partial \pi_i^r}\frac{\partial \pi_i^r}{\partial c_i} = \frac{\partial M_i}{\partial c_i} - \frac{\partial \Phi^r}{\partial c_i} \tag{9}$$

This decision problem resembles that of the level contribution norm, with the key difference being that since c_i is now divided by e_i, the marginal reduction in punishment from increasing contributions is lower. This fact unambiguously reduce the incentives to increase contributions, and we predict that

Prediction VI: With a proportional enforcement norm, the increases in contribution, if any, will be lower than under a level enforcement norm.

A proportional enforcement norm results in some interesting consequences with respect to punishment. For instance, someone who exert effort above average and contributes above average may still be punished. Another consequence is that, ceteris paribus, the lower the effort the greater the incentives to reduce effort. This fact stems from the fact with lower effort, the denominator is more receptive to a change in effort.[8] This implies that low effort individuals has greater incentives to reduce effort rather than contribution, and under certain circumstances may actually benefit from reducing *both*. The implication of this line of reasoning is that free-riders have strong incentives to reduce both effort and potentially contribution to the point where reducing punishment would not render greater utility. But so the imperfectly conditional contributors through their innate preferences, and considering that Fischbacher and Gächter (2010) finds that a majority of individuals ($\approx 55\%$) are such imperfect cooperators, we conjecture that

Prediction VII: With a proportional enforcement norm, a nontrivial proportion of sessions will exhibit a tendency for both effort and contribution reduction.

Prediction VII': These two predictions combined indicate that a proportional enforcement norm will on average be less socially effective. That is, achieve a lower increase in the absolute level of contributions while also reducing effort.

This result is driven by punishment aversion, which is a disutility. Since a disutility is driving down the level of the public good, and prohibits compensation through private

[8] consider for instance someone who earns 20 and contributes 15. Her ratio is equivalent to someone who earn 8 and contributes 6. Yet a one unit reduction of effort has a much greater impact on the ratio for the latter than for the former

earnings, the punishment mechanism could be welfare destructive in. However, this is not to say that a proportional norm is necessarily ineffective. The actual outcomes are highly dependent on the real preferences for wealth, moral alignment and punishment aversion.

Endogenous choice of Enforcement Norm

As argued by Reuben and Riedl (2012), there is no a priori reason to believe that individuals manages to coordinate on a norm, nor what norm may prevail. In their study, they found that the several groups managed to achieve full efficiency, and that most converged on some proportional norm. Muehlbacher and Kirchler (2009) found that the propensity to contribute was decreasing in the level of effort and Antinyah et al (2014) found that those who exerted effort for their endowment showed a lower propensity to punish. These findings indicate that effort causes individuals to put less weight on their other-regarding preferences. These various findings suggest that when individuals are able to choose whether to enforce a level norm or a proportional norm, both are in force to various extents.

Moreover, their results also indicates that the punishment mechanism is effective. One possible way that these two norms interact is that they are adhered to by different types. If one takes into account a general tendency for a self-serving bias, a plausible outcome is that those with less than average endowments adhere to a proportional norm, which would give those above average incentive to increase their contribution and reduce their effort. On the other hand, those with above average endowments would benefit most from adhering to a level enforcement norm, which in turn would give those with lower contributions incentives to increase their contributions. The impact of such a distribution of norms would be a tendency for convergence on an effort- and contribution level that is welfare enhancing. However, the tendency to reduce other-regarding preferences indicate that the punishment mechanism could be less efficient than other studies suggest. Therefore, we conjecture that

Conjecture I: When individuals are able to choose norm, they adhere most to that which serves their status best. Therefore, those with high effort enforce a level contribution norm, while those with low effort enforce a proportional contribution norm.

Conjecture II: In a nontrivial proportion of sessions, individuals approach the socially effective outcome of full cooperation. However, there will also be a significant proportion of sessions in which the punishment mechanism will not be socially effective.

