

Growth and Inequality in Public Good Games ^{*}

Simon Gaechter[†]

University of Nottingham,
CESifo and IZA

Friederike Mengel^{‡,§}

University of Essex
and Maastricht University

Elias Tsakas[¶]

Maastricht University

Alexander Vostroknutov^{||}

Maastricht University
and European University at St.Petersburg

July 25, 2014

Abstract

In a novel experimental design we study public good games with dynamic interdependencies. More precisely, each agent's income at the end of a period serves as her endowment in the following period. In this setting growth and inequality arise endogenously allowing us to address new questions regarding their interplay and effect on cooperation levels. We find that, in stark contrast to standard public good experiments, contributions are increasing over time even in the absence of punishment possibilities. In both treatments (with and w/o punishment) inequality and group income are positively correlated for poor groups (with below median income), but negatively correlated for rich groups (with above median income). There is very strong path dependence: inequality in early periods is strongly negatively correlated with group income in later periods. These results give new insights into why people cooperate and should make us rethink previous results from the literature on repeated public good games regarding the decay of cooperation in the absence of punishment.

^{*}We thank Ingvild Almas, Martin Duwfenberg, Tore Ellingsen, Drew Fudenberg, Santiago Oliveros and Marie-Claire Villeval as well as participants in seminars and conferences at Essex, Maastricht and Lund (GIRLS 2014) for helpful comments. Friederike Mengel thanks the European Union (grant PERG08-GA-2010-277026) for financial support. Simon Gaechter gratefully acknowledges support from the European Research Council grant ERC-AdG 295707 COOPERATION.

[†]School of Economics, University of Nottingham, University Park Campus, Nottingham NG7 2RD, United Kingdom.

[‡]Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, e-mail: fr.mengel@gmail.com

[§]Department of Economics (AE1), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

[¶]Department of Economics (AE1), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

^{||}Department of Economics (AE1), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

1 Introduction

Across disciplines, public good games have been used as the standard model to study a wide array of social dilemma situations. Those include joint ventures (Grossman and Shapiro, 1986), R&D cooperation (Cozzi, 1999; Kamien et al., 1992), political action funds of special interest groups or parties (Dawes et al., 1986), multilateral foreign aid or effort provision in work teams (Ostrom, 1990; Hamilton et al., 2003; Tirole, 1986). But also pricing or market sharing agreements by firms (Green and Porter, 1984) as well as many economic activities in the family (Becker, 1981) can be thought of as instances of cooperation in public good games. One feature that is common to many of these examples is that there are dynamic interdependencies: not only will the same set of people meet again, but previous outcomes affect future endowments (both in terms of the stock of physical and social capital).

In this paper we present a novel experimental design which enables us to study such dynamic interdependencies in public good provision. We focus on two key aspects. First, if contributions today are high, then there will be higher wealth in the next period (*growth*). Second, heterogeneity in contributions today creates *inequality* in endowments in the following period. The standard design avoids these intertemporal effects by giving each participant a new endowment in each period.¹ A typical result for these studies is that while some people do contribute positive amounts to the public good, contributions rapidly decline to zero as the game is repeated (see e.g. Isaac et al. (1984); Andreoni (1995) or Fischbacher and Gaechter (2010) among many others). It has also been shown that introducing the possibility of punishment stabilizes contributions (Ostrom et al., 1992; Fehr and Gaechter, 2000; Andreoni et al., 2003).

One question arises immediately: is the threat of punishment needed to stabilize contributions if there are dynamic interdependencies or is the possibility of *growth* sufficient to sustain cooperation? Theoretically, the unique Nash equilibrium of our public good game will prescribe zero contributions in each period just as in the standard setting. However, if participants do cooperate, exponential wealth increase can be obtained. Focusing on growth and inequality also inspires new questions. If, for example, one member of an R&D collaboration fails to put in sufficient resources or if one member of a drug cartel does not respect market sharing agreements, then typically the shirking group member will be richer and more powerful in the following period. This leads to a second important question: how does such endogenous *inequality* affect contributions and hence growth?²

Growth and inequality can also interact with the possibility of punishment in different ways. The threat of punishment can lead to higher growth if it induces higher contributions as in the standard design. On the other hand, punishment that is executed on the outcome path can induce a multiplier effect which can hamper growth severely. Maybe more interestingly, punishment can interact with inequality in non-trivial ways. For example, the lowest contributor in a group will typically become the richest group member and be largely “immune” to punishment by her much poorer group members. This happens because any punishment that poor group members can afford is small relative to the richest group member’s wealth. Rich group members, on the other hand, might be able to punish others harshly at a relatively low cost to themselves. This endogenously created *asymmetry of punishment* possibilities translates income inequality into inequality in power: something that can often be observed in real-life social dilemma situations.³

¹By standard design we refer to the many studies of linear voluntary contribution experiments, where the one shot game is repeated for some number of periods. See Ledyard (1995) for a survey of the earlier literature and Chaudhury (2011) or Chapter 6 in Plott and Smith (2009) for a survey of more recent results.

²The relationship between inequality and growth is an open question also in the macroeconomic literature and in development economics (see e.g. Barro (2000); Forbes (2000); or Persson and Tabellini (1991) among many others). In 2012, for example, Afghanistan, Belarus and Germany had about the same Gini coefficient, but very different wealth levels (according to <http://data.worldbank.org/indicator/SI.POV.Gini/>).

³History books are full of examples where “shirking group members” have assumed power by exploiting asym-

In our experiment participants are matched in fixed groups of four people to play the public good game. As in most other public good experiments we focus on the most challenging social dilemma situations, where the unique Nash equilibrium prescribes zero contributions by all group members, but where efficiency would require group members to contribute their entire endowment. In our baseline treatments participants interact for ten periods. Each participant's income at the end of a period constitutes their endowment for the next period. We also conduct punishment treatments, where in each period the public good game is followed by a punishment stage with the standard 3:1 technology, where participants can deduct three tokens from others at a cost of one token to themselves. Participants can use their entire wealth to punish others but are not allowed to set others' wealth to below zero.

Our main findings can be summarized as follows. Even in the absence of punishment, contribution levels are reasonably high (between 20 – 60% of endowments) and, unlike in the standard setting, there is no decay of cooperation. In terms of the realized potential for growth and the level of inequality, there is a lot of variation across groups. Individual earnings range between 2 Euros and 241 Euros.

In stark contrast to the standard setting, (the possibility of) punishment leads on average to a *reduction* in wealth. This is despite the fact that people tend to contribute more in the treatment with punishment. The reason for the lower wealth levels with punishment lies in the use of reciprocal and anti-social punishment. Especially in poor groups (with below median wealth) we find substantial evidence of anti-social punishment, where shirkers use their higher wealth and consequently higher power to punish as a means to harshly sanction poorer group members at relatively low cost to themselves (relative to their wealth). The high levels of reciprocal and anti-social punishment are what keeps wealth levels low in some groups and in some cases even leads to a total destruction of all wealth. Additional treatments, where we artificially eliminate endogenous inequality and hence asymmetric punishment possibilities (but keep growth possibilities) demonstrate that it is indeed the endogenous asymmetry in punishment possibilities in this setting that undermines the effectiveness of punishment. The fact that punishment leads to lower wealth both for the individual who punishes and for the group as a whole raises questions about the evolutionary origins of punishment.⁴

While the possibility of punishment leads to a reduction in wealth, it also reduces inequality on average. This is despite the inequality-increasing presence of anti-social punishment. In terms of the relationship between inequality and growth we find the following results. There is strong path dependence. Inequality in period 2 is strongly negatively correlated with wealth in period 10. In particular, a 1% increase in inequality will lead to a 0.2 – 0.5% decrease in wealth depending on the treatment. Inequality and growth are positively related across groups with below median wealth and negatively related across groups with above median wealth.

To our knowledge our experiment is the first to study public good games with dynamic interdependencies and the possibility of asymmetric punishment.⁵ Some authors have studied

metric punishment opportunities. John Rockefeller reportedly became rich by buying up all barrel production equipment to prevent others from producing oil. Later Standard Oil literally sent out thugs to raid the premises of competitors as a form of punishment (Josephson, 1962). In an interesting study of upper-level drug-dealing Adler (1985) discusses endogenously arising asymmetries in punishment possibilities. Johnson and Earle (1987) discuss historical examples of cooperation and punishment among North-American Indians.

⁴Recently Dreber et al. (2008) made a similar point arguing that punishment is maladaptive in repeated cooperation games with a short horizon. Gaechter et al. (2008) have shown, however, that punishment in the standard setting is strictly beneficial for the group with a longer time horizon. In our setting, by contrast, in many groups with punishment (almost) all wealth is destroyed after 10-15 periods. Those groups will be unable to recover even if the horizon is very long.

⁵In an unpublished MA thesis Huck (2006) has conducted a dynamic public good game without the possibility of punishment. He also didn't analyze growth and inequality but, like us, he finds a pattern of increasing contributions (in terms of absolute amounts) and no endgame effect.

settings where there are dynamic interdependencies to public good provision in a different sense. Duffy et al. (2007) studied threshold public good games with multiple contribution rounds, where, theoretically, “completion equilibria” (with positive contributions) do exist. They find that, as in the standard setting, contributions do decline over time (see also Croson and Marks (1998) among others). Several authors link public good games over time via explicit reputation mechanisms (e.g. Milinski et al., 2002). None of these studies though focuses on the core dynamic interdependency, which is that the level of cooperation and hence the share of the efficient allocation realized at any point in time affects endowments in subsequent periods. Guererik et al. (2013) have recently studied a setting where past wealth affects future endowments. They use a non-linear exchange rate, though, which effectively eliminates the possibility of exponential growth and contains inequality. Consequently their results are much more similar to those obtained in the standard setting.

Other authors have studied the effects of *exogenous* income inequality. In the model by Bergstrom et al. (1986), where positive contributions can be part of an equilibrium, public good provision increases with inequality in the income distribution. This prediction has been experimentally tested by Chan et al. (1996) who find that group behaviour conforms with the theoretical prediction. Other authors have found that exogenous income inequality decreases contributions (van Dijk et al., 2002; Ostrom et al., 1994). Chan et al. (1999) found no effect of income inequality on contributions. Reuben and Riedl (2013) also find that without punishment there is no effect of income inequality on contributions, while with punishment participants contribute proportionally to their endowments, leading to increased contributions. All these papers deal with one-shot games and *exogenously* imposed inequality.

Our paper is also related to the entire literature on the standard setting of repeated public good games surveyed in Chaudhury (2011). The effectiveness of punishment to increase contributions within this setting has been demonstrated by Ostrom et al. (1992), Fehr and Gaechter (2000) or Andreoni et al. (2003) among others. In contrast to these findings, contributions in our setting do not decay over time even without the possibility of punishment. Several authors have found that punishment can lead to lower payoffs to the individual who punishes and sometimes also to the group if the horizon of the game is short (Fehr and Gaechter, 2002; Herrmann et al., 2008; Egas and Riedl, 2008; Dreber et al., 2008). Under a long horizon, though, punishment in the standard setting has been found to be strictly beneficial (Gaechter et al., 2008). In our setting punishment is detrimental both under a 10-period and 15-period horizon. We do not study longer horizons, but since many groups in the treatment with punishment destroy (almost) all wealth by period 10, these groups will be unable to recover even with a very long horizon.

This paper is organized as follows. In section 2 we present the experimental design. Section 3 discusses the theoretical predictions and summarizes our research questions. Section 4 contains our main results. In Section 5 we discuss the mechanisms underlying these results. Section 6 concludes. An Online Appendix contains screenshots (Online Appendix A), experimental instructions and the questionnaire (Online Appendix B), proofs of the theoretical predictions (Online Appendix C) and several additional results, tables and figures (Online Appendix D-H).

2 Experimental Design

In our experiment participants play a sequence of dynamically interdependent public good games in two basic treatments: (1) Treatment NOPUNISH in which participants only play the public good game; (2) Treatment PUNISH where, after each period, participants can also subtract tokens from other members of their group at a cost. We now describe these treatments in turn.

Treatment NOPUNISH At the beginning of the experiment participants are randomly matched into groups of 4, which stay the same throughout the experiment. Before period 1, each participant is given 20 tokens as initial endowment. Participants then, in each period, can divide their tokens into two accounts: a *private* account and a *group* account. The private account is for their personal use only and other participants cannot influence or benefit from the tokens in the private account. The private account has a return of 1. The group account is different: all 4 participants in the group can place their tokens into the group account. The group account has a return of 1.5 and after the return is calculated, the tokens put into the group account are equally divided among all 4 participants. Therefore, if participant i has N_i^t tokens *before* period t and she places a_i^t tokens into the group account,⁶ then, at the *end* of period t , she will have $N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1..4} a_j^t$ tokens. In the NOPUNISH treatment, this amount also corresponds to the number of tokens *before* period $t + 1$ (the endowment at $t + 1$):

$$N_i^{t+1} = N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1..4} a_j^t.$$

Therefore, in period $t + 1$, the number of tokens that each participant can invest depends on the choices of all group members in period t . This is what makes our set-up different from the standard setting where the amount of tokens before each period is fixed and the earnings from the previous periods cannot be used for investment into the public good. Before the next period starts all participants observe the screen that summarizes the information from the current period (Figure A.1 in Online Appendix A). In particular, each participant observes the amount placed into the group account by each individual member of her group. The identities of Other1, Other2 and Other3 (see Figure A.1 in Online Appendix A) stay fixed throughout the experiment.

Treatment PUNISH The PUNISH treatment is the same as the NOPUNISH treatment with the difference that, after all participants in the group have placed their tokens into the private and group accounts, they have a possibility to subtract tokens from other members of the group. Participants observe the same information as in the NOPUNISH treatment (see Figure A.2 in Online Appendix A). Participants can then subtract tokens from each other group member individually. The cost of subtraction is $\frac{1}{3}$ of the tokens subtracted. No participant can lose more tokens than she has and no one can subtract more than 3 times their own number of tokens (i.e. spend more than they have on punishment).⁷

Let p_i^j denote the cost of punishment that participant i incurs after punishing participant j by $3p_i^j$ tokens. The maximal loss that participant i can afford is $\sum_{j \neq i} p_i^j \leq W_i^t := N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1..4} a_j^t$. The amount of tokens that participant i has before period $t + 1$ is then

$$N_i^{t+1} = \max\{0, W_i^t - \sum_{j \neq i} (p_i^j + 3p_j^i)\}.$$

After the punishment phase each participant observes the following information before the next period starts (see Figure A.3 in Online Appendix A): the amount of tokens that each other member of the group has subtracted from herself; the total amount of tokens subtracted from each member of the group; amount of tokens each group member has at the end of the period, which serves as the endowment in the next period.

⁶The rest of the tokens automatically go to the private account.

⁷If participants' punishment plans are infeasible, e.g. because in the aggregate they would set a player's wealth below zero, then only player 1 and 2's plans are executed if feasible (as labeled in the software). If those are also infeasible then only player 1's plan is executed.

Additional Treatments As we outlined above the dynamic interdependencies in this setting create two types of effects: (i) they create the possibility for endogenous growth and (ii) they create endogenous inequality and hence asymmetries in the power to punish others. To isolate the impact of these two forces we ran additional treatments, where we artificially eliminate growth (treatments NOPUNISH-NOGROWTH and PUNISH-NOGROWTH) but keep endogenous inequality or where we artificially eliminate inequality but keep growth (treatments NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY). We will discuss these treatments in detail in Section 5.

Length variations In addition to the punishment variation, we added a second treatment variation which changes the number of repetitions of the public good game. We conducted sessions with 10 repetitions and sessions with 15 repetitions to understand how a longer horizon might affect behaviour and hence get some insights into how participants might perceive this environment strategically.⁸ Due to the (potentially) exponential increase of earnings over time in our setting, we limited the long horizon to 15 periods. For most of the paper we will focus on the first 10 periods in both the 10 and 15 period sessions, but we will look at how different these sessions are in Section 5. All details of all treatments were known to all participants.⁹ Table 1 summarizes the treatment structure and the number of independent observations, participants and sessions for each treatment.

	15 periods	10 periods	Overall
W/o Punishment (NOPUNISH)	15 (60,2)	23 (92,3)	38 (152,5)
With Punishment (PUNISH)	15 (60,2)	21 (84,3)	36 (144,5)
No Inequality w/o punish (NOPUNISH-NOINEQUALITY)	-	24 (96,3)	24 (96,3)
No Inequality with punish (PUNISH-NOINEQUALITY)	-	14 (56,3)	14 (56,3)
No Growth w/o punish (NOPUNISH-NOGROWTH)	-	29 (116,4)	29 (116,4)
No Growth with punish (PUNISH-NOGROWTH)	-	23 (92,3)	23 (92,3)

Table 1: Number of Independent Observations (Participants, Sessions).

Participants answered a sequence of questions at the end of the experiment (see Online Appendix B). We report summary statistics on our participants' characteristics in Tables 10 and 11 in Online Appendix F, where also further analysis of the questionnaire data can be found. In all treatments the amount that participants received at the end of the experiment was equal to €2 show up fee plus the amount of tokens after the last period converted into Euros. 1 token was equal to 0.05 Euros. All experiments were run at the BEE-Lab at Maastricht University in September - November 2012 (treatments PUNISH and NOPUNISH). The additional treatments were run in May-June 2014. Average earnings were 11.94 Euros with a minimum of 2 Euros and a maximum of 241.65 Euros. No other sessions apart from those reported here were conducted and there were no pilot studies. Instructions for the two treatments can be found in Online Appendix B.

3 Theoretical Background and Research Questions

In order to show that our setting is no different in terms of basic incentives from the standard setting, we briefly discuss theoretical predictions. We focus on the standard solution concepts for

⁸Gaechter et al. (2008) have found that the length of the horizon can affect contributions in the standard setting, particularly in treatments with punishment. The length variations allows us to see whether this is true in our setting as well.

