

Growth and Inequality in Public Good Provision ^{*}

Simon Gächter[†]

University of Nottingham,
CESifo and IZA

Friederike Mengel[‡]

University of Essex

Elias Tsakas[§]

Maastricht University

Alexander Vostroknutov[¶]

University of Trento

February 24, 2017

Abstract

In a novel experimental design we study public good games with dynamic interdependencies, where each agent's wealth at the end of period t serves as her endowment in $t+1$. In this setting growth and inequality arise endogenously allowing us to address new questions regarding their interplay and effect on cooperation. We find that amounts contributed are increasing over time even in the absence of punishment possibilities. Variation in wealth is substantial with the richest groups earning more than ten times what the poorest groups earn. Introducing the possibility of punishment does not increase wealth and in some cases even decreases it. In the presence of a punishment option inequality in early periods is strongly negatively correlated with group income in later periods, highlighting negative interaction effects between endogenous inequality and punishment.

^{*}We thank Ingvild Almas, Benjamin Beranek, Tim Cason, Martin Dufwenberg, Tore Ellingsen, Drew Fudenberg, Santiago Oliveros, Ernesto Reuben and Marie-Claire Villeval as well as two anonymous Reviewers and participants in seminars and conferences at Essex, Maastricht, Lund (GIRLS 2014) and Toulouse (EEA ESEM 2014) for helpful comments. Friederike Mengel thanks the European Union (grant PERG08-GA-2010-277026) for financial support. Simon Gächter gratefully acknowledges support from the European Research Council grant ERC-AdG 295707 COOPERATION and the ESRC network for Integrated Behavioural Science (NIBS, ES/K002201/1).

[†]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

[¶]Center for Mind/Brain Sciences, University of Trento, Via delle Regole 101, 38123 Mattarello, Italy.

1 Introduction

1.1 Motivation

Social dilemmas, where collective and private interests are in conflict, abound in economic and social life. Public good games have been used across disciplines as the standard tool 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 that can be modeled with 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 interact 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 that captures such dynamic interdependencies. Our design builds on what has become the workhorse model to study public good provision in experiments (see e.g. Isaac et al. (1984); Andreoni (1995) or Fischbacher and Gächter (2010) among many others): participants are matched in fixed groups of four people to play the public good game for 10 or 15 periods. As in most other public good experiments we focus on the most challenging social dilemma situations, where the unique subgame perfect Nash equilibrium prescribes zero contributions by all group members, but where efficiency requires group members to contribute their entire endowment. We also conduct experiments where each group member can punish other group members by reducing their first stage earnings at a cost to themselves (Ostrom et al., 1992; Fehr and Gächter, 2000; Andreoni et al., 2003).

The key difference to previous research using this “standard” design is that each participant’s wealth at the end of a period constitutes their endowment for the next period, whereas in the “standard design” endowments are allocated exogenously and tend to be the same in each round.¹ In our design, endowments are created endogenously, which leads to dynamic interdependencies.

We focus on two important implications of introducing these interdependencies. First, if overall 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. Growth and inequality can interact with the possibility of punishment in different ways. The threat of punishment can lead to higher growth if it induces higher contributions, but punishment 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. In particular, rich group members can be largely “immune” to punishment by poorer group members, if the punishment that poor group members can afford is too small relative to the richest group member’s wealth. As all endowment can be used for punishment, rich group members, on the other hand, might be able to punish others harshly at a relatively low cost to themselves. This asymmetry of punishment possibilities translates wealth inequality into inequality in power to punish. In addition, in unequal groups richer group members will typically be free-riders implying that punishment power could be in the “wrong hands”. This raises the question of whether punishment will be as effective in increasing contributions and group income as it has been in settings without these dynamic interdependencies.

Our main findings can be summarized as follows. Even in the absence of punishment contri-

¹Throughout the paper we will use the expression “standard design” to 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.

bution and wealth levels display a strictly increasing trend over time. 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. The Gini coefficient assumes the full range between 0 (equal wealth of all group members) and 1 (one group member appropriates the entire wealth) in the experiment.

Punishment (or the possibility thereof) does *not* increase wealth. This is true in both the 10 and 15-period variations despite the fact that people tend to contribute more in the treatment with punishment in the 10-period variations. We find evidence for two mechanisms behind this result: (i) in groups where inequality is high (above median) there is more anti-social than pro-social punishment, i.e. shirkers punish contributors more than vice versa and (ii) much of this punishment happens in early periods implying that resources are taken away exponentially.²

While the possibility of punishment does not increase wealth and in some cases even strictly decreases it, it also does not increase inequality on average. This is true despite the inequality-increasing presence of anti-social punishment. Analysis of data from Herrmann et al. (2008) shows that, in a comparable standard setting, punishment increases both wealth and inequality. In terms of the relationship between inequality and growth we find that inequality in period 2 is strongly negatively correlated with wealth in period 10 in the treatment with punishment possibilities. In particular, a 1% increase in inequality in period 2 leads to a $\approx 0.5\%$ decrease in wealth in period 10. Inequality and growth are positively related across groups with below median wealth and negatively related across groups with above median wealth.

In interpreting these results, it is important to note how the setting we introduce differs from the standard public good game described above. There are two main differences: (i) there is no consumption until the last round in our setting, i.e. one's entire wealth can be reinvested at the end of the period and (ii) endowments are endogenous, i.e. determined by previous outcomes. R+D cooperation often displays these features, but also the evolution of societies could be viewed under this lens. In the standard setting, by contrast, there is full consumption, i.e. no wealth can be reinvested and endowments are exogenous (and stationary). Volunteering, e.g. at a food bank, seems a good example falling into this category. Other types of volunteering, such as in natural conservation or archiving, are examples involving stationary exogenous endowments, but no or little consumption. The case of full consumption with endogenous endowments describes the one-shot game. Finally, note that many applications, such as infrastructure investments, multilateral foreign aid or pricing agreements will fall somewhere in between these extremes with some, but not full consumption and with partially endogenous endowments.

1.2 Literature Review

To our knowledge our experiment is the first to study public good provision with dynamic interdependencies and endogenously arising asymmetric punishment possibilities.³ As such it contributes to studies of public goods with dynamic interdependencies more broadly. Those include Battaglini et al. (2014) who study the Markov perfect equilibrium dynamics in the provision of a durable public good over time where there is consumption in each period. They find evidence of significant under-provision relative to the interior equilibrium. Duffy et al. (2007) studied

²History books report many examples where “shirking group members” have assumed power by exploiting asymmetric punishment opportunities. Standard Oil reportedly sent out thugs to raid the premises of competitors as a form of punishment (Josephson, 1962). Adler (1985) discusses endogenously arising asymmetries in punishment possibilities in a study of upper-level drug-dealing and Johnson and Earle (1987) among North-American Indians.

³In an unpublished Master thesis Huck (2006) has conducted a dynamic public good game without the possibility of punishment and related contributions to personality characteristics elicited in a questionnaire. He also did not analyze growth and inequality but, like us, he finds a pattern of increasing contributions (in terms of absolute amounts) and no endgame effect.

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). Noussair and Soo (2008) and Cadigan et al. (2011) study dynamic public good settings where the current return from the public good depends on past contributions. Other studies link public good games over time via explicit reputation mechanisms (e.g. Milinski et al., 2002). Guererck et al. (2017) have recently studied a setting where games are linked via endowments. They use a non-linear exchange rate, though, which effectively eliminates the possibility of exponential growth and contains inequality. Consequently their results are more similar to those obtained in the standard setting.