One dimension that our analysis cannot answer fully is how the enforcement norms affect punishment intensity. With results, we would have investigated this directly through a presentation of the data. However, lacking this opportunity, we instead incorporate such analysis in the form of two questions to be empirically investigated in the results section.

Question I: How will punishment behavior vary across enforcement environments? In particular, in which group is punishment most severe?

Question II: Which environment will yield the highest average payoff net of punishment?

Experimental Design

Overview

In this experiment, subjects play a repeated linear public goods game with random reshuffling of players between each round. They play for 10 rounds and prior to each round, they perform a task that determines their endowment in the subsequent round. Our experimental design is based on a 2×3 factorial design, where we vary punishment opportunity and the availability of enforcement norms. In the first enforcement environment, denoted (REL), subjects can only punish according to a proportional norm. In the second environment, denoted (ABS), subjects can only punish according to a level norm and in the last environment, denoted (FULL), they are given full information as to other player's behavior and may thus choose what norm to enforce. Each environment has both a Punishment (P) and No Punishment (NP) treatment, leaving us with 6 treatments in total. The experiment will be carried out over separate sessions with an target of a sum total 80 subjects in each punishment treatment and 40 subjects in each non-punishment treatment.[9]

The Earnings Decision

The task is the first phase in each period and the outcome determines the endowment in the subsequent round of the public goods game. We model the task after Ariely et al (2009), where subjects alternately press two keys on a computer keyboard, "X" and "Z" repeatedly for 1 minute. There is no consensus in the literature on how the task should be constructed (Cherry et al., 2005; Antinyan et al. 2014). We settled on a task where the effort is wholly physical, but only mildly strenuous, to avoid systematic biased in endowment biases towards personal attributes[10]. A crucial moment in this study is variation in the number of presses subjects achieves and a worry is the risk of low variation. Since this cannot be ruled out a

[9] See results section for further discussion on sample size.

[10] In this regard, our choice of task differs from previous literature. Antinyan et al. (2014) use a tournament approach where lowest performing subject is ruled out of the public goods game. In Cherry et al. (2005) participants are allocated to different endowment tiers based on number of correctly answered GMAT questions. In this design, such approaches are problematic. We strive to avoid confounding our results with player attrition and subject characteristics such as intelligence.

priori, we will conduct a pilot study for the task separately, from which we may calibrate the endowment conversion parameter. Nevertheless, we have strong reasons to believe a considerable variation will be present; based on the distribution of presses in Ariely et al (2009), the range of clicks for a 5-minute duration is approximately 240 - 900. Linearly interpolation suggest a range of approximately 50-200 successful presses.

The performance in this task determines the endowment in the subsequent round of the public goods game. For each completed pair, subjects earn 0.5 tokens which should generate an endowment span of 25 - 100 tokens each round. Each token is worth 0.1 SEK, and thus the expected range of endowment in real money terms is 2.5 to 10 SEK each round. Based on typical outcomes in (homogeneous) linear public goods games, this suggest an average earning of about 100-150 SEK from the experiment.[11]

The Contribution Decision

We employ a standard linear public goods game with a stranger protocol. This ensures that the evolution of contribution, effort and punishment behavior is driven by intrinsic preferences and beliefs about other participants general behavior. In each session, the participants are randomly matched into groups of four who play one round of the public goods game. To ensure a social dilemma, the MPCR is set at 0.4, which is the same as in Fehr and Gächter (2000). When punishment is available, subjects are allowed to allocate three punishment tokens to a co-player at a personal cost of one token[12]. To limit intertemporal considerations, the amount of punishment cannot exceed that period's payoff before punishment and individuals cannot earn negative payoff. In the case of punishment, the payoff for individual i is given by

$$\pi_{it} = max\Big[0,\ 0.5e_i - c_{it} + 0.4\sum_{j=1}^{4} c_{jt} - \sum_{\substack{j=1 \\ j\neq i}}^{4} p_{ijt} - \sum_{\substack{j=1 \\ j\neq i}}^{4} 3p_{jit}\Big]\ ,\ \forall t, i \tag{10}$$

$$\text{with} \quad c_{it} \ \leq\ 0.5e_{it} \quad \& \quad \sum_{\substack{j=1 \\ j\neq i}}^{4} p_{ijt} \ \leq\ 0.5e_{it} - c_{it} + 0.4\sum_{j=1}^{4} c_{jt}$$