⁹There was one session that was intended to be 15 periods, but where a computer crashed in period 11. We added this session as a 10 period session to the data, but none of the results are affected if we exclude this session.

dynamic games with observable actions, i.e., Nash equilibrium (NE) and subgame perfect equilibrium (SPE) under the standard textbook assumptions of self-interest and common knowledge of rationality. First, observe that the games that describe our treatments are not repeated games, as the set of available actions at each subgame depends on the moves that have been already realized. Therefore, unlike in the standard setting, we cannot directly apply the wide range of well-known results on finitely repeated games. Nevertheless, it turns out that both solution concepts lead to very similar predictions as in the standard setting. On the equilibrium path all players contribute 0 in all periods, and moreover in the punishment treatment no player ever punishes. As is the case with the standard setting, positive contributions can be sustained under different assumptions. Particularly reputational models, that invoke the existence of behavioural types, such as e.g. Kreps et al. (1982) or Steiner (2007) would, as in the standard setting, induce positive contributions. The purpose of stating the following result is not to argue that it will be a good prediction of behaviour, but rather to demonstrate why basic incentives are no different here from the standard setting.

In this section we will limit ourselves to state the result informally and give an intuition. A formal statement of the result as well as proofs can be found in Online Appendix C. Our first result shows that every Nash equilibrium is such that all players contribute 0 in every period in both our games. Moreover, in our public good with punishment all players punish 0 at all histories on the equilibrium path.

Proposition 1 *Consider the public good game with growth as defined above.*

- (i) *In any Nash equilibrium in the public good game without punishment all players contribute zero on the equilibrium path.*
- (ii) *In any Nash equilibrium in the public good game with punishment all players contribute zero and punish zero on the equilibrium path.*

This proposition immediately implies that contributions will be zero in any subgame perfect equilibrium as well. Furthermore, it is rather trivial to show that the subgame perfect equilibrium is unique for both our games, implying that from a game-theoretic point of view these predictions are rather strong. We refer to a formal statement of this proposition as well as proofs to Section C in the Online Appendix. The intuition of the result is as follows. No matter how punishing the strategies of opponents are, there is no way to induce positive contributions in the last period of play in a Nash equilibrium. People will always have an incentive to deviate and contribute nothing in the last period. Given that any NE strategy will prescribe zero contributions in the last period irrespective of the previous history of play, the same argument can be made for the one before last period etc. Hence, in terms of basic incentives our setting is no different from the standard setting.

Of course we know from previous public good experiments that this prediction is often not an accurate prediction for behaviour in the lab. Typical results obtained in the previous literature are (i) that contributions are initially positive, but then decline over time eventually reaching zero (Isaac et al. (1984); Andreoni (1995); Fischbacher and Gächter (2010)) and that (ii) the threat of punishment is needed and effective in sustaining contributions over the horizon of the game (Ostrom et al. (1992); Fehr and Gächter (2000); Andreoni et al. (2003)). Hence, based on the existing literature, we can formulate the following research questions.

Q1a Do contributions converge to zero over time?

Q1b Is (the threat of) punishment needed to sustain positive contributions?

New questions we want to address with our study are questions regarding the effects of growth and inequality that previous studies have been unable to address. The first question asks whether (the possibility of) punishment leads to higher growth. Based on the evidence in previous public good experiments one might conjecture that this is the case, especially if punishment is needed to sustain positive contributions. On the other hand, because punishment does destroy resources which could otherwise be used productively in the following period, there might also be a negative effect of punishment on growth.

A similar question arises with respect to inequality. On the one hand one may conjecture that the possibility of punishment reduces inequality, because free-riders might be deterred via the threat of punishment leading to more equal outcomes. On the other hand, the possibility of punishment can also lead to more inequality. The reason is that if someone manages to free-ride successfully, i.e. without being punished, that person will have higher wealth in future periods and hence more possibilities to punish others at a relatively low cost to himself. Hence, the possibility of punishment could strengthen existing inequalities because it makes shirking individuals more powerful. This in turn could undermine the effectiveness of punishment.

Finally, one can ask about the relation between inequality and growth. Is it the “rich” or the “poor” groups that will be more unequal? How does initial inequality affect growth and hence final wealth? While the former question refers to the relation of growth and inequality across groups, the latter is a question about path dependency, i.e. how do initial differences in inequality translate into long run differences in growth. Both types of question can be addressed with our design. The previous discussion is summarized in the following set of research questions.

Q2a How do growth and inequality affect the effectiveness of punishment?

Q2b What is the relation between inequality and growth?

In the following section we will address these questions in turn. We start with Question 1 (Section 4.1) and then move to Question 2 in Sections 4.2 and 4.3.

4 Results

This section comprises our main results. For the remainder of this section we will aggregate our 15-period and 10-period variations and consider the first 10 periods in the former. We will test whether the two variations differ in the first 10 periods in Section 5. This will also give us some insights into how participants react strategically to this environment with growth and inequality.

4.1 A first look at contributions

We start by comparing behaviour in our experiment against two insights from the standard repeated public good game: the decay of cooperation over time and the necessity, but also effectiveness of punishment in sustaining positive contributions (Q1a and Q1b).

Figure 1 shows the average amount of tokens participants contributed over time. Contributions are clearly non-zero and are increasing over time in both treatments. Without punishment participants contribute about 10 tokens (50% of their endowment) in the first period and then steadily increase this amount over time. This is in stark contrast with the standard setting, where contributions without punishment decline to zero over time.

Note that, in our context, increasing contributions over time imply that participants will have increasing endowments over time. One may ask how contributions evolve over time as a share of endowment. This question does not arise in the standard setting, where, since the endowment is constant, the evolution of contributions over time is the same irrespective of whether it is

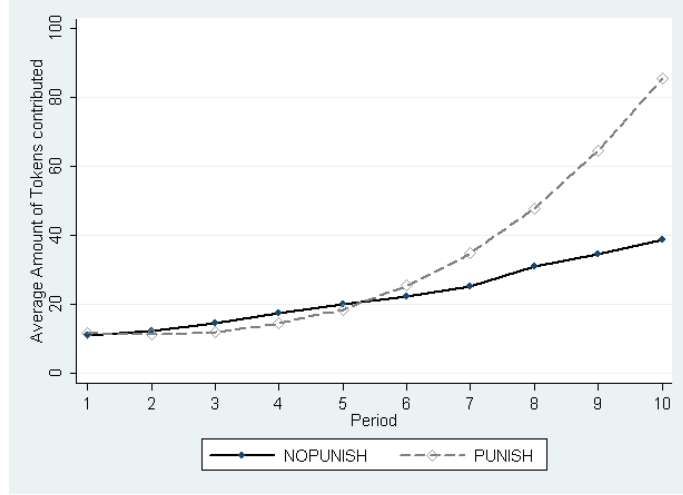


Figure 1: The average amount of tokens contributed over time in treatments NOPUNISH (Without Punishment) and PUNISH (With Punishment).

measured in absolute terms or as a share of endowment. This question raises two issues (i) what is the right way to measure shares? (should e.g. the same weight be given to someone contributing 1 out of 1 tokens as to someone contributing 400 out of 700?) and (ii) what is the quantity we should be interested in? We discuss these questions in detail in Online Appendix D. Figure D.1 in Online Appendix D plots the share of overall income contributed over time for both treatments. It shows that contributions in treatment NOPUNISH measured in this way are first slowly decreasing and then (starting from period 6) constant at a level of about 38% of income.

What happens when we introduce the possibility of punishment? As in the standard setting, contributions seem to be higher in treatment PUNISH compared to NOPUNISH (Figure 1). However, here they start to differ only starting from period 7 and not from period 1 on as is the case in the standard setting. We ran a random effects OLS regression where we regress the amount of tokens contributed on period, a treatment dummy and an interaction term (with standard errors clustered by matching group). The regression confirms strictly increasing contributions in treatment NOPUNISH ($p = 0.056$) as well as in PUNISH ($p = 0.044$) and contributions which increase more strongly in PUNISH compared to NOPUNISH over time ($p = 0.103$). Upon omitting the marginally insignificant interaction term (period*PUNISH) we find no difference in *average* contributions between PUNISH and NOPUNISH ($p = 0.306$) (see also Table 15 in Section 5). Given the only marginal insignificance of the interaction term (period*PUNISH) we do not want to rule out, though, that contributions are higher in PUNISH compared to NOPUNISH in later periods, which is suggested by Figure 1. In fact, as a share of endowment contributions are higher in treatment PUNISH (at a level of about 60% of endowments), which is largely due to the fact, however, that many groups in treatment PUNISH have very low group income.

To sum up, the possibility of growth seems to discipline contributions and the threat of punishment only has a small effect in our setting. Unlike in standard repeated public good experiments, there is no decay of cooperation over time even in the treatment without punishment.

Result 1: Contributions over time *Contributions are positive and non-decreasing over time even in the absence of punishment possibilities.*

4.2 Growth and Inequality across treatments

In this subsection we compare the levels of growth and inequality across our two treatments.

Growth We start by analyzing growth. To measure growth we define a variable “group income” which sums the endowments of all participants in a given group at the beginning of each period. In period 1 group income will be 80 in all groups by construction. The maximal group income that can be reached in period 10 (if everyone contributes their entire endowment in each period) is approximately 3075 tokens or €153.¹⁰ The minimal group income that can be reached (if no one ever contributes anything) is 80 in treatment NOPUNISH and 0 in treatment PUNISH (if people punish until all endowment is gone). Table 2 shows some summary statistics regarding group income.

	Group Income		Gini coefficient	
	NOPUNISH	PUNISH	NOPUNISH	PUNISH
Mean	239.92	166.11	0.14	0.13
Median	184	80	0.12	0.05
Std.Dev.	224.58	308.20	0.11	0.21
Observations	380	360	380	360
$t = 10$ Mean	470.31	407.02	0.19	0.15
$t = 10$ Median	304	129	0.21	0.03
$t = 10$ Std.Dev.	438.02	633.57	0.12	0.25
$t = 10$ Max	2060	2074	0.46	1
$t = 10$ Min	131	0	0.00	0
Observations	38	36	38	36

Table 2: Summary Statistics on the variables **group income** and **Gini coefficient** .

Irrespective of the treatment, groups do achieve growth on average. Period 10 group income is substantially above 80 on average. The large standard deviation indicates, though, that by far not all groups achieve growth. There are several groups where group income does not rise (substantially) above 80 ever. In fact in some groups in treatment PUNISH group income even reaches zero as participants destroy all wealth via punishment and counter-punishment. Figures E.1-E.2 in Online Appendix E illustrate group income (and Gini coefficients) over time in different matching groups. Figure D.2 in Online Appendix D shows the distribution of period 10 individual incomes in the two treatments. Both the average and median group income are higher in treatment NOPUNISH compared to treatment PUNISH. These results indicate that the threat of punishment is not needed to sustain high contributions in public good games with growth. On the contrary, the existence of a punishment possibility seems to hamper growth on average. Note also that all standard deviations are much higher in the treatments with punishment. Hence across group variation—the difference in group income between the most and least successful groups—is highest in that treatment. Figure 2(a) shows how average group income evolves across time in both treatments. It can be seen that income is steadily increasing over time in both treatments on a remarkably smooth path. Table 3 shows that treatment differences in time-trends are not significant (column (1)), but on average group income is lower in NOPUNISH compared to PUNISH (column (2)).

Inequality. As measure of inequality we use the Gini coefficient as defined in Deaton (1997). The smallest possible value the Gini coefficient takes is zero (if all four group members own one fourth of the group income) and the largest possible value it takes is one (if one group member appropriates the entire group income). Table 2 shows some summary statistics regarding the Gini coefficient. Overall, values of the Gini coefficient seem relatively “low” at least in comparison to the Gini coefficients of countries.¹¹ In terms of treatment comparisons, the means are similar

¹⁰After period 15 the maximum is approximately 35000 tokens or €1751.

¹¹In 2011 e.g. Norway, Japan and the Czech Republic have some of the lowest Gini coefficients with values of 0.24. The highest Gini coefficients are found in Namibia, South Africa and the Seychelles with 0.63-0.65. Also

	group income		Gini	
	(1)	(2)	(3)	(4)
period	42.62*** (7.20)	39.07*** (6.46)	0.014*** (0.002)	0.008*** (0.002)
PUNISH	-33.69 (33.73)	-73.80** (38.84)	0.060*** (0.021)	-0.012 (0.024)
period X PUNISH	-7.29 (13.04)		-0.013*** (0.005)	
constant	5.51 (19.93)	25.02 (29.28)	0.066*** (0.006)	0.102 (0.014)
Observations	2960	2960	2960	2960
Groups	74	74	74	74
ρ	0.3980	0.3975	0.3316	0.3260
VCE robust S.E.	Yes	Yes	Yes	Yes

Table 3: Random Effects OLS regression of group income (columns (1) and (2)) or Gini coefficient (columns (3) and (4)) on period and treatment dummy. Significance at the 1,5,10 percent level is denoted by ***, **, *, respectively.

across treatments, but the median Gini coefficient tends to be much higher in treatment NOPUNISH. Both a two-sided ranksum test on all data ($p < 0.0001$) as well as regressions in Table 3 support this impression. In column (3) of Table 3 the treatment dummy is positive, but has a negative interaction with period, showing that after around period 4 Gini coefficients tend to be higher in treatment NOPUNISH. Figure 2(b) illustrates these results. The figure also suggests that the Gini coefficient in treatment PUNISH follows a cyclical pattern. This phenomenon is driven by groups with lower group income, as we discuss below (see also Figure 3(d)).

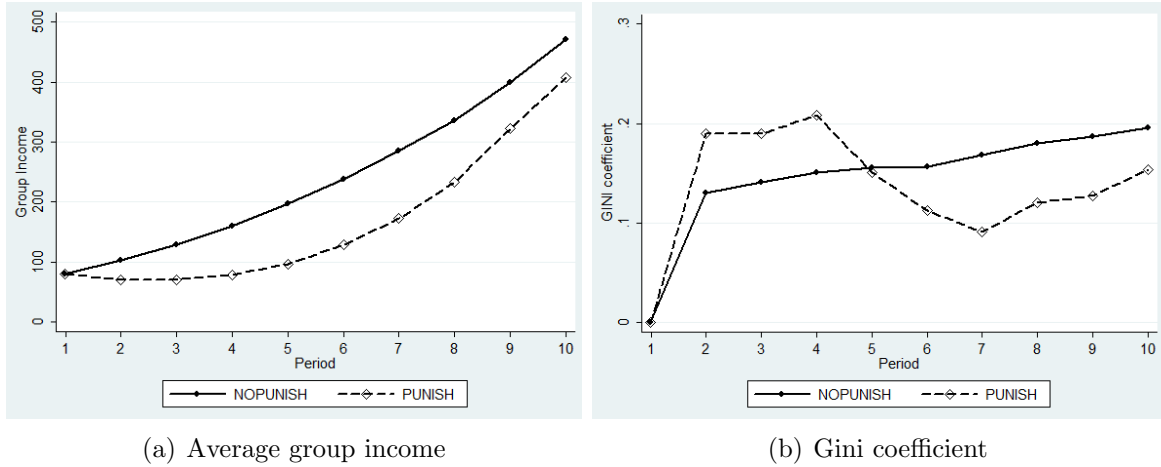


Figure 2: Average group income and Gini coefficient over time in the two treatments.

Hence, unlike in the standard setting, it seems that in the public good game with (endogenous) growth and inequality, the threat of punishment is not needed to sustain high contributions and steady wealth accumulation. Dynamic interdependencies alone seem to be more effective than the possibility of punishment in securing high wealth. Introducing the possibility of punishment even leads to lower group income on average. Below we will see that this is because there are substantial amounts of punishment *on* the outcome path. On the other hand, we have also seen that the possibility of punishment tends to help contain long run inequality.

some quite poor countries have low Gini coefficients, such as Afghanistan or Belarus with 0.27, which is lower than e.g. Germany's coefficient of 0.28. Data from Worldbank Development Research Group (<http://data.worldbank.org/indicator/SI.POV.Gini/>).

	group income	group income	Gini	Gini
PUNISH	-22.54 (59.80)	-111.55*** (21.29)	-0.039** (0.017)	0.010 (0.039)
constant	317.05*** (30.72)	162.80*** (5.92)	0.122*** (0.012)	0.170*** (0.020)
Observations	1440	1520	1440	1520
Groups	36	38	36	38
ρ	0.173	0.386	0.231	0.292
group income	above median	below median	above median	below median
VCE robust S.E.	Yes	Yes	Yes	Yes

Table 4: Random Effects OLS regression of group income (first two columns) and Gini coefficient (last two columns) on treatment dummy. Data from periods 1-10. Sample split by median group income's. Significance at the 1,5,10 percent level is denoted by ***, **, *, respectively.

Splitting the Sample by Median Group Income. We have already seen before that across-group variation tends to be quite high. Hence it seems reasonable to split our sample by the median group income and compare “successful” (above median group income) and “unsuccessful” (below median group income) groups. We split the sample by group income in period 10, which is 304 without punishment and 129 with punishment (see Table 2).¹²

Table 4 shows the results of a regression of group income and the Gini coefficient on a constant and the treatment dummy for treatment PUNISH. The difference to the regressions in Table 3 is that the sample is split by median group income as explained above. Let us first focus on groups with a below median group income. In these groups group income is higher in treatment NOPUNISH compared to PUNISH (second column in Table 4). There is no difference in average inequality between treatments in poor groups (last column in Table 4). In rich (above median) groups, there is no significant difference between the two treatments in terms of group income. However, in treatment PUNISH, inequality is significantly lower for those groups on average.

The dynamics of group income and inequality are very similar for above median groups, as Figure 3 illustrates.¹³ For below median groups the dynamics of group income are similar as well (Figure 3(b)). Figure 3(d) shows, however, that the dynamics of inequality over time is quite different across treatments in below median groups. In treatment NOPUNISH inequality is steadily increasing over time at a slow rate, while in treatment PUNISH the Gini coefficient seems to follow a different pattern. After increasing initially, it decreases sharply in periods 5, 6 and 7 and then starts to increase again. This pattern reflects cycles of reciprocal punishment. Depending on who punishes (shirkers or high contributors) inequality increases or decreases. Groups that follow an extreme pattern of punishment and reciprocal counter-punishment end up with both zero group income and zero inequality. Once those groups have reached zero inequality the average Gini coefficient follows a similar path as in the NOPUNISH treatment, albeit at a lower level of inequality.