Our results also contribute to research on the impact of inequality on public good provision. Most existing literature has studied the effects of *exogenous* income inequality. Chan et al. (1996) experimentally test a prediction by Bergstrom et al. (1986), where public good provision increases with inequality in the income distribution in an equilibrium with positive contributions. They find that group behavior conforms with the theoretical prediction. Other authors have found that exogenous income inequality decreases contributions (van Dijk et al., 2002; Ostrom et al., 1994) or found no effect (Chan et al., 1999). Reuben and Riedl (2013) 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. Sadrieh and Verbon (2006) study exogenous inequality in the Bergstrom et al. (1986) setting with the possibility of growth. They find that exogenous variations of inequality are mostly neutral to growth.⁴ The latter result contrasts with our finding that endogenous inequality is negatively related to growth in the presence of punishment. One possible reason for the difference between these results is that, as discussed above, with endogenous inequality the power to punish tends to lie “in the wrong hands” (those of free-riders).

Our paper also contributes to the literature on the standard setting of repeated public good games surveyed in Chaudhury (2011). One insight emerging from this literature is that societies can maintain high levels of contributions and wealth if there is a threat of punishment to free-riders (Ostrom et al., 1992; Fehr and Gächter, 2000, 2002; Andreoni et al., 2003).⁵ While some studies have found that punishment can lead to lower payoffs if the horizon of the game is short (Egas and Riedl, 2008; Dreber et al., 2008), under a long horizon punishment in the standard setting has been found to be strictly beneficial (Gächter et al., 2008). Punishment does not increase wealth in our setting neither under a 10-period nor a 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. Some studies of the standard setting investigate how an increased marginal per capita rate of return affects contributions (Isaac and Walker, 1988; Goeree et al., 2002). There is some relation of these studies to our setting, since the possibilities of exponential growth provides increased incentives to contribute.

Finally, the dynamic setting we study also relates to other dynamic games studied in economics. In the common pool resource game (CPR, Ostrom et al. (1994)) players extract from a resource in each period with the non-extracted part growing at a fixed rate. There are various differences between this setting and our dynamic public good game (apart from the frame). One is that whatever is contributed in our public good game is mechanically shared among all participants. How the benefits of non-extraction are shared depends on players’ strategies. On the

⁴In a trust game with limited growth Greiner et al. (2011), however, find that both exogenous and endogenous variation in income affect growth.

⁵Cason and Gangadharan (2015) note, though, that the impact of punishment is weaker in nonlinear social dilemmas.

other hand what is withheld in our public good game in one period can be contributed in the next. What is extracted in the CPR game, however, cannot be reverted to the pool in future periods. As a consequence of these various differences, both games have different equilibria. The CPR game has an interior stationary Markov equilibrium where players extract too much relative to the efficient benchmark (Mailath and Samuelson, 2005). By contrast, in our dynamic public good game the only equilibrium has zero contributions in all rounds (Section 3).

Herr et al. (1997) study two variations of the CPR game, one where extraction by one player increases the cost of extraction by others in the current period only, and one where extraction increases the costs for all future periods. While both versions have interior equilibria, the latter type of externality should worsen the CPR problem, i.e. lead to lower efficiency. Experimental behaviour is in line with equilibrium benchmarks. Botelho et al. (2014) study a game that lies in between a classic CPR and Centipede game, in that they study a CPR game which ends as soon as the resource stock falls below a certain level. They show that increasing uncertainty about that level increases extraction and hence the inefficiency. Battaglini et al. (2014) discussed above can be viewed as a public good study within the CPR paradigm.

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, experimental instructions and the questionnaire, proofs of the theoretical predictions and several additional results, tables and figures.

2 Experimental Design

In our experiment participants play a sequence of dynamically interdependent public good games in two main 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 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. Participants are indexed $i \in I = \{1, 2, 3, 4\}$. Before period 1, each participant is given 20 tokens as initial endowment. In each period participants can divide their tokens into two accounts: a *private* account and a *group* account. The private account, which has a return of 1, is for their personal use only and other participants cannot influence or benefit from the tokens in the private account. 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 c_i^t tokens into the group account, then the rest of the tokens automatically go to the private account and at the *end* of period t , she will have $N_i^t - c_i^t + \frac{1.5}{4} \sum_{j=1.4} c_j^t$ tokens. In the NOPUNISH treatment, this amount also corresponds to the number of tokens *before* period $t + 1$ (the endowment at $t + 1$):

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

Therefore, in period $t + 1$, the number of tokens that each participant can invest depends on the choices of all group members in previous periods (endogenous endowments). 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 previous periods cannot be used for investment into the public good (exogenous stationary endowments). At the end of each period participants observe information about endowments and contributions of all group members (Figure A.1 in Online Appendix A).

Treatment PUNISH The PUNISH treatment is the same as the NOPUNISH treatment with the difference that, after all participants in the group have allocated their tokens and observed others' allocations (as in NOPUNISH), they have a possibility to subtract tokens from individual members of the group. The cost of subtraction is $\frac{1}{3}$ of the tokens subtracted.

More specifically, let $p_{i,j}^t$ denote the cost of punishment that participant i incurs after punishing participant j by $3p_{i,j}^t$ tokens at period t , and denote by $W_i^t := N_i^t - c_i^t + \frac{1.5}{4} \sum_{j=1}^4 c_j^t$ the amount of pre-punishment tokens of player i . The amount of tokens that participant i has after punishment and hence at the start of period $t + 1$ is then

$$N_i^{t+1} = \min_{J \in \mathcal{J}_i^t} \left\{ W_i^t - \sum_{j \in J} 3p_{j,i}^t \right\} - \sum_{j \neq i} p_{i,j}^t$$

where \mathcal{J}_i^t contains all nonempty subsets $J \subseteq I$ such that (a) $i \notin J$, (b) if $k \in J$ then $j \in J$ for all $j \in \{1, \dots, k\} \setminus \{i\}$, and (c) $W_i^t - \sum_{j \in J} 3p_{j,i}^t \geq 0$. This formulation accounts for the fact that punishments which would set a player's wealth below zero are not executed.⁶ 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 subtracted from her; the total amount of tokens subtracted from each member of the group; the amount of tokens each group member has at the end of the period, which serves as the endowment in the next period. Labels for group members on these screens are randomized across periods.

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 better understand 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 more 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 periods and sessions with 15 periods to understand how a longer horizon might affect behaviour and hence get some insights into how participants might perceive this environment strategically.⁷ Due to the (potentially) exponential increase of earnings over time in our setting,

⁶If 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. Since subject labels are randomized by the software this does not create imbalances between subjects. Note also that all players have to pay p_i^j for all punishments intended irrespective of whether they are feasible and of whether punishing sets their own payoffs to zero. Since participants' earnings can still become negative if they punish someone else, this restriction does not affect the set of subgame perfect Nash equilibria. In imposing this restriction we follow the literature on the standard setting (see e.g. Fehr and Gächter (2000)).

⁷Gächter et al. (2008) have found that the length of the horizon can affect contributions in the standard setting, particularly in treatments with punishment.

we limited the long horizon to 15 periods.⁸ Participants were informed about all these details of the design at the beginning of the experiment.

Other Details Participants answered a sequence of questions at the end of the experiment (see Online Appendix B). We report summary statistics on these characteristics in Tables 14 and 15 in Online Appendix F, where 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 a €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 and used z-tree (Fischbacher, 2007). Table 8 in Appendix D shows the exact order of sessions for the main treatments. In total 656 participants took part in our experiment: 152 in treatment NOPUNISH (38 groups), 144 in treatment PUNISH (36 groups) and the remainder in one of the additional treatments. Table 7 in Appendix D summarizes the treatment structure and the number of independent observations, participants and sessions for each treatment. 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.

3 Theoretical Background and Research Questions

In this section we briefly discuss the theoretical predictions and then present our research questions. We focus on the standard solution concepts for 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 SPE leads 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. The purpose of stating these results is not to argue that they will be good predictions of behaviour, but rather to give the reader a sense of the incentive properties of the public good game with dynamic interdependencies. In this section we will state the result informally. Formal statements and proofs can be found in Online Appendix C.

