Experiments on punishment and cooperation

PhD Thesis

Ruslan Shichman





School of Economics
May 2018

Acknowledgements

I would like to thank my supervisors Professor Simon Gaechter and Professor Martin Sefton for guiding me through this journey that started back in 2012. It was a great honor to be supervised by them. I appreciate their time, patience, the knowledge they shared with me and their inputs. They not only assisted me with specific challenges in my research but also helped me in changing my general approach. Especially, I want to thank Professor Gaechter for being willing to assist me whenever I popped into his office.

Also, special thanks go to Professor Jonathan Tan for his comments on my chapters; Dr Benjamin Beranek for helping me with my first sessions and his help on many other occasions; Professors Robin Cubit for his inputs during various seminars and for taking care of the bureaucracy during my voluntary suspension; Professor Chris Starmer and other members of The Network for Integrated Behavioural Science for funding my last experiment. Orestis Kopsacheilis, Valeria Burdea, and Depi Alempaki thank you for the assistance with innumerous experimental sessions.

I am thankful to Professor Avi Weiss for laying the foundations for the current thesis. Professor Daniel Levy for all the conversations we had. Professor Joseph Deutsch and Bar-Ilan Department of Economics for the support.

I could not finish this project without my family. My dear parents Roman and Serafima who have always been there for me. My loving wife Natalia, who is my muse, my inspiration. Also, I appreciate the support of all other members of my family in Israel, Germany, Ukraine, and the USA.

I also wish to thank all my friends. Special thanks go to Lent family for

making me feel home away from home, and Derrick Abudu for assisting me with the submission of the thesis.

My gratitude goes to Professor Jacob Rowe and all the staff of the Hematology department at the Shaare Zedek Medical Center, as well as Dr. Israel Henig and the members of staff at the Hematology department of the Rambam Health Care Campus for taking care of me and making my return to the UK possible.

Finally, I would like to express my deepest gratitude towards the University of Nottingham, Economic and Social Research Council and the British people for granting me the opportunity to study and work in the UK.

Contents

Acknowledgements **i**

Introduction **1**

Chapter 1: Cooperation, Punishment and Domination **3**
1.1 Introduction 5
1.2 Related literature 9
1.3 The experiment 12
1.3.1 Experimental designs 12
1.3.2 Recruitment and procedures 16
1.4 Experimental results 17
1.4.1 Contributions to the public account 17
1.4.2 Punishment 21
1.4.3 Earnings 27
1.4.4 Privilege in contrast to Alpha 29
1.4.5 Bankruptcies 31
1.4.6 Pivotal participants in contrast to others 33
1.5 Discussion and conclusions 38
1.6 Bibliography 41
Appendix A: Instructions 45
Appendix B: Regression Tables 56

Chapter 2: Cooperation, Punishment and Domination With Endogenous Selection to Pivotal Positions **60**
2.1 Introduction 62
2.2 Related literature 64
2.3 The experiment 66
2.3.1 Experimental designs 66
2.3.2 Recruitment and procedures 70
2.4 Experimental results 71
2.4.1 Ball catching task 71

2.4.2 Contributions . . . 73
2.4.3 Punishment . . . 77
2.4.4 Earnings . . . 80
2.4.5 Privilege in contrast to Alpha . . . 82
2.4.6 Bankruptcies . . . 84
2.4.7 Pivotal participants in contrast to others . . . 85
2.4.8 Comparison with Chapter 1 . . . 86
2.5 Discussion and conclusions . . . 87
2.6 Bibliography . . . 90
Appendix A: Instructions . . . 92
Appendix B: Regression Tables . . . 104

Chapter 3: Punishment: getting even or sending a message? 109
3.1 Introduction . . . 111
3.2 Related literature and research questions . . . 113
3.3 Experimental designs . . . 117
3.3.1 Baseline treatment . . . 118
3.3.2 Study 1 - Fate treatment . . . 119
3.3.3 Study 2 - Awareness treatment . . . 120
3.4 Recruitment, procedures and sample . . . 122
3.4.1 Recruitment and procedures . . . 122
3.4.2 Attrition and attention . . . 123
3.4.3 Samples . . . 125
3.5 Experimental results . . . 125
3.5.1 Study 1 - Fate . . . 125
3.5.2 Study 2 - Awareness . . . 137
3.6 General Discussion . . . 147
3.7 Bibliography . . . 149
Appendix . . . 155

Conclusions 166

Introduction

Assuming rationality of profit maximising agents, various economic models made specific and testable predictions about human behaviour. These models suggested that costly punishment should be exercised only if future material gains compensate for the incurred costs. The experimental evidence, however, demonstrates that individuals are willing to pay the price to affect the well-being of others even when no pecuniary or reputational benefits can be derived.

The robustness of the phenomenon gave rise to a whole strand of literature. New theoretical models were developed to explain the motives behind such behaviour. Moreover, further experimental inquiries have been undertaken both to understand the motivation and other aspects of demand for punishment, such as responsiveness to the price.

Economists became interested in punishment not solely due to the inability of standard models to accommodate the evidence obtained in laboratories, but also because of its potential in explaining high cooperation levels in humans societies. It has been demonstrated that under certain conditions the presence of decentralised punishment opportunities in a public goods game can stave off the decay in cooperation levels and even increase them.

The current work consists of four original studies on punishment, which or-

ganised in three chapters. The first two chapters include systematic studies of the effects of asymmetries in returns from cooperation and punishment power on cooperation levels in a public goods game with an endogenous number punishment stages in each period. One of the treatments allows for multiple concurrent asymmetries to model the presence of dominant individuals that disproportionately benefit from cooperation. In Chapter 1 we demonstrate the negative effects asymmetries have on cooperation and other key parameters. The random assignment to the positions of power is replaced by a contest in Chapter 2. The data suggest that the contest did not change substantially the willingness of participants to accept inequalities in cooperative part of the game.

In Chapter 3 two original studies demonstrate that punishment is being used as a means of communication by a considerable number of participants. The second study of this chapter also shows that individuals are ready to pay not only to inform others about the punishment they assigned but also about lack of punishment or even willing to hide own decisions in one-shot interactions.

Chapter 1

Cooperation, Punishment and Domination

Abstract

In this paper, we systematically study the effect of asymmetries in returns and punishment power on cooperation in public goods games with an endogenous number of punishment stages and random designation to favourable positions. Our results suggest that in the presence of an even distribution of punishment power, inequality in returns leads to lower contributions, lower earnings and higher propensity to engage in prolonged conflicts. Redistribution of punishment power, favouring individuals with higher returns from cooperation, suppresses the tendency to feud and renders high levels of cooperation more likely but has little effect on average contributions, punishment spending or earnings. In settings characterised by equal gains from cooperation, heterogeneity in punishment power leads to lower contributions and earnings, as well as increased punishment spending, in the last periods of the experiment. We conclude that asymmetries are detrimental for cooperation in small groups.

1.1 Introduction

The importance of cooperation for the prosperity of our species cannot be underestimated. Although in some cases individuals may seek personal benefit at the expense of the group, cooperation is omnipresent even when personal and collective interests seem to be at odds with each other.

A growing experimental literature attempts to address this dilemma by utilising variations of the public goods game (PGG). Results show that a few individuals are willing to cooperate even when personal and group interests do not coincide (for literature reviews see Fehr and Fischbacher (2003) and Chaudhuri (2011)). This goodwill gradually evaporates in the presence of selfish individuals, unless freeloaders are allowed to be penalised, as was shown in the seminal work of Fehr and Gächter (2000) and numerous studies that followed[1].

Some studies share a number of basic settings, which include uniform distributions of benefits from cooperation, also known as marginal per capita returns (hereafter MPCR), and punishment efficiency, which is the ratio between the cost of punishment and its effect. Although complete equality might be considered as a natural benchmark for an experimental design, such egalitarian institutions are not common nowadays. One can find them in hunter-gatherer bands and some tribes at the outskirts of human civilisation.

Among the salient characteristics of such groups is equality in political power and strong sharing norms, which lead to an equal distribution of wealth among their members. Moreover, similar to the settings of various PGGs with punishment opportunities, egalitarian groups do not rely on centralised authority for

[1] Other mechanisms, such as pre-play communication (Isaac and Walker, 1988), have been successful in boosting cooperation in PGG, but are of no interest to us in this study.

enforcement of norms, but rather on decentralised (peer) punishment.

Leaders in such groups are constantly held in check and sanctioned by the rank and file for actions that can be interpreted as a power grab. For instance, giving direct orders to a fellow member of a group can be sufficient to trigger retaliation, not to mention more abusive behaviour (Boehm, 2009, p.43). The absence of coercive centralised authority allows individuals to keep their autonomy but also leaves them vulnerable to intra-group feuds. The devastating potential of such conflicts can be learned from the story of Fayu people of Papua, who at some point lost about 80% of their population due to blood feuds (Diamond, 1998, p. 266).

Unlike egalitarian groups, most contemporary societies and many societies of the past feature inequalities along certain dimensions. Some studies changed the aforementioned experimental settings to probe into the effects asymmetries might have on cooperation. None, however, allowed for multiple concurrent asymmetries.

In the animal kingdom, it is not uncommon to find groups led by strongest individuals, who get the lion's share of resources. Higher benefits from cooperation and lower costs of administering punishment may render enforcement of pro-social behaviour beneficial, under the assumption of personal fitness maximisation. For example, chimpanzee alpha males intervene to pacify intra-group conflicts (Boehm, 2009, p. 26) and compensated through disproportional rate of reproduction (Wroblewski et al., 2009).

Even though human behavioural patterns are more complicated than those of primates, the same principle, promotion of cooperation by the strongest and compensation for the effort through the appropriation of a disproportionate share of resources, may apply to our species. It is not clear, however, whether the

hierarchical organisation of societies is due to our nature or in spite of it. Some suggest that our social nature has been profoundly affected by the egalitarian institutions that dominated during the Upper Palaeolithic (Boehm, 1997).

The general goal of this study is to look into the effects of inequalities in punishment efficacy and benefits on cooperation under experimental settings that make within period feuds possible. More specifically, we aim at establishing whether dominant individuals ("alphas") can promote higher levels of cooperation and increase the total welfare by containing feuds in small groups.

For our purposes, we use four treatments varying in a distribution of gains from cooperation and punishment power. In the baseline treatment, the distributions of benefits from cooperation and punishment efficacy are uniform. Two other treatments are characterised by asymmetry in one of the two parameters, so one group member either receives more from the public good or is more efficient in reducing others' earnings. The fourth treatment models the presence of an "alpha" individual by endowing a single group member with both advantages.

To make intra-period feuds possible we follow Nikiforakis and Engelmann (2011). The number of punishment rounds is not restricted ex-ante. Participants can punish each other within the same period so long as they and the targeted individuals possess enough Experimental Currency Units (hereafter ECU).

Our study not only replicates certain results but also adds substantially to the existing economic literature on cooperation and enforcement. To our knowledge, no other study has looked at the effect of multiple concurrent asymmetries on cooperation. In addition, we explore the effects of asymmetry in punishment power in the presence of feuding opportunities.

The findings suggest that egalitarian institutions are superior in achieving high

levels of cooperation and welfare in small groups. However, the extent and the significance of differences with other treatments vary. On the one hand, in line with previous reports (Nikiforakis et al., 2012; Reuben and Riedl, 2013) uneven allocation of MPCR in combination with an even distribution of punishment efficiency had negative consequences. In the current study, it led to substantially lower contributions, earnings and a stronger tendency for prolonged conflicts. On the other hand, the negative effect of inequality in punishment power was limited to the final periods of the game. Finally, the presence of the "alphas" had a negative effect on cooperation levels and welfare relative to the symmetric treatment, but differences are only weakly significant.

The comparison of the treatments with asymmetric MPCR demonstrates that "alphas" did not succeed in overcoming the tensions arising from those discrepancies. On average cooperation and earnings levels were not markedly different. Nonetheless, in their presence prolonged conflicts were less likely, the proportion of groups that maintained high levels of cooperation for a considerable number of periods has gone up, but also some individuals went bankrupt after losing all their earnings.

The rest of the paper is organized as follows. Sections 2 contains the related literature. Sections 3 presents the experimental design, data, and procedures. Section 4 includes the statistical analysis. Finally, discussion and conclusions presented in Section 5, which is followed by the Appendices A and B.

1.2 Related literature

This paper relates to the seminal experimental studies by Yamagishi (1986), Ostrom et al. (1992) and Fehr and Gächter (2000), who demonstrated the ability of individuals to sustain cooperation in PGG in the presence of sanctioning opportunities. The experimental design used by Fehr and Gächter (2000) is characterised by a complete symmetry. Specifically, players receive equal endowments at the beginning of each period, as well as equal returns from the public good and punishment efficacy. Also in every period, each can make one punishment decision, so that any further retaliation is possible in the following period if the composition of groups and identification numbers remain the same.

Later studies changed some parameters of the above design. Among them Cinyabuguma et al. (2006) and Nikiforakis (2008), who introduced an extra punishment stage. The motives for the change and the implementation differed. In the former, the authors looked into the possibility that penalisation of punishing behaviour could diminish the occurrence of punishment against cooperators. Every three periods participants were shown a list of the amounts by which other subjects had reduced the earnings of below-average contributors, of average contributors, and of above-average contributors. Cinyabuguma et al. (2006) report about less punishment against cooperators, along with a slight improvement in cooperation levels and efficiency in the treatment with the second order punishment.

Nikiforakis (2008) checked the robustness of the result obtained by Fehr and Gächter (2000). He introduced a second punishment stage in every period and provided individuals with full information about contribution and punishment decisions of others. The author demonstrates that introduction of counter-punishment

impaired groups ability to sustain high cooperation levels, as they fell over time.

The restriction on the number of punishment stages was relaxed even further by Denant-Boemont et al. (2007) with five punishment stages. The result is lower contributions and welfare, compared to the treatment with a single round of sanctions.