Hence, if groups manage to be successful with punishment, then they achieve equally high group income at a lower cost in terms of inequality. The latter is true on average, but in particular also for later periods (i.e. in the medium or long run) as Figure 3 illustrates. If groups under punishment are unsuccessful, then their group income is lower, while they do not manage to be more equal than groups without punishment. In sum, punishment *does not* ensure higher levels of contributions (and hence group income) overall. However, it is able to reduce inequality in successful groups. Punishment can also have detrimental effects: it can lead to (a total)

¹²In principle there is another way to split the sample by group income. Across all periods, the median group income is 184 without punishment and 80 with punishment. Splitting the sample according to those medians, the split is imbalanced in terms of periods and a given group can be on either side depending on the period. This is why we split by period 10 group income.

¹³Figures D.4 and D.5 in the Online Appendix show the regression versions of Figure 3.

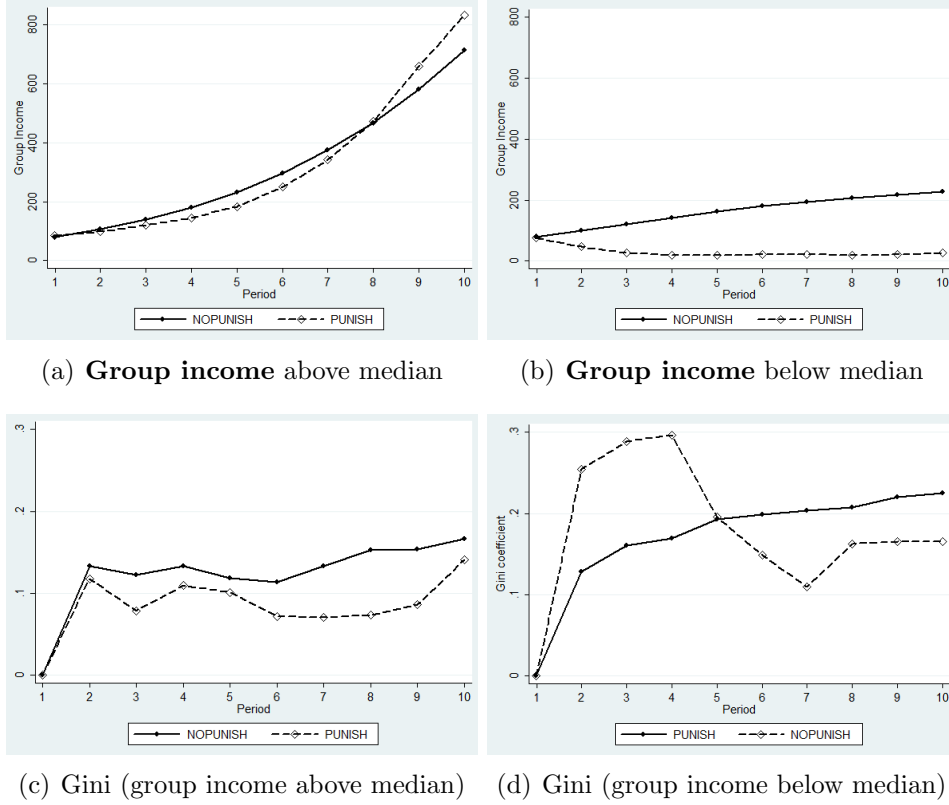


Figure 3: Group Income Gini over time in the two treatments in successful (above median group income) and unsuccessful (below median group income) groups.

destruction of wealth in some groups. Overall, dynamic interdependencies seem to be more effective than the possibility of punishment in securing high wealth.

Result 2: Growth and Inequality Across Treatments. *On average (across all groups) both wealth (group income) and inequality (Gini) are higher in treatment NOPUNISH. The differences in wealth are driven by groups with below median group income.*

4.3 The Relation between Growth and Inequality

Table 5 shows the correlation between period 10 group income and Gini coefficients for both treatments and for above and below median groups. For successful groups (with above median group income) this correlation tends to be negative, while it tends to be positive for unsuccessful groups (in treatment NOPUNISH the latter correlation is only marginally statistically significant).

Figure 4 illustrates this result. If we aggregate below and above median groups, then *overall* Gini coefficient and group income are essentially uncorrelated (NOPUNISH: $\rho = -0.2372$, PUNISH: $\rho = 0.2949^*$). What is noticeable is that in treatment PUNISH virtually all groups that achieve very high levels of group income in period 10, have a Gini coefficient of almost zero.

	Without Punishment	With Punishment
above median group income	-0.2721**	-0.4558***
below median group income	0.1994*	0.5824***

Table 5: Spearman correlation coefficients. Correlation between $t = 10$ group income and Gini-coefficient. Sample split by $t = 10$ group income. ***, **, * significance at 1,5,10 percent level.

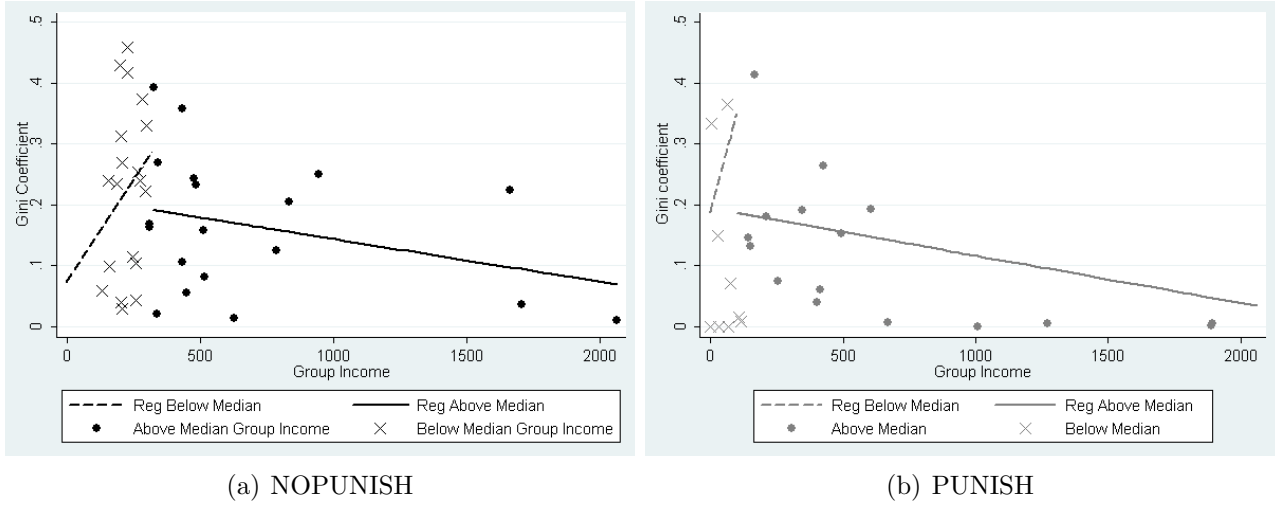


Figure 4: Correlation between the group income and Gini in period 10 only for the 38 (36) groups in treatment NOPUNISH (PUNISH). Each point represents one group with their period 10 group income and Gini coefficient. Dots are above median groups and crosses below median groups. Lines are fitted values from linear regression of Gini on group income. In the graph for treatment PUNISH some below median groups with a Gini coefficient of 1 are omitted from the graph (not the regression) for expositional clarity.

These correlations suggest that in order to achieve moderate levels of group income, groups need to incur some inequality. However, groups that achieve extremely high wealth are characterized by lower inequality. These relationships are particularly strong and statistically significant in treatment PUNISH. Remember also that in Result 1 we have seen that the possibility of punishment depresses group income especially in poor (below median) groups but achieves lower inequality. Taken together these results suggest that the possibility of punishing shirkers initially can have two very different effects: (i) it can lead to a vicious cycle of punishment and counter-punishment which leads to equality by destroying nearly all and sometimes all wealth or (ii) it can induce a deterrent effect where the threat of punishment suffices to ensure very high contributions by all group members leading to very high wealth and very low inequality.

Path Dependence. A different way of approaching the question of how growth and inequality are related is to study the extent of path dependency *within* groups, i.e. to ask whether groups with a higher group income (Gini coefficient) early on in the game will also end up with a higher group income eventually.

	9	8	7	6	5	4	3	2
Group Income								
NOPUNISH	0.99***	0.97***	0.96***	0.93***	0.87***	0.82***	0.77***	0.62***
PUNISH	0.99***	0.99***	0.98***	0.96***	0.94***	0.91***	0.91***	0.83***
Gini coefficient								
NOPUNISH	-0.22***	-0.13*	-0.25***	-0.26***	-0.33***	-0.21***	-0.27***	0.05
PUNISH	0.23***	0.25***	0.17**	-0.06	-0.18**	-0.40***	-0.47***	-0.47***

Table 6: Path Dependence: Correlation of group income in period 10 with group income (Gini) in period $X=2, \dots, 9$.

Table 6 presents evidence in this respect. It shows the correlation between group income in period 10 and the group income or the Gini coefficient in all previous periods 2 to 9 (since all groups have the same group income and Gini coefficient at the beginning of period 1 this

	Without Punishment	With Punishment
above median groups: group income in per 2	106	92
below median groups: group income in per 2	100	49
above median groups: Gini in per 2	0.13	0.11
below median groups: Gini in per 2	0.13	0.27

Table 7: group income and Gini-coefficient in period 2 of above and below median groups (Sample split by $t = 10$ group income).

correlation is zero for period 1). Path dependence is huge and tends to be stronger in the treatment with punishment. Initial inequality is highly detrimental to final group income in both treatments, but more so in the treatment with punishment. (The positive correlation between period 7,8,9 inequality and period 10 group income in treatment PUNISH is due to the fact that by then several groups have zero group income and inequality).

Table 7 illustrates the extent of path dependence. It shows period 2 group income and Gini coefficients of the groups which end up above or below the median in terms of group income in period 10. For treatment NOPUNISH the differences are not large. Successful (above median income) groups have a group income in period 2 of about 106, while eventually unsuccessful groups have a group income of about 100 on average. In terms of the Gini coefficient, both types of groups have a Gini coefficient of 0.13 on average. In treatment PUNISH, on the other hand, the differences are very large. Groups that are eventually successful have a group income in period 2 of about 92, while unsuccessful groups have a group income of only about 49. In terms of inequality, eventually successful groups have a Gini coefficient of about 0.11, while unsuccessful groups have a Gini coefficient of 0.27, more than twice as high. Hence, successful and unsuccessful groups already differ after the first round of the public good game is played.

One question about the path dependence result is whether it just reflects a stable distribution of “contribution types” or whether there is something more fundamental to it in the sense that *the same* people are more likely to end up with a much lower group income if initial inequality is high. The current evidence (in particular the treatment difference) suggests that the latter might be the case.

Additional support for this conjecture can also be derived from our post-experimental questionnaire. Table 12 in the Online Appendix shows the average amount in Euros that participants decide to donate to Medics without Borders at the end of the experiment. We find that participants from groups with a group income above the median do *not* contribute more on average than those from groups with below median group income. This is despite the fact that participants from groups with above median group income earn 178 tokens on average in period 10 (189 in treatment PUNISH), while those from groups with below median group income earn only 56 tokens (23 tokens) on average in period 10. This evidence suggests that participants in groups with above median group income are *not* per se more altruistic than others. We also elicited 14 other personality characteristics as well as a measure of risk aversion in the questionnaire. Tables 13-14 in the Online Appendix show that none of them is able to explain the variation in group income or Gini that we find in both treatments.

- Result 3: Relation between Growth and Inequality.**
1. *In both treatments group income and Gini coefficient are positively correlated for poor groups (below median group income) and negatively correlated for rich groups (above median group income). Both effects are stronger in treatment PUNISH.*
 2. *There is large path dependence: (a) Initial group income is strongly positively correlated with eventual group income. (b) Initial inequality is highly detrimental to final group income in treatment PUNISH and to a lesser extent so in treatment NOPUNISH.*

Before we conclude this section, let us point to potential links to other literatures. The sign of the relation between inequality and growth as well as the causal link between the two has been at the center of a big debate in macroeconomics and development (see e.g. Barro, 2000; Forbes, 2000; Persson and Tabellini, 1991, among many others). Most of these authors find either a negative relation or no significant relation at all. In our context the relation depends on the wealth of the group. For very poor groups in our data inequality and group income are positively related, while they are negatively related for richer groups. This is reminiscent of the famous Kuznets curve (Kuznets (1955)) which claims an inverse u-shaped relationship between growth and inequality. The connection between our setting and the fate of countries is far too loose to draw definite conclusions from our results on this question. Our results suggest, however, that there may be interesting links, between the level of social capital (cooperation, trust), the level of inequality and the level of growth of societies.¹⁴

In another strand of literature, economic historians have long argued about the importance of institutional lock in, where institutions are broadly understood as both formal constraints, such as rules and laws and informal constraints, such as norms of behaviour, conventions or codes of conduct (North, 1994). Our study provides one powerful example of how a society (group) can get locked into dysfunctional behavioural norms, where, if initial behavior leads to very high inequality, it becomes almost impossible to escape a path of punishment and counter-punishment that eventually leads to a destruction of all wealth.

To sum up the evidence presented in Section 4, we have seen that participants in this setting are able to achieve significant levels of cooperation and wealth and they are able to do so even in the absence of punishment possibilities. In fact, introducing the possibility of punishment even leads to lower levels of wealth. This is in stark contrast to previous insights from the standard literature on repeated public good games.

5 Discussion and Additional Results

This section is devoted to gaining a better understanding of where the differences between our setting and the standard setting originate. Remember that the dynamic interdependencies we introduce have two consequences: (i) they create the possibility for endogenous growth and (ii) they create endogenous inequality and hence asymmetries in the power to punish others. To understand the effects of these two forces we ran additional treatments, where we artificially eliminate growth (treatments NOPUNISH-NOGROWTH and PUNISH-NOGROWTH) or inequality. We discuss the former in Section 5.1. In Section 5.2. we discuss our new treatments where we artificially eliminate inequality (treatments NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY). In Section 5.3. we provide some additional results on punishment. At the end of this section we compare short (10 period) and long (15 period) horizon games in order to understand how participants perceive this environment strategically (Section 5.4).

5.1 “Eliminating Growth Possibilities”

In treatments NOPUNISH-NOGROWTH and PUNISH-NOGROWTH we artificially eliminate growth by re-normalizing all endowments at the end of a period s.t. they sum to 80. More precisely, while in treatment NOPUNISH the following relation holds for endowments $N_i^{t+1} = N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1, \dots, 4} a_j^t$, we now re-normalize

¹⁴Knack and Keefer (1997) and Zak and Knack (2001) have already identified a relationship between trust and growth across different countries. Our setting and results point to a potential mechanism behind this relationship.

$$N_i^{t+1} = \frac{80 * \left(N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1, \dots, 4} a_j^t \right)}{\sum_{j=1, \dots, 4} \left(N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1, \dots, 4} a_j^t \right)}.$$

We do the analogous normalization for the punishment version (PUNISH-NOGROWTH).

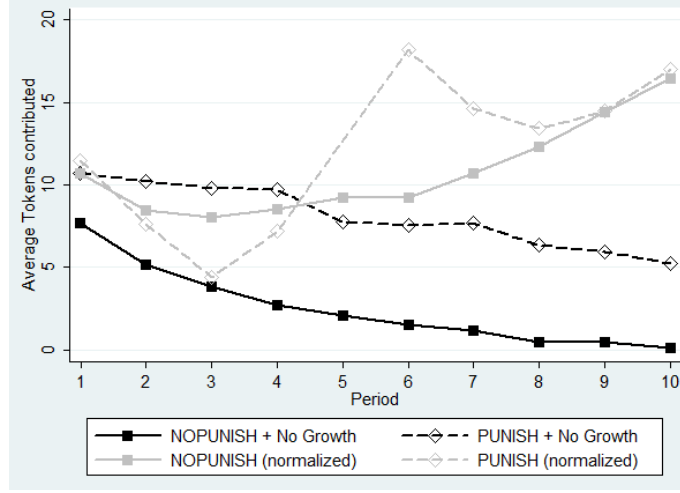


Figure 5: “Artificially eliminating growth”: average contributions over time in treatments NOPUNISH, PUNISH (both normalized), NOPUNISH-NOGROWTH and PUNISH-NOGROWTH.

Figure 5 shows the average amount of tokens contributed over time in treatments NOPUNISH, PUNISH, NOPUNISH-NOGROWTH and PUNISH-NOGROWTH. To ensure a fair comparison, we also normalize contributions in NOPUNISH and PUNISH to reflect the percentage of a non-normalized endowment this person contributed. The figure shows that the possibility of exponential growth is crucial to sustain high and increasing contributions. In treatment NOPUNISH-NOGROWTH, where the possibility of growth is eliminated, contributions decline to zero as in the standard setting. Contributions in PUNISH-NOGROWTH are higher than in NOPUNISH-NOGROWTH, but they are also declining over time. As we will see in the next section the lower effectiveness of punishments is likely due to endogenous inequality and hence punishment possibilities. In stark contrast to the NO-GROWTH treatments, normalized contributions in NOPUNISH and PUNISH are increasing over time. The statistical significance of these treatment differences is confirmed in the first column of regression table 8.

Result 4: Growth Possibilities *The possibility of exponential growth is crucial to sustain positive and increasing contributions over time. If this possibility is eliminated contributions decrease to zero as in the standard setting.*

5.2 “Eliminating Inequality”

We now turn to treatments NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY, where we artificially eliminate inequality. In these treatments we redistribute all earnings at the beginning of each period, s.t. endowments are equal for all players. Hence each player receives

$$N_i^{t+1} = \frac{\sum_{i=1, \dots, 4} \left(N_i^t - a_i^t + \frac{1.5}{4} \sum_{j=1, \dots, 4} a_j^t \right)}{4}$$

at the beginning of period $t + 1$. To maintain the character of the public good game as a social dilemma (and to ensure that the unique equilibrium is still zero contributions in each period),

	norm contribution	contribution
PUNISH	4.915 (5.019)	18.895** (8.660)
NOPUNISH-NOGROWTH	−25.470*** (4.975)	
PUNISH-NOGROWTH	−19.908*** (5.072)	
NOPUNISH-NOINEQUALITY		18.766** (9.026)
PUNISH-NOINEQUALITY		43.212*** (14.072)
constant	18.191*** (4.966)	13.587*** (2.727)
N	6320	6560
Clusters	158	164
R^2	0.0330	0.0461

Table 8: (Normalized) contributions regressed on treatment dummies. Simple OLS regression. Standard errors clustered by matching group. Baseline is treatment NOPUNISH. ***, **, * significance at 1,5,10 percent level.

we implemented an additional change in this treatment. In particular we paid a randomly drawn period rather than the sum of all ten periods.