Proposition 1 *Consider the public good game with growth as defined above. The unique SPE of both the game without punishment and with punishment is such that all players contribute 0 throughout the entire game (both on and off the equilibrium path).*

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) have the potential to induce positive contributions in either setting, but also to drive a wedge between the two settings in terms of the extent of cooperation found. More precisely, following Kreps et al. (1982), let us assume that there are two types of agents: (i) “standard” rational and self interested types and (ii) conditional cooperators who start out by contributing 10 tokens and then match at each time t the minimal contribution in their group at $t - 1$. We assume that all agents are of type (i) and that each agent has a minimal doubt that other agents may be of type (ii), i.e. conditional cooperators. More precisely we denote by μ the

⁸There 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 primary findings are affected if we exclude this session.

probability with which agents believe that all other agents are conditional cooperators. We also denote by T the length of the game, which in our experiment is either 10 or 15 periods. We can then state the following result.

Proposition 2 *Assume $\mu < \frac{10}{35}$. All sequential equilibria involve positive contributions on the outcome path*

- *if and only if $\mu > \hat{\mu}_S := \frac{50}{20T+5}$ in the standard setting and*
- *if and only if $\mu > \hat{\mu}_G := \frac{6.25}{4 \cdot 1.5^T - 2.25}$ in our setting.*

For any fixed $\mu < \frac{10}{35}$, the equilibrium always involves strictly more periods of positive contributions in our setting.

The proof of this result can be found in Online Appendix C. Because of the incentives provided by exponential growth, even a small doubt that some agents may be conditional cooperators suffices to induce positive contributions on the equilibrium path. In the standard setting, by contrast, such doubts have to be substantial ($\mu > 0.19$) to even get one period of positive contributions. In fact, if $0.03 < \mu < 0.19$ and $T = 10$, then positive contributions can only be sustained in the setting with growth. For any given μ , more periods of contributions will be observed in the setting with growth. And, across the periods where contributions are positive, the amount contributed is constant in the standard setting and increasing in the setting with growth (for a related result in Centipede games, see McKelvey and Palfrey (1992)).

We next summarize our research questions. First and foremost, we are interested in whether participants contribute positive amounts and in how much wealth is created in this game.

Q1 *Are there sustained positive contributions in our setting? How much wealth is created?*

Apart from wealth creation, we are also interested in the amount of inequality created both within and across groups. Of particular interest is the relation between inequality and growth. The sign of the relation between inequality and growth of societies as well as the causal link between the two has been at the center of a debate in macroeconomics and development (Barro, 2000; Forbes, 2000; Persson and Tabellini, 1991). Hence, while Q1 asks whether societies can see positive contributions and increasing wealth over time, our next question Q2 asks how much inequality is generated in this process and how inequality affects growth. Our setting allows us to address this question both within and across groups. In particular, we ask whether “rich” or “poor” groups will be more unequal and how initial inequality affects growth and hence final wealth.

Q2 *How much inequality is generated in our setting? What are the consequences of endogenous inequality for growth?*

Last, we are interested in the effects of punishment in our setting. Punishment has been shown to be effective in the standard setting in securing high contributions and wealth (Gächter et al., 2008). In our setting, however, punishment does destroy resources which could otherwise be used productively in the following period. Punishment in the dynamic setting can have additional adverse effects that operate via endogenously created inequality. Because earnings are carried over across periods, free-riders in our setting will have more resources to punish others than contributors. 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.

Q3 *How does the possibility of punishment affect contributions, growth and inequality?*

4 Results

This section comprises our main results. Subsection 4.1 focuses on Question 1, i.e. on how much participants contribute and how much wealth is generated in this setting. Subsection 4.2 focuses on Question 2, i.e. how much inequality is endogenously created. We study the effect of punishment (Q3) within each of these subsections. In Section 5 we discuss additional results and possible mechanisms behind our results.

4.1 Provision of the Public Good and Wealth Creation

We start by discussing contributions. Panel (a) in Figure 1 shows the average amount of tokens participants contributed over time. Contributions are clearly non-zero and are increasing over time in all treatments. Even without punishment participants contribute about 10 tokens in the first period and then steadily increase this amount over time. Contributions flatten out towards the end of the experiment and there is even a statistically significant drop in period 15 in the long horizon game. Note that increasing contributions over time imply that participants have increasing endowments over time. Hence increasing contributions do not necessarily imply that participants contribute increasing shares of their endowments. Panel (b) in Figure 1 shows the share of overall endowments contributed over time in all main treatments. In NOPUNISH participants contribute around 55% of their endowment in period 1. This amount steadily decreases between periods 2-8 and is roughly constant afterwards at a level of about 35% of endowments. In the 15-period games shares follow a similar pattern until period 10, then start to decrease again afterwards.

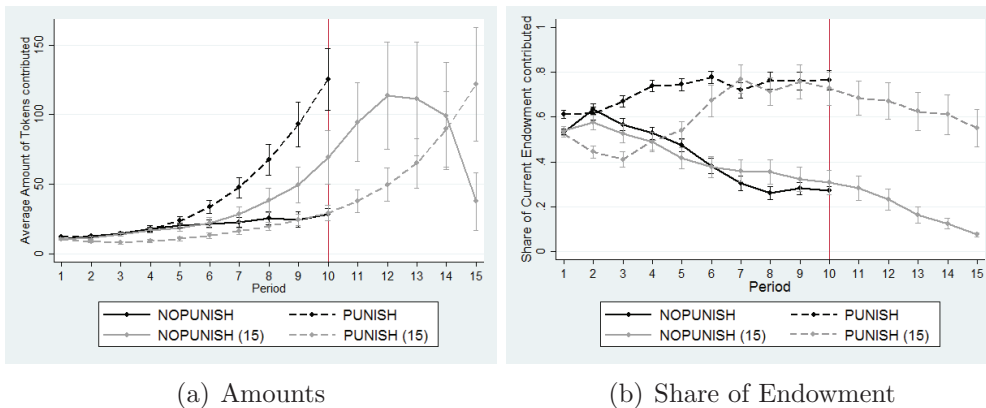


Figure 1: The average amount of tokens contributed over time in treatments NOPUNISH (Without Punishment) and PUNISH (With Punishment) in the 10-period and 15-period variations. Bars indicate one standard error of the mean.

Panel (a) in Figure 1 presents a stark contrast to typical 10-period public good games in the standard setting, where contributions are decreasing over time (see e.g. Figure 3 in (Fehr and Gächter, 2000)).⁹ It should be kept in mind, though, that in the standard setting with stationary exogenous endowments, the dynamics of contributions over time is the same irrespective of whether it is measured in absolute terms or as a share of endowment. In Fehr and Gächter (2000), for example, contributions equal around 50-60 percent of endowments in Period 1 and decrease to around 10 percent of endowments by Period 10. Panel (b) in Figure 1 shows that in our setting contributions stabilize at a higher level ($\approx 35\%$) in the 10-period games even if

⁹Decreasing patterns in public good games are found for MPCR's roughly in the range of the MPCR of our one shot game of 0.375. For higher MPCR's (e.g. 0.75) contributions are still decreasing, but more weakly so (Isaac and Walker, 1988; Zelmer, 2003).

measured as share of endowments contributed. One possible explanation is that the dynamic incentives with the possibility of exponential growth have a similar effect to that of increasing MPCR’s in the standard setting (Isaac and Walker, 1988; Goeree et al., 2002).¹⁰ This higher level of shares contributed seems not to be sustainable in the longer run, though. By period 15 a similar share of endowments is contributed as in the standard setting after 10 periods (10-20%).