Where p_{ijt} and p_{jit} denotes the punishment individual i allocates to individual j and vice versa. In the No-Punishment treatment they are set to zero by default. After subjects have allocated their contributions, they are shown their individual payoff and some information regarding the behavior of others. This information varies between treatment according to

[11] The range of total earnings after 10 rounds is 25 to 100 SEK. If the mean endowment is 65 SEK, average contribution level is 50% and MPCR in the public goods game is 0.4, subjects will have earned about 100-150 when having finalized the experiment.

[12] The punishment mechanism is the same as one of the mechanisms in Nikiforakis and Normann (2008), who found the 1:3 ratio efficient enough to yield increasing contribution levels over rounds.

the following table;

Environment	**Other-regarding Information**	**Variable**
REL	Co-player's contribution relative to her income from the task, i.e. contribution ratio.	γ_j
ABS	Co-player's contribution in absolute terms. That is, their level contribution in tokens.	c_j
FULL	Co-player's earnings, contribution and contribution ratio.	e_j, c_j, γ_j

Regarding the information of punishment behaviours we follow Fehr and Gächter (2000), where subjects are only shown the aggregate punishment received (and a summary of their own punishment allocation to others)[13].

Implementation

We employ a double blind protocol, where neither the experimenter nor the subjects are informed as to their treatment group nor do know of other treatment groups. This is to limit experiment expectations effects. In short, each subject arrive at the experiment location, check in and proceed to a computer terminal. Before the experiment starts, subjects receive information regarding the experiment (see Appendix A for details) outlining the task and the public goods game. Subjects are informed about the relationship between the task and the public goods game, that they will repeat each round for a unspecified number of times, between which the group in the public goods game is randomly reshuffled.

Due to availability and ease of logistics, we will perform the experiment on a student population. Although the student population may not be representative of the Swedish population at large, Croson (2007) shows that results obtained on student populations are typically replicable on other populations. Moreover, students are typically the subject pool population in behavioural economics (Croson, 2007 and Cubitt et al, 2012). We will recruit students from Stockholm University, the Royal Institute of Technology, Karolinska Institutet and Södertörn College. Care will be taken to restrict number of students with an educational background in economics, as there are indications that such students have an increased tendency to free-ride than other populations (Ames et al, 1981).

The experiment would be designed using the o-Tree program (Chen, Schonger and Wickens, 2014) and conducted at computer terminals at each institution's PC-lab. The coordination with subjects before meeting would be handled through ORSEE (see Greiner

[13]If we were to allow for more information it is likely that the punishment behaviour would change as it opens up for anti-social behavior (Nikiforakis et al, 2008 and 2013)

2004) in order to easily manage the logistics and also to depersonalize interaction with the subjects. Instructions would be displayed on-screen and read out loud prior to commencing the experiment. While the treatment designs in any one treatment is not complex, we ensure that subjects have understood the game by letting them fill out a form with 10 control questions, in some groups prior to the experiment, in other after the experiment. This procedure has been used previously by, among others, Nikiforiakis et al (2013) and Gächter and Fehr (2010), where no respondent answered incorrectly. We therefore expect subjects to fully understand the experiment at hand.

Results

Discussion

In the experimental literature, a frequent environment in which the dynamics of social dilemmas has been studied is the public goods game. Yet one of the most important features of social dilemmas has been left almost untouched - the fact that individuals (and collection of individuals) choose their private endowment in expectation of their returns from public goods. This has large and important implications for the efficacy of enforcement mechanisms such as private punishment, or central mechanism such as taxation. By only studying the contribution decision, we are left only with the frames of some the most intriguing pictures.