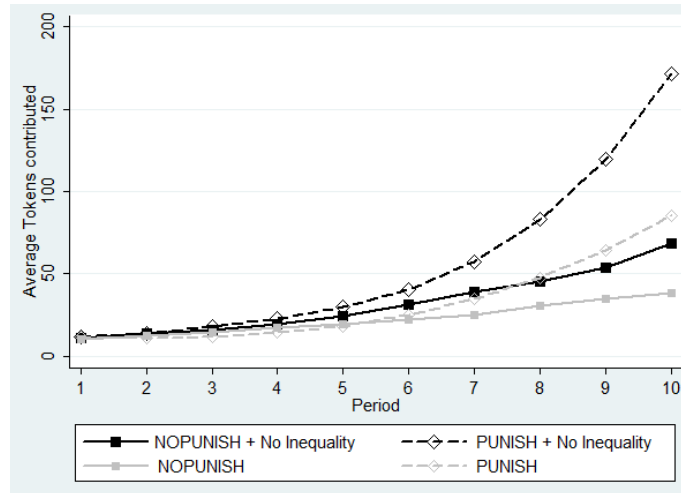


Figure 6: “Artificially eliminating inequality”: average contributions over time in treatments NOPUNISH, PUNISH, NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY.

Figure 6 shows the average contributions over time in treatments NOPUNISH, PUNISH, NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY. It can be seen that inequality leads to lower contributions in both the treatments with and without punishment. Furthermore, if inequality is eliminated then contributions are much higher with punishment (PUNISH-NOINEQUALITY) than without (NOPUNISH-NOINEQUALITY) and this difference can be observed across all periods. Table 8 (second column) confirms the statistical significance of these findings. Removing inequality also reverses a result we have observed in Table 2, that group income is lower with punishment than without. If inequality is removed, then average group income in period 10 is higher in treatment PUNISH-NOINEQUALITY (772 tokens) compared to NOPUNISH-NOINEQUALITY (602 tokens).

Result 5: Inequality *Removing endogenous inequality leads to higher contributions. Group*

income and contributions are higher with than without punishment possibilities if and only if inequality is removed.

It seems that endogenous inequality in endowments with the associated inequality in the power to punish undermines the effectiveness of punishment we are used to from the standard setting. In the following subsection we take a closer look at how people punish in treatment PUNISH to see whether we can find evidence of anti-social punishment where shirkers exploit their higher punishment power to punish non-shirkers.

5.3 Anatomy of Punishment

Our strategy in this subsection is to compare above and below median groups in treatment PUNISH to understand which patterns of punishment lead to low group income in this treatment and are hence crucial for the low effectiveness of punishment in this setting.

Figure 7 reveals an interesting pattern. It shows the amounts of tokens participants use to punish over time. In groups with above median group income the absolute amounts used to punish tend to remain stable or increase over time, while they tend to decrease in below median groups. Particularly striking is the fact that, in terms of amounts, the major difference between above and below median groups seems to lie in how much they punish in the first two periods of the game. This indicates that much of the difference between successful and unsuccessful groups in PUNISH can be traced back to differences in punishment behaviour in the first two periods of the game. In successful groups, there is also an interesting peak in punishment one period before the game ends. This suggests that some participants may tolerate some degrees of free-riding because of the detrimental effect that punishment can have on growth and wait to punish others harshly at the end of the game. While in below median groups most punishment happens in the beginning of the game, above median groups punish at the end.¹⁵

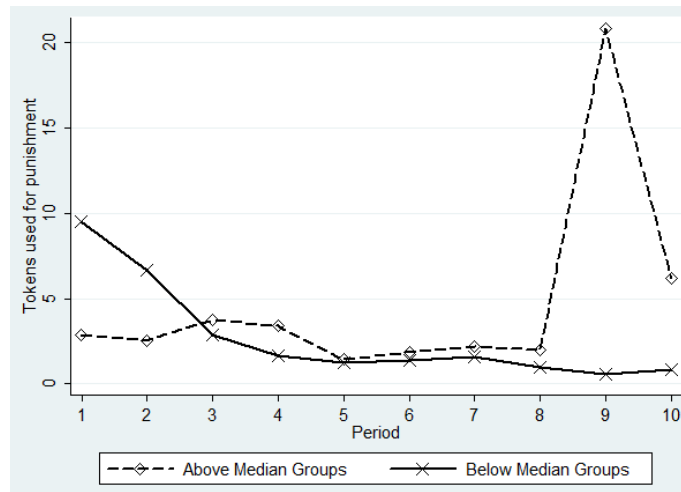


Figure 7: Average Amount of Tokens used to punish over time.

We next ask the question under which conditions punishment is “*social*” and when it is “*anti-social*”. We look at social and antisocial punishment because Herrmann et al. (2008) have found that anti-social punishment strongly undermines successful cooperation, whereas social punishment fosters cooperation. Punishment by player i is *social* if i punishes a player who has contributed a lower share of his income to the public good than i herself. Punishment by player i

¹⁵Relatedly Fudenberg and Pathak (2010) have shown that punishment can sustain cooperation even if it is only observed at the end of a session.

is *antisocial* if i punishes a player who has contributed a higher share of his income to the public good than i herself.

Table 17 shows that there is at least twice as much social punishment than anti-social punishment on average. Both types of punishment are higher if inequality is high (social: Spearman $\rho = 0.2146^{***}$; anti-social: $\rho = 0.0913^{***}$). The relation between group income and punishment is not statistically significant on average according to a Spearman test. The reason turns out to be that in rich groups (group income > 80) punishment is negatively correlated with group income (social: $\rho = -0.0870^{**}$; anti-social: $\rho = -0.1206^{***}$), while it is positively correlated for poor groups (social: $\rho = 0.2153^{***}$; anti-social: $\rho = 0.2487^{***}$). This is reminiscent of the correlation between group income and the Gini coefficient that we discovered earlier. Since we discovered the latter correlation also in treatment NOPUNISH, we conclude that a high Gini coefficient is causal to punishment.

	Social Punishment	Antisocial Punishment
all	1.57 (14.39)	0.69 (3.44)
Gini > 0.05	1.18 (2.29)	1.28 (10.72)
Gini < 0.05	1.99 (20.43)	0.58 (2.23)
group income > 80	2.40 (21.19)	0.61 (4.62)
group income < 80	0.52 (1.72)	0.48 (1.87)
Gini > 0.05 + per 6-10	1.06 (2.60)	2.16 (14.42)
Gini < 0.05 + per 6-10	2.36 (27.13)	0.19 (1.06)
group income > 80 + per 6-10	3.13 (27.45)	0.68 (5.33)
group income < 80 + per 6-10	0.16 (0.62)	0.22 (1.29)

Table 9: Mean (SD) of amount of tokens of social and anti-social punishment.

Overall, we can summarize our results on punishment as follows. There is about twice as much social compared to anti-social punishment. In successful groups (above median group income) there is more social punishment, while in unsuccessful (below median group income) groups there is more antisocial punishment. High inequality increases both types of punishment. Unsuccessful groups punish at the beginning of the game, while successful groups withhold their punishment until the end. There is also substantial evidence of richer participants that punish the poorest participants at a low cost to themselves (see Figure G.2 in Online Appendix G). In successful groups (above median group income) overall punishment is negatively correlated with group income, while it is positively correlated for unsuccessful (below median group income) groups.

5.4 Comparison of Long and Short Horizon games

Before we conclude we compare behaviour in the short horizon (10 period) and long horizon (15 period) games. This comparison can give some insights into how participants perceive this environment strategically and whether and how they adjust their contributions as the time horizon (and hence the scope for growth) increases. In Online Appendix H we provide additional results on the longer run dynamics in the 15 period variations.

Figure 8 compares the evolution of group income and Gini over time for the long and short horizon games in treatment NOPUNISH. In terms of group income, there is no discernible difference between the long and short versions. However, while the Gini coefficient tends to increase in both cases, it seems to do so at a faster rate in the short-horizon games. One possible interpretation is that participants are reluctant to let inequality get too big early on in the long horizon games, because of the detrimental and long-lasting effect it can have on growth. This is intuitive given the extent of path dependence observed in Table 6. Less aggressive shirking and

withholding contributions as a form of punishment must balance each other out to achieve about the same level of growth as in the short-horizon games, but lower levels of inequality.

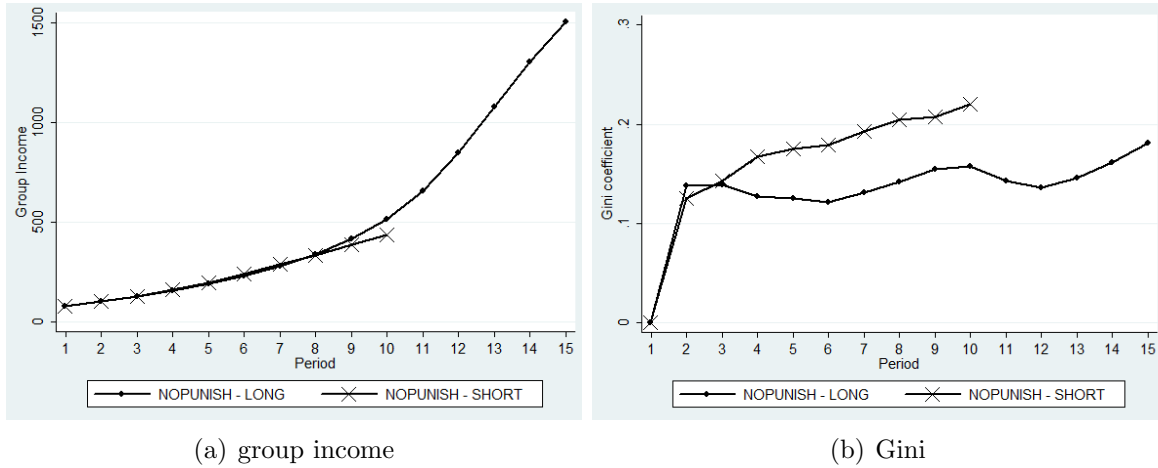


Figure 8: Average group income and Gini coefficient over time in the two treatments. Comparison of long and short horizons in treatment NOPUNISH.

Figure 9 compares the evolution of group income and Gini over time for the long and short horizon games in treatment PUNISH. Again, there does not seem to be much difference in the evolution of group income. In both, the short and long horizon games, the Gini coefficient does follow the same cyclical pattern observed in Figure 2, however cycles seem to be longer in the long horizon games. The intuition could be similar. Participants might be more reluctant to punish asymmetrically if such punishment increases inequality because of the detrimental effect of inequality on growth.

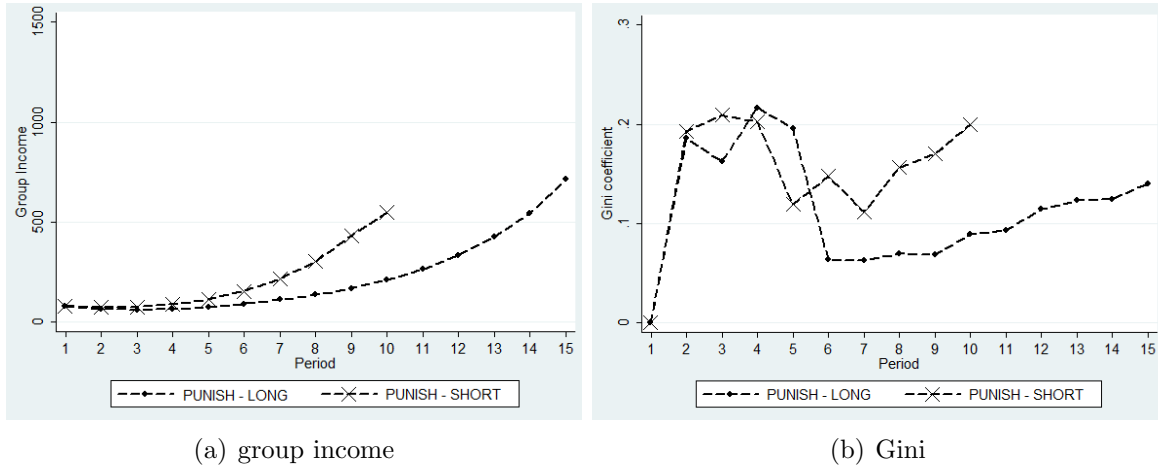


Figure 9: Average group income and Gini coefficient over time in the two treatments. Comparison of long and short horizons in treatment PUNISH.

Overall it seems that participants do not react very strongly to the different time horizons. They seem to be a bit more careful with letting inequality get too large if the horizon is longer. In Online Appendix H we compare the long and short horizon games more extensively. There we show that there are not many differences in terms of the amounts contributed as the length of the horizon increases (Figure H.4). This is another difference to the standard setting (see e.g. Gaechter et al. (2008)). We also show that our main results regarding growth, inequality and treatment comparisons hold true irrespective of the length of the game (Figures H.1 - H.3).

6 Conclusions

We studied a public good game in a novel design, which allows for endogenous growth and inequality. More precisely, each agent's income at the end of a period serves as her endowment in the next period. Results are very different from the received literature on standard public good games. There is no decay of cooperation over time even in the absence of punishment possibilities and punishment leads to lower group income. Punishment is able, however, to curb inequality. Across groups, group income and inequality are negatively related for rich groups and positively related for poor groups. Within groups we observe huge path dependence. Inequality in early periods of the game is strongly negatively correlated with group income at the end of the experiments.

Those results show that people are able to establish persistent cooperation in a setting that shares one key feature with many real-life interactions: the fact that past behaviour matters for future endowments. They should also make us rethink our intuition based on the standard setting about the effectiveness of punishment to sustain cooperation. Punishment has mostly negative consequences in our setting. One advantage of punishment that emerges is its effectiveness to curb inequality in some cases.

References

- Adler, P. (1985). *Wheeling and Dealing - An Ethnography of Upper-Level Drug Dealing and Smuggling Community*. Columbia University Press.
- Andreoni, J. (1995). Cooperation in public goods experiments: Kindness or confusion? *The American Economic Review* 85(4), 891–904.
- Andreoni, J., W. Harbaugh, and L. Vesterlund (2003). The carrot or the stick: rewards, punishment and cooperation. *The American Economic Review* 93(3), 893–902.
- Barro, R. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth* 5, 5–32.
- Becker, G. (1981). *A treatise on the family*. Cambridge.: Harvard University Press.
- Bergstrom, T., L. Blume, and H. Varian (1986). On the private provision of public goods. *Journal of Public Economics* 29, 25–49.
- Chan, K., S. Mestelman, R. Moir, and R. Muller (1996). The voluntary provision of public goods under varying income distributions. *The Canadian Journal of Economics* 29(1), 54–69.
- Chan, K. S., S. Mestelman, R. Moir, and R. Muller (1999). Heterogeneity and the voluntary provision of public goods. *Experimental Economics* 2(1), 5–30.
- Chaudhury, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* 14, 47–83.
- Cozzi, G. (1999). R+d cooperation and growth. *Journal of Economic Theory* 86(1), 17–49.
- Croson, R. and M. Marks (1998). Alternative rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics* 67(2), 195–220.
- Dawes, R., J. Orbell, R. Simmons, and A. J. van de Kragt (1986). Organizing groups for collective action. *The American Political Science Review* 80(4), 1171–1185.
- Deaton, A. (1997). *Analysis of Household Surveys*. Baltimore, MD.: Johns Hopkins University Press.
- Dreber, A., D. Rand, D. Fudenberg, and M. Nowak (2008). Winners don’t punish. *Nature* 452, 348–351.
- Duffy, J., J. Ochs, and L. Vesterlund (2007). Giving little by litte: Dynamic voluntary contribution games. *Journal of Public Economics* 91(9), 1708–1730.
- Egas, M. and A. Riedl (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B - Biological Sciences* 275, 871–878.
- Fehr, E. and S. Gaechter (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90, 341–369.
- Fehr, E. and S. Gaechter (2002). Altruistic punishment in humans. *Nature* 415, 137–140.
- Fischbacher, U. and S. Gaechter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review* 100(1), 541–556.

- Forbes, K. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review* 90(4), 869–997.
- Fudenberg, D. and P. Pathak (2010). Unobserved punishment supports cooperation. *Journal of Public Economics* 94(1), 78–86.
- Gaechter, S., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science* 322, 1510.
- Green, E. and R. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica* 52(1), 87–100.
- Grossman, G. and C. Shapiro (1986). Research joint ventures: An anti-trust analysis. *Journal of Law, Economics and Organization* 2(2).
- Guererk, O., B. Rockenbach, and I. Wolff (2013). The effects of punishment in dynamic public-good games. mimeo.
- Hamilton, B. H., J. A. Nickerson, and H. Owan (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy* 111, 465–497.
- Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science* 319, 1362–1367.
- Huck, T. (2006). Dynamisches public good experiment. Master’s thesis, Universitaet Erfurt.
- Isaac, R. M., J. M. Walker, and S. Thomas (1984). Divergent evidence on free-riding: An experimental examination of possible explanations. *Public Choice* 43(2), 113–149.
- Johnson, A. and T. Earle (1987). *The Evolution of Human Societies: From Foraging Group to Agrarian State*. Stanford University Press.
- Josephson, M. (1962). *The Robber Barons*. Mariner Books.
- Kamien, M. I., E. Muller, and I. Zang (1992). Research joint ventures and r+d cartels. *The American Economic Review* 82(5), 1293–1306.
- Knack, S. and P. Keefer (1997). Does social capital have an economic payoff?: a cross country investigation. *The Quarterly Journal of Economics* 112(4), 1251–1288.
- Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoner’s dilemma. *Journal of Economic Theory* 27, 245–252.
- Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review* 45(1), 1–28.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton University Press.
- Milinski, M., D. Semmann, and H.-J. Krambeck (2002). Reputation helps solve the ‘tragedy of the commons’. *Nature* 415, 424–426.
- North, D. (1994). Performance through time. *The American Economic Review* 84(3), 359–368.