	<i>Wealth</i>		<i>Ranksum Test</i>	
	NOPUNISH	PUNISH	Higher Ranksum	p-value
$t = 10$ Mean	439.69	540.30		
$t = 10$ Median	299	143		
$t = 10$ Std.Dev.	345.39	793.17	NOPUNISH	0.0928
$t = 10$ Max	1663	2724		
$t = 10$ Min	158	0		
Observations	23	21		
$t = 15$ Mean	1503.46	713.06		
$t = 15$ Median	422	248		
$t = 15$ Std.Dev.	2873.79	1279.54	NOPUNISH	0.3720
$t = 15$ Max	8687	5096		
$t = 15$ Min	155	0		
Observations	15	15		

Table 1: Summary Statistics on variable **wealth**.

Possibly of more interest are the implications contributions have for wealth generation and growth. To measure growth we define a variable “wealth” which sums the endowments of all participants in a given group at the beginning of the following period, i.e. wealth in t would be $\sum_{i \in I} N_i^{t+1}$ using the notation from Section 3. Before the start of period 1 wealth will be 80 in all groups by construction. The maximal wealth that can be reached in period 10 (if everyone contributes their entire endowment in each period) is approximately 3075 tokens or €153. After period 15 the maximum is approximately 35000 tokens or €1751. The minimal wealth that can be reached (if no one ever contributes anything) is 80 in NOPUNISH and negative in PUNISH (if someone punishes others and is heavily punished in the same period). Table 1 shows some summary statistics regarding wealth. Groups do achieve growth on average. While there is clearly growth, groups do not realize the maximal potential efficiency. In the 10-period games groups reach on average a level of 439.69 tokens out of 3075 maximally possible or 14.29%. There is large heterogeneity with the richest group reaching 1663 tokens and hence more than 10 times more than the poorest group.

Figure 2 shows the dynamics of wealth over time. Panel (a) focuses on all groups, panel (b) on those with above median wealth in Period 10 (“successful” groups) and panel (c) on those with below median wealth in Period 10 (“unsuccessful” groups). Average wealth is increasing across periods (see coefficients β_1 and β_4 in Tables 9 and 10 in Appendix D) and is substantially above 80 once period 10 is reached (Table 1). Not all groups achieve growth, however. There are several groups where wealth does not rise (substantially) above 80. Figures E.1-E.2 in Online Appendix E illustrate wealth (and Gini coefficients) over time in different matching groups.

Result 1: Contributions and Wealth creation *Amounts contributed are positive and increasing over time even without punishment. Wealth is growing over time, but there is large variation across groups with the richest group earning more than 10 times more than the poorest after 10 periods and more than 55 times more after 15 periods.*

¹⁰In a CPR game Herr et al. (1997) observe a pattern of relatively constant extraction over time. They find extraction rates that are close to the prediction of the interior subgame perfect Nash equilibrium. In our setting, the SPE is not interior, but prescribes zero contributions instead, while the contribution levels we find experimentally are substantially above that. Kimbrough and Vostroknutov (2015) study a dynamic CPR game where for sufficiently high regrowth rates the resource will be provided indefinitely in equilibrium.

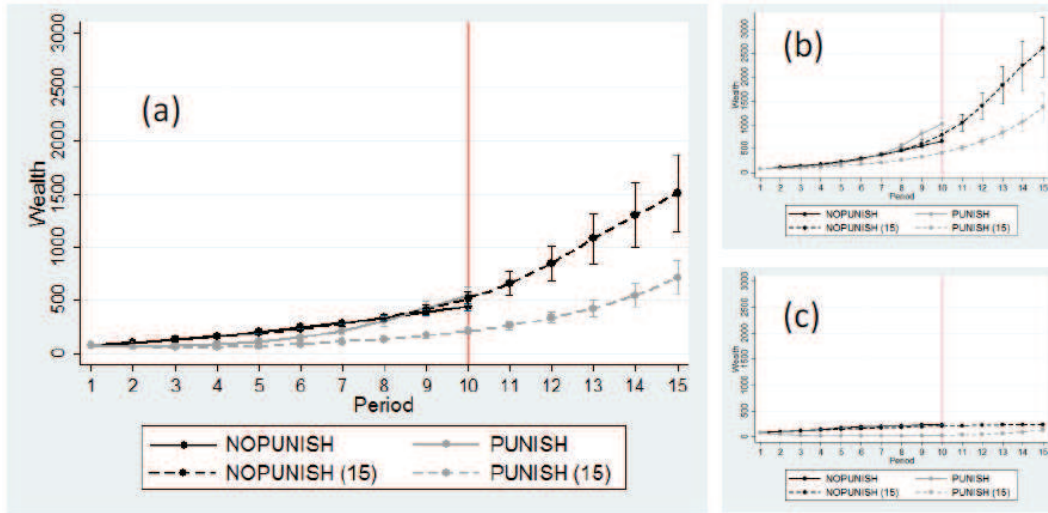


Figure 2: Average wealth over time across treatments. Panel (a) all groups, (b) groups with above median wealth and (c) groups with below median wealth. Bars indicate one standard error of the mean.

Sample	Wealth					
	10 period games				15 period games	
	(1) All	(2) Below Median	(3) Above Median	(4) All	(5) Below Median	(6) Above Median
PUNISH	107.0 (89.2)	-213.0*** (17.68)	361.3 (280.2)	-790.4** (395.6)	-101.3 (101.3)	-1,239 (1,408)
Constant	439.7*** (71.46)	238.3*** (12.51)	659.4*** (117.9)	1,503** (732.2)	239.7*** (23.86)	2,609* (1,269)
Observations	44	22	22	30	15	15
R-squared	0.008	0.873	0.074	0.033	0.061	0.049

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: OLS regression of wealth on treatment dummy. Period 10 (15) data only.

Effects of Punishment on Contributions and Wealth

We next consider the effect of punishment on contributions and wealth. Absolute contributions end up higher in PUNISH compared to NOPUNISH (Panel (a) in Figure 1), but they start to differ only from period 7 onwards in the 10-period games and in the 15-period games they are higher only in the very last period. In terms of shares contributed (Panel (b) of Figure 1), in PUNISH contributions are roughly constant across all periods at a level of about 60%, though there is a slight decrease towards the end of the 15 period games. This is above the share at which groups stabilize in treatments without punishment.

However, groups in PUNISH are also poorer. Median wealth is higher in NOPUNISH compared to PUNISH (Table 1) and the difference in mean ranks is statistically significant in the 10-period games according to a one-sided ranksum test ($p < 0.0001$ for all periods; $p = 0.0928$ for $t = 10$ data only). To assess the statistical significance of differences in means we run OLS regressions where we regress wealth on a treatment dummy for PUNISH (Table 2). These regressions show that differences in means are only significant for below median groups in the 10-period games which earn on average 213 tokens less in PUNISH compared to NOPUNISH. Several of these groups, in fact, end up with zero income in PUNISH. In the 15-period games differences are statistically significant across all groups with mean wealth in period 15 being about 790 tokens lower in PUNISH compared to NOPUNISH. The fact that punishment seems more harmful the longer the horizon stands in contrast to results obtained in the standard setting by e.g. Gächter

et al. (2008). Columns (1) and (2) of Table 9 in Appendix D assess differences in time trends using a linear (column (1)) and square (column (2)) polynomial in period. While wealth is increasing in both treatments ($\beta_1, \beta_1 + \beta_3$), the linear trend in column (1) is no different across treatments (β_3). The square polynomial does reveal differences in time trends, though. While wealth is initially increasing at a lower rate in PUNISH (β_3), it increases at a faster rate in later period (β_5). Figure 2 illustrates these results. It shows that wealth is lower in PUNISH compared to NOPUNISH across all periods except for the last three periods in the 10-period games.