Finally, Nikiforakis and Engelmann (2011) carried this trend to its logical conclusion by not restricting the number of punishment stages ex-ante, so participants were allowed to punish one another so long as conflicting parties hand enough resources. The authors conclude that the threat of feud eruption within a period substantially lowered the level of altruistic punishment and consequently the levels of cooperation.

A similar setup, but for an extended horizon and with a more efficient punishment mechanism (30 periods and 1/3 punishment cost-effect ratio versus 10 periods and 1/2 punishment cost-effect ratio in the above study) produced higher levels of cooperation relative to the treatment without punishment opportunities (Engelmann and Nikiforakis, 2014). The earnings, however, were not markedly different.

Nikiforakis et al. (2012) shed light on the effects of feuds in the presence of normative conflicts, which according to the authors arise in case multiple behavioural rules coexist. The normative conflict was modelled through the introduction of asymmetric returns from the public good. Their experimental design included two such treatments, one had a single punishment stage per period, and another one featured an endogenous number of punishment stages. Also, an asymmetric treatment was used as a benchmark for evaluation of the impact of normative conflicts in the presence of feuding opportunities. The authors demonstrate that differences

in MPCR amplify the likelihood of a feud to erupt and decrease earnings, despite similar contribution levels.

The negative effects of asymmetric returns are corroborated by Reuben and Riedl (2013). They studied the influences of asymmetries in endowments and returns on cooperation. A standard symmetric PGG with a group size of 3 was used as the reference point for three original asymmetric treatments. In one of them, the returns of a lucky subject were increased from 0.5 to 0.75. In consonance with the previous study, heterogeneity in returns had adverse consequences, as reflected by the weakly significant increase in contributions of players with low returns after punishment opportunities were introduced.

Nikiforakis et al. (2010), O'Gorman et al. (2009), Carpenter et al. (2012) and Nosenzo and Sefton (2012), changed the allocation of punishment power within groups. In the first paper, authors found the results produced by symmetric and asymmetric punishment institutions to be very similar on the aggregate level. At the individual level, higher punishment efficiency was associated with higher earnings and higher punishment spending.

The other three papers are in disagreements about the relative performance of decentralised punishment institutions and those with a solitary punisher. The former are superior according to Carpenter et al. (2012) and Nosenzo and Sefton (2012), but O'Gorman et al. (2009) show that similar levels of cooperation can be sustained at a lower cost in the presence of a sole punisher. The differences in results could stem from variation in the assignment process to the special role. O'Gorman et al. (2009) reassigned the role at the beginning of every new period, whereas in the other two studies one individual was randomly chosen to hold this position for the whole game.

The literature is unanimous on the adverse effect of asymmetries in MPCR on cooperation, but not on the effects of asymmetries in punishment. To the best of our knowledge, no previous papers examined the effect of multiple concurrent asymmetries or asymmetries in punishment power in the presence of opportunities to feud. In the next section, we present the experimental design we use to close these gaps and to shed more light on the effects of the aforementioned asymmetries on cooperation.

1.3 The experiment

1.3.1 Experimental designs

Among numerous papers that used PGG with punishment opportunities Nikiforakis and Engelmann (2011), Nikiforakis et al. (2012) and Engelmann and Nikiforakis (2014) had the most profound influence on our design. Its general structure is standard, with each period consisting of two parts. The first part includes a single voluntary contribution stage (s=1), with four participants ($i \in \{1, ..., 4\}$) in each group simultaneously deciding how to split their 20 ECU endowment. They choose how much to allocate towards the public account (c_i) and how much to keep in their private accounts. Any point contributed to the first increases the earnings of each group member by α_i. The earnings of a player i after this phase are given by the following payoff function

$$\pi_i^1 = 20 - c_i + \alpha_i \sum_{j=1}^{4} c_j.$$

.

After observing own earnings subjects enter the second phase. They learn the exact contributions of other members and are enabled to assign costly punishment points[2].

To allow the eruption of intra-period feuds the number of punishment stages is not set ex-ante. Therefore, so long as participants keep punishing, further stages are entered, and players are informed about punishment decisions of the others in a previous stage. By restricting the punishment to integers, we ensure that the number of stages is finite[3].

In general, at any stage $s \geq 2$ a participant i, can decrease the earnings of a subject j, by assigning punishment points (p_{ij}^s), if and only if $\pi_i^{s-1} \geq 1$ and $\pi_j^{s-1} > 0$. Each assigned punishment point reduces the earning of a punisher by 1 ECU and of the punished subject by β_i ECU. The reduction is restricted, so sanctions cannot drive the earnings of a targeted individual below zero. Only if earnings left after deductions due to punishment by others are insufficient to cover the cost of own punishment, negative earnings can arise. To summarise, earnings of a subject i at any stage, beyond stage one, can be expressed by

[2] Participants can assign 0 points to abstain from punishing others.

[3] The number of punishment stages cannot exceed 64 in treatments with asymmetric punishment power or 33 in other treatments. Participants are not provided with these numbers but have enough information to calculate them.

$$\pi_i^s = max\{0, \pi_i^{s-1} - \sum_{j=1, j\neq i}^{4} \beta_j * p_{ji}^s\} - \sum_{j=1, j\neq i}^{4} p_{ij}^s$$

If no punishment is meted out at a particular stage, the period ends, and a new one begins. The same is true if by the end of a stage all group members have earnings below 1 ECU or three group members have non-positive earnings since such circumstances render further retaliation impossible. The experiment is 30 periods long to account for a long horizon welfare implications, previously discussed by Gächter et al. (2008) and Engelmann and Nikiforakis (2014).

Treatments

Table 1.1 Experimental design.

Treatment	MPCR	Effectiveness of punishment
Symmetric	{0.4, 0.4, 0.4, 0.4}	{3, 3, 3, 3}
Privilege	{0.7, 0.3, 0.3, 0.3}	{3, 3, 3, 3}
Strongman	{0.4, 0.4, 0.4, 0.4}	{9, 1, 1, 1}
Alpha	{0.7, 0.3, 0.3, 0.3}	{9, 1, 1, 1}

MPCR - marginal per capita returns from contributions to the public account; Effectiveness of punishment - a reduction effect of each punishment point on the earnings of a punished individual. The average MPCR and punishment efficiency in all treatments are 0.4 and 3 respectively. The numbers in the curly brackets show the respective distributions for groups in each treatment. The first number in a set is a parameter of "Participant 1" player.

Table 1.1 summarises our two by two experimental design. At the top of the list are parameters of the *Symmetric* treatment (ST), with equal distributions of

returns and punishment power. In the second and the third rows are the *Privilege* treatment (PT) with asymmetric returns and the *Strongman* treatment (SMT) with unequal distribution of punishment efficacy. The *Alpha* treatment (AT), as a model of groups dominated by strong punishers, incorporates both asymmetries, with a single group member benefiting from both higher MPCR and higher punishment efficacy.

The overall efficiency ($\sum_1^4 \alpha_i = 1.6$) is constant across treatments, *id est* total earnings attainable by a group in a single period cannot exceed 128 ECU regardless of the treatment. In the *Symmetric* treatment, which follows the *Rich Punishment* in Engelmann and Nikiforakis (2014), the gains from the "project" are dispensed equally between group members ($\{\alpha_i = 0.4\} \ \forall \ i$). In the *Privilege* treatment, which is parallel to the main treatment of Nikiforakis et al. (2012), any subject being randomly assigned the ID of 'Participant 1' enjoys higher MPCR, α_1=0.7, while others get just 0.3.

Both treatments above are characterised by the uniform distribution of punishment efficacy. More specifically at the cost of 1 ECU, earnings of a targeted individual are decreased by 3 ECU. In contrast, in the *Strongman* treatment, the pivotal ('Participant 1') players by sacrificing 1 ECU can lower the earnings of a punished individual by 9 ECU. Taking into account previously reported evidence (Nikiforakis et al., 2010), we keep the average punishment effectiveness constant across treatments. Therefore in the *Strongman* treatment, a punishment point assigned by a non-pivotal player reduces the income of a targeted subject by 1 ECU. While the distribution of punishment power in this treatment is skewed, the allocation of benefits is the same as in the *Symmetric* treatment.

In the *Alpha* treatment, one individual in every group enjoys higher returns

from cooperation and higher punishment efficiency. Therefore the parameters in this treatment are $\{\alpha_1 = 0.7, e_1 = 9\}$ for a subject with ID of 'Participant 1' (hereafter "alpha") and $\{\alpha_i = 0.3, e_i = 1\}\ \forall\ i \neq 1$.

1.3.2 Recruitment and procedures

The experiment was conducted in the Center for Decision Research and Economics (CeDEx) lab at the University of Nottingham using z-Tree (Fischbacher, 2007). 176 students[4] (11 groups per treatment) from the University of Nottingham were randomly recruited using ORSEE (Greiner, 2015). Three sessions were run in June 2015; the rest were conducted in late October and in the first half of November 2015.

Upon arrival at the lab, participants drew a number to choose their computer stations. As soon as all the participants were seated, the rules were read aloud by the experimenter. After everyone successfully answered all comprehension questions the 30-period experiment began.

On average, the experimental sessions lasted two and a half hours. The payoff was comprised of a participation fee (£3), the initial capital of 150 ECU (to cover potential losses) and summation of earnings in all rounds. At the end of the experiment, the earnings in ECU were converted to GBP at the exchange rate of 1 ECU = 2 pence and paid in cash. On average participants left the lab with approximately £20.5.

[4] Out of all participants 66 were male. Their share in each treatment was not substantially different (Pearson's χ^2 test: $p = 0.264$).

1.4 Experimental results

In this section, we present the analyses of the main experimental variables. We begin by examining differences between treatments in contributions to the public good. Next, we explore the differences in punitive behaviour, followed by the analysis of participants' earning. Further analyses provide insights into differences between *Privilege* and *Alpha*, bankruptcies in our experiment, as well as a comparison between the behaviour of the pivotal individuals and others.

To test the significance of detected discrepancies we mainly rely on two-tailed Mann-Whitney (hereafter MW) test with group averages as independent observations and multilevel mixed random effects regressions. In the latter, we regressed the main experimental variables, such as contributions to the public account, on treatment dummies (the *Symmetric* treatment is the base category), time variables ("Periods" for periods 11-30 and "Periods3" for the last three periods) and the interactions between them.

1.4.1 Contributions to the public account

Result 1. *First-round contributions are significantly higher in treatments with the symmetric MPCR.*

Figure 1.1 presents the evolution of average contributions in all four treatments over the 30 periods of the study. Examination of the diagram suggests that the first-round average contributions in treatments with the symmetric MPCR were higher than in the other two treatments (11.30 ECU, 11.39 ECU for *Symmetric* and *Strongman* respectively, versus 8.16 ECU for *Alpha* and 6.66 ECU for *Privilege*).

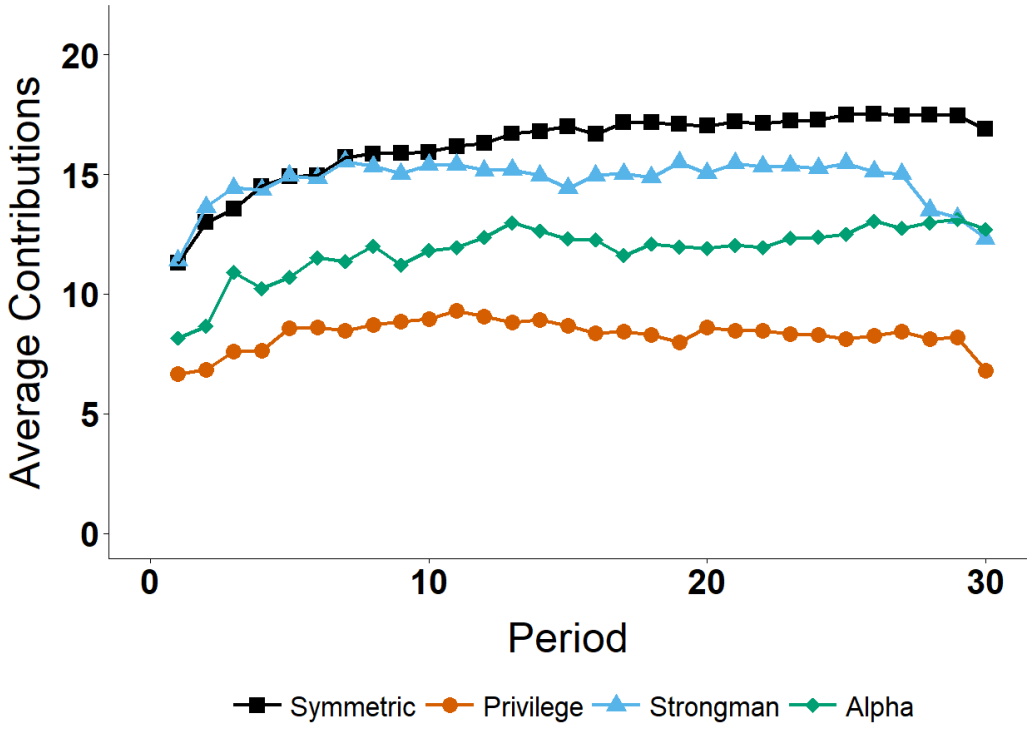


Figure 1.1: Evolution of average contributions by treatment.

According to a Kruskal-Wallis test with individual decisions as independent variables[5], first-period contribution choices do not come from the same distribution ($p = 0.006$). MW tests attribute this result to the highly significant differences between the treatments with symmetric returns from cooperation and the *Privilege* treatment (MW test on PT = ST: $p = 0.007$; PT = SMT: $p = 0.002$) and to differences with the *Alpha* treatment (MW test on AT = ST: $p = 0.075$; AT = SMT: $p = 0.044$).