- Ostrom, E. (1990). *Governing the Commons - The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.
- Ostrom, E., R. Gardner, and J. Walker (1994). *Rules, Games and Common Pool Resources*. Ann Arbor: University of Michigan Press.
- Ostrom, E., J. Walker, and R. Gardner (1992). Covenants without a sword: Self-governance is possible. *The American Political Science Review* 86, 104–115.
- Persson, T. and G. Tabellini (1991). Is inequality harmful for growth? theory and evidence. *American Economic Review* 48, 600–621.
- Plott, C. R. and V. Smith (Eds.) (2009). *Handbook of Experimental Economics Results*. North-Holland.
- Reuben, E. and A. Riedl (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior* 77, 122–137.
- Steiner, J. (2007). A trace of anger is enough: On the enforcement of social norms. *Economics Bulletin* 8, 1–4.
- Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics and Organization* 2(2), 181.
- van Dijk, F., J. Sonnemans, and F. van Winden (2002). Social ties in a public good experiment. *Journal of Public Economics* 85, 275–299.
- Zak, P. J. and S. Knack (2001). Trust and growth. *Economic Journal* 111, 295–321.

This is the Online Appendix

Table of Contents

A: Screenshots
B: Questionnaire Experimental Instructions
C: Proof of Theoretical Results
D: Additional Analyses Section 4
E: Matching Groups
F: Questionnaire Data
G: Additional Analyses Section 5
H: Questionnaire Experimental Instructions

A Screenshots

	Tokens at the beginning of this period	Tokens placed in group account	Share of group account return	Tokens in the end of the period
You	9	5.00	4.87	9
Other1	11	2.00	4.87	14
Other2	12	6.00	4.87	11
Other3	1	0.00	4.87	6

Figure A.1: Information the participants see after each period in NOPUNISH treatment.

	Tokens at the beginning of this period	Tokens placed in group account	Share of group account return	Total tokens before subtraction	Number of tokens to subtract
You	9	5.00	4.87	9	
Other1	11	2.00	4.87	14	
Other2	12	6.00	4.87	11	
Other3	1	0.00	4.87	6	

Figure A.2: The screen shot of the punishment stage.

	Tokens at the beginning of this period	Tokens placed in group account	Share of group account return	Total tokens before subtraction	You subtracted	You got subtracted by others	Sum of subtractions by others	Tokens in the end of the period
You	9	5.00	4.87	9			4.00	3
Other1	11	2.00	4.87	14	1.00	1.00	4.00	9
Other2	12	6.00	4.87	11	2.00	0.00	4.00	6
Other3	1	0.00	4.87	6	3.00	3.00	6.00	0

Figure A.3: Information available to the participants after the punishment phase.

B Questionnaire and Experimental Instructions

B.1 Questionnaire

The following questions were asked after both PUNISH and NOPUNISH treatments.

- What is your gender?
- What is your nationality?
- What is your year of birth?
- What is your field of studies?
- For how many years have you been studying at university?

Suppose you have a hypothetical choice between a bet and a sure outcome. What would you choose in the following cases:

- €10 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €20 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €30 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €40 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €50 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €60 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance
- €70 Euro or 100 Euro with 50% chance and €0 Euro with 50% chance

Personality questions: indicate how strongly you agree with the following statements (1 means disagree strongly, 7 agree strongly).

- I am a quick thinker
- I get easily offended
- I am very satisfied with myself
- I am very dependent on others
- Generally speaking, I am happy
- Work plays a very important role in my life
- Family plays a very important role in my life
- Friends play a very important role in my life
- Religion plays a very important role in my life
- Politics plays a very important role in my life
- Generally, most people can be trusted
- In the long run, hard work brings a better life
- The government should take responsibility that people are better provided for
- Incomes should be made more equal

In addition the participants were asked if they would be willing to donate some of their earnings to Doctors without Borders.

B.2 Instructions NOPUNISH Treatment

General information

You are about to participate in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in cash at the end of the experiment

This set of instructions is for your private use only. During the experiment you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then we will come to your seat and answer your questions. Any violation of this rule excludes you immediately from the experiment and all payments. The funds for conducting this experiment were provided by the Marie Curie Reintegration Grant from the EU.

Throughout the experiment you will make decisions about amounts of tokens. At the end of the experiment all tokens you have will be converted into Euros at the exchange rate 0.05 Euro for 1 token and paid you in cash in addition to the show-up fee of 2 Euros.

During the experiment all your decisions will be treated confidentially. This means that none of the other participants will know which decisions you made.

Experimental Instructions

The experiment will consist of 10 decision making periods. At the beginning of the experiment, you will be matched with 3 other people in this room. Therefore, there are 4 people, including yourself, participating in your group. You will be matched with the same people during the entire experiment. None of the participants knows who is in which group.

Before the first period you, and each other person in your group, will be given the endowment of 20 tokens.

At the beginning of the first period you will be asked to allocate your endowment between a private account and a group account.

The tokens that you place in the private account have a return of 1 at the end of the first period. This means that at the end of the first period your private account will contain exactly the amount of tokens you put into the private account at the beginning of the period. Nobody except yourself benefits from your private account.

The tokens that you place in the group account are summed together with the tokens that the other three members of your group place in the group account. The tokens in the group account have a return of 1.5. Every member of the group benefits equally from the group account. Specifically, the total amount of tokens placed in the group account by all group members is multiplied by 1.5 and then is equally divided among the four group members. Hence, your share of the group account at the end of the first period is

$$1.5 * (\text{sum of tokens in the group account}) / 4$$

Your endowment at the beginning of the second period will be equal to the amount of tokens contained in your private account at the end of the first period plus your share of the group account at the end of the first period.

At the beginning of the second period you will be again asked to allocate the endowment that you have at the beginning of the second period between a private account and a group account. Both the private and the group account work in exactly the same manner as in the first period, namely, they have the same returns.

The structure of the experiment at all subsequent periods is identical: your endowment at the beginning of each period is equal to the amount of tokens in your private account at the end of the previous period plus your share of the group account at the end of the previous period.

At the end of each period, you will be informed about

- The endowment all four group members had at the beginning of the period
- How much each group member allocated to the group account and to their respective private accounts.

- Your share of the group account (remember it is the same for all group members).

All other participants will receive exactly the same information.

Your total income in the end of the experiment is equal to the amount of tokens in your private account and your share of the group account at the end of period 10. At the end of the experiment there will be a short questionnaire for you to fill in.

B.3 Instructions PUNISH Treatment

General information

You are about to participate in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in cash at the end of the experiment

This set of instructions is for your private use only. During the experiment you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then we will come to your seat and answer your questions. Any violation of this rule excludes you immediately from the experiment and all payments. The funds for conducting this experiment were provided by the Marie Curie Reintegration Grant from the EU.

Throughout the experiment you will make decisions about amounts of tokens. At the end of the experiment all tokens you have will be converted into Euros at the exchange rate 0.05 Euro for 1 token and paid you in cash in addition to the show-up fee of 2 Euros.

During the experiment all your decisions will be treated confidentially. This means that none of the other participants will know which decisions you made.

Experimental Instructions

The experiment will consist of 10 decision making periods. Each period consists of two stages. At the beginning of the experiment, you will be randomly matched with 3 other people in this room. Therefore, there are 4 people, including yourself, participating in your group. You will be matched with the same people during the entire experiment. None of the participants knows who is in which group.

Before the first period you, and each other person in your group, will be given the endowment of 20 tokens.

At the first stage of the first period you will be asked to allocate your endowment between a private account and a group account.

The tokens that you place in the private account have a return of 1 at the end of the first stage. This means that at the end of the first stage your private account will contain exactly the amount of tokens you put into the private account at the beginning of the first stage. Nobody except yourself benefits from your private account.

The tokens that you place in the group account are summed together with the tokens that the other three members of your group place in the group account. The tokens in the group account have a return of 1.5. Every member of the group benefits equally from the tokens in the group account. Specifically, the total amount of tokens placed in the group account by all group members is multiplied by 1.5 and then is equally divided among the four group members. Hence, your share of the group account at the end of the first stage of the first period is

$$1.5 * (\text{sum of tokens in the group account}) / 4$$

In the second stage of the first period you will be asked to react to the decisions made during the first stage of the first period. At this point, you will already know the decisions taken by each group member at the first stage. You will decide whether you want to subtract tokens from any other group member or not. The members that you decide to subtract tokens from will lose the amount of tokens you chose. Subtracting tokens from someone else is costly for you too. The following table illustrates the relation between your cost in tokens and the amount of tokens that are taken away from the member of your group:

Tokens subtracted	Cost for you
3	1
6	2
9	3
...	...
$3y$	y

You may subtract different amounts of tokens from different group members. Other group members will be able to subtract tokens from you as well. You lose the sum of tokens that other three group members decided to subtract from you. Any group member including you can only lose maximum the amount of tokens he or she has.

At the beginning of the second period your endowment will be equal to the amount of tokens contained in your private account at the end of the first stage of the first period plus your share of the group account at the end of the first stage of the first period, minus your cost for subtracting others' tokens and minus the amount of tokens subtracted from you by other members.

At the first stage of the second period you will be again asked to allocate the endowment that you have at the beginning of the second period between a private account and a group account. Both the private and the group account work in exactly the same manner as in the first period, namely, they have the same returns. At the second stage of the second period you will be asked to react to the decisions made during the first stage of the second period in exactly the same manner as in the first period.

The structure of the experiment at all subsequent periods is identical: your endowment at the beginning of each period is equal to the amount of tokens in your private account at the end of the first stage of previous period, plus your share of the group account at the end of the first stage of the previous period, minus your cost from subtracting other members' tokens at the second stage of the previous period, minus the amount of tokens subtracted from you by other members at the second stage of the previous period.

At the end of each period, you will be informed about

- The endowment all four group members had at the beginning of the period
- How much each group member allocated to the group account and to their respective private accounts
- Your share of the group account (remember it is the same for all group members)
- How many tokens each member subtracted from you.

All other participants will receive exactly the same instructions.

Your total income in the end of the experiment is equal to the amount of tokens left after last subtraction in your private account and your share of the group account at the end of period 10. At the end of the experiment there will be a short questionnaire for you to fill in.

C Proofs of Section 3

In this section we establish the theoretical results discussed informally in section 3. We start with some notation and then prove the result of zero contributions and zero punishment on the equilibrium path of the unique Nash equilibrium.

C.1 Notation and preliminaries

Players and histories. Let $I = \{1, \dots, 4\}$ is the set of players, with typical elements i and j . Furthermore, assume that H denotes the (finite) set of nodes – else called histories or information sets – with typical element h . In games without punishment, every history corresponds to

a contribution stage, henceforth called contribution history. On the other hand, in a game with punishment we distinguish between histories that correspond to a contribution stage and histories that correspond to a punishment stage, with the latter being called punishment histories. Therefore, from a game-theoretic point of view, a game with punishment contains many more histories than a game without punishment.

Strategies. Let A_i^h be the finite set of moves that player i has at $h \in H$, with typical element a_i^h . More specifically, if h is a contribution history,

$$A_i^h := \{0, \dots, N_i^h\}$$

is the set of possible contributions that player i can make at h , with N_i^h denoting the number of tokens in i 's private account upon reaching the history h . If, on the other hand, h is a punishment history,

$$A_i^h := \{ (p_i^j)_{j \neq i} \in \mathbb{N}^3 : \sum_{j \neq i} p_i^j \leq N_i^h \}$$

is the set of possible punishments with $3p_i^j$ denoting the amount that player i subtracts from j 's private account, and p_i^j the corresponding cost that player i has to incur in order to punish player j .¹⁶ For notation simplicity, let

$$c_i := \sum_{j \neq i} p_i^j$$

denote i 's cost from punishing other players, and

$$p_i := 3 \sum_{j \neq i} p_j^i$$

the amount subtracted from i 's account due to being punished by others. As usual, $A^h := A_1^h \times \dots \times A_4^h$ denotes the set of action profiles at h . Obviously, A_i^h depends on the amount of tokens that player i has accumulated in her private account so far. This is for instance why our public good game without punishment is not a repeated game, as opposed to the standard case where $A_i^h = \{1, \dots, 20\}$ for all $h \in H$. Moreover, let $A_i := \prod_{h \in H} A_i^h$ and $A := A_1 \times \dots \times A_4$ denote i 's set of strategies and the set of strategy profiles respectively.

Paths of play. Let H_t denote the set of histories at time t , e.g., $H_1 := \{h_1\}$ contains the unique history that corresponds to the normal form game played at period 1, else called the root of the game. Each strategy profile $a \in A$ induces a sequence of histories $(h_1, \dots, h_T) \in H_1 \times \dots \times H_T$, henceforth called a path of play, or simply a path. In this case, $(a^{h_1}, \dots, a^{h_T})$ is called the realization of a . In other words, the realization of a strategy profile is the observed behavior. We say that some history $h \in H$ is on the path (h_1, \dots, h_T) , if there exists some $t = 1, \dots, T$ such that $h_t = h$. Otherwise, we say that it is off the path.

A sequence $(h_t, \dots, h_T) \in H_t \times \dots \times H_T$ of histories is called a conditional path if there exists a path $(h_1, \dots, h_t, \dots, h_T)$. For any history h'_t , possibly off the path induced by a , the strategy profile a induces a path (h'_t, \dots, h'_T) conditional on h'_t . This corresponds to the behavior we would observe if the history h'_t was reached and every player moved according to the strategy profile a at all histories after the t -th round.

¹⁶Obviously, player i 's total cost from punishing cannot exceed the number of tokens in her private account at the corresponding history.

Payoff functions. We define player i 's payoff function $u_i : A \rightarrow \mathbb{R}$ in our *public good game without punishment* by

$$u_i(a) = 20 - \sum_{t=1}^T a_i^{h_t} + \frac{r}{4} \sum_{t=1}^T \sum_{j=1}^4 a_j^{h_t} \quad (1)$$

where (h_1, \dots, h_T) is the path induced by a , and r denotes the returns of the public good. Moreover, we denote player i 's payoff conditional on the history h'_t by

$$u_i(a|h'_t) = N_i^{h'_t} - \sum_{\tau=t}^T a_i^{h'_\tau} + \frac{r}{4} \sum_{\tau=t}^T \sum_{j=1}^4 a_j^{h'_\tau} \quad (2)$$

where (h'_t, \dots, h'_T) is the path induced by a conditional on h'_t . Notice that h'_t may in principle be off the path (h_1, \dots, h_T) induced by a .

Likewise, we define the payoff function $u_i^p : A \rightarrow \mathbb{R}$ in our *public good game with punishment*. First, notice that the number of realized histories is even, alternating between a contribution and a punishment history, i.e., $T = 2R$, where R is the number of rounds in our game. Then, i 's payoff function is given by

$$u_i^p(a) = 20 - \sum_{t=1}^R a_i^{h_{2t-1}} + \frac{r}{4} \sum_{t=1}^R \sum_{j=1}^4 a_j^{h_{2t-1}} - \sum_{t=1}^R \min \{ (c_i^{h_{2t}} + p_i^{h_{2t}}), N_i^{h_{2t}} \} \quad (3)$$

where (h_1, \dots, h_T) is the path induced by a , and r denotes the returns of the public good. Finally, we denote player i 's payoff conditional on the history h'_t by

$$u_i^p(a|h'_t) = N_i^{h'_t} - \sum_{\tau=t}^R a_i^{h'_{2\tau-1}} + \frac{r}{4} \sum_{\tau=t}^R \sum_{j=1}^4 a_j^{h'_{2\tau-1}} - \sum_{\tau=t}^R \min \{ (c_i^{h'_{2\tau}} + p_i^{h'_{2\tau}}), N_i^{h'_{2\tau}} \} \quad (4)$$

where (h'_t, \dots, h'_T) is the path induced by a conditional on h'_t . Notice that h'_t may in principle be off the path (h_1, \dots, h_T) induced by a .

Obviously the difference between the payoff functions in the two games lays in the fact that without punishment $T = R$ instead of $T = 2R$, and furthermore the fact that in the punishment game there is an additional term that captures the costs of punishing and being punished. Notice that the later cannot exceed the number of tokens contained in the private account at that history. Note that in all our cases, we assume $r = 1.5$.

C.2 Predicted behavior: Results and Proofs

We say that a strategy $a_i \in A_i$ is a best response to $a_{-i} \in A_{-i}$, and we write $a_i \in BR_i(a_{-i})$, whenever $u_i(a_i, a_{-i}) \geq u_i(b_i, a_{-i})$, and respectively $u_i^p(a_i, a_{-i}) \geq u_i^p(b_i, a_{-i})$, for all $b_i \in A_i$. The strategy profile a is a Nash equilibrium whenever $a_i \in BR_i(a_{-i})$ for every $i \in I$.

Proposition 2 *Consider the public good game with growth as defined above.*

- (i) *Let $a \in A$ be a Nash equilibrium in the public good game without punishment, and let (h_1, \dots, h_T) denote the path induced by a . Then, $a_i^{h_t} = 0$ for every $i \in I$ and for all $t = 1, 2, \dots, T$.*
- (ii) *Let $a \in A$ be a Nash equilibrium in the public good game with punishment, and let (h_1, \dots, h_T) denote the path induced by a . Then, $a_i^{h_t} = 0$ for every $i \in I$ and for all $t = 1, 3, \dots, T-1$. Moreover, $c_i^{h_t} = 0$ for every $i \in I$ and for all $t = 2, 4, \dots, T$.*

Proof of Proposition 2. We prove (i) and notice that (ii) is proven by following the same steps. Suppose that there exists some $t = 1, \dots, T$ and some player $i \in I$ such that $a_i^{h_t} > 0$.

Firstly, we show that $t \neq T$: Observe that otherwise player i can profitably deviate from a by choosing $b_i \in A_i$, where $b_i^h = a_i^h$ for all $h \in H \setminus \{h_T\}$ and $b_i^{h_T} = a_i^{h_T}$. Likewise, we show that $a_j^{h_T} = 0$ for all $j \in I$.