In this paper, we proposed an analysis of cooperative behavior when private earnings was an endogenous decision. We studied how a popular enforcement mechanism (at least among experimenters) affected behavior by altering individuals beliefs. A key finding was that a proportional enforcement norm created incentives for free-riding through lower effort and could lead to a socially undesirable outcome. We designed an experiment to empirically test the effects of various enforcement norms and how they interacted when individuals where free to choose themselves. In lieu of results, we conjectured that a form of self-serving biases could lead to a selective adoption of norms, with the surprising outcome that such self-seeking behavior could bring about a socially effective outcome.

In deriving our predictions, we inevitably needed to make abstractions, some of which will have a material bearing on the outcome. A key component of our analysis was the imperfect conditionality in contribution that a majority of individuals seem to exhibit (Fischbacher and Gächter 2010). A critical extension in the current context would be add the earnings dimension to this type of analysis. We would expect that this imperfect conditionality holds for effort as well, in which case the effect would be compounded. Equally important was the gradual updating of beliefs, which induce a "momentum of change"; if the punishment mechanism managed to facilitate an initial change in actual contribution,

belief updating would have individuals revise their beliefs of future contribution in the direction of the deviation, which in turn could cause further upward or downward revisions by its own momentum. However, the general belief updating process we considered here was a weighted average of past contributions and past beliefs of contributions. Most likely, a richer understanding of changes in behaviour could be attained by allowing a belief updating process that accounts explicitly for other mechanisms, in particular anticipated punishment. To our minds, properly understanding how individuals revise their beliefs and act upon them according to intrinsic preferences is a crucial in gaining insights into the public goods dilemma.

When people can free-ride by reducing their effort, a critical property of any enforcement mechanism is its effect on effort incentives, as a poorly designed mechanism may have large detrimental effects. A mechanism of particular relevance for policy making is taxation. Our results indicate that with progressive taxation, individuals have considerable incentives to reduce effort. A promising avenue to understand people's innate reactions to central enforcement is through experimental studies of public goods games.

Another feature of our analysis that suggest a promising line of research was the fact that the distribution of beliefs and player types exerted a strong influence on the outcome. By more sophisticated empirical methods one could uncover clues to the tipping points where cooperation may be socially effective. In this paper we have merely scratched the surface of a type of social dilemma that is pervasive in modern society and features in a large range of high-stake cooperation scenarios. By better understanding how people choose their effort in participation of returns from public goods we may understand what type of institutions create socially preferable outcomes.

References

[1] J. Andreoni. "Why free ride?. Strategies and learning in public goods experiments". In: *Journal of Public Economics* 3 (), pp. 291–304.

[2] Armenak Antinyan, Luca Corazzini, and Daniel Neururer. "Public Good Provision, Punishment and the Endowment Origin: Experimental Evidence". In: *Punishment and the Endowment Origin: Experimental Evidence (March 26, 2014)." Marco Fanno" Working Paper* 169 (2014).

[3] D. Ariely, A. Bracha, and S. Meier. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". In: *American Economic Review* 99.1 (2009), pp. 544–555.

[4] A. Chaudhuri. "Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature". In: *Experimental Economics* 14.1 (2011), pp. 47–83.

[5] Todd L Cherry, Stephan Kroll, and Jason F Shogren. "The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab". In: *Journal of Economic Behavior & Organization* 57.3 (2005), pp. 357–365.

[6] Matthias Cinyabuguma, Talbot Page, and Louis Putterman. "Can second-order punishment deter perverse punishment?" In: *Experimental Economics* 9.3 (2006), pp. 265–279.

[7] Jeremy Clark. "House money effects in public good experiments". In: *Experimental Economics* 5.3 (2002), pp. 223–231.