Result 2. *Average contributions in the Symmetric treatment were higher than in the other three treatments.*

The trends in Figure 1.1 and Figure 1.2 suggest that an adjustment was taking place during the initial periods of the game, as participants learned the institutions and checked which strategies work for them. Therefore, in further analyses, we use data from the last 20 periods and regard the first 10 periods as a learning phase.

The second column in Table 1.2 (Panel A) gives an overview of contributions in the last 20 periods. Both Table 1.2 and Figure 1.1 suggest that during these periods the highest average contributions were observed in the *Symmetric* treatment (17.06 ECU); average contributions in *Strongman* were slightly lower (14.82 ECU), but higher than in the treatments with asymmetries in returns (8.40 and 12.39 ECU for *Privilege* and *Alpha* respectively).

A Kruskal-Wallis test with average contributions across relevant periods strongly rejects the null hypothesis that contributions come from the same distribution. By performing MW tests on group level averages, we establish the significance of differ-

[5] The observations are independent as contribution decisions were made before receiving any feedback on decisions of others.

ences between the *Symmetric* treatment and treatments with asymmetric returns (MW test on ST = PT: $p = 0.004$; ST = PT: $p = 0.040$). Except for the gap between the *Strongman* treatment and the *Privilege* treatment (MW test on SMT = PT: $p = 0.026$), all other differences are not significant (for p values see the second column in Panel B of Table 1.2).

Statistical insignificance of some of the above variations can be attributed to the inadequacy of the MW test to deal with panel data[6]. Therefore, to fully benefit from the information gathered during the sessions, we turn to regression analysis, namely multilevel mixed random effects regression which allows for individual and group differences, as well as for treatment-specific residuals.

The coefficients reported in Table B1, confirm that transfers to the public account were substantially higher in the *Symmetric* treatment than in the *Privilege* treatment over the 20 periods (Wald test on "Privilege Treatment (PT)": $p = 0.042$; "PT × Periods": $p < 0.001$; "PT × Periods" + "PT × Periods3": $p = 0.347$). Also, contributions in the *Symmetric* treatment were higher relative to the *Strongman* treatment, due to a small difference in trends, which substantially increased during the last periods (Wald test on "Strongman Treatment (SMT)": $p = 0.833$; "SMT × Periods": $p = 0.008$; "SMT × Periods" + "SMT × Periods3": $p < 0.001$). Finally, the *Symmetric* treatment outperformed the *Alpha* treatment, due to a small, but significant (at 10% level), difference in trends in periods 11 to 17 and despite the markedly better performance of the latter during the final periods (Wald test on "Alpha Treatment (AT)": $p = 0.224$; "AT × Periods": $p = 0.057$; "AT × Periods" + "AT × Periods3": $p = 0.044$).

Considering the above results, along with those presented in one of the section

[6] For an example and discussion see Harrison (2007).

below, where treatments with asymmetric returns are compared, we argue that asymmetries in punishment power in our study have mixed effects on cooperation levels. In the presence of equal returns from cooperation, they undermine groups prospects to sustain cooperation towards the end of the game. However, they have a slight positive effect on cooperation levels over time, when returns from cooperation are different (see Result 6.).

Table 1.2 Main variables across treatments (periods 11-30).

Treatment	Contributions	Punishment	Income Reduction	Earnings
Panel A (group-level std. dev. in parenthesis)				
Symmetric (ST)	17.06	0.10	0.39	29.85
	(6.41)	(0.17)	(0.67)	(4.07)
Privilege (PT)	8.40	0.66	2.48	22.56
	(7.98)	(1.85)	(3.24)	(5.93)
Strongman (SMT)	14.82	0.74	2.72	26.23
	(7.44)	(1.52)	(4.45)	(7.74)
Alpha (AT)	12.39	0.81	3.81	23.67
	(8.37)	(1.63)	(7.75)	(10.68)
Panel B: Statistical comparison (Mann-Whitney, two sided, p values)				
ST-PT	0.004	0.053	0.053	0.003
ST-SMT	0.171	0.148	0.113	0.125
ST-AT	0.040	0.117	0.142	0.059
PT-SMT	0.026	0.789	0.894	0.066
PT-AT	0.123	0.510	0.489	0.358
SMT-AT	0.554	0.741	0.792	0.693

1.4.2 Punishment

In this part of our analysis, we turn our attention to the spending on sanctions, the overall decrease in incomes it caused (sum of points spent on retribution in

addition to the reduction they caused) and the frequency of prolonged conflicts. Figure 1.2 shows that the highest levels of punishment spending took place in the first round (or first three rounds in case of the *Privilege* treatment). During the last two-thirds of the experiment, average spending on retaliation stabilised around 0 (with sporadic spikes) in the *Symmetric* treatment, at a slightly higher level for most periods in the *Strongman* treatment and it kept fluctuating in treatments with unequal MPCR.

Result 3. *(i) Participants in the Symmetric treatment on average spent considerably less on punishment than participants in the Strongman and the Privilege treatments.*
(ii) The overall income reduction due to punishment was the lowest in the Symmetric treatment.

The third column of Table 1.2 summarises trends in the last 20 periods - participants in the *Symmetric* treatment on average spent the least on sanctions (0.10 ECU), followed by participants in the *Privilege* treatment (0.66 ECU), the *Strongman* treatment (0.74 ECU) and the *Alpha* treatment (0.81 ECU). Using MW tests for group averages, we can reject the null hypothesis (the level of punishment spending is the same in both treatments) only for the gap between the *Symmetric* and the *Privilege* treatments (MW test $p = 0.053$).

The coefficients reported in the second column of Table B1 confirm that the average level of spending on retaliation in the *Symmetric* treatment was substantially lower than in the *Privilege* treatment, tough the gap was shrinking over time (Wald test on "Privilege Treatment (PT)": $p = 0.001$; "PT $\times$ Periods": $p <$

0.001; "PT × Periods" + "PT × Periods3": $p = 0.693$). Furthermore, the results suggest that for most rounds the levels of spending in the *Symmetric* and the *Strongman* treatments were statistically indistinguishable (Wald test on "Strong Treatment (SMT)": $p = 0.527$; "ST × Periods": $p = 0.503$). Only in the final rounds, the number of assigned punishment points was substantially higher in the latter (Wald test on "SMT × Periods" + "SMT × Periods3": $p < 0.001$). Finally, the introduction of multiple asymmetries did not affect the punishment spending considerably, as the variation between the *Symmetric* and the *Alpha* treatments is insignificant (Wald test on "Alpha Treatment (AT)": $p = 0.220$; "AT × Periods": $p = 0.979$; "AT × Periods" + "AT × Periods3": $p = 0.286$).

Due to the differences in the distribution of punishment efficiency, similar aggregate levels of punishment spending could lead to different consequences, but this is not the case according to MW tests. The nonparametric tests suggest that the only (weakly) significant difference in income reduction is between the *Symmetric* treatment and the *Privilege* treatment (MW test on ST = PT: $p = 0.053$; for the remaining p values see the fourth column of panel B in Table 1.2).

The insights about treatments' relative performance obtained from the regression analysis (column 3 in Table B1) are similar to those we got from the regression in column 2. The exceptions include the weakly significant level and last periods trends differences between *Symmetric* and *Alpha* (Wald test on "Alpha Treatment (AT)": $p = 0.081$; "AT × Periods": $p = 0.938$; "AT × Periods" + "AT × Periods3": $p = 0.093$).

The above results point at a considerable variability in punishment spending and income loss. Therefore, many of the differences that seem to be large are either weakly significant or insignificant.

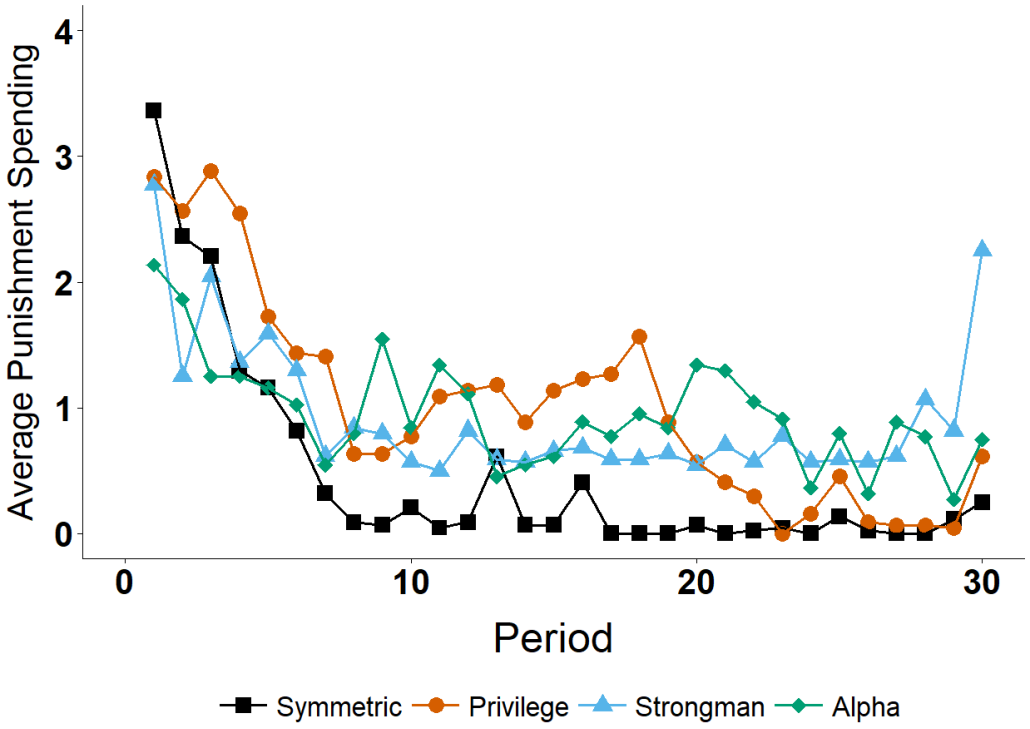


Figure 1.2: Evolution of average punishment spending by treatment.

Result 4. *(i) The share of rounds with prolonged intra-period conflicts in the Privilege treatment was higher than in any other treatment.*
(ii) Inter-period feuds were absent in the Symmetric treatment.

Before proceeding with the analysis, we need to define two types of feuds. The term intra-period feud, in our study, refers to cases with at least three punishment stages that involve at least one individual in both roles, the punisher and the punished, in the same period. The definition of inter-period feuds is straightforward - conflicts that spilled over into later periods. Although they might be easier to define, the inter-period feuds are more difficult to trace, since fixed identification numbers that enable delayed retaliation and discrepancies between contributions of clashing individuals make it very difficult to disentangle the motives driving the aggression. Therefore, we regard incidents of second stage punishment in the presence of equal contributions as occurrences of inter-period feuds, when they can be regarded as a delayed response to previous punishment.

The data show that in most periods in the analysis no punishment points were meted out (74% of all the periods in the *Alpha* treatment, 75% in the *Privilege* and the *Strongman* treatments, 92% in the *Symmetric* treatment). Moreover, frequently when punishment took place, punished individuals decided not to retaliate (52% of periods with punishment in the *Privilege* treatment, 70% in the *Alpha* treatment, 77% in the *Strongman* treatment and 85% in the *Symmetric* treatment). Consequently, the vast majority of periods were over within the first two punishment stages. Nonetheless, some participants engaged in conflicts that lasted longer.

Tendencies for prolonged within period conflicts varied across treatments and

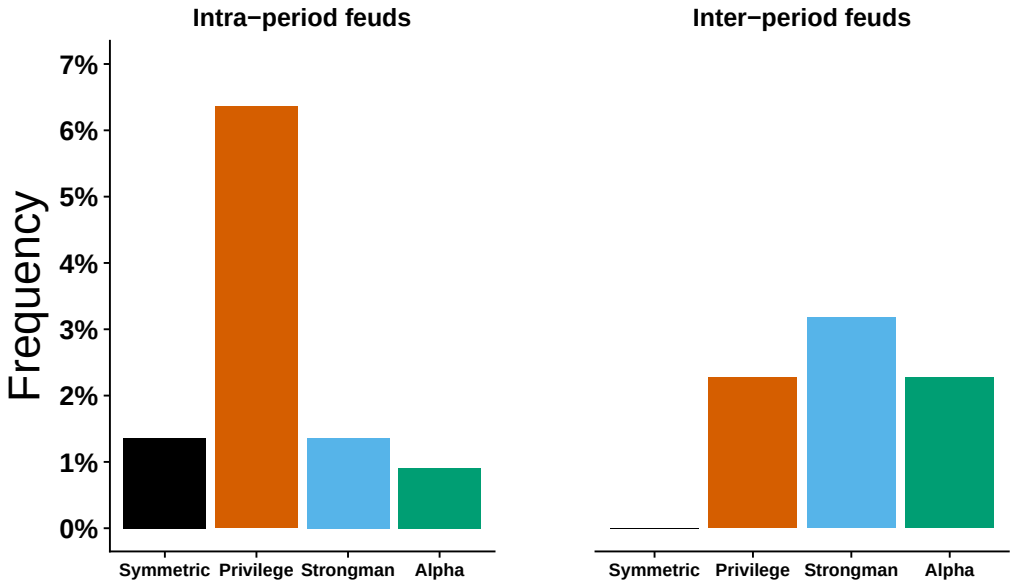


Figure 1.3: Frequency of feuds (out of the total number of periods).

some of them can be classified as feuds in accordance with our definition. The frequencies of intra-period feuds are displayed in the left bar chart of Figure 1.3, which suggests that they were the most frequent in the *Privilege* treatment.

We fit a logistic regression with group-level random effects and clustered standard errors by treatment to compare between tendencies to feud (the dependent variable gets 1 if in a certain period a feud took place and 0 otherwise). The results reported in column 1 of Table B2, corroborate that feuds were more likely to erupt in the *Privilege* treatment than in the *Alpha* treatment (Wald test on "Alpha Treatment": $p = 0.026$), the *Strongman* treatment (Wald test on "Strongman Treatment": $p = 0.048$), and somewhat more likely than in the *Symmetric* treatment (Wald test on "Symmetric Treatment": $p = 0.082$).