Secondly, we show that $t \neq T - 1$: Suppose instead that $t = T - 1$. Let h'_T be the direct successor of h_{T-1} if every $j \neq i$ plays according to a_j and i plays 0 (instead of $a_i^{h_{T-1}}$) at h_{T-1} , i.e., formally $h'_T = \eta_T(0, a_{-i}^{h_{T-1}})$. Then, define the strategy $b_i \in A_i$ such that $b_i^h = a_i^h$ for all $h \in H \setminus \{h_{T-1}, h'_T\}$ and $b_i^{h_{T-1}} = 0$ and $b_i^{h'_T} = 0$. Now, observe that b_i is a profitable deviation from a_i given a_{-i} , since

$$\begin{aligned}
u_i(a) &= 20 - \sum_{t=1}^T a_i^{h_t} + \frac{r}{4} \sum_{t=1}^T \sum_{j=1}^4 a_j^{h_t} \\
&= 20 - \sum_{t=1}^{T-1} a_i^{h_t} + \frac{r}{4} \sum_{t=1}^{T-1} \sum_{j=1}^4 a_j^{h_t} && (\text{since } a_j^{h_T} = 0 \text{ for all } j \in I) \\
&< 20 - \sum_{t=1}^{T-1} b_i^{h_t} + \frac{r}{4} \sum_{t=1}^{T-1} \left(b_i^{h_t} + \sum_{j \neq i} a_j^{h_t} \right) && (\text{since } b_i^{h_{T-1}} < a_i^{h_{T-1}}) \\
&= 20 - \sum_{t=1}^{T-1} b_i^{h_t} + \frac{r}{4} \sum_{t=1}^{T-1} \left(b_i^{h_t} + \sum_{j \neq i} a_j^{h_t} \right) - b_i^{h'_T} && (\text{since } b_i^{h'_T} = 0) \\
&\leq 20 - \sum_{t=1}^{T-1} b_i^{h_t} + \frac{r}{4} \sum_{t=1}^{T-1} \left(b_i^{h_t} + \sum_{j \neq i} a_j^{h_t} \right) - b_i^{h'_T} + \frac{r}{4} \left(b_i^{h'_T} + \sum_{j \neq i} a_j^{h'_T} \right) \\
&= u_i(b_i, a_{-i}).
\end{aligned}$$

Likewise, we show inductively that $t \neq T - k$ for all $k = 2, \dots, T - 1$, which contradicts our initial assumption, thus completing the proof. ■

We say that a strategy $a_i \in A_i$ is a best response to $a_{-i} \in A_{-i}$ conditionally on h , and we write $a_i \in BR_i(a_{-i}|h)$, whenever $u_i(a_i, a_{-i}|h) \geq u_i(b_i, a_{-i}|h)$, and respectively $u_i^p(a_i, a_{-i}|h) \geq u_i^p(b_i, a_{-i}|h)$, for all $b_i \in A_i$. The strategy profile a is a subgame perfect equilibrium whenever $a_i \in BR_i(a_{-i}|h)$ for every $i \in I$ and every $h \in H$.

The following result, which follows from the previous proposition, shows that all players contribute 0 at all rounds on the subgame perfect equilibrium path, in both our games, and furthermore in our public good with punishment all player punish equal to 0 at all punishment histories on the subgame perfect equilibrium path.

Corollary 3 *Consider the public good game with growth as defined above.*

- (i) *Let $a \in A$ be a subgame perfect equilibrium in the public good game without punishment, and let (h_1, \dots, h_T) denote the path induced by a . Then, $a_i^{h_t} = 0$ for every $i \in I$ and for all $t = 1, 2, \dots, T$.*
- (ii) *Let $a \in A$ be a subgame perfect equilibrium in the public good game with punishment, and let (h_1, \dots, h_T) denote the path induced by a . Then, $a_i^{h_t} = 0$ for every $i \in I$ and for all $t = 1, 3, \dots, T - 1$. Moreover, $c_i^{h_t} = 0$ for every $i \in I$ and for all $t = 2, 4, \dots, T$.*

It is rather trivial to show that the subgame perfect equilibrium is unique for both our games, implying that from a game-theoretic point of view these predictions are rather strong.

D Additional Analyses Section 4

In this section we provide some additional tables and graphs regarding our main results presented in Section 4.

Share of Endowment contributed Figure D.1 shows the shares of endowments contributed over time in both treatments. The figures shows that in treatment NOPUNISH participants contribute around 55% of their endowment in period 1 (i.e. around 10 tokens). This amount steadily decreases until period 5 and its roughly constant afterwards at a level of about 38% of endowments. (Note that constantly contributing 38% of endowments requires exponentially increasing contributions in terms of tokens.) In treatment PUNISH contributions are roughly constant across all 10 periods at a level of about 60%.

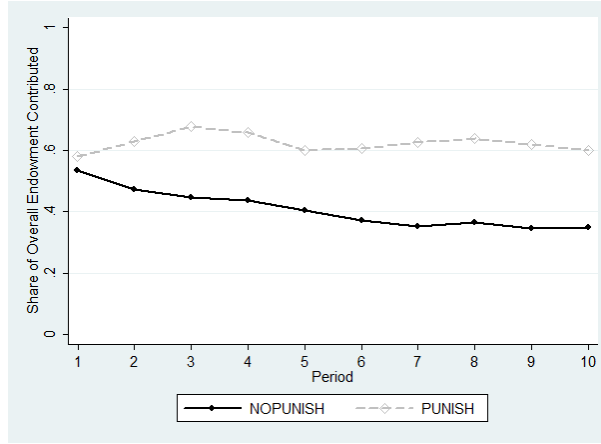


Figure D.1: Share of Endowments contributed (overall amounts contributed divided by overall earnings) in treatments NOPUNISH and PUNISH.

We mentioned in the main text that there are alternative ways to measure shares. As an alternative one could e.g. compute individual shares contributed and then average across these. This method has the following problem. Assume in one group (call it group A) one member has 900 tokens and contributes 600. 3 members have 2 tokens each and contribute nothing. The share contributed is 0.22, while 600 tokens are contributed. In another group (call it group B) all 4 members have 1 token each and two of them contribute it and two don't. The share is 0.5, while only 2 tokens are contributed. We probably wouldn't want to call group B more cooperative than group A, but the share contributed would indicate so. While the share computed in Figure D.1 does not have this problem for *within* treatment analysis (since we sum all contributions and divide them through the sum of all endowments), the across treatment comparison between NOPUNISH and PUNISH does suffer from this problem, albeit (because of the aggregation of many groups) in a mitigated form.

This brings us to a second question, which is which quantity we really should be interested in: the share of endowments contributed or the amount of tokens contributed? We have decided to focus on amounts contributed in the main text and discuss this choice here in more detail. Which measure is more appropriate depends, of course, a lot on the question we are interested in. If we ask which society is wealthier we should prefer absolute amounts rather than shares, as the example of group A and B above illustrates. If we ask which person is more cooperative, the person contributing 1 out of 1 token or the person contributing 600 out of 900, the question becomes less clear. Clearly the person with an endowment of 1 token cannot contribute more than the 1 token he or she has. This endowment is not exogenous, however, and contributing 1 token is essentially costless and riskless for our participant, while contributing 600 tokens (at

the risk of getting only 225 back) does potentially involve a real cost for our participants. The question we are ultimately interested in is at which rate do contributions have to grow to keep pace with growing income such that there is successful and meaningful provision to the public good. It is impossible to answer this question and it should be kept in mind that as contributions grow faster income will grow faster as well.

All these considerations taken together have led us to adopt the absolute amount of tokens as the right measure for contributions in this context. As we have discussed above in the standard setting, since the endowment is constant, the evolution of contributions over time is the same irrespective of whether it is measured in absolute terms or as a share of endowment. Hence all these questions do not arise in the standard setting.

Cumulative Distribution of Earnings Figure D.2 shows the distribution of period 10 individual earnings across treatments NOPUNISH and PUNISH. It can be seen that around 30 percent of individuals earn nothing in the punishment condition, 80 percent are worse off compared to treatment NOPUNISH and the top 20 percent are richer in PUNISH compared to NOPUNISH (even though we have seen above that the latter difference is not statistically significant). This underscores the asymmetry in power that punishment induces in this setting with endogenous growth and inequality.

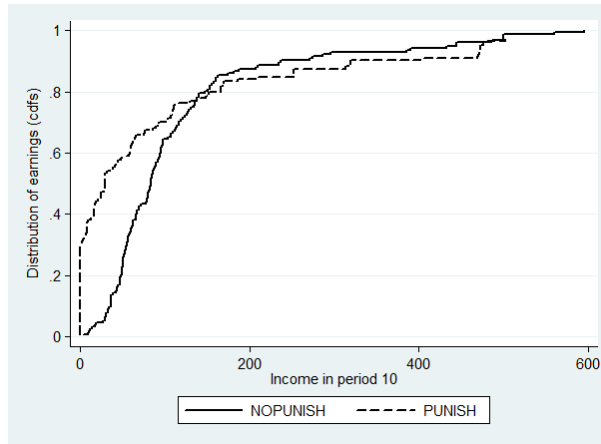
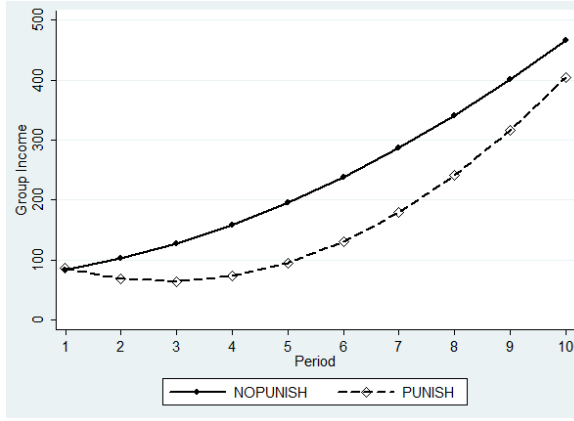


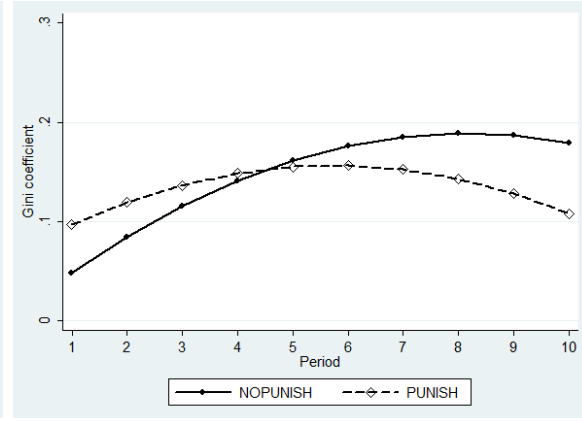
Figure D.2: Distribution of period 10 individual earnings in treatments NOPUNISH and PUNISH.

Regression Graphs on Group Income and Gini coefficient Figures D.3-D.5 correspond to Figures 2 and 3 in the main text with the difference that the Figures here show predicted values from random effects OLS regressions.

Qualitatively there does not seem to be a substantial difference between the data and regression versions of these graphs. Group income is higher in treatment NOPUNISH compared to PUNISH overall and particularly in unsuccessful groups (with below median group income). The dynamics of the Gini coefficient also differ between treatments NOPUNISH and PUNISH with the coefficient being steadily increasing in NOPUNISH and displaying a more cyclical pattern in treatment PUNISH. All these patterns are in line with what we have seen in Section 4.

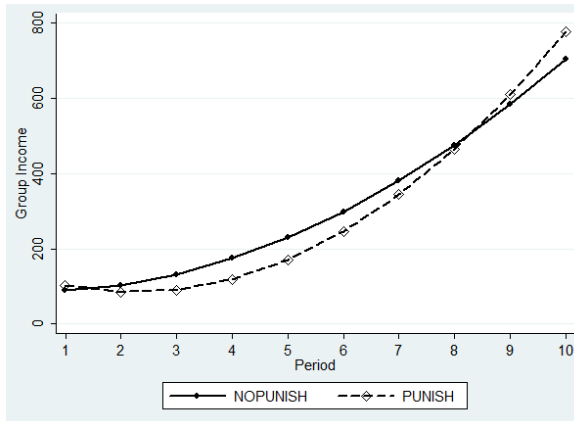


(a) group income

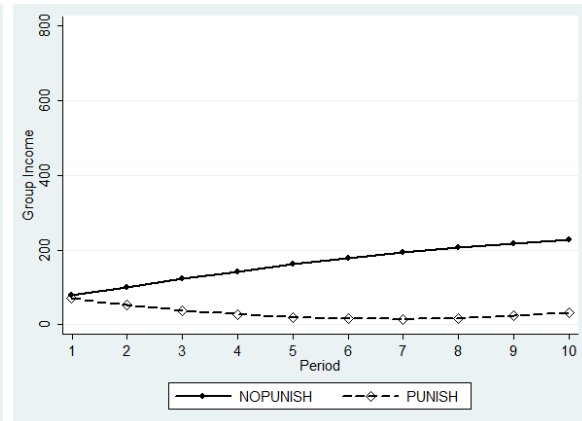


(b) Gini

Figure D.3: Group income and Gini coefficient over time in the two treatments. Predicted values from random effects OLS regressions.

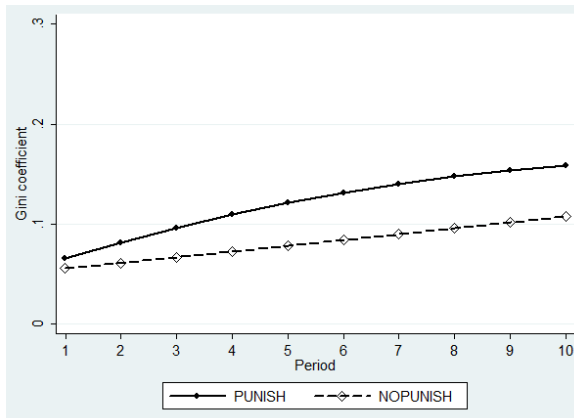


(a) group income above median

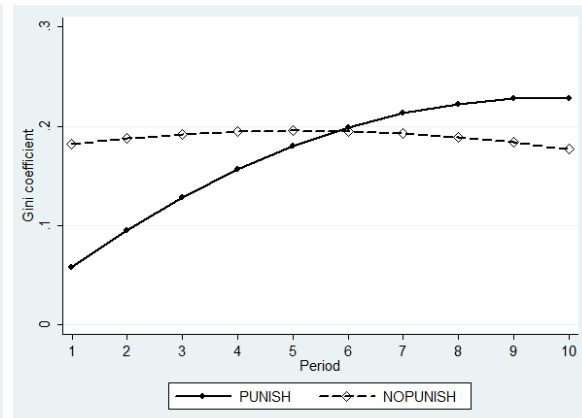


(b) group income below median

Figure D.4: Group income over time in the two treatments in successful (above median group income) and unsuccessful (below median group income) groups. Predicted values from random effects OLS regressions. Note the different scale on the y-axis.



(a) group income above median



(b) group income below median

Figure D.5: Gini over time in the two treatments in successful (above median group income) and unsuccessful (below median group income) groups. Predicted values from random effects OLS regressions.

E Matching Groups

Figures E.1-E.2 show the evolution of group income and Gini coefficient over time for the six poorest and six richest matching groups in each treatment as measured by $t = 10$ group income. Graphs on additional matching groups are available upon request. In treatment NOPUNISH (Figure E.1) the evolution of both indicators is relatively smooth. In treatment PUNISH (Figure E.2) an interesting phenomenon can be observed. In groups where the Gini coefficient rises sharply in early periods (e.g. groups 201 or 208), there is so much punishment that group income ends up being zero.

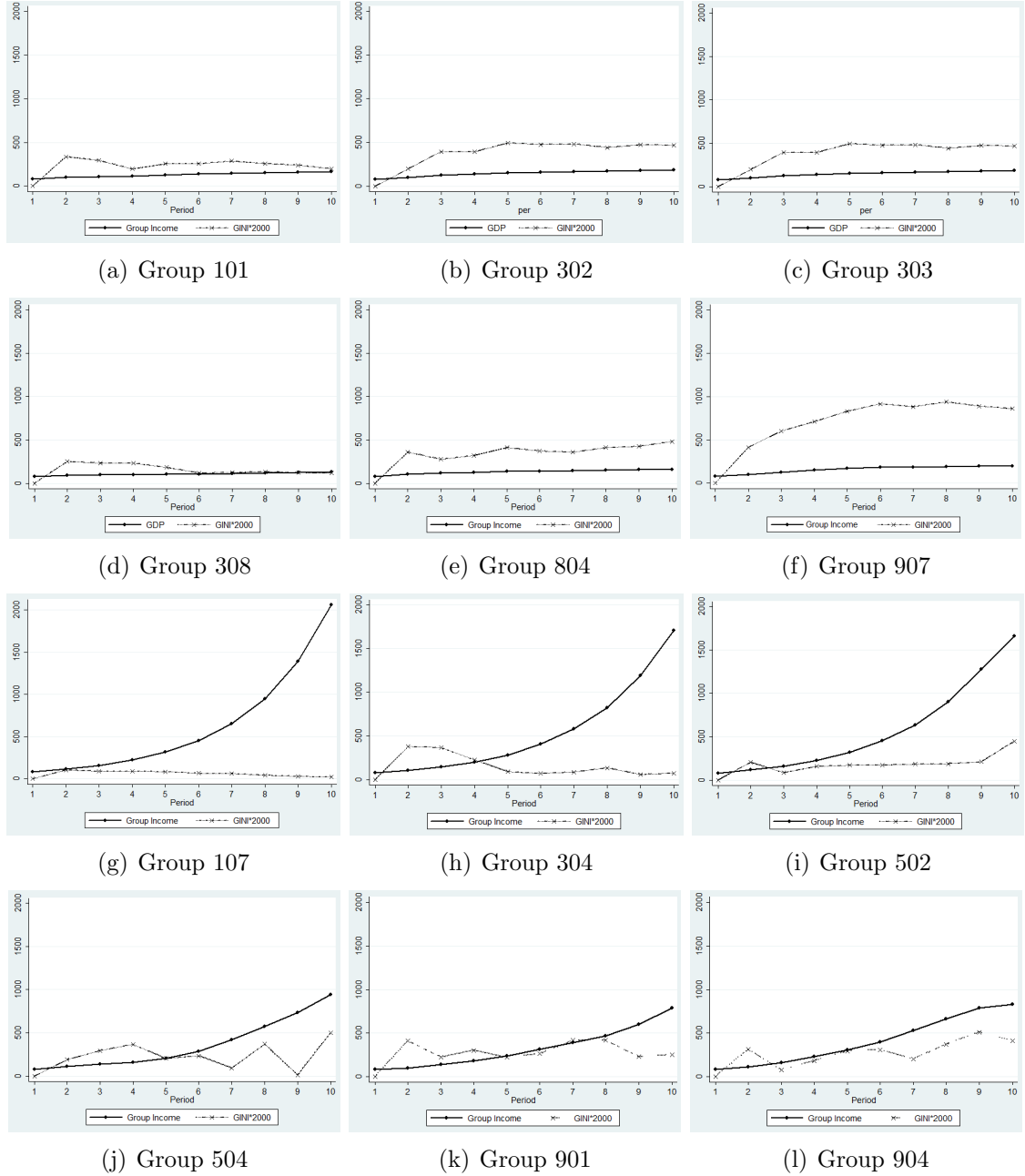


Figure E.1: Group income and Gini coefficient across the six poorest (panels (a)-(f)) and six richest ((g)-(l)) matching groups (as measured by $t = 10$ group income). Treatment NOPUNISH. Gini coefficient is multiplied by 2000 to be on the same scale as group income.