[8] Rachel Croson. *The use of students as participants in experimental research*. Tech. rep. Working paper, 2007.

[9] Martin Schonger Daniel L. Chen and Chris Wickens. *oTree - An Open-Source Platform for Laboratory, Online and Field Experiments*. Tech. rep. www.otree.org/oTree.pdf: Working paper, 2014.

[10] Dirk Engelmann and Nikos Nikiforakis. "In the long-run we are all dead: On the benefits of peer punishment in rich environments". In: (2013).

[11] J. Falkinger et al. "A simple mechanism for the efficient provision of public goods: Experimental evidence". In: *American Economic Review* 90.1 (2000), pp. 247–264.

[12] Ernst Fehr and Simon Gächter. "Fairness and retaliation: The economics of reciprocity". In: *The journal of economic perspectives* (2000), pp. 159–181.

[13] Urs Fischbacher. "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental economics* 10.2 (2007), pp. 171–178.

[14] Urs Fischbacher and Simon Gächter. "Social preferences, beliefs, and the dynamics of free riding in public good experiments". In: *American Economic Review* 100.1 (2010), 541—556.

[15] Urs Fischbacher, Simon Gächter, and Ernst Fehr. "Are people conditionally cooperative? Evidence from a public goods experiment". In: *Economics Letters* 71.3 (2001), pp. 397–404.

[16] Simon Gächter, Elke Renner, and Martin Sefton. "The long-run benefits of punishment". In: *Science* 322.5907 (2008), pp. 1510–1510.

[17] Philip Graves. "A Note on the Design of Experiments Involving Public Goods". In: *CESifo Working Paper, volume=, number=3187, pages=, year=2010, publisher=* ().

[18] Ben Greiner. "The online recruitment system orsee 2.0-a guide for the organization of experiments in economics". In: *University of Cologne, Working paper series in economics* 10.23 (2004), pp. 63–104.

[19] Glenn W Harrison. "House money effects in public good experiments: Comment". In: *Experimental Economics* 10.4 (2007), pp. 429–437.

[20] S.D. Levitt and J.A. List. "What do laboratory experiments measuring social preferences reveal about the real world?" In: *Journal of Economic Perspectives* 21.2 (2007), pp. 153–174.

[21] Gerald Marwell and Ruth E Ames. "Economists free ride, does anyone else?: Experiments on the provision of public goods, IV". In: *Journal of public economics* 15.3 (1981), pp. 295–310.

[22] Stephan Muehlbacher and Erich Kirchler. "Origin of endowments in public good games: The impact of effort on contributions." In: *Journal of Neuroscience, Psychology, and Economics* 2.1 (2009), p. 59.

[23] Nikos Nikiforakis. "Feedback, punishment and cooperation in public good experiments". In: *Games and Economic Behavior* 68.2 (2010), pp. 689–702.

[24] Nikos Nikiforakis. "Punishment and counter-punishment in public good games: Can we really govern ourselves?" In: *Journal of Public Economics* 92.1 (2008), pp. 91–112.

[25] Nikos Nikiforakis and Hans-Theo Normann. "A comparative statics analysis of punishment in public-good experiments". In: *Experimental Economics* 11.4 (2008), pp. 358–369.

[26] M. Poppe. "The specificity of social dilemma situations". In: *Journal of Economic Psychology* 26.3 SPEC. ISS. (2005), pp. 431–441.

[27] Ernesto Reuben and Arno Riedl. "Enforcement of contribution norms in public good games with heterogeneous populations". In: *Games and Economic Behavior* 77.1 (2013), pp. 122–137.

[28] Stefan Traub. "Bardsley, N., R. Cubitt, G. Loomes, P. Moffat, C. Starmer, and R. Sugden (eds): Experimental economics—rethinking the rules". In: *Journal of Evolutionary Economics* 22.2 (2012), pp. 389–393.

Appendix A