Overall, these results indicate that the threat of punishment is not needed to sustain high contributions in public good games with growth and does not increase wealth. Instead the dynamic evolution of the public good drives contributions and wealth. This is in stark contrast to what we know from other social dilemma games, where punishment has been shown to be successful in both raising contributions (in Fehr and Gächter (2000)’s study on the standard setting by around 600%) as well as wealth *in later periods* (in Fehr and Gächter (2000)’s study from period 4 onwards (Result 8 in Fehr and Gächter (2000))). In Section 5 we discuss in more detail why punishment is less effective in our setting. There we also analyze data from an experiment by Herrmann et al. (2008) conducted in the standard setting and show that in this case punishment leads to unambiguously higher wealth.

Result 2: Effect of Punishment on Wealth creation *The possibility of punishment does not increase wealth. In the long horizon games (15 periods), average wealth is even lower in PUNISH compared to NOPUNISH.*

4.2 Inequality

In this subsection we focus on the amount of inequality created endogenously in our setting. 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 wealth) and the largest possible value it takes is one (if one group member holds the entire wealth). Table 1 shows some summary statistics regarding the Gini coefficient. The period 10 Gini coefficient ranges between 0 and 0.43 in NOPUNISH with a median of 0.22 and assumes the full range between 0 and 1 in PUNISH with a median of 0.03. The period 15 Gini coefficient ranges between 0.03 and 0.49 in NOPUNISH and between 0 and 0.67 in PUNISH.

	<i>Gini coefficient</i>		<i>Ranksum test</i>	
	NOPUNISH	PUNISH	Higher Ranksum	p-value
<i>t</i> = 10 Mean	0.21	0.22		
<i>t</i> = 10 Median	0.22	0.03		
<i>t</i> = 10 Std.Dev.	0.12	0.35		
<i>t</i> = 10 Max	0.43	1	NOPUNISH	0.0395
<i>t</i> = 10 Min	0.00	0		
Observations	23	21		
<i>t</i> = 15 Mean	0.18	0.14		
<i>t</i> = 15 Median	0.10	0.05		
<i>t</i> = 15 Std.Dev.	0.16	0.20		
<i>t</i> = 15 Max	0.49	0.67	NOPUNISH	0.1508
<i>t</i> = 15 Min	0.03	0		
Observations	15	15		

Table 3: Summary Statistics on the variable **Gini coefficient**. Period 10 (15) data only.

Figure 3 illustrates the dynamics of the Gini coefficient over time. The figure shows that in NOPUNISH inequality is sharply increasing from the initial value of zero and then increases more slowly between periods 2-10 to reach a level of ≈ 0.2 in period 10 (β_2 and β_4 in column (2) of Tables 11 and 12 in Appendix D). Interestingly, these patterns almost perfectly mimic the trends in inequality identified in a recent paper by Nishi et al. (2015) (see their Figure 2) who study

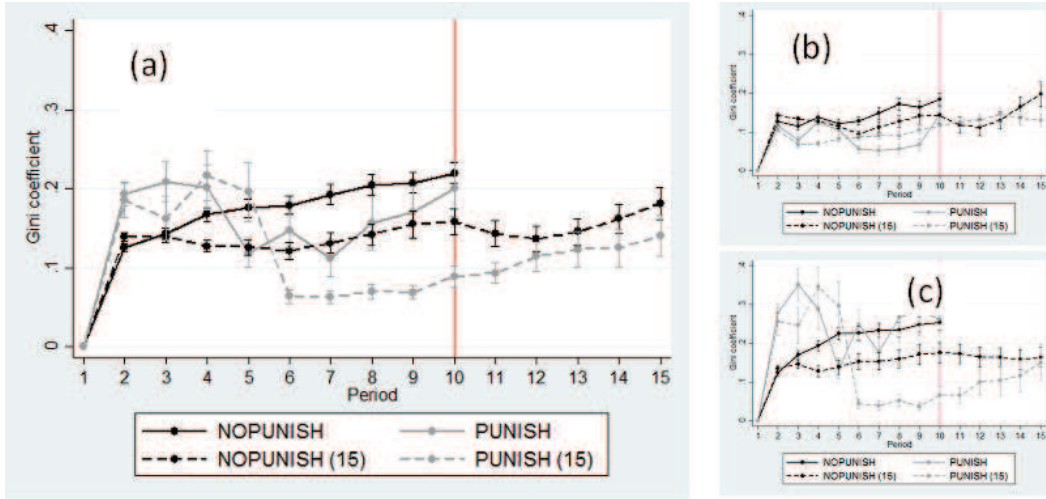


Figure 3: Average Gini coefficient over time across treatments. Panel (a) all groups, (b) groups with above median wealth and (c) groups with below median wealth. Bars indicate one standard error of the mean.

the effect of visibility of wealth (and inequality) in a cooperation game played on a network. A longer horizon seems to lead to slower growth in inequality, but by period 15 Gini coefficients are not statistically different from those reached in period 10 of the shorter horizon games. Panel (b) shows groups with above median wealth, which display lower levels of inequality than those with below median wealth illustrated in Panel (c) (ranksum test $p < 0.001$ all periods, $p = 0.1096$ period 10, $p > 0.1$ period 15).

Sample	Gini coefficient					
	(1) All	10 period games		(4) All	15 period games	
		(2) Below Median	(3) Above Median		(5) Below Median	(6) Above Median
PUNISH	-0.020 (0.035)	0.008 (0.128)	-0.040 (0.062)	-0.041 (0.033)	-0.013 (0.110)	-0.068 (0.072)
Constant	0.220*** (0.024)	0.253*** (0.038)	0.184*** (0.029)	0.181*** (0.023)	0.163*** (0.053)	0.197*** (0.064)
Observations	44	22	22	30	15	15
R-squared	0.002	0.000	0.019	0.013	0.001	0.055

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: OLS regression of Gini coefficient on treatment dummy. Period 10 data only.

Result 3: Inequality. Mean Gini coefficients are increasing over time reaching ≈ 0.2 in the last period. There is substantial variation in Gini coefficients across groups, particularly in PUNISH where the Gini assumes the full range between 0 and 1 in the 10-period games.

Effects of Punishment on Inequality

We next study the effect of punishment on inequality. Mean Gini coefficients are similar across treatments and there are no statistically significant differences in mean Gini coefficients between NOPUNISH and PUNISH in period 10 or 15, respectively (see Table 4). By contrast, analysis conducted with data from a standard public good game in Section 5.4 shows that in this setting punishment leads to more inequality. In fact, in our case the median Gini coefficient tends to be even higher in NOPUNISH compared to PUNISH (Table 3). The difference in mean ranks is statistically significant (one-sided ranksum test $p = 0.0395$) only in the 10-period games. There is

also more variation in PUNISH (one-sided variance ratio test on period 10 data only, $p < 0.0001$; period 15 data: $p < 0.0001$) where the Gini coefficient assumes the full range between 0 and 1 in the 10-period games and between 0 and 0.67 in the 15-period games.

Figure 3 illustrates these results as well as differences in time trends. Across all groups (Panel (a)) the dynamics of inequality seem similar across treatments with somewhat more volatility in PUNISH. Columns (1)-(2) in Table 11 show that there are few statistically significant differences in these time trends. The dynamics of inequality over time seem also quite similar in above median groups, where the difference is mostly one of means (Panel (b)). In below median groups, however, time trends are quite different across treatments (Panel (c)). In NOPUNISH inequality is steadily increasing over time at a slow rate, while in PUNISH the Gini coefficient seems to follow a different pattern. After increasing initially, it decreases sharply around periods 4-7 and then starts to increase again. One possible interpretation is that this pattern reflects cycles of reciprocal punishment. Depending on who punishes (shirkers or high contributors) inequality increases or decreases. We discuss differences in punishment behaviour between below and above median groups in more detail in Section 5.3.