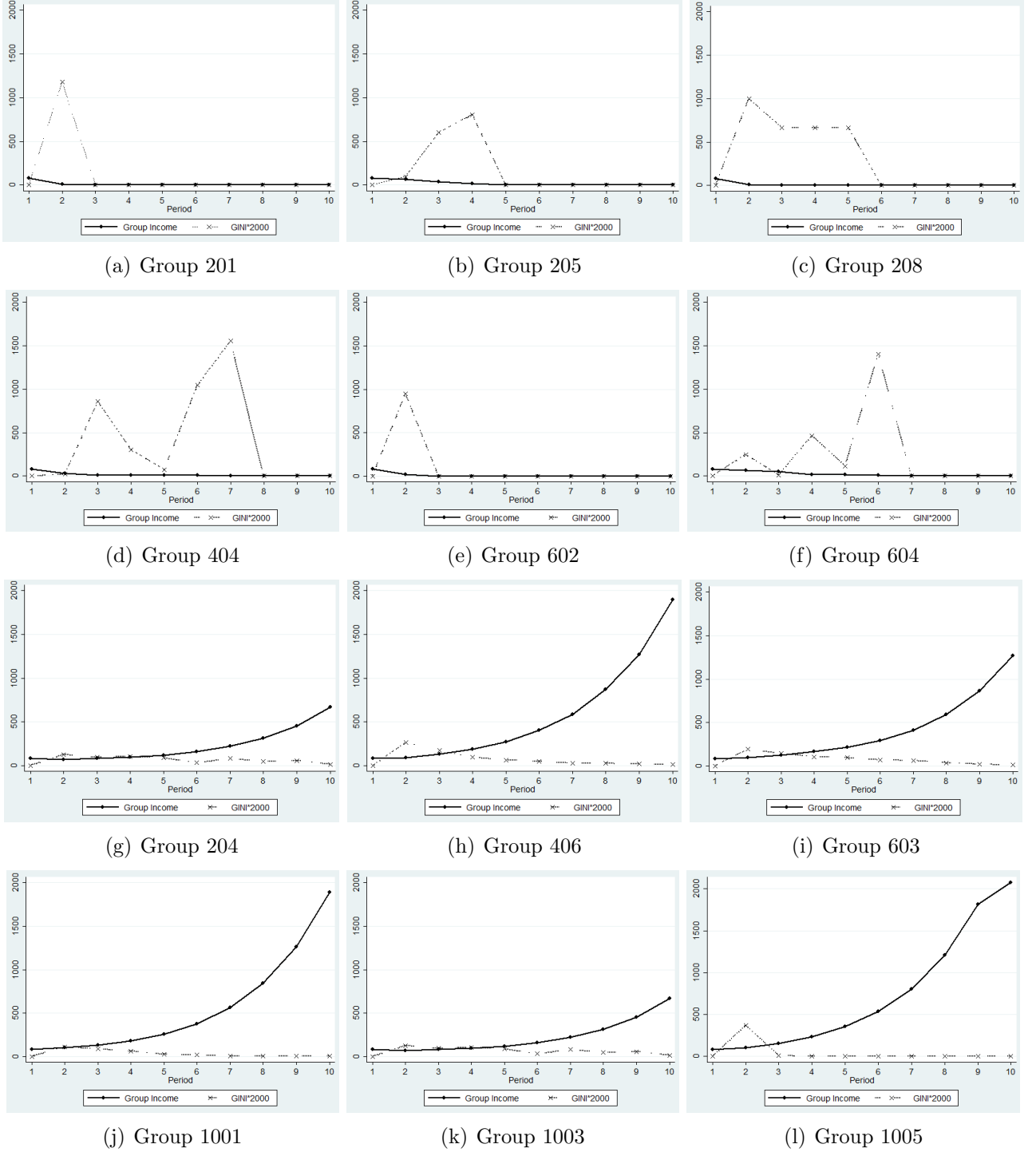


Figure E.2: Group income and Gini coefficient across the six poorest (panels (a)-(f)) and six richest ((g)-(l)) matching groups (as measured by $t = 10$ group income). Treatment PUNISH. Gini coefficient is multiplied by 2000 to be on the same scale as group income.

F Questionnaire Data

In this section we summarize the results from our post-experimental questionnaire. All questions can be found in Online Appendix B. Before we discuss the responses we should mention that due to computer problems in some of the sessions our questionnaire data are incomplete. Those problems were exogenous to session and participant characteristics, so our collected data should be representative. However, the reader should be aware that they don't contain all our partici-

pants. We were able to collect full questionnaires from 124 out of 152 participants in treatment NOPUNISH and 84 out of 144 participants in treatment PUNISH. Table 10 summarizes the key characteristics of our participants. About half of them are female. Also about half are German, but there are also significant percentages of Dutch, Western European (Belgium, Luxemburg, France or UK) and Eastern European participants. They are on average 21.5 years old in both treatments. The youngest participant was 17 and the oldest 35. Around 40 percent of them are business students and almost all others students from other fields (very few non-students). They have spent on average 2 years at university. Our risk aversion measure has full support in our sample and there are no significant treatment differences in the distributions of any of the variables mentioned in Table 10.

	NOPUNISH	PUNISH
Gender (Share female)	0.42	0.52
Share German	0.56	0.43
Share Dutch	0.13	0.10
Share BEL/LUX/FRA/UK	0.11	0.13
Share Eastern Europe	0.10	0.20
Average Age (Range)	21.5(18, 35)	21.5(17, 28)
Share Business	0.41	0.40
Share Economics	0.20	0.12
Share European Studies	0.07	0.12
Share Psychology	0.08	0.05
Years studied (Range)	2.1(0, 10)	2.0(0, 5)
Risk Aversion (Range)	3.39(0, 7)	3.19(0, 7)

Table 10: Summary Statistics Questionnaire Data. Only Nationality Categories and Fields of Study with more than 10 percent answers are mentioned explicitly. The variable risk aversion can take values from 0 to 7, where 0 is most risk averse and 7 least risk averse.

Table 11 summarizes the responses to the personality questionnaire. Again the distribution of answers is very similar across treatments.

	NOPUNISH	PUNISH
Q1 I am a quick thinker	5.18	5.40
Q2 I get easily offended	3.58	3.66
Q3 very satisfied	5.07	5.16
Q4 very dependent	2.67	2.75
Q5 generally happy	5.71	5.75
Q6 work important	4.77	4.89
Q7 family important	5.67	6.03
Q8 friends important	6.01	6.05
Q9 religion important	2.47	2.26
Q10 politics important	3.65	3.60
Q11 most people trusted	3.72	3.88
Q12 hard work better	5.48	5.44
Q13 government responsible	4.29	4.45
Q14 incomes equal	3.78	3.76

Table 11: Summary Statistics Questionnaire Data, Mean Reply to Personality Characteristics Questions of the form “How strongly do you agree to the following statements?” 1 - disagree strongly, 7 - agree strongly. The exact statements can be found in Online Appendix B.

We then regress our measures of growth (group income) and inequality (Gini) in the two treatments on the questionnaire data. We use simple OLS regressions of group income and Gini in period 10 on individual questionnaire data and we cluster standard errors by matching group. Table 13 shows the results for treatment NOPUNISH. Overall our questionnaire measures

have a hard time to explain the variation in group income and Gini and almost all of them are insignificant. There might be somewhat of a gender effect in treatment NOPUNISH. In particular group income seems to be lower in groups with more women. Strong agreement to the statement “Friends play an important role in my life” seems to predict somewhat higher group income in treatment NOPUNISH. Both of these results should be interpreted with care, though, since we regress on quite a large set of variables. The overall message seems to be that our questionnaire data *cannot* explain the variation in group income and Gini coefficient.

	NOPUNISH	PUNISH
above median group income	2.31 (2.18)	1.70 (1.42)
below median group income	2.24 (1.53)	1.46 (3.17)

Table 12: Average Donation (Std. Dev.) in Euros to Medics without Borders.

Table 14 shows the results of the analogous regression for treatment PUNISH. Here the result is even clearer. None of the variables seems systematically able to explain any of the variation in group income or Gini observed in this treatment. There is a significant coefficient on risk aversion, indicating that higher risk aversion of group members might lead to higher group income in these treatments. This effect would be intuitive if risk averse participants react more strongly to the threat of punishment, but it disappears once we stop controlling for the personality characteristics.

Finally we have a look at how much our participants decide to donate to Medics without Borders. Table 12 shows the average donation in Euros to medics without Borders. Participants in treatment PUNISH seem to donate somewhat less than participants in treatment NOPUNISH. We compare the distribution of donations using a two-sided ranksum test where we treat each individual donation as an independent observation. The two treatments are significantly different ($p = 0.0432$) on aggregate and if we restrict to below median groups ($p = 0.0134$), but not restricted to above median groups ($p = 0.4711$).

More interestingly, though, participants from groups with group income above the median do *not* seem to contribute more on average than those from groups with below median group income. There is no significant difference in treatment NOPUNISH ($p = 0.9195$) and a marginally significant difference in treatment PUNISH ($p = 0.0506$).

This is despite the fact that participants from groups with above median group income earn 178 tokens on average in period 10 (189 in treatment PUNISH), while those from groups with below median group income earn only 56 tokens (23 tokens) on average in period 10. This evidence suggests hence that participants in groups with above median group income are *not* per se more altruistic than others.

	(group income)	(group income)	(Gini)	(Gini)
gender	−147.55** (70.58)	−151.23** (58.84)	−0.04* (0.02)	−0.02 (0.02)
age	−21.46 (21.89)	−17.36 (19.57)	−0.00 (0.00)	−0.00 (0.00)
risk aversion	−38.45 (26.31)	−35.06 (23.62)	0.00 (0.01)	0.00 (0.01)
Q1	−29.48 (19.14)		−0.00 (0.01)	
Q2	−2.33 (15.53)		0.00 (0.00)	
Q3	−8.93 (42.65)		−0.00 (0.01)	
Q4	−10.22 (24.77)		0.00 (0.00)	
Q5	28.45 (41.43)		0.02 (0.01)	
Q6	−6.60 (26.88)		0.00 (0.00)	
Q7	−30.41 (25.03)		−0.00 (0.00)	
Q8	96.04** (38.20)		−0.00 (0.01)	
Q9	−15.25 (25.27)		0.00 (0.00)	
Q10	46.12 (28.45)		−0.02* (0.01)	
Q11	11.35 (16.87)		0.00 (0.00)	
Q12	3.88 (24.16)		−0.00 (0.00)	
Q13	−60.60 (42.82)		0.00 (0.01)	
Q14	8.16 (12.98)		−0.00 (0.00)	
constant	43100.80 (43656.74)	35194.95 (39033.02)	3.91 (10.07)	1.02 (8.02)
Observations	124	124	124	124
Groups	31	31	31	31
R^2	0.1387	0.0607	0.0780	0.0110
VCE robust S.E.	Yes	Yes	Yes	Yes

Table 13: OLS regression of period 10 group income and Gini coefficient on questionnaire characteristics. Treatment NOPUNISH

	(group income)	(group income)	(Gini)	(Gini)
gender	-40.17 (169.29)	-47.60 (204.07)	-0.04 (0.04)	-0.03 (0.05)
age	-7.17 (26.43)	0.10 (25.60)	0.01 (0.01)	0.01 (0.02)
risk aversion	124.04** (58.87)	73.43 (47.51)	0.00 (0.02)	0.00 (0.02)
Q1	-2.91 (62.92)		0.01 (0.01)	
Q2	-40.90 (28.99)		0.02 (0.02)	
Q3	-30.15 (50.43)		-0.00 (0.01)	
Q4	-52.27 (102.48)		-0.02 (0.02)	
Q5	-67.12 (68.65)		-0.01 (0.02)	
Q6	117.22 (75.48)		-0.00 (0.01)	
Q7	5.52 (81.70)		0.02 (0.02)	
Q8	-53.67 (61.58)		-0.00 (0.01)	
Q9	74.55 (55.88)		0.01 (0.01)	
Q10	-14.53 (40.04)		-0.01 (0.01)	
Q11	34.14 (41.89)		0.00 (0.02)	
Q12	-53.44 (107.01)		0.00 (0.02)	
Q13	-41.32 (34.60)		0.01 (0.01)	
Q14	3.71 (81.51)		-0.02 (0.02)	
constant	15173.64 (52764.84)	67.15 (50911.71)	-20.02 (35.82)	-28.47 (40.83)
Observations	84	84	84	84
Groups	21	21	21	21
R^2	0.1430	0.0243	0.1213	0.0352
VCE robust S.E.	Yes	Yes	Yes	Yes

Table 14: OLS regression of period 10 group income and Gini coefficient on questionnaire characteristics. Treatment PUNISH

G Additional Analysis Section 5

G.1 Contribution Behavior

We collect additional evidence on contribution behaviour.

Dynamics of Contributions. We start by analyzing the dynamics of individual contributions to the public good over time. Figure G.1 shows the average amount of tokens that participants contribute to the public good over time across the two treatments. Figure 1(a) is the same as Figure 1, which we already discussed above, with the only difference that the y -axis ranges now from 0 to 200 to enable a better comparison of levels across the three subfigures. Figures 1(b) and 1(c) show the amount of tokens contributed by above median (1(b)) and below median groups (1(c)), respectively.

The Figure shows that participants tend to contribute more in treatment PUNISH compared to NOPUNISH especially in later periods. In below median groups participants contribute very little in both treatments and do not increase (nor decrease) their contributions over time. Contributions are increasing in both treatments in above median groups, though, and on average with a steeper slope in treatment PUNISH. There is no evidence of an “endgame effect.”¹⁷ This presents a significant difference from the standard results, where contributions tend to zero over time. In our setting, on the other hand, participants (in successful groups and on average) contribute an increasing amount over time. Hence, participants seem to react in a very different way to this environment compared to the standard repeated public good game.

Note that, since group income is allowed to grow over time in our setting, contributions are not necessarily increasing as a share of endowment. If we look at shares of endowments contributed they are about constant on average and range between 50-60 % in PUNISH. They range between 20-50 % in NOPUNISH. In above median groups they range between 60-80% of group income in PUNISH and between 30-60 % in NOPUNISH. In below median groups they range between 30-50 % in PUNISH and 10-40 % in NOPUNISH.

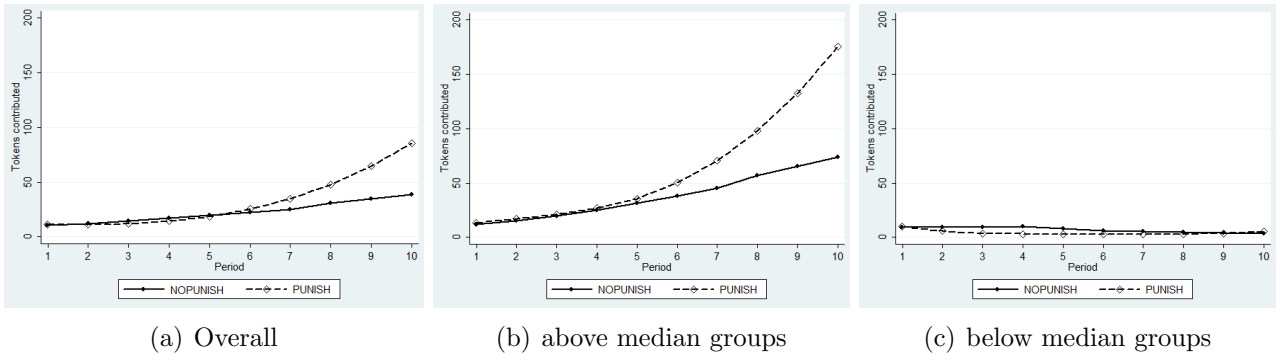


Figure G.1: Tokens contributed over time.

Treatment Differences. Table 15 shows the results of random effects OLS regressions where we regress individual contributions on a treatment dummy ($\delta_{PUNISH} = 1$ for treatment PUNISH) and a constant. Column (1) shows the regression for the entire sample (periods < 11). In

¹⁷Remember, though, that we have aggregated data from the 10 and 15 period games in this session. Hence part of the latter observation may stem simply from the fact that there is no drop in contributions in the long (15 period) games. Figure H.4 in the Online Appendix corresponds to Figure 1(a), but separate the short (10 period) and long (15 period) horizon games. There is no endgame effect in treatment PUNISH and none in the short horizon games in treatment NOPUNISH. However, there seems to be an endgame effect in the long horizon games.

	(1)	(2)	(3)	(4)	(5)	(6)
δ_{PUNISH}	9.88 (9.66)	22.19 (16.16)	-2.42** (1.09)	0.16 (12.82)	10.66 (10.09)	15.11 (13.99)
constant	22.59*** (5.02)	38.27*** (8.69)	6.91*** (0.56)	19.19*** (4.53)	23.91*** (5.56)	10.71*** (0.42)
Observations	2960	1480	1480	595	2428	532
Groups	74	37	37	45	74	74
ρ	0.359	0.256	0.268	0.468	0.465	0.12
group income	all	above median	below median	any	any	any
Gini coefficient	all	any	any	> 0.2	> 0	= 0
VCE robust S.E.	Yes	Yes	Yes	Yes	Yes	Yes

Table 15: Individual contributions regressed on treatment dummy ($\delta_{PUNISH} = 1$ for treatment PUNISH). Samples split by median group income and by size of Gini coefficient. Random Effects OLS regression.