While intra-period feuds were most frequent in the *Privilege* treatment, Figure 1.3 suggests that the *Strongman* treatment had the highest incidence of inter-

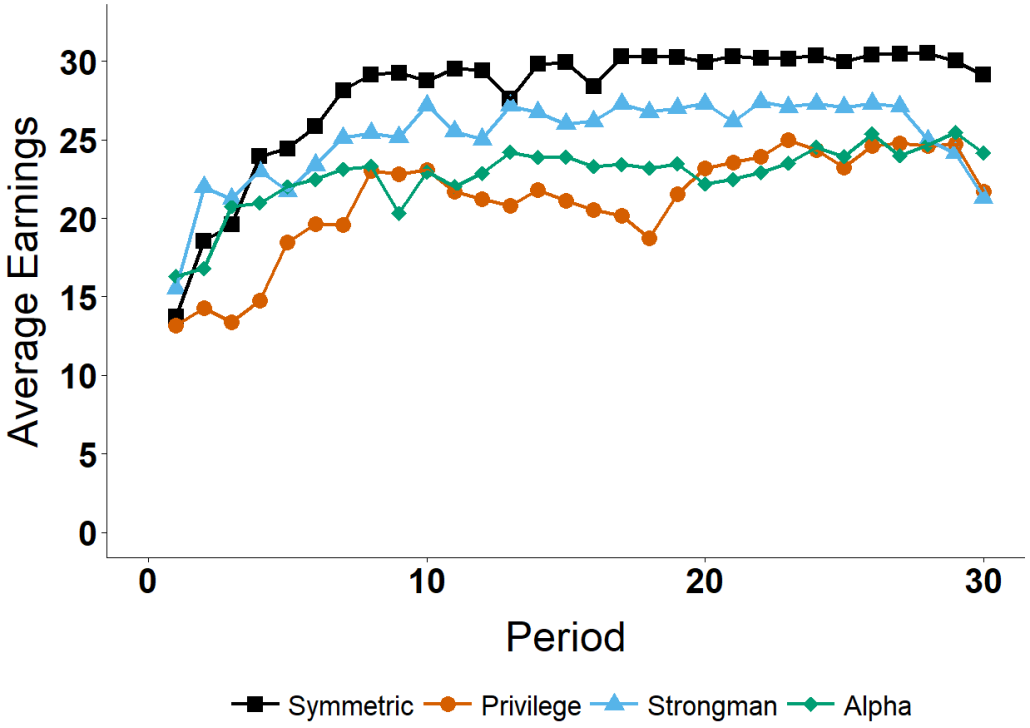


Figure 1.4: Evolution of average earnings by treatment.

period feuds (7 cases). No cases of inter-period feuds were observed in the *Symmetric* treatment, 5 in the *Privilege* and the *Alpha* treatments. The differences are not substantial as the estimated logistic model is statistically insignificant (see column 2 of Table B2).

1.4.3 Earnings

Result 5. *Earnings in the Symmetric treatment were higher than in other treatments.*

The discrepancies analysed in previous subsections translated into differences in earnings that are summarised in the last column of Panel A in Table 1.2. Both

Table 1.2 and Figure 1.4 show that the *Symmetric* treatment was characterised by the highest average level of earnings (29.85 ECU), the second highest in the *Strongman* treatment (26.23 ECU), followed by the *Alpha* (23.67 ECU) and the *Privilege* treatments (22.56 ECU).

The MW tests for group averages over the periods in the analysis confirm that average earnings were higher in the *Symmetric* treatment relative to the *Privilege* treatment (MW test on ST = PT: $p = 0.003$). The difference between the *Symmetric* and the *Alpha* treatments is only weakly significant (MW test on ST = AT: $p = 0.059$), and the hypothesis of equal earnings in the *Symmetric* and the *Strongman* treatments cannot be rejected at 10% level (MW test $p = 0.125$).

The regression analysis (for coefficient see column five in Table B1) supports most of the insights obtained using the nonparametric test. According to the estimated model the incentives set in the *Symmetric* treatment on average lead to substantially higher earnings than in the *Privilege* treatment (Wald test on "Privilege Treatment (PT)": $p < 0.001$; "PT × Periods": $p < 0.001$; "PT × Periods" + "PT × Periods3": $p = 0.451$). Also, regression analysis points out the similarity of average earnings in the *Symmetric* and the *Strongman* treatments throughout the game (Wald test on "Strongman Treatment (SMT)": $p = 0.347$; "SMT × Periods": $p = 0.829$). Only in the last periods as a consequence of the drop in average contributions, accompanied by surging demand for punishment, profits in the *Strongman* treatment dropped markedly relative to the *Symmetric* treatment (Wald test on "SMT × Periods" + "SMT × Periods3": $p < 0.001$). Lastly, there is (weak) evidence about higher earnings in the *Symmetric* treatment relative to the *Alpha* treatment (Wald test on "Alpha Treatment (AT)": $p = 0.055$; "AT × Periods": $p = 0.690$; "AT × Periods" + "AT × Periods3": $p = 0.020$).

1.4.4 Privilege in contrast to Alpha

This part of the analysis is dedicated to the comparison between the *Privilege* treatment and the *Alpha* treatment. By comparing these two treatments, we aim at checking whether the presence of "alphas" allowed groups to suppress the tensions arising from asymmetric returns.

Result 6. *On average participants contributed slightly more in the Alpha treatment.*

The postestimation analysis of the model presented in column 1 of Table B1 indicates that the gap in contributions between the two treatments was statistically significant (Wald on test on "Privilege Treatment (PT)" = "Alpha Treatment (AT)": $p = 0.420$; "PT × Periods" = "AT × Periods": $p = 0.017$; "PT × Periods" + "PT × Periods3" ="AT × Periods" + "AT × Periods3": $p = 0.012$), but not substantial (on average 0.82 ECU per period).

Result 7. *There is weak evidence on lower average punishment spending and overall loss of earnings due to punishment in the Alpha treatment.*

On the one hand, the estimated level of punishment spending and its overall impact on earnings after the first 10 periods are lower in the *Alpha* treatment, but the differences are only weakly significant (Wald test on "Privilege Treatment (PT)" = "Alpha Treatment (AT)" p values are 0.066 and 0.091, columns 2 and 3 respectively). On the other hand, the negative trends in punishment spending were stronger in the *Privilege* treatment (Wald test on "PT × Periods" = "AT × Periods": $p = 0.003$; "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.254$). Same is true for the descent in punishment driven loss

of earnings (Wald test on "PT × Periods" = "AT × Periods": $p < 0.001$; "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 1.000$).

Result 8. *On average participants in Privilege earned slightly more than in Alpha.*

Ultimately, the tendencies in contributions and punishment shaped participants' earnings. Regression analysis indicates that average earnings did not vary substantially after the first ten periods (Wald test on "Privilege Treatment (PT)" = "Alpha Treatment (AT)": $p = 0.116$). As the game proceeded further, average earnings in the *Privilege* treatment outstripped those in the *Alpha* treatment (Wald test "PT × Periods" = "AT × Periods": $p < 0.001$). In the final three rounds trends changed, as the average earnings in the *Privilege* treatment were falling, while in the *Alpha* treatment the positive trend line became even steeper (Wald test on "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.021$).

The insignificance of differences in the intercepts leads to the conclusion that on average individuals in the *Privilege* treatment earned more in contrast to their peers in the *Alpha* treatment. The estimated average income gap is about 1.88 ECU per period.

Result 9. *The Alpha treatment is characterised by a higher proportion of groups successful in reaching high levels of cooperation.*

Only in a single group with a "privileged" individual full contributions were observed for 15 periods or more (out of 20 we analyse) and in another one for just one period. By contrast, 4 groups in the *Alpha* treatment managed to sustain full cooperation for at least three fourth of the analysed periods. Another two

groups managed to do the same for 4 and 8 periods respectively. As a result, 6 out of 11 groups in the *Alpha* treatment had average earnings greater than 30 ECU, in comparison with a single group in the *Privilege* treatment[7]. By performing a MW test, we find this difference in proportions to be statistically significant ($p = 0.025$).

The above result requires clarification since earlier we showed that the earnings gap on the treatment level was modest and in favour of the *Privilege* treatment. The explanation lies in two groups in the *Alpha* treatment with exceptionally low average earnings (3.24 and 5.03 ECU). They can be easily spotted in Figure 1.5, due to their proximity to the x-axis and the warm colours pointing at extreme levels of punishment that contributed towards unprecedented bankruptcies of some individuals[8]. Additional information about these groups is presented in the following subsection. At this point, it is worth noting that no bankruptcies occurred in the *Privilege* treatment.

Further visual inspection of Figure 1.5 suggests that earnings in the *Alpha* treatment were more volatile. This observation is reflected in the *Alpha* treatment's greater variances (114.11) relative to the *Privilege* treatment (35.20), although the difference is not statistically significant at 10% level (Leven's test $p = 0.333$).

1.4.5 Bankruptcies

None of the previous studies with the endogenous number of punishment stages (Engelmann and Nikiforakis, 2014; Nicklisch and Wolff, 2011; Nikiforakis and Engelmann, 2011; Nikiforakis et al., 2012) report on subjects who lost all of their

[7] 32 ECU is the maximum attainable average profit for a group in our setting.

[8] A subject was defined as bankrupt if at some point in the experiment the total accrued earnings, including the initial capital, were supposed to turn negative.

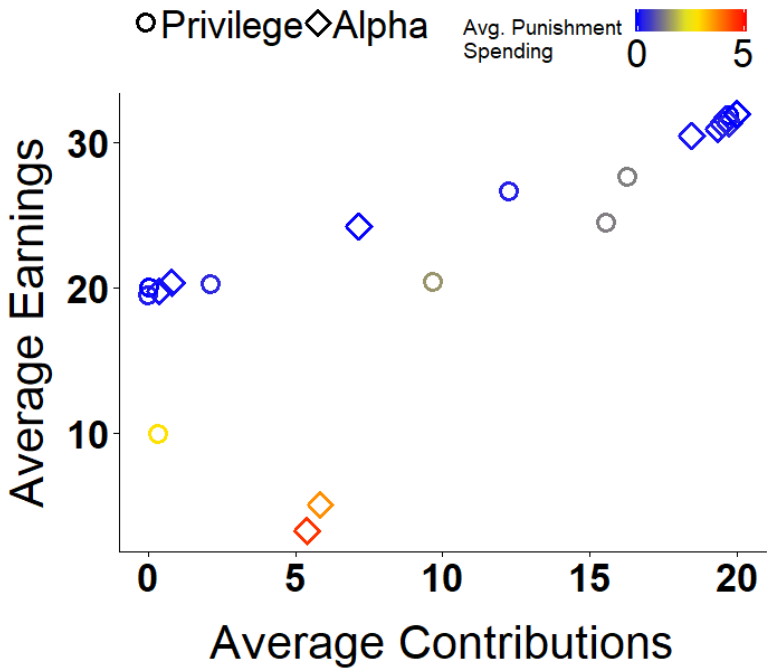


Figure 1.5: Groups' average contributions, earnings and punishment spending in ECU.

earnings after amassing negative profits in multiple periods. In our study, with the introduction of asymmetry in punishment power, three subjects did go bankrupt.

Two of the bankrupt subjects took part in the *Alpha* treatment sessions and another one in the *Strongman* treatment. In groups with the "alpha" individuals, contribution levels were not high, but most group members contributed their full endowment at various points of the experiment. In the *Strongman* treatment group, low contributions by the individual with the higher punishment efficiency decreased the average contributions over time, and by 17th period, all group members contributed nothing towards the public good (in later rounds small contributions between 1 and 3 ECU were recorded).

The bankruptcies, however, cannot be attributed to the low levels of cooperation per se. Other groups in the same treatments failed in reaching high levels

of cooperation, but the consequences were not that devastating. Furthermore, as was mentioned in the previous subsection, no participants went bankrupt in the *Privilege* treatment group with similar levels of punishment and even lower average contributions (see Figure 1.3).

Apparently, what distinguishes these groups from others is the commitment of some subjects to retribution, even when facing a disproportionally stronger opponent. The bankrupt participants in the *Alpha* treatment spent on average 7.3 and 6.6 ECU per period throughout the whole game, whereas participants in other groups spent on average to 0.21 ECU (SD = 1.08). The bankrupt subject of the *Strongman* treatment spent on average 15.8 ECU in contrast to the average of 0.5 ECU (SD = 2.97) in the remaining groups in that treatment.

We can only hypothesise about the reasons behind such behaviour. It is doubtful that the bankrupt subjects misunderstood the instructions since none of them had any difficulties in answering control questions.

One potential explanation might lie in the domain of strong preferences for justice, delivered in the form of equalisation of pain (Frijda, 1994; Heider, 2013). The inequality in punishment power allowed stronger participants to maintain relatively high profits after they exchanged blows. Therefore, the rank and file could afflict a comparable amount of damage only through excessive expenditure on punishment, which exposed them to negative earnings.

1.4.6 Pivotal participants in contrast to others

We finalise the analysis by comparing the performance of individuals with higher capacities to other co-players. For this purpose, we estimate similar regressions as

Table 1.3 Main variables by treatment and role (periods 11-30).