Result 4: Effect of Punishment on Inequality. *Mean inequality (Gini) is not statistically different in PUNISH compared to NOPUNISH in period 10 (15). There is more across-group variation in inequality in PUNISH.*

5 Discussion and Additional Results

In this section we discuss some of the potential mechanisms underlying our main results. Section 5.1 discusses results on the relationship between growth and inequality. In Section 5.2. we will try to tease apart the effect of exponential growth opportunities from the effect of endogenously created inequality. In Section 5.3, we provide some additional results on punishment, which should help us get a deeper understanding of why punishment is less effective here than in the standard setting. Finally, in Section 5.4 we use data from experiments conducted by Herrmann et al. (2008) to study wealth and inequality in the standard setting.

5.1 The Relation between Growth and Inequality

We study the relationship between growth and inequality across and within groups starting with across-group inequality. In NOPUNISH the two variables are essentially uncorrelated. The Spearman correlation coefficient is -0.1840 in above median groups and -0.0630 in below median groups, neither of which is statistically different from zero. In PUNISH, by contrast, there is a substantial and statistically significant relationship between these two variables. In above median groups the correlation is negative (Spearman correlation coefficient $\rho = -0.6091^{**}$): higher wealth comes with lower inequality. In fact, most groups that achieve very high levels of wealth in period 10, have a Gini coefficient of almost zero. For below median groups, the correlation is positive ($\rho = 0.4683^*$): higher wealth comes with higher inequality. As a consequence inequality is highest in groups with intermediate levels of wealth in PUNISH. Figure D.1 in Appendix D illustrates the correlation between period 10 wealth and Gini coefficients for both treatments. The same patterns hold in the 15-period games, but, possibly due to lower power, there is no statistical significance.

These differences between above and below median groups appear already as early as in the second period of the game. For NOPUNISH both successful (above median wealth in period 10) and unsuccessful (below median wealth in period 10) groups have a Gini coefficient of 0.13 on average in period 2. In the 15-period games these numbers are 0.13 and 0.14, respectively. In PUNISH, on the other hand, there are substantial differences. Groups that are eventually

successful have a Gini coefficient of about 0.11 in period 2 (0.10 in the 15-period games), while unsuccessful groups have a Gini coefficient of 0.27 (0.25 in 15-period games), more than twice as high. Hence, successful and unsuccessful groups already differ after the first round of the public good game.

This observation motivates us to study the extent of path dependency *within* groups. Table 5 presents evidence in this respect. It shows the correlation between wealth in period 10 (period 15) and the wealth or the Gini coefficient in previous periods. Path dependence is evident. Early period wealth is strongly correlated with late period wealth. The Spearman correlation coefficient between period 2 and 10 income is 0.48*** in NOPUNISH and even 0.82*** in PUNISH. Maybe more interestingly, also early period inequality is highly detrimental to final wealth, but only in PUNISH. The correlation between inequality in period 2 and wealth in period 10 is -0.47^{***} , which is substantial. In NOPUNISH, by contrast, early period inequality (in periods 2-4) is not negatively correlated with wealth. In this treatment the negative correlation appears only from period 5 onwards. The fact that there is no negative correlation between period 7,8,9 inequality and period 10 wealth in PUNISH is due to the fact that by then several groups have zero wealth and inequality. Dropping these groups restores the negative correlation. Comparing 15 period wealth and early period inequality (bottom panel of Table 5) shows similar patterns.

	<i>Wealth in period ...</i>							
	9	8	7	6	5	4	3	2
NOPUNISH	0.99***	0.98***	0.98***	0.93***	0.85***	0.76***	0.75***	0.48***
PUNISH	0.99***	0.99***	0.98***	0.97***	0.93***	0.90***	0.89***	0.82***
	<i>Gini coefficient in period ...</i>							
	9	8	7	6	5	4	3	2
NOPUNISH	-0.29***	-0.16*	-0.35***	-0.46***	-0.34***	-0.03	0.03	0.11
PUNISH	0.03	0.11	-0.06	-0.43***	-0.17*	-0.32***	-0.50***	-0.47***
	<i>Gini coefficient in period ... (15 period games)</i>							
	14	13	12	11 ...	5	4	3	2
NOPUNISH	-0.12	-0.13	-0.12	-0.16	-0.19	0.07	-0.02	0.18
PUNISH	0.31*	-0.50***	-0.41***	-0.53***	-0.56***	-0.50***	-0.55***	-0.46***

Table 5: Correlation of wealth in period 10 with wealth (Gini) in period 2,...,9. Bottom panel: correlation of wealth in period 15 with Gini in periods 11-14 and 2-5, respectively.

One question about these findings is whether they just reflect 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 wealth if initial inequality is high. The fact that the negative correlation between early period inequality and wealth is only observed in PUNISH, suggests that this is not just a mechanical effect of having different distributions of “contribution types” across groups.

Additional support for this view can be derived from our post-experimental questionnaire. Table 16 in Online Appendix F 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 above median groups do *not* contribute more on average than those from groups with below median wealth. This is despite the fact that participants from groups with above median wealth earn 178 tokens on average in period 10 (189 in treatment PUNISH), while those from groups with below median wealth earn only 56 tokens (23 tokens) on average in period 10. This suggests that participants in groups with above median wealth 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 17-18 in Online Appendix F show that none of them is able to explain the variation in wealth or inequality that we observe.

- Result 5: Relation between Growth and Inequality.** 1. *In PUNISH wealth and Gini coefficient are positively correlated for poor groups (below median wealth) and negatively correlated for rich groups (above median wealth) in the 10-period games. These correlations are weak and not statistically different from zero in NOPUNISH.*
2. (a) *Early period wealth is positively correlated with eventual wealth in all treatments.*
(b) *Early period inequality is negatively correlated with final wealth only in PUNISH.*

Taken together both the findings on across- as well as within- group correlation point to a detrimental role of inequality if there are punishment possibilities. One of the reasons, hence, why punishment is not as effective in this setting seems to be that people react strongly to inequality. The findings are also indicative as to why we observe such substantial variation across groups both in terms of wealth and inequality. Groups in which initial behavior leads to high inequality seem to get locked into a path of punishment and counter-punishment (interpreted as experimental equivalents of conflict) that eventually leads to a destruction of all wealth. We will see more evidence of such behaviour in Section 5.3. In fact there is a strand of literature, where economic historians point out the importance of institutional lock in with institutions 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 an example of how a society (group) can get locked into dysfunctional behavioural norms.

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 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 wealth 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 too loose to draw any conclusions. 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.¹¹

5.2 Eliminating inequality and growth possibilities

In this subsection we further tease apart the importance of endogenous inequality and growth by studying treatments where we shut down one of these channels exogenously. In treatments NOPUNISH-NOGROWTH and PUNISH-NOGROWTH we artificially eliminate growth by re-normalizing all endowments at the beginning 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

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

This means that each participant's endowment ranges between 0 and 80 in each round and the sum of endowments in a group equals 80 in each period. This 10 period game hence, can be viewed as a public good game played in period 10, where the endowment each participant

¹¹Knack 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.

has is determined by the game played in periods 1-9. The unique SPNE is zero contributions in each period as in our main treatments. Note that this structure, while derived from our treatment NOPUNISH via a single change creates a situation where everyone contributing no longer Pareto dominates nobody contributing in periods 1-9. Consequently, participants may have different motives for contributing in these treatments compared to our main treatments and the results should be interpreted with this in mind. We do the analogous normalization for the punishment version (PUNISH-NOGROWTH). We had 116 participants (29 groups) in treatment NOPUNISH-NOGROWTH and 92 participants (23 groups) in treatment PUNISH-NOGROWTH.