Columns (2) and (3) we have split the sample according to the $t = 10$ median group income's as they can be found in Table 2. Column (2) considers only groups above and column (3) below the median group income. If we first look at the coefficient on the constant, we see that in treatment NOPUNISH contributions are about 38 tokens for groups above the median and about 7 tokens for groups below the median. Above median groups hence contribute more than 5 times more than below median groups on average. The difference is highly statistically significant. Again, in stark contrast to the standard setting, people do *not* contribute more in PUNISH compared to NOPUNISH on average. In below median groups they even contribute somewhat less (column (3)).

In Columns (4)-(6) we restrict the sample to three different levels of the Gini coefficient (above 0.2 in column (4), strictly above 0 in column (5) and exactly zero in column (6).) We do not split the sample by median group income, since the median Gini coefficient in treatment PUNISH is only 0.03 and one would find mostly groups with zero group income below that median. In treatment NOPUNISH comparison of columns (4), (5) and (6) shows that contributions are higher the lower the level of inequality. There are no significant treatment differences. In Online Appendix G.1 we study some of the determinants of contributions. Regressing contributions on the level of inequality, group income or similar presents severe endogeneity issues, which we address in detail in the Online Appendix. We find that participants' contributions increase if others contributed more in the previous period, indicating the presence of reciprocal behaviour, and decrease with increases in inequality. Overall, the results confirm earlier impressions that punishment is not needed to sustain positive contributions in this setting. Contributions are *not* higher in PUNISH compared to NOPUNISH. In the next section we study punishment behaviour.

G.2 Punishment Behaviour

In this section we collect additional evidence on punishment. Table 16 shows the correlation coefficients between how much participants punish and four other variables: how much they were punished in the past (reciprocal punishment), how much others contributed in shares of their income and tokens (social vs antisocial punishment) and the Gini coefficient in the previous period. We consider the mean across all ten periods for each individual, hence using individuals as independent observations. Individuals that were punished more do punish more in the next period (*reciprocal* punishment). This effect is sizeable and significant. Particularly in below median groups such reciprocal punishment seems to play an important role. Participants also punish less if others contributed more in the previous period and more the higher the level of inequality.

	share punished	share contr. others	tokens contr. others	Gini $t - 1$
entire sample	0.7358***	-0.5019***	-0.7328***	0.3716***
above median group income	0.0531	0.0476	-0.1392	0.1497
below median group income	0.8057***	-0.5818***	-0.6019***	-0.1375

Table 16: Spearman correlation coefficients between the mean share of income used for punishment in $t=1,..10$ and mean of lagged variables. “Share punished” indicates the share of endowment a participant lost in the previous period due to punishment. “Share contr. others” indicates the share of their endowment others contributed in the past period. “Tokens contr. others” the amount of tokens contributed by others in the previous period and “Gini $t-1$ ” the Gini coefficient from period $t - 1$.

Table 17 shows how much social and anti-social punishment there is in the data depending on group income, the size of the Gini coefficient and the rank of a participant within a group.¹⁸

Asymmetric Punishment. Is there evidence that richer group members (typically shirkers) tend to coerce poorer participants into contributing? The punishment can be implemented by spending a small share of their (large) endowments on substantially punishing poorer participants. Similarly, we can ask whether poorer participants tend to ruin themselves, spending a large share of their (small) endowments to inflict relatively little punishment on the richest participants. To these ends we define the following ratio $ASY = \frac{p_i^j/N_i}{3*p_i^j/N_j}$, where $\frac{p_i^j}{N_i}$ is the share of endowment that i spends on punishing j and $\frac{3*p_i^j}{N_j}$ the share of endowment j loses due to being punished by i . Obviously if $p_i^j > 0$, then this reduces to $\frac{N_j}{3*N_i}$ which is exactly equal to $\frac{1}{3}$ if $N_i = N_j$. The latter denotes the case of symmetric punishment. If $ASY > \frac{1}{3}$, then a poor punisher is losing a larger share of their endowment than the punished person, while if $ASY < \frac{1}{3}$, then a relatively rich punisher loses a smaller share of their endowment than the punished person. We set ASY to zero if $p_i^j = 0$.

Figure G.2 shows the extent of asymmetric punishment over time (Figure 2(a)) and as a function of the Gini coefficient (Figure 2(b)). The first thing to note is that punishment is often asymmetric and sometimes participants are spending a large share of their endowment to reduce the opponent’s endowment by just a bit. The other extreme is also very common, where a punisher inflicts a large reduction in endowment on somebody else at a very low cost to themselves. Figure 2(a) illustrates that there are no significant time trends. The variable ASY

¹⁸If the sum of person i ’s punishment plans exceeds her income, we only sum her plans for the first two other players (as labeled in the software). If that sum still exceeds her income we only use her punishment plan for the first player. If that also exceeds her income we set her share of punishment equal to 1. This is also how the software implemented unfeasible punishment plans.

	Social Punishment	Antisocial Punishment
all	1.57 (14.39)	0.69 (3.44)
Gini> 0.05	1.18 (2.29)	1.28 (10.72)
Gini< 0.05	1.99 (20.43)	0.58 (2.23)
group income> 80	2.40 (21.19)	0.61 (4.62)
group income< 80	0.52 (1.72)	0.48 (1.87)
periods 1-5	1.37 (3.31)	0.91 (2.27)
periods 6-10	2.02 (22.41)	0.48 (4.38)
Gini> 0.05 + per 1-5	1.26 (3.21)	0.20 (0.73)
Gini< 0.05 + per 1-5	1.53 (3.43)	1.08 (3.22)
group income> 80 + per 1-5	1.35 (3.51)	0.51 (1.86)
group income< 80 + per 1-5	0.91 (2.35)	0.77 (2.31)
Gini> 0.05 + per 6-10	1.06 (2.60)	2.16 (14.42)
Gini< 0.05 + per 6-10	2.36 (27.13)	0.19 (1.06)
group income> 80 + per 6-10	3.13 (27.45)	0.68 (5.33)
group income< 80 + per 6-10	0.16 (0.62)	0.22 (1.29)
asocial=0	1.54 (15.30)	-
asocial> 0	1.82 (3.42)	-
social=0	-	0.66 (3.76)
social> 0	-	0.79 (2.23)
poorest	1.58 (3.03)	0.47 (1.96)
not poorest	1.57 (15.39)	0.72 (3.61)
richest	1.06 (2.54)	0.89 (2.53)
not richest	1.66 (15.52)	0.66 (3.57)

Table 17: Mean (SD) of amount of tokens of social and anti-social punishment. Periods 1-10.

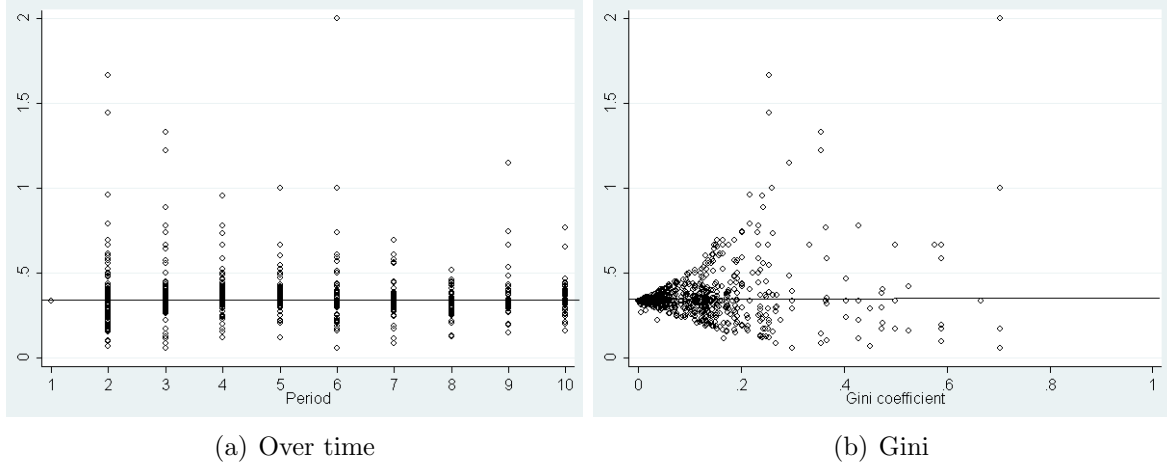


Figure G.2: Asymmetric Punishment. The horizontal line indicates the case of symmetry (where $ASY = \frac{1}{3}$).

is negatively correlated with inequality (Spearman $\rho = -0.2638^{***}$), meaning that in unequal groups we observe more often punishment where a “rich” punisher punishes a poorer group member at relatively low cost to themselves. It seems that in these groups some shirkers coerce contributors into contributing large shares of their endowment by punishing them harshly at low costs to themselves. This is illustrated in Figure 2(b).

H Robustness 15 period games

In this section we revisit our previous results from Section 4, but focus uniquely on the 15 period games. In this way we can check whether the results are robust to considering the longer time frame. Figure H.1 compares the evolution of the group income and Gini coefficients over time across the two treatments. It is the analogue to Figure 2, which focuses on the first ten periods of both the longer (15-period) and shorter (10-period) games. The two figures look remarkably similar and there is no detectable change in trend after period 10.

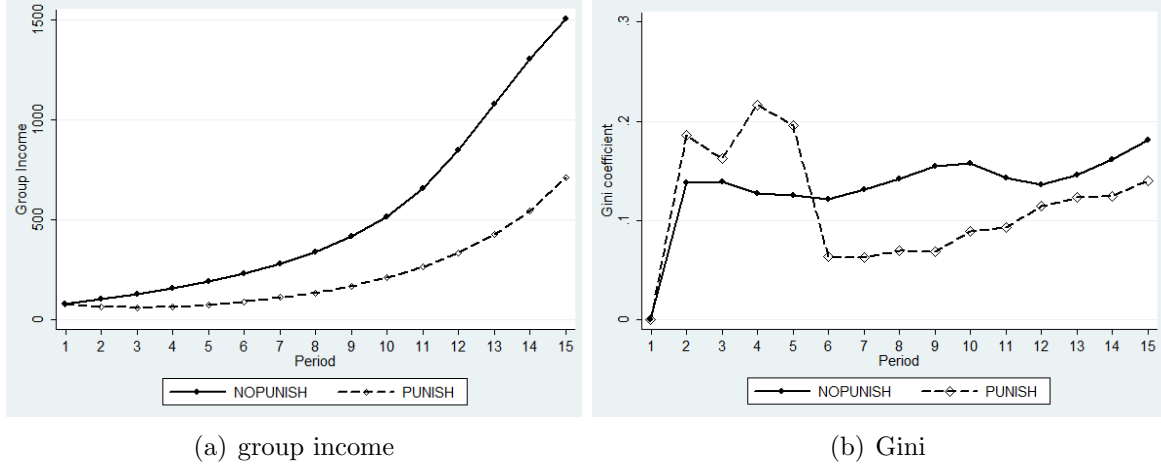


Figure H.1: Average group income and Gini coefficient over time in the two treatments. Only 15 period games are considered

We then split the sample by median group income. In treatment NOPUNISH the median group income in period 15 is 422 and in treatment PUNISH it is 248. Figure H.2 corresponds to the first two graphs in figure 3 in the main text and compares the group income over time in the two treatments separately for successful (above median group income) and unsuccessful (below median group income) groups. The two figures H.2 and 3 are remarkably similar. There are “big” differences in group income between the two treatments for below median groups and smaller differences for above median groups. The latter difference seems to be somewhat clearer in the long (15 period) games. Average group income in treatment PUNISH is always smaller than in treatment NOPUNISH.

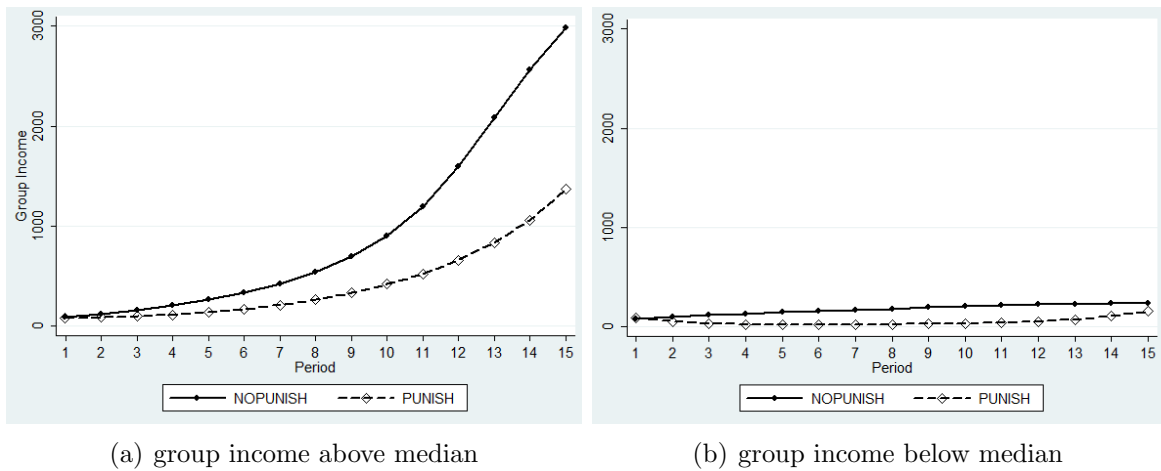


Figure H.2: group income over time in the two treatments - sample split by median group income in period 15. Only 15 period games are considered

Figure H.3 corresponds to the last two graphs in figure 3 and compares the group income over time in the two treatments separately for successful (above median group income) and unsuccessful (below median group income) groups. Again the two figures are pretty similar, though the treatment differences are maybe not quite as pronounced over the longer horizon.

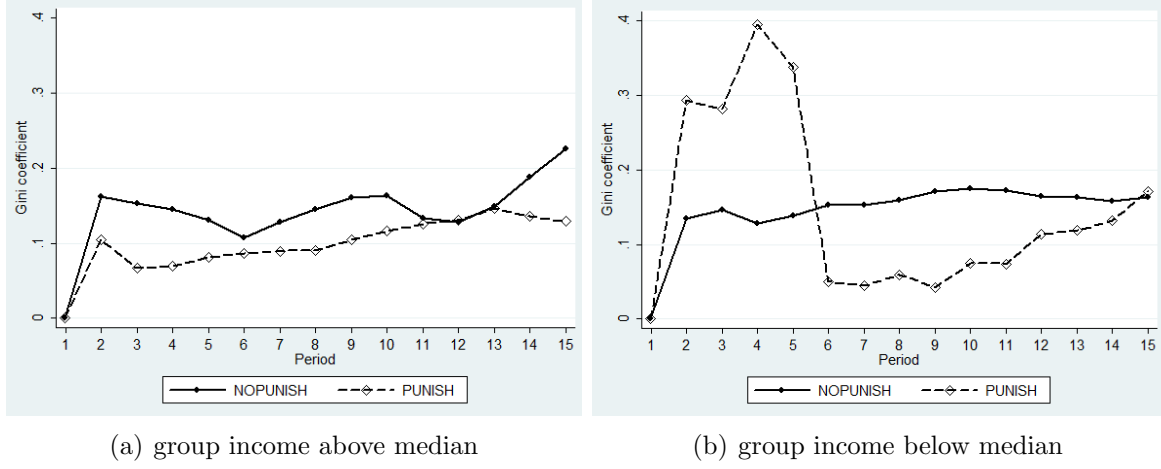


Figure H.3: Gini over time in the two treatments - sample split by median group income in period 15. Only 15 period games are considered

Figure H.4 studies contributions over time in the short and long horizon games. In the short horizon games the picture is very similar to that we have seen in Figure 1. Contributions are non-decreasing over time. Starting from about period 6, they seem to be higher in treatment PUNISH compared to NOPUNISH, but the average contributions are not statistically different. There are no endgame effects. In the long horizon games the picture looks somewhat different. Contributions here are even higher in NOPUNISH compared to PUNISH. Interestingly, also, in the long horizon game there seems to be an endgame effect in treatment NOPUNISH. Contributions start to be constant, rather than increasing, starting from about period 12 and are sharply decreasing in the last period. They are still significantly above zero, though, even in period 15.

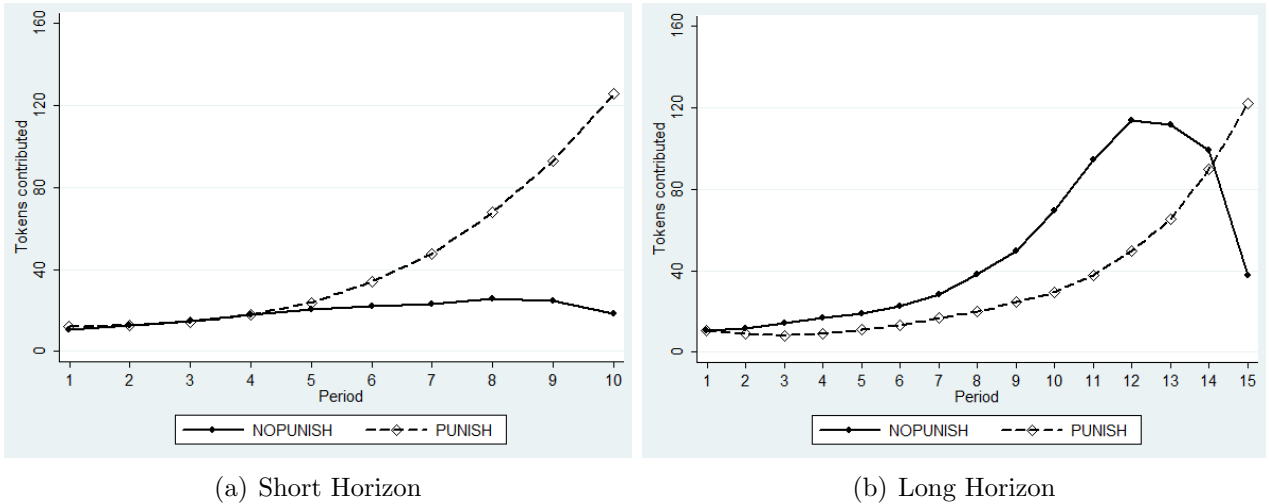


Figure H.4: Tokens contributed over time. Games with short (10 period) horizon vs games with long (15 period) horizon.