Treatment	Contributions	Punishment	Income Reduction	Earnings
Symmetric	17.06	0.10	0.39	29.85
	(6.41)	(0.17)	(0.67)	(4.07)
Privilege - Participant 1	9.95	0.69	2.23	31.35
	(8.81)	(1.07)	(3.26)	(14.65)
Privilege - Rank and File	7.88	0.65	2.56	19.63
	(7.83)	(0.86)	(3.26)	(3.85)
Strongman - Participant 1	14.23	0.72	2.87	26.61
	(7.26)	(1.11)	(6.05)	(9.84)
Strongman - Rank and File	15.02	0.74	2.66	26.11
	(7.56)	(1.81)	(4.08)	(7.30)
Alpha - Participant 1	12.86	1.25	2.98	38.86
	(8.33)	(2.61)	(6.19)	(18.47)
Alpha - Rank and File	12.23	0.67	4.08	18.61
	(8.48)	(1.32)	(8.28)	(9.16)

above. This time, however, with group-level random effects only and interactions of "Participant 1" dummy with the independent variables included in the earlier models (for coefficients check Table B3). Although the special ID is meaningless in the *Symmetric* treatment, we do not omit this treatment. Instead, the "ST" dummy takes the same value for all observations regardless of the ID.

Result 10. *Pivotal subjects on average contributed more than others only in the Privilege treatment.*

Table 1.3 breaks down the average contributions, punishment spending, the loss of earnings and earnings by treatment and role. The second column demonstrates that the *Privilege* treatment is characterised by the greatest difference in average contributions of "Participant 1" players and others.

According to the estimated model (column 1 in Table B3), the differences in contribution levels, for the "alpha" and "strongman" individuals relative to other group members were neither statistically different after the first 10 periods (Wald tests on "AT - Participant 1": $p = 0.810$; "SMT - Participant 1" = "SMT - Rank and file": $p = 0.463$), nor in the course of the game (Wald tests on "AT - Par.1 × Periods": $p = 0.415$; "AT - Par.1 × Periods" + "AT - Par.1 × Periods3": $p = 0.576$; "SMT - Par.1 × Periods" = "SMT - R&F × Periods": $p = 0.901$; "SMT - Par.1 × Periods" + "SMT - Par.1 × Periods3" = "SMT - R&F × Periods" + "SMT - R&F × Periods3": $p = 0.300$). In contrast, the average contributions level after the first 10 periods of the "privileged" was significantly higher relative to other group members (Wald test on "PT - Participant 1" = "PT - Rank and file": $p < 0.001$), the gap persisted over time, but became narrower at the end of the game (Wald tests on "PT - Par.1 × Periods" = "PT - R&F × Periods": $p = 0.341$; "PT - Par.1

× Periods" + "PT - Par.1 × Periods3" = "PT - R&F × Periods" + "PT - R&F × Periods3": $p = 0.043$).

Result 11. *Pivotal participants neither spent more on retribution nor lost more ECU due to the presence of punishment opportunities.*

None of the differences are sufficiently strong to suggest that on average subjects with the ID of "Participant 1" spent more on punishment in any of the treatments (the respective Wald test p values for the *Alpha* treatment, the *Privilege* treatment, and the *Strongman* treatment are 0.487, 0.901 and 0.393 for the intercepts; 0.935, 0.878 and 0.425 for the slope of the trend line; 0.716, 0.378 and 0.559 for the trends in the final 3 periods). When it comes to the income reduction due to punishment the bottom line is the same, as we find no marked differences when comparing pivotal participants to the rank and file (the respective Wald test p values for the *Alpha* treatment, the *Privilege* treatment, and the *Strongman* treatment are 0.586, 0.553 and 0.408 for differences between intercepts; 0.959, 0.635 and 0.392 for the slopes of the trend lines; 0.967, 0.561 and 0.643 for the slope in last 3 periods).

Result 12. *(i)"Alphas" and "privileged" on average earned more than the other members of their groups.*
(ii)Average earnings of "alphas" exceeded those of the participants in the Symmetric treatment.

According to the results earnings of pivotal individuals in the *Alpha* and the *Privilege* treatments are substantially higher than those of the other players in their groups (Wald tests on "AT - Participant 1": $p < 0.001$; "PT - Participant 1"

= “PT - Rank and File”: $p < 0.001$). Because the evolution of average earnings over time was not substantially different for the special players (Wald tests on “AT - Par.1 × Periods”: $p = 0.871$; “AT - Par.1 × Periods” + “AT - Par.1 × Periods3”: $p = 0.525$; “PT - Par.1 × Periods” = “PT - R&F × Periods”: $p = 0.569$; “PT - Par.1 × Periods” + “PT - Par.1 × Periods3” = “PT - R&F × Periods” + “PT - R&F × Periods3”: $p = 0.578$) we conclude that both “alphas” and “privileged” enjoyed higher average earnings than their co-group members. “Strongmen”, however, did not translate their advantage in punishment efficiency into higher earnings (Wald tests on “SMT - Participant 1” = “SMT - Rank and File”: $p = 0.176$; “SMT - Par.1 × Periods” = “SMT - R&F × Periods”: $p = 0.277$; “SMT - Par.1 × Periods” + “SMT - Par.1 × Periods3” = “SMT - R&F × Periods” + “SMT - R&F × Periods3”: $p = 0.922$).

The above results suggest that the pivotal participants in treatments with asymmetric returns on average enjoyed higher earnings than other co-players. Nonetheless, only for “alphas” the average level of earnings exceeded that in the *Symmetric* treatment (Wald tests on “AT - Participant 1” = “ST”: $p = 0.013$; “AT - Par.1 × Periods” = “ST × Periods”: $p = 0.776$; “AT - Par.1 × Periods” + “AT - Par.1 × Periods3” = “ST × Periods” + “ST × Periods3”: $p = 0.165$). “Privileged” on the other hand had a similar level of earnings as the subjects in the *Symmetric* treatment (Wald tests on “PT - Participant 1” = “ST”: $p = 0.727$; “PT - Par.1 × Periods” = “ST × Periods”: $p = 0.232$; “PT - Par.1 × Periods” + “PT - Par.1 × Periods3” = “ST × Periods” + “ST × Periods3”: $p = 0.868$).

1.5 Discussion and conclusions

In this paper, we used two by two experimental design, based on a linear public goods game with punishment opportunities, to probe into the influence of asymmetries on cooperation. Participants in this study interacted for 30 periods in fixed groups of 4 and were able to punish and counter punish so long as both punishers and punished had sufficient resources. The endogeneity of punishment stages, along with fixed identification labels, allowed for intra- and inter-period feuds to arise.

Our results indicate that the introduction of both asymmetries in MPCR and punishment efficacy negatively affected groups' ability to cooperate but to various extents. While negative effects of asymmetry in punishment power were restricted to the final interactions of the experiment, the heterogeneity in MPCR substantially impaired groups' ability to reach high levels of cooperation from the very first period of the game (even before receiving any feedback on others' actions).

Considering the previous literature these results support reports about adverse effects of asymmetries in MPCR (Nikiforakis et al., 2012; Reuben and Riedl, 2013), but are at odds with Nikiforakis et al. (2010) who suggest that symmetric and asymmetric institutions produce similar results. Moreover, we demonstrate that in the presence of feuding opportunities the more efficient punishers did not obtain higher earning and did not spend more on punishment, as reported by Nikiforakis et al. for setting that lacks feuding opportunities.

Transformation of the "privileged" into "alphas", through bundling up the higher returns from cooperation and the more efficient punishment mechanism, lead to greater variability in performance of groups; consequently, the differences with the

Symmetric treatment were only weakly significant, but they demonstrate that on average participants in the *Alpha* treatment contributed less, lost more resources to punishment and on average enjoyed lower earnings. From the personal perspective, however, "alphas" benefited markedly from the unequal distribution of resources, as the level of earnings they managed to keep exceeded that in the *Symmetric* treatment.

As with any experimental study generalisations of our results should be made with a high degree of caution and alternative interpretations of the results should be considered. With this in mind, our results suggest that cooperation in small groups would be better promoted by egalitarian institutions, which keep the tensions low and allow swift enforcement of norms through decentralised punishment.

The above observation is probably not robust to an increase in groups' size. Although such changes were not part of our design, it is reasonable to suggest that as groups grow, it becomes more challenging to monitor the actions of all members and to reconcile different interests to avoid conflicts. It is in the presence of conflicts that concentration of punishment power is more likely to produce high levels of cooperation.

Our results suggest that disparity in punishment power is critical in differentiating between the performances of treatments with asymmetric returns that model a normative conflict. Considering the damage capacity of even the most primitive tools to human bodies, a creation of substantial differences in punishment power would require a sufficient level of division of labour and resources for recruitment of professional bodyguards and soldiers. If these conditions are not satisfied the rank and file should be able to keep in check their leaders and maintain egalitarian institutions.

Finally, the random allocation to the positions of power allowed us to circumvent the self-selection to those positions and to check the influence of different institutions on cooperation. In reality, however, top places within a hierarchy are not occupied by random individuals. Therefore, we need to understand to what extent the random assignment to the advantageous positions affected behaviour in our study. In the next chapter, we examine the behaviour in the same four treatments, after replacing the random nomination to the pivotal positions with a real effort task.

1.6 Bibliography

Boehm, C. (1997). Impact of the human egalitarian syndrome on Darwinian selection mechanics. *The american naturalist*, 150(S1):S100—-S121.

Boehm, C. (2009). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.

Carpenter, J., Kariv, S., and Schotter, A. (2012). Network architecture, cooperation and punishment in public good experiments. *Review of Economic Design*, 16(2-3):93–118.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Cinyabuguma, M., Page, T., and Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9(3):265–279.

Denant-Boemont, L., Masclet, D., and Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic theory*, 33(1):145–167.

Diamond, J. M. (1998). *Guns, germs, and steel: A short history of everybody for the last 13,000 years*. Random House.

Engelmann, D. and Nikiforakis, N. (2014). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, pages 1–17.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Frijda, N. H. (1994). The lex talionis: On vengeance. *Emotions: Essays on emotion theory*, pages 263–289.

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907):1510.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Harrison, G. W. (2007). House money effects in public good experiments: Comment. *Experimental Economics*, 10(4):429–437.

Heider, F. (2013). *The psychology of interpersonal relations*. Psychology Press.

Isaac, M. and Walker, J. (1988). Isaac Walker P88 Group Size Effects in Public Goods ProvisionThe Voluntary Contribution Mechanism. *The Quartely Journal Of Economics*, 103(1):179–199.

Nicklisch, A. and Wolff, I. (2011). Cooperation Norms in Multiple-Stage Punishment. *Journal of Public Economic Theory*, 13(5):791–827.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1):91–112.

Nikiforakis, N. and Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization*, 78(3):319–332.

Nikiforakis, N., Normann, H.-T., and Wallace, B. (2010). Asymmetric enforcement of cooperation in a social dilemma. *Southern Economic Journal*, pages 638–659.

Nikiforakis, N., Noussair, C. N., and Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, 96(9):797–807.

Nosenzo, D. and Sefton, M. (2012). Promoting Cooperation: The distribution of reward and punishment power. Technical report, CeDEx Discussion Paper Series.

O'Gorman, R., Henrich, J., and Van Vugt, M. (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings. Biological sciences / The Royal Society*, 276(1655):323–9.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a Sword: Self-Governance Is Possible. *American Political Science Review*, 86(02):404–417.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Wroblewski, E. E., Murray, C. M., Keele, B. F., Schumacher-Stankey, J. C., Hahn, B. H., and Pusey, A. E. (2009). Male dominance rank and reproductive success in chimpanzees, Pan troglodytes schweinfurthii. *Animal Behaviour*, 77:873–885.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1):110–116.

Appendix A: Instructions

Please note that in the instructions below, we mark treatment-specific text as follows: [Symmetric], {Privilege}, |Strongman| ,<Alpha>. The actual instructions only contained the relevant text with no markings.

You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions.

The instructions were distributed to all participants. **Please do not communicate with the other participants during the experiment.** Should you have any questions please raise your hand.

During the experiment we shall not speak of Pounds, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECU. At the end of the experiment the total amount of ECU you have earned will be converted to £ at the rate of **1 ECU = 2 p** and will be immediately paid to you in cash. Your **£6** show up fee consists of two parts. **£3** will be paid by the end of the experiment regardless of your results. Another **£3** will be used to cover potential negative earnings that might arise during the experiment.

At the beginning of the experiment the participants will be randomly divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of each group will remain the same throughout the experiment.**

The experiment will last 20 periods. In the beginning of the experiment, after you are randomly divided into groups, **each participant in your group will be given a number for identification** (i.e. Participant 1, Participant 2, Participant 3, and Participant 4). **Each participant will keep this number throughout the experiment.** No one will learn the participants' identities after the experiment.

Each of the 20 periods is divided into a number of stages.

The first stage

At the beginning of each of the 20 periods each participant will receive 20 ECU. In the following, we shall refer to this amount as the "endowment". Your task in the first stage is to decide how to use your endowment. **You have to decide how many of the 20 ECU you want to contribute to a project** (from 0 to 20) **and how many of them to keep for yourself**. The consequences of your decision are explained in detail below.

Once all the participants have chosen their contribution to the project you will be informed about the group's total contribution, your earnings from the project and your earnings at the end of the first stage. Your earnings at the end of the first stage are calculated using the following formula. If you have any difficulties do not hesitate to ask us.

Earnings at the end of stage 1 = Endowment of ECU - Your contribution to the Project + α*Total contribution to the Project

---

This formula shows that your earnings at the end of the first stage consist of two parts:

1. The ECU which you have kept for yourself (endowment – contribution)

2. The earnings from the project, which equal to the α share of the group's total contribution.**The α [\|is 0.4 for all four members of the group\|] <{for Participant 1 is 0.7 and 0.3 for the other three members of the group}>.**

The earnings of each group member from the project are calculated in the same way. Suppose the sum of the contributions of all group members is 60 ECU. In this case, [\|earnings of each group member from the project are 0.4*60=24 ECU\|] <{Participant 1 earns 0.7*60=42 ECU from the project, while the other members of the group receive 0.3*60=18 ECU}>. If the total contribution to the project is 9 ECU, then earnings are [\|0.4*9=3.6 ECU\|] <{0.7*9=5.6 ECU and 0.3*9=2.7 ECU respectively}>.

You always have the option of keeping the ECU for yourself or contributing them to the project. Each ECU that you keep raises your end of period earnings by 1 ECU. Supposing you contributed this ECU to the project instead, then the total contribution to the project would rise by 1 ECU. Your earnings from the project would thus rise by [\|**0.4***1 ECU\|] <{α*1 ECU}>. The earnings of the other group members would also rise [\|by 0.4 ECU\|] <{ in accordance with their α }>, so that the total earnings of the group from the project would be increased by 1.6 ECU. Your contribution to the project therefore also raises the earnings of the other group members. On the other hand you also gain earnings for each ECU contributed by the other members to the project. In particular, for each ECU contributed by any member, you earn [\|0.4 ECU\|] <{0.7 ECU in case you are Participant 1 and 0.3 otherwise}>.

The second stage

At the second stage you will be informed about how much each group member contributed individually to the project at the first stage. At this stage you can **reduce the earnings of each member of your group by distributing points or leave unaltered by assigning 0 points**. The other group members can also reduce your earnings if they wish to.

To reduce another participant's earnings you will have to distribute points. Each point you distribute will cost you 1 ECU and will reduce the earnings of the person you assign it to by β ECU. **[{The β is 3 for all four members of the group (each assigned point reduces the earnings of a targeted participant by 3 ECU).}] <|The β for Participant 1 is 9 (each point that Participant 1 assigns, reduces the earnings of a targeted participant by 9 ECU) and $\beta = 1$ for the other three members of the group (each point that other participants assign, reduces the earnings of a targeted participant by 1 ECU).|>** By choosing 0 points you will leave the earnings of a participant unchanged.

Example: [{Suppose that you are Participant 1 and you assign 2 points to Participant 2. This costs you 2 ECU and reduces Participant 2's earnings by 2*3 = 6 ECU.}]

<|Suppose that you are Participant 1 and you assign 2 points to Participant 2. This costs you 2 ECU and reduces Participant 2's earnings by 9*2 = 18 ECU. In case you are not Participant 1, 2 points still cost you 2 EUCs, but reduce the earnings of a targeted individual by 1*2 = 2 EUC.|>

Your total earnings from the two stages are therefore calculated as follows:

Total earnings (in ECU) at the end of stage 2 =

Earnings from the 1^{st} stage – β*Points you receive – Points you assign

---

Please note that your payoff for a round can be reduced to 0 in case the total amount of points to be deducted, due to the received points, is equal to or exceeds your earnings in ECU. Only if your earnings are insufficient to cover the cost of the points you assigned to others, your earnings become negative. Once your earnings are negative or zero, you no longer can assign points to others and no further points can be assigned to you. This is true for any stage of any round.

For example assume that your earnings from previous stage are 30 ECU. In the current stage you received [{4 points from Participant 2 and additional 8 points from Participant 1}] <|9 points from Participant 2 and additional 3 points from Participant 1|>. Therefore, your earnings should be reduced by [{3*4+3*8=36 ECU}] <|9*1+3*9=36 ECU|>. According to the rules, because 36>30 your earnings will be reduced to zero, provided you allocated no points in this round. If, in addition, you allocated 5 points in total, your earnings will fall to -5 by the end of this stage, as a result you will not be able to reduce the earnings of others and others will not be allowed to reduce yours till the end of the round.

If your earnings are negative at the end of the period, this will be covered by the points you accumulated in the previous stages and by the 150 ECU that we gave you in the beginning. It is very unlikely, but possible, that due to negative payoffs in multiple periods you will run out of ECU. Under such circumstances you will have to wait till

the end of the session and be paid the second half of your participation fee.

If none of the members of your group distributes points then the period finishes and the next period begins again with stage one. The period also finishes if none of the members of your group are allowed to assign further points. This will be explained below. Otherwise, a third stage will follow.

The third stage

In the third stage, you will be informed of the points that each person in your group assigned to you and the other members in your group. Similarly, the other members of your group will be informed about how many points you assigned to each of them. Then you can again **reduce or leave unaltered the earnings of each member of your group**. The other group members can also reduce your earnings if they wish to.

Note that in order for someone to be able to assign points at this stage he or she must have at least 1 ECU (i.e. the cost of one point). Additionally, if one member has non-positive earnings, the others will not be able to reduce her earnings further.

The costs of assigning points, as well as the earnings reduction caused by each point remain the same as before.

Since a group member can only assign points if her current earnings are at least 1 ECU, the period ends when the earnings of all members of a group has been reduced to less than 1 ECU. Furthermore, since it is not allowed to assign points to a member who already has zero or negative earnings, the period ends if three members have non-positive earnings, because the fourth member of the group cannot reduce their

earnings further; hence the third stage will only be entered if after the second stage at least one group member has 1 ECU left; at least another one has retained earnings greater than 0 and points were assigned in the second stage.

Fourth stage and beyond

Next stages are the **same as stage 3**. **After being informed of the points distributed in your group you will be able to assign further points.** The costs of assigning points, as well as the earnings reduction caused by each point remain the same as before. The fourth, fifth, or any further stage will only be entered if at least one member of the group assigned points in the previous stage.

When does a period end?

A period ends and a new one begins when one of the following occurs.

- No points are distributed in a given stage.
- Points are distributed, but no participants would be allowed to assign any more points if another stage followed. This can happen if nobody in a group has at least 1 ECU left or if three of the individuals have non-positive earnings.

[{**Example:** Assume that after the third stage, the earnings are as follows:

Participant 1: 2 ECU

Participant 2: 10 ECU

Participant 3: 3 ECU

Participant 4: 2 ECU

Assume that in the fourth stage Participant 4 assigns 2 points to Participant 3 and no more points assigned. Hence the earnings are

Participant 1: 2 ECU

Participant 2: 10 ECU

Participant 3: 0 ECU (= max (3 ECU – 3*2 ECU, 0)

Participant 4: 0 ECU (= 2 ECU – 2 ECU)

Since points were assigned in stage four and two participants have positive earnings, a fifth stage follows. If in the fifth stage Participant 2 assigns 2 points to Participant 1 and Participant 1 assigns 1 point to Participant 2, the earnings are

Participant 1: -1 ECU (= max (2 ECU – 3*2 ECU, 0) – 1 ECU)

Participant 2: 5 ECU (= max (10 ECU - 3*1 ECU, 0) – 2 ECU)

Participant 3: 0 ECU

Participant 4: 0 ECU}]

<|**Example:** Assume that after the third stage, the earnings are as follows:

Participant 1: 2 ECU

Participant 2: 12 ECU

Participant 3: 3 ECU

Participant 4: 2 ECU

Assume that in the fourth stage Participant 3 assigns 3 points to Participant 2 and no more points assigned. Hence the earnings are

Participant 1: 2 ECU

Participant 2: 12 ECU

Participant 3: 0 ECU (= 3 ECU – 3 ECU)

Participant 4: 0 ECU (= max (2 ECU – 3*1 ECU, 0)

Since points were assigned in stage four and two participants have positive earnings, a fifth stage follows. If in the fifth stage Participant 2 assigns 2 points to Participant

1 and Participant 1 assigns 1 point to Participant 2, the earnings are

Participant 1: -1 ECU (= max (2 ECU – 2*1 ECU, 0) – 1 ECU)

Participant 2: 1 ECU (= max (12 ECU - 1*9 ECU, 0) – 2 ECU)

Participant 3: 0 ECU

Participant 4: 0 ECU|>

The period now ends, because only Participant 2 can assign more points, but because all other participants have non-positive earnings, Participant 2 is not allowed to assign points to any of them.

If you have any further questions please raise your hand and one of the supervisors will come to help you.

Control Questionnaire

1. Each group member has an endowment of 20 ECU. Nobody (including you) contributes any ECU to the project. How high are:

   (a) Your earnings after the first stage, if you are Participant 1?

   (b) Your earnings after the first stage, if you are Participant 3?

2. Each group member has an endowment of 20 ECU. You contribute 20 ECU to the project. All other group members contribute 20 ECU each to the project. What are:

   (a) Your earnings after the first stage, if you are Participant 1?

   (b) Your earnings after the first stage, if you are Participant 4?

4. Each group member has an endowment of 20 ECU. The other three group members contribute together a total of 30 ECU to the project. What are:

   (a) Your earnings after the first stage if you contribute 0 ECU to the project, if you are Participant 1?

   (b) Your earnings after the first stage if you contribute 0 ECU to the project, if you are Participant 2?

5. Your earnings from the first period 25 ECU. How much will your earnings be at the end of the second stage be if:

   (a) You receive 2 points from Participant 2, but do not assign any yourself?

   (b) You receive 2 points from Participant 1 and assign 3 points yourself?

6. *True or false:* Each participant will be given an identification number which will allow others in their group to follow their actions in the experiment.

7. *Yes or NO?* Assume you assign 2 points to another group member, no one else in your group assigns any points and all members in your group have positive earnings. Will another stage follow?

8. *Yes or NO?* Assume no member of your group assigns points including you. Will another stage follow?

Appendix B: Regression Tables

Table B1 Main variables as a function of treatment and time.

	(1) Contrib.	(2) Pun. Spending	(3) Loss of Earnings	(4) Earnings
Privilege Treatment (PT)	−6.320**	1.989***	7.341***	−11.133***
	(3.105)	(0.588)	(2.135)	(3.186)
Strongman Treatment (SMT)	−0.656	0.346	2.443	−2.969
	(3.120)	(0.548)	(2.114)	(3.158)
Alpha Treatment (AT)	−3.798	0.748	3.612*	−6.043*
	(3.122)	(0.610)	(2.071)	(3.148)
Periods (11-30)	0.071***	−0.014**	−0.055***	0.097***
	(0.011)	(0.007)	(0.019)	(0.019)
Periods3 (28-30)	−0.254***	0.092*	0.366***	−0.519***
	(0.084)	(0.051)	(0.139)	(0.143)
PT × Periods	−0.114***	−0.071***	−0.263***	0.194***
	(0.020)	(0.018)	(0.047)	(0.047)
SMT × Periods	−0.066***	0.009	−0.022	−0.009
	(0.025)	(0.014)	(0.045)	(0.042)
AT × Periods	−0.048*	−0.001	−0.003	−0.016
	(0.025)	(0.019)	(0.039)	(0.040)
PT × Periods3	−0.011	0.119	0.431	−0.438
	(0.144)	(0.129)	(0.346)	(0.349)
SMT × Periods3	−0.762***	0.341***	1.129***	−1.525***
	(0.183)	(0.102)	(0.328)	(0.312)
AT × Periods3	0.397**	−0.142	−0.445	0.656**
	(0.187)	(0.143)	(0.288)	(0.297)
Constant	15.677***	0.352	1.400	28.001***
	(2.190)	(0.362)	(1.409)	(2.187)
σ(group)	7.186***	0.811***	4.271***	6.043***
	(0.776)	(0.177)	(0.516)	(0.917)
σ(subject)	1.525***	1.524***	2.869***	7.605***
	(0.109)	(0.101)	(0.208)	(0.479)
σ(residual ST)	1.568***	0.953***	2.607***	2.681***
	(0.038)	(0.023)	(0.064)	(0.066)
σ(residual PT)	2.192***	2.221***	5.942***	5.972***
	(0.054)	(0.054)	(0.145)	(0.146)
σ(residual SMT)	3.045***	1.649***	5.568***	5.195***
	(0.074)	(0.041)	(0.138)	(0.127)
σ(residual AT)	3.137***	2.502***	4.712***	4.883***
	(0.077)	(0.061)	(0.116)	(0.120)
Observations	3520	3520	3520	3520
Wald chi2(11)	140.35***	73.66***	105.11***	150.73***

Columns 1 - 4 report the coefficients of multilevel mixed random effects models, which allow to control for individual and group differences. The respective dependent variables are contributions, punishment spending, overall loss of earnings due to punishment and earnings. The independent variables include treatment dummies with appropriate names; 'Periods' is a time variable for periods 11-30; 'Periods3' is a time variable for periods 28-30; interaction variables are also included. The Symmetric treatment is the omitted category. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2 Propensity to feud.

	(1)	(2)
Symmetric Treatment	-1.669^{*}	empty
	(0.960)	–
Strongman Treatment	-1.586^{**}	0.878
	(0.803)	(2.695)
Alpha Treatment	-2.042^{**}	1.215
	(0.918)	(2.525)
Periods (11-30)	-0.727^{**}	0.143
	(0.351)	(0.400)
Periods^2	0.017^{*}	-0.005
	(0.009)	(0.011)
Constant	3.964	-7.764
	(3.169)	(3.120)
Observations	880	660
Wald chi^2	16.68^{***}	3.05

Columns 1 - 2 report the coefficients of logistic regressions with group-level random effects and clustered standard errors fitted to the incidence of within period and inter-period feuds respectively. The independent variables include treatment dummies with appropriate names; 'Periods' is a time variable for periods 11-30; 'Periods^2' is a time variable squared. The Privilege treatment in the baseline category. Robust standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3 Main experimental variables as function of treatment, time and role

	(1) Contrib.	(2) Pun. Spending	(3) Loss of Earnings	(4) Earnings
ST	3.727	−0.591	−3.863*	11.167***
	(3.145)	(0.662)	(2.155)	(3.300)
PT - Participant 1	−0.291	1.463	2.599	9.839***
	(3.223)	(0.920)	(2.727)	(3.979)
PT - Rank and File	−3.361	1.376*	3.772	−3.234***
	(3.166)	(0.749)	(2.341)	(3.527)
SMT - Participant 1	2.472	0.365	−2.848	10.358***
	(3.275)	(1.048)	(2.911)	(3.761)
SMT - Rank and File	3.270	−0.448	−0.943	7.478**
	(3.183)	(0.803)	(2.414)	(3.447)
AT - Participant 1	−0.285	0.627	−1.002	20.496***
	(1.184)	(0.901)	(1.837)	(2.488)
Periods (11-30)	0.011	−0.012	−0.057	0.086
	(0.300)	(0.023)	(0.046)	(0.063)
Periods3 (28-30)	0.213	−0.081	−0.074	−0.005
	(0.220)	(0.168)	(0.342)	(0.463)
ST × Periods	0.061*	−0.002	0.002	0.011
	(0.032)	(0.024)	(0.050)	(0.066)
PT - Part.1 × Periods	−0.084*	−0.077*	−0.225**	0.149
	(0.048)	(0.040)	(0.098)	(0.130)
PT - R&F × Periods	−0.043	−0.071**	−0.273***	0.224**
	(0.037)	(0.030)	(0.068)	(0.091)
SMT - Part.1 × Periods	−0.001	−0.021	0.054	−0.086
	(0.056)	(0.047)	(0.111)	(3.158)
SMT - R&F × Periods	−0.007	0.017	−0.045	0.031
	(0.041)	(0.033)	(0.074)	(0.083)
AT - Part.1 × Periods	0.049	−0.004	−0.005	−0.020
	(0.060)	(0.045)	(0.093)	(0.125)

Continued on the next page.

Table B3 - continued from the previous page.

	(1)	(2)	(3)	(4)
ST × Periods3	−0.467**	0.172	0.440	−0.514
	(0.236)	(0.175)	(0.369)	(3.148)
PT - Part.1 × Periods3	−0.894**	0.468	0.538	−0.521
	(0.353)	(0.295)	(0.723)	(0.959)
PT - R&F × Periods3	−0.339	0.232	0.982*	−1.096
	(0.272)	(0.219)	(0.502)	(0.670)
SMT - Part.1 × Periods3	−1.525***	0.399	1.770**	−2.005**
	(0.414)	(0.350)	(0.817)	(0.828)
SMT - R&F × Periods3	−1.130***	0.552**	1.502***	−2.050***
	(0.300)	(0.244)	(0.548)	(0.609)
AT - Part.1 × Periods3	−0.277	0.122	−0.022	−0.566*
	(0.441)	(0.336)	(0.684)	(0.926)
Constant	11.950***	0.943*	5.262***	16.840***
	(2.257)	(0.558)	(1.636)	(2.479)
σ(group)	7.224***	1.090***	4.488***	6.043***
	(0.772)	(0.236)	(0.489)	(0.917)
σ(residual ST)	1.579***	0.950***	2.590***	2.647***
	(0.038)	(0.023)	(0.062)	(0.066)
σ(residual PT)	2.578***	2.276***	5.965***	7.863***
	(0.062)	(0.055)	(0.143)	(0.146)
σ(residual SMT)	3.279***	2.872***	6.945***	6.429***
	(0.079)	(0.069)	(0.167)	(0.154)
σ(residual AT)	3.573***	2.272***	5.547***	7.510***
	(0.086)	(0.065)	(0.133)	(0.180)
Observations	3520	3520	3520	3520
Wald chi2(20)	253.80***	60.01***	103.00***	1673.28***

Columns 1 - 4 report the coefficients of random effects models, which allow to control for group differences. The respective dependent variables are contributions, punishment spending, overall loss of earnings due to punishment and earnings. The independent variables include the interaction terms of treatment dummies with the role dummy (1: contest winner/pivotal player; 0: otherwise); 'Periods' is a time variable for periods 11-20; 'Periods3' is a time variable for periods 18-20; interaction variables with the time variables are also included. 'Symmetric - Rank and File' is the omitted category. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

Cooperation, Punishment and Domination With Endogenous Selection to Pivotal Positions

Abstract

In the previous chapter, we demonstrated that asymmetries in returns from cooperation and punishment efficiency negatively affect cooperation in a public goods game with an endogenous number of punishment stages. The concentration of advantageous parameters in the hands of one individual did not improve the average performance of groups substantially. We speculated that some results could be driven by the random appointment to the advantageous positions. In this paper, the random mechanism is replaced with a contest, to check whether the acquisition of advantages can result in a greater acceptance of inequalities and increase the performance of asymmetric treatments relative to the symmetric. The results indeed demonstrate such an improvement, however, it was driven mainly by a decline in contributions of the symmetric treatment, rather than by a positive change in key indicators of other treatments. This observation serves as evidence of the fragility of symmetric institutions.

2.1 Introduction

In the previous chapter, we explored the effect of asymmetries in marginal returns from cooperation (hereafter MPCR) and punishment efficacy on cooperation, retribution, and welfare, which we measure by participants' earnings. A higher degree of resemblance to the real world, in contrast to many previous papers (for example Fehr and Gächter (2000), Cinyabuguma et al. (2006) and Reuben and Riedl (2013)), was achieved through endogenisation of punishment stages in each period, so that participants could punish one another so long as rival parties had sufficient resources (Nikiforakis and Engelmann, 2011).

In line with previous literature (Nikiforakis et al. (2012) and Reuben and Riedl (2013)) introduction of heterogeneity in MPCR in groups with a uniform distribution of punishment efficiency had negative consequences. The effect in our first study was even stronger than previously reported, since not only average earnings were affected, but also the contributions. Furthermore, in concordance with Nikiforakis et al. (2012) the presence of inequalities amplified the chance of within period feuds to erupt.

Also, according to the results, the inequality in punishment efficacy in the presence of equal returns from cooperation did not markedly affect the performance during most parts of the experiment. However, tensions escalated towards the end of the game, as contributions decreased and the demand for punishment soared.

In one of the treatments, both advantages were concentrated in the hands of a single player to model the presence of dominant individuals that are not only stronger than others but also appropriate the greatest part of resources. This setting had a mild effect on aggregate levels of the most variables relative to the treatment with the sole asymmetry in returns. Nonetheless, the treatment with multiple concurrent

asymmetries was characterised by high variance on the group level, as some managed to reach high levels of cooperation, whereas other hit rock bottom with two individuals from different groups losing all their earnings. The high variance rendered differences with the symmetric treatment weakly significant and granted a slight advantage in average contributions over the treatment with uneven distribution of MPCR.

One of the key parameters in the previous design is the random assignment to the advantageous roles. It is clear, however, that in reality, one's position within a hierarchy is not random, but rather determined by competition. The ways and contexts for setting someone's rank might vary across societies. Nonetheless, different forms of competition, have something in common, a success in a contest leads to the establishment of control over additional resources. Therefore, the random assignment to advantageous positions might be deemed inequitable, hence could have negatively affected contributions and punishment levels in treatments with asymmetries.

In the current study, we use a contest based on the ball catching task (Gächter et al., 2016) to determine the identity of a group member endowed with the ID of "Participant 1" and advantages associated with it. The winner enjoys a higher MPCR in the *Privilege* treatment (PT), higher punishment efficacy in the *Strongman* treatment (SMT) and both in the *Alpha* treatment (AT). Participants in the *Symmetric* treatment, with uniform distributions of MPCR and punishment efficacy, also compete, but no extra benefits are awarded to the winner, only the ID of "Participant 1".

The goal of the new design is to probe into the effects a pre-play contest on the dynamics in public goods games with an endogenous number of punishment stages. Specifically, we are interested in checking whether such a procedure instills a sense of deservedness, making cooperation more prevalent in treatments with asymmetries relative to the *Symmetric* treatment, especially when higher benefits from cooperation

are backed by a higher punishment efficacy. We hypothesise that if people perceive that an advantageous position is deservingly earned by one's superior performance, then it might justify higher rewards in the following section of the game.

After considering the evidence from the current and the previous chapters, we conclude that the contest did not instill a greater sense of "Participant 1" players being worthy of the extra benefits in the eyes of their peers. Moreover, it had some negative effects in the form of an augmented tendency to feud in treatments other than the *Privilege* treatment and increased tensions that resulted in prompt bankruptcies. Finally, it had a substantial adverse effect on the relative performance of the *Symmetric* treatment.

The rest of the paper is organised as follows. Section 2 presents related literature. Section 3 includes the experimental design, data, and procedures. The statistical analysis is included in Section 4. Finally, discussion and conclusions presented in Section 4, which is followed by the Appendices.

2.2 Related literature

It has been demonstrated that earned entitlements affect further interactions in bargaining games. For example, Hoffman and Spitzer (1982, 1985) showed that earned advantageous positions lead to less equal splits, in contrast to random allocation. According to Güth and Tietz (1986) when positions in the ultimatum game are auctioned independently, the offers to the second mover decrease significantly. Hoffman et al. (1994) report about lower proposals to the second movers in the ultimatum game, with no increase in rejection rate, once allocation of roles is made via a contest. Also, Hoffman et al. (1994) point at a similar tendency in the dictator game, as allo-

cators tend to be less generous in splitting the endowment when the right to make the decision is acquired through a contest.

A pre-play tournament was used by Nikiforakis et al. (2012) to award higher benefits from cooperation in a public goods game with an endogenous number of punishment stages per period. In their asymmetric treatments, two participants who scored highest in an encryption task enjoyed higher MPCR in the second phase of the experiment (10-period public goods game with an endogenous number of punishment stages in each period). The authors explain that the rationale for using the real effort task is to induce the sense of entitlement for higher MPCR and as a result to decrease the willingness to accept the norm of equal earnings. Nikiforakis et al. (2012) showed that in *Asymmetric* treatment earnings were lower and prolonged conflicts were more likely in contrast to the *Symmetric* treatment, whose participants had to encrypt 30 words to qualify for the 10 periods of the public goods experiment with a uniform distribution of MPCR.

The evidence from bargaining games seems to suggest that both contest winners and losers mostly agree on the former's right for certain benefits. In contrast, Nikiforakis et al. (2012) argue that the sense of entitlement is induced in winners, while those who did not win disregard the outcomes of the contest, as a result, tensions grow further. Nonetheless, considering the negative effect of asymmetry in MPCR reported by Reuben and Riedl (2013) under random allocation of extra benefits and the results reported in Chapter 1, the effect of the contest is not definite.

2.3 The experiment

This paper closely related to the one in the previous chapter hence adds to the literature on costly enforcement in public goods games, associated with the prominent works of Yamagishi (1986), Ostrom et al. (1992) and Fehr and Gächter (2000). The closest study to ours is Nikiforakis et al. (2012), who used a real effort task to assign higher MPCR in the public goods setting with punishment opportunities and an endogenous number of punishment stages in a period.

2.3.1 Experimental designs

Part 1 - The Contest

Each experimental session in this study comprises two parts. In the first part, participants compete in the ball catching task (Gächter et al., 2016) for the ID of "Participant 1" and the associated benefits, after receiving full information about the experiment. Therefore, upon entering the contest phase, they know about the benefits that will be conferred on them in case of a victory.

Subjects are acquainted with the task by completing a practice round before the actual contest takes place. In both rounds, participants gain points by catching balls with a paddle, which can be moved by clicking the "Left" and "Right" buttons. 4 points are awarded for every ball caught, and 1 point is lost for every move of the paddle.

After each round of the ball catching task, participants learn about points they earned, costs incurred and their net score, but not of others' performance. The feedback screen of the actual contest is followed by a question about own effort to win. Subjects are asked to evaluate their effort by entering a number between 0 (completely disagree) and 10 (completely agree) to state their (dis)agreement with the following

statement "I want to win and tried my best to score high".

On the final screen, participants are shown their ID number, along with a message about their success or failure in winning the contest. Winners get the ID of "Participant 1". The remaining IDs are allocated randomly, so no additional information about their performance in the ball catching task is provided.

Part 2 - Public Goods Game

The second part of the experiment is almost identical to the experiment in the previous chapter. The only difference between the two is the length of the game. After considering the pros and cons of the long horizon sessions (30 periods), due to our focus being mainly on the relative performance of different treatments, the number of periods was reduced to 20. All other details and parameters remain unchanged.

In the first stage (s=1) of a period group members simultaneously split their 20 Experimental Currency Unit (ECU) endowment between the public good (c_i) and their private accounts. For every point contributed to the former α_i ECU are added to the private account of a participant i ($i \in 1, ..., 4$). The payoff function below can be used to calculate the earnings of any participant by the end of the stage

$$\pi_i^1 = 20 - c_i + \alpha_i \sum_{j=1}^{4} c_j.$$

After observing own earnings, subjects proceed to the second stage (s=2). In this stage, they are shown the exact contributions of others and are allowed to decrease

others' earnings by assigning costly deduction points. Each point costs 1 ECU and decreases the earnings of a punished by β_i ECU. A number of deduction points assigned by a subject i can range between 0 and π_i^1. Their effect is restricted so that earnings of a punished individual cannot decrease below 0. Negative earnings are possible if points left after deductions due to punishment by others are insufficient to cover the cost of points assigned by the subject.

As soon as all group members finish assigning deduction points stage 2 ends. Participants are presented with a summary that includes earnings at the beginning of a stage, cost of punishment points assigned to others, the total number of reduction points received, their effect and the final earnings.

Further stages allow the assignment of additional punishment points, after receiving full information of others' punishment decisions[1]. Any stage beyond stage 2 is entered only if punishment points were meted out in a previous stage, one group member keeps at least 1 ECU upon entering that stage, and another one has strictly positive earnings.

To summarise, the earnings of a participant i in $s \geq 2$ can be expressed by

$$\pi_i^s = max\{0, \pi_i^{s-1} - \sum_{j=1, j\neq i}^{4} \beta_j * p_{ji}^s\} - \sum_{j=1, j\neq i}^{4} p_{ij}^s$$

The design allows for both feuds within a period, as the number of punishment stages is endogenous and between periods feuds, since the composition of groups and

[1] The number of punishment stages cannot exceed 64 in treatments with asymmetric punishment power or 33 in other treatments. Participants are not provided with these numbers but have enough information to calculate them.

participants' IDs are fixed.

Treatments

Table 2.1 Experimental design.

Treatment	MPCR	Effectiveness of punishment	No. of groups
Symmetric (ST)	$\{0.4, 0.4, 0.4, 0.4\}$	$\{3, 3, 3, 3\}$	16
Privilege (PT)	$\{0.7, 0.3, 0.3, 0.3\}$	$\{3, 3, 3, 3\}$	17
Strongman (SMT)	$\{0.4, 0.4, 0.4, 0.4\}$	$\{9, 1, 1, 1\}$	16
Alpha (AT)	$\{0.7, 0.3, 0.3, 0.3\}$	$\{9, 1, 1, 1\}$	17

MPCR - marginal per capita returns from contributions to the public account; Effectiveness of punishment - a reduction effect of each punishment point on the earnings of a punished individual. The average MPCR and punishment efficiency in all treatments are 0.4 and 3 respectively. The numbers in the curly brackets show the respective distributions for groups in each treatment. The first number in a set is a parameter of "Participant 1" player.

Table 2.1 provides a summary of our experimental design. All treatments share the general settings described in the previous two section but differ in the distribution of main parameters. At the top of the list are parameters of the *Symmetric* treatment (ST), with equal distributions of returns and punishment power. In the second and the third rows are the *Privilege* treatment (PT) with asymmetric returns and the *Strongman* treatment (SMT) with unequal distribution of punishment efficacy. The *Alpha* treatment (AT), as a model of groups dominated by strong individuals, incorporates both asymmetries, with a single group member benefiting from both higher MPCR and higher punishment efficacy.

The equal distribution of MPCR in the *Symmetric* and the *Strongman* treatments mean that for every point transferred to the public account the earnings of any group member increase by 0.4 ECU. In the other two treatments, winners of the contest ($i = 1$) get 0.7 ECU for every point in the public account, while others receive just

0.3 ECU. The differences in MPCR are calibrated, so cross treatment efficiency is constant ($\sum_1^4 \alpha = 1.6$), meaning that the maximum number of ECU obtainable for a group in a single period cannot exceed 128.

The distribution of punishment efficacy is uniform in the *Symmetric* and the *Privilege* treatments, so a point assigned by any participant reduces the earnings of a targeted individual by 3 ECU. Things are different for participants in the *Strongman* and the *Alpha* treatments, as contest winners enjoy the effect-cost ratio of 9, while others by sacrificing 1 ECU can cause a reduction by the same amount. Following Nikiforakis et al. (2010) the average punishment efficacy stable across treatments.

2.3.2 Recruitment and procedures

The experiment was conducted in the Center for Decision Research and Economics (CeDEx) lab at the University of Nottingham using z-Tree (Fischbacher, 2007). With the help of ORSEE (Greiner, 2015) 264 students (66 groups of four) from the University of Nottingham were randomly recruited to participate in the experiment[2]. Four sessions were run in November 2016, and another nine were conducted in February 2017.

Upon arrival at the lab, participants drew a number to choose their computer stations. As soon as all the participants were seated, the instructions were read aloud by the experimenter. After everyone successfully answered all comprehension questions, participants proceeded to the contest.

On average, the experimental sessions lasted two hours. The payoff for an individual participant was comprised of the participation fee (£3), the summation of

[2] Out of all participants, 148 were female. Their share in various treatments did not vary substantially (Pearson's χ^2 test: p value = 0.370.

earnings in all rounds and the initial capital of 150 ECU. At the end of the experiment, the earnings in ECU were converted to GBP at the exchange rate of 1 ECU = 2 pence and paid in cash. On average participants earned approximately £15.

2.4 Experimental results

To be consistent with the analysis in the previous chapter we regard the first 10 periods as an adaptation phase, during which individuals familiarise themselves with the institutions and other co-players. This phase is characterised by high levels of average punishment spending, which decline as participants proceed to later periods of the game. Therefore, the comparison is mostly based on the performance in periods 11-20.

We examine the differences between treatments mainly by using a Kruskal-Wallis, and two-tailed Mann-Whitney (henceforth MW) tests with group averages across all analysed periods as independent observations, as well as by estimating multilevel mixed random effects regressions, with contributions, punishment spending, income reduction, and earnings as dependent variables. The independent variables include treatment dummies (the *Symmetric* treatment is the base category), time variables ("Periods" for periods 11-20 and "Periods3" for periods 18-20) and interactions between them.

Additional specifications and other models are presented further below.

2.4.1 Ball catching task

Before proceeding with the analysis of the public good phase we look at the numbers from the ball catching task. On average it was necessary to obtain a net score of 151.50

to win the contest in the *Symmetric* treatment, 145.06 for the victory in *Privilege*, 148.69 in *Strongman* and 144.47 to get the position of "alpha". The respective net score values for individuals who did not win were 126.04, 115.90, 111.60 and 116.37.

According to a Kruskal-Wallis (KW) test the cross treatment differences between winners' average scores are statistically insignificant (KW test $p = 0.573$), but substantially higher in the *Symmetric* treatment among the non-winners (KW test $p = 0.011$; M-W test on ST = PT: $p = 0.031$; ST = SMT: $p = < 0.001$; ST = AT: $p = 0.044$). The above figures demonstrate that objectively, requirements for a victory did not differ considerably between the treatments.

Subjectively, however, individuals exerted less effort or were less interested in winning in the *Symmetric* treatment. On average, when asked to state the degree of interest in winning, participants in the *Symmetric* treatment entered a lower number (8.12) than participants in the *Privilege* (9.04), the *Strongman* (8.77) or the *Alpha* (9.12) treatments. The significance of the differences between the *Symmetric* treatment and other treatments is confirmed by the nonparametric tests (KW test $p = 0.002$; M-W test on ST = PT: $p = 0.002$; ST = SMT: $p = 0.009$; ST = AT: $p < 0.001$).

The fact that individuals were less interested in winning a contest that was not affecting the second phase of the game is not surprising. It is the relatively small size of the gap that is not obvious. The average number entered by participants in *Symmetric* is pretty high despite the absence of any real benefits[3]. This fact can be

[3] Niederle and Vesterlund (2007) report about gender differences in competitive behaviour, which they mainly attribute to male overconfidence and differences in preferences for entering and performing in a competition. Our data provide some insights into the differences in preferences of male and female subjects too. We find (weak) evidence that male participants were more interested in winning, as indicated by their answers to the post contest question (KW test $p = 0.061$). Moreover, males on average caught more balls (KW test $p = 0.002$) and moved the paddle more often (KW test $p = 0.004$). Regression analyses confirm those results and also point at a weakly significant difference in the task's net score in favour of men.

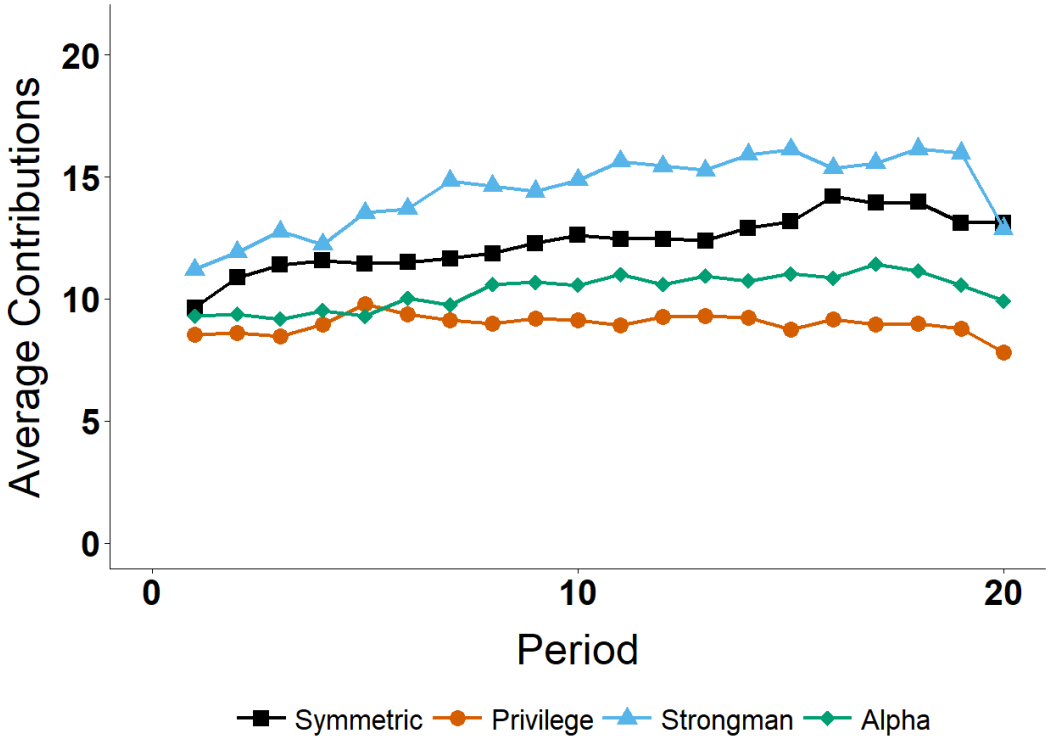


Figure 2.1: Evolution of average contributions by treatment.

considered as evidence for the competitive mood induced by the contest, which could be affecting the course of the second part of the experiment.

2.4.2 Contributions

Result 1. *On average contributions in the first round were not markedly different between the treatments.*

Figure 2.1 suggests that first period contributions were the highest in the *Strongman* treatment (11.20 ECU), followed by *Symmetric* (9.64 ECU), *Alpha* (9.29 ECU) and *Privilege* (8.54 ECU). The null hypothesis that the contributions are identically distributed across treatments cannot be rejected according to a Kruskal-Wallis test

($p = 0.186$).

Result 2. *(i) Average contributions in the Strongman treatment were higher than in the other treatments.*
(ii) Participants in treatments with symmetric MPCR on average contributed more.

Table 2.2 summarises the main variables in the last 10 periods. The second column of Panel A and the visual inspection of Figure 2.1 demonstrate that the relative level of average contributions in the first period serves as a good predictor of performance in further periods, as on average the *Strongman* treatment was characterised by higher average contributions (15.43 ECU) than the *Symmetric* treatment (13.17 ECU), the *Alpha* treatment (10.82 ECU) or the *Privilege* treatment (8.92 ECU).

According to a Kruskal-Wallis test, the average contributions come from different distributions ($p = 0.032$). MW tests, whose p values can be found in the second column of Panel B in Table 2.2, attribute this result to gaps between the *Symmetric* and the *Privilege* treatments (MW test $p = 0.044$), the *Strongman* and the *Privilege* treatments (MW test $p = 0.012$), as well as the *Strongman* and the *Alpha* treatments (MW test $p = 0.032$).

To fully benefit from the panel structure of the data we estimate a multilevel mixed random effects model with treatment-specific residuals. From coefficients reported in column one of Table B2 we infer that the level of contributions after the first 10 periods in the *Strongman* treatment was the highest, but only the difference with the *Symmetric* treatment is significant at %5 level (Wald test on "Strongman Treatment (SMT)": $p = 0.033$; "Privilege Treatment (PT)" = "Strongman Treatment (SMT)": $p = 0.057$; "Strongman Treatment (SMT)" = "Alpha Treatment (AT)": $p = 0.057$).

When it comes to trends, there is weak evidence on the gap between *Strongman* and *Symmetric* was getting smaller as the game proceeded (Wald test on "SMT × Periods": $p = 0.071$), while other gaps persisted over time (Wald test on "SMT × Periods" = "PT × Periods": $p = 0.361$; "SMT × Periods" = "AT × Periods": $p = 0.832$).

The endgame effect is reflected in the decreasing contributions during the final three periods in all treatments (Wald test on "Periods" + "Periods3": $p = 0.016$; "PT × Periods" + "PT × Periods3": $p = 0.999$; "SMT × Periods" + "SMT × Periods3": $p = 0.131$; "AT × Periods" + "AT × Periods3": $p = 0.816$). The differences between the *Strongman* treatment and treatments with asymmetric MPCR during this phase of the game are not statistically significant (Wald test on "PT × Periods" + "PT × Periods3" = "SMT × Periods" + "SMT × Periods3": $p = 0.164$; "AT × Periods" + "AT × Periods3" = "SMT × Periods" + "SMT × Periods3": $p = 0.201$).

The data also show that by the end of the first 10 periods the differences in contribution levels between the *Symmetric* treatment and the treatments with the asymmetry in MPCR were not substantial (Wald test on "Privilege Treatment (PT)": $p = 0.845$; "Alpha Treatment (AT)": $p = 0.647$). Nonetheless, contributions in the former were higher, as indicated by the steeper slope of the trend line (Wald test on "PT × Periods": $p = 0.002$; "AT × Periods": $p = 0.016$).

Finally, the amounts transferred to the public account in the *Privilege* treatment do not differ substantially from those transferred in the *Alpha* treatment (Wald test on "Privilege Treatment (PT)" = "Alpha Treatment (AT)": $p = 0.802$; "PT × Periods" = "AT × Periods": $p = 0.412$; "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.836$).

Table 2.2 Main variables across treatments (periods 11-20).

Treatment	Contributions	Punishment	Income Reduction	Earnings
Panel A (group-level std. dev. in the parenthesis)				
Symmetric (ST)	13.17	0.73	2.82	25.09
	(5.74)	(1.38)	(5.18)	(6.04)
Privilege (PT)	8.92	0.79	2.90	22.45
	(7.08)	(1.31)	(4.64)	(6.19)
Strongman (SMT)	15.43	1.22	4.42	25.08
	(4.28)	(2.18)	(7.62)	(8.60)
Alpha (AT)	10.82	1.08	4.18	22.43
	(6.15)	(1.90)	(5.88)	(7.33)
Panel B: Statistical comparison (Mann-Whitney, two sided, p values)				
ST-PT	0.044	0.626	0.664	0.179
ST-SMT	0.521	0.152	0.136	0.752
ST-AT	0.309	0.227	0.139	0.245
PT-SMT	0