Table 13 (column (1)) in Appendix D shows the results of regression where we regress normalized contributions on treatment dummies. The baseline is NOPUNISH. To ensure a fair comparison, we also normalize contributions in NOPUNISH and PUNISH. Normalized contributions in these treatments are computed by multiplying the share of actual endowment contributed with normalized endowments. Maybe unsurprisingly, given the absence of even social incentives to contribute across periods 1-9 of the games without growth, contributions are substantially lower and close to zero.

In treatments NOPUNISH-NOINEQUALITY and PUNISH-NOINEQUALITY 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 reduce incentives to contribute, we implemented an additional change in this treatment. In particular we paid a randomly drawn period rather than the last period.¹² Of course paying a randomly selected period could affect behaviour *per se*, a possibility we cannot rule out. To this extent, the results from this treatment should be read as suggestive and as complementing the evidence from Sections 5.1 and 5.3 rather than yielding definite conclusions on the role of inequality. We had 96 participants (24 groups) in NOPUNISH-NOINEQUALITY and 56 participants (14 groups) in PUNISH-NOINEQUALITY.

The second column in Table 13 (Appendix D) shows that inequality leads to lower contributions in both the treatments with and without punishment. Eliminating inequality increases mean contributions by 26 tokens without punishment and by 42 tokens with punishment. Furthermore, if inequality is eliminated then mean contributions are ≈ 24 tokens higher with punishment (PUNISH-NOINEQUALITY) than without (NOPUNISH-NOINEQUALITY), a difference that is statistically significant ($\chi^2 < 0.001$). It seems that endogenous inequality in endowments with the associated inequality in the power to punish undermines the effectiveness of punishment. Since we exogenously remove inequality in the treatments discussed here, the results from this section support a causal interpretation of the effect of inequality on wealth.

5.3 Anatomy of Punishment

Both sections 5.1 and 5.2 have pointed to a negative role of inequality for contributions and wealth particularly in treatment PUNISH. To understand why this is the case it is helpful to

¹²Paying the last period in this treatment (as we did in the other treatments) would mean that there are incentives to contribute across all periods, making this setting difficult to compare to our other treatments. Paying a randomly selected period reduces this incentive. In particular, with a randomly paid period there are no incentives to contribute in the last period. In the 9-th period there is a tradeoff between increasing the endowment available in period 10 by contributing and increasing the 9th period payoff by not contributing. This tradeoff is resolved towards not contributing for periods 9,8,7 and 6. In earlier periods, however, there are incentives to contribute in order to increase the endowment available in future rounds.

study punishment patterns more closely.

We first compare above and below median groups in treatment PUNISH to understand which patterns of punishment lead to low wealth in this treatment and are hence crucial for the ineffectiveness of punishment in this setting. Figure 4 reveals an interesting pattern in this regard. It shows the amounts of tokens participants use to punish over time. In groups with above median wealth the absolute amounts used to punish tend to remain stable or increase over time (OLS coefficient: 1.402* (10 periods); 0.098 (15 periods)),¹³ while they tend to decrease in below median groups (OLS coefficient: -0.830^{***} (10 periods); -0.332^{***} (15 periods)). 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 (two-sided ranksum test $p < 0.0001$). 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 while they wait to punish others harshly at the end of the game. This seems intuitive because of the detrimental effect that punishment can have on growth. While in below median groups most punishment happens in the beginning of the game, above median groups punish at the end.¹⁴

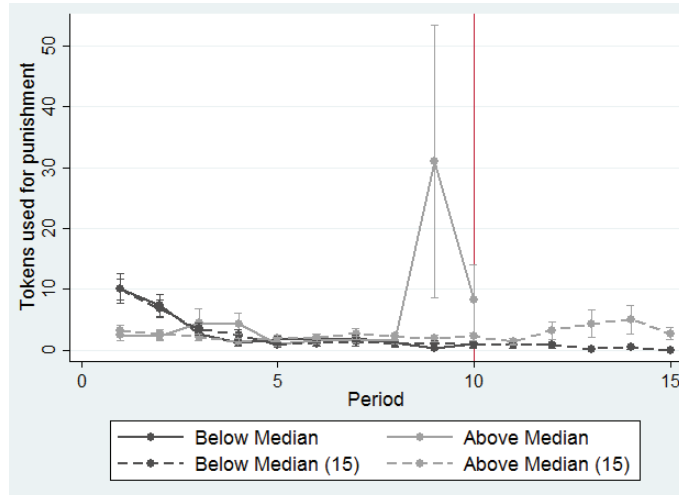


Figure 4: Average Amount of Tokens used to punish over time by groups with above median wealth (dashed line) and groups with below median wealth (solid line).

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

Table 6 shows that there is more pro-social punishment than anti-social punishment (one-sided ranksum test, $p < 0.0001$). Anti-social punishment is higher if inequality is high (pro-social: Spearman $\rho = 0.0155$ ($\rho = 0.1231^{***}$, 15 period games); anti-social: $\rho = 0.0777^{**}$ ($\rho = 0.1515^{***}$, 15 period games)). Anti-social punishment is not significantly related to wealth ($\rho = -0.0109$; $\rho = 0.1184$ in 15 period games), but there is more pro-social punishment in high-wealth compared to low-wealth groups ($\rho = 0.1371^{***}$; $\rho = 0.2302^{***}$ in 15 period games).

¹³We regress total amount spent on (intended) punishment on a constant and a variable period ranging from 1,...,10 (15, respectively).

¹⁴Relatedly, Fudenberg and Pathak (2010) have shown that punishment can sustain cooperation even if it is only observed at the end of a session.

	<i>10 period games</i>	
	Pro-social Punishment	Antisocial Punishment
all	1.91 (18.73)	0.65 (4.07)
Gini > 0.04	1.14 (3.24)	1.52 (12.03)
Gini < 0.04	2.68 (26.22)	0.45 (2.10)
wealth > 80	3.10 (26.84)	0.62 (5.25)
wealth < 80	0.44 (1.58)	0.46 (1.87)
	<i>15 period games</i>	
	Pro-social Punishment	Antisocial Punishment
all	1.24 (5.01)	0.65 (2.83)
Gini > 0.08	1.25 (4.23)	0.84 (3.55)
Gini < 0.08	1.23 (5.55)	0.50 (2.09)
wealth > 158	2.25 (8.26)	0.62 (3.84)
wealth < 158	0.77 (2.14)	0.66 (2.22)

Table 6: Mean (SD) of amount of tokens of pro-social and anti-social punishment.

5.4 Wealth and Inequality in the Standard Setting

Before we conclude, we have a brief look at data from the standard setting. We focus on wealth and inequality, as those are not commonly reported measures in the standard setting. In particular, we use data from experiments Herrmann et al. (2008) conducted in Bonn. Bonn and Maastricht are at only 120km driving distance and a large share of students at both Universities come from the lower middle Rhine and upper lower Rhine area (between Koblenz and Dueseldorf) in Germany. This should make subject pools approximately comparable. Just as us Herrmann et al. (2008) conducted 10 period games with an endowment of 20 tokens (in their case exogenously given each period). They also conducted NOPUNISH and PUNISH treatments with the 3:1 technology that we employ. The return on contributing in the one-shot game was 0.375 in our experiments and at 0.4 slightly higher in Herrmann et al. (2008).

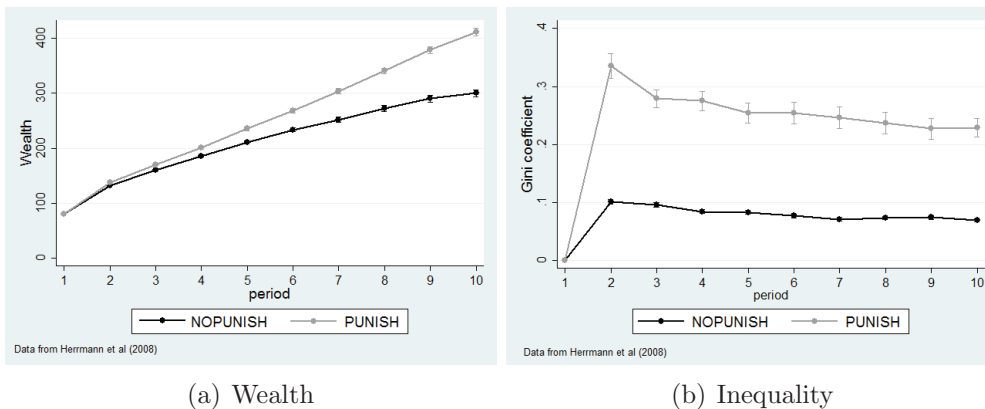


Figure 5: Wealth and Inequality in data from Herrmann et al. (2008)'s Bonn experiments. Error bars show standard errors of the mean.

Figure 5 shows wealth and inequality (Gini coefficient) over time in Herrmann et al. (2008)'s experiments. We computed wealth by adding income of all group members in each period and subtracting 80 in each period 2,...,10. If all group members contribute zero in each period (and do not punish) wealth thus computed is 80 in each period, just as in our setting. If all group members contribute their full endowment in each period (and do not punish) maximal wealth in period 10 is 3075 in our setting and 1280 in Herrmann et al. (2008). Figure 5 shows that both wealth and inequality are substantially higher under the punishment condition in Herrmann et al. (2008)'s experiments (t-test period 10 sample means: $p < 0.01$). The effect of punishment, hence, differs from what we observe in the dynamic setting, where punishment, if at all, decreases wealth and inequality. Note also, that we do not observe the cyclical pattern in inequality identified in the PUNISH treatments in our setting.

In summary, section 5.1 demonstrates a negative correlation between early period inequality and wealth in period 10 in treatment PUNISH. The negative impact of inequality on contributions and wealth is confirmed with some caveats in Section 5.2 where we find that artificially removing inequality increases contributions and wealth both with and without punishment, but with a particularly strong effect in the punishment treatments. Section 5.3 provides some insights into why inequality is so detrimental in the punishment condition. There are two main effects: (i) in groups where inequality is high (above median) there is more anti-social than pro-social punishment, implying that shirkers punish contributors more than vice versa and (ii) much of this punishment happens in early periods in unsuccessful groups implying that resources are taken away exponentially. All these channels contribute to the negative impact of punishment possibilities on wealth creation in the dynamic setting. The absence of inequality in endowments in the standard setting seems to make wealth inequality less salient. This can potentially explain why punishment does not trigger the adverse affects identified in the dynamic setting in the standard setting (Section 5.4).

6 Conclusions

We studied public good games with dynamic interdependencies, where each agent's wealth at the end of a period serves as her endowment in the following period. We found that contributions are increasing over time even in the absence of punishment possibilities. The possibility of punishment does not increase wealth. These results suggest that in settings with a strong dynamic component societies achieve cooperation via the incentives provided by the dynamic evolution of the public good and not so much via the threat of punishment. Across groups, inequality in early periods is strongly negatively correlated with wealth in later periods.

These results show that people are able to establish persistent cooperation in a setting that shares one key feature with many real-life interactions: past behaviour matters for future endowments. They also point to the limits of punishment in securing high contributions and wealth.

The results highlight the importance of incorporating the dynamic aspects, present in many real-life interactions, explicitly in experimental designs. In this paper we have done so within the context of cooperation and public good provision. We should emphasize, though, that we view this design as complementary to the standard design where contributions occur from stationary exogenous budgets. Both settings have natural applications outside the lab. Public goods are often created by providing effort, as e.g. in the case of volunteering. In these cases, clearly, endowments cannot easily increase over time. In many other settings, such as those discussed in the Introduction, provision of public goods creates wealth from which future public goods can be provided. The evolution of societies might be viewed through this lens. We have seen that the two settings can yield quite different conclusions. In the standard setting with stationary budgets and full consumption punishment has shown to be very effective at raising contributions and wealth (in the medium run). In the dynamic setting studied here, the effectiveness of punishment seems much more limited. It is not successful in raising wealth in the 10-period games and is even detrimental to wealth in the games with longer horizon.

Future research should aim at getting a deeper understanding of the mechanisms behind these differences and adapt this setting to other contexts where dynamic interdependencies are likely to play an important role. Studying intermediate settings, with some but not full consumption and only partially endogenous endowments should also be of interest for future research.

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.
- Battaglini, M., S. Nunnari, and T. Palfrey (2014). The dynamic free rider problem: A laboratory study. mimeo.
- Battigalli, P. and M. Siniscalchi (1999). Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory* 88, 188–230.
- Battigalli, P. and M. Siniscalchi (2002). Strong belief and forward induction reasoning. *Journal of Economic Theory* 106, 356–391.
- 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.
- Botelho, A., A. Dinar, L. C. Pinto, and A. Rapoport (2014). Time and uncertainty in resource dilemmas: equilibrium solutions and experimental results. *Experimental Economics* 17, 649–672.
- Cadigan, J., P. Wayland, P. Schmitt, and K. Swope (2011). An experimental dynamic public goods game with carryover. *Journal of Economic Behavior and Organization* 80, 523–531.
- Cason, T. and L. Gangadharan (2015). Promoting cooperation in nonlinear social dilemmas through peer punishment. *Experimental Economics* 18, 66–88.
- 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. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90, 341–369.
- Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature* 415, 137–140.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischbacher, U. and S. Gächter (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.
- Gächter, S., E. Renner, and M. Sefton (2008). The long-run benefits of punishment. *Science* 322, 1510.
- Goeree, J. K., C. Holt, and S. K. Laury (2002). Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *Journal of Public Economics* 83, 255–276.
- Green, E. and R. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica* 52(1), 87–100.
- Greiner, B., A. Ockenfels, and P. Werner (2011). The dynamic interplay of inequality and trust - an experimental analysis. *Journal of Economic Behavior and Organization* 81(2), 355–365.
- 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 (2017). The effects of punishment in dynamic public-good games. working paper 106 Thurgau Institute of Economics.
- 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.
- Herr, A., R. Gardner, and J..M.Walker (1997). An experimental study of time-independent and time-dependent externalities in the commons. *Games and Economic Behavior* 19, 77–96.
- 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. and J. Walker (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics* 103(1), 179–199.
- 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.
- Kimbrough, E. and A. Vostroknutov (2015). The social and ecological determinants of common pool resource sustainability. *Journal of Environmental Economics and Management* 72, 38–53.
- 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.
- Mailath, G. and L. Samuelson (2005). *Repeated Games and Reputation*. Oxford University Press.
- McKelvey, R. and T. Palfrey (1992). An experimental study of the centipede game. *Econometrica* 60(4), 803–836.
- Milinski, M., D. Semmann, and H.-J. Krambeck (2002). Reputation helps solve the 'tragedy of the commons'. *Nature* 415, 424–426.
- Nishi, A., H. Shirado, D. G. Rand, and N. Christakis (2015). Inequality and visibility of wealth in experimental social networks. *Nature* 526.
- North, D. (1994). Performance through time. *The American Economic Review* 84(3), 359–368.
- Noussair, C. and C. Soo (2008). Voluntary contributions to a dynamic public good: Experimental evidence. *Economics Letters* 98, 71–77.
- 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.
- Sadrieh, A. and H. Verbon (2006). Inequality, cooperation and growth: an experimental study. *European Economic Review* 50(5), 1197–1222.
- 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.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics* 6, 299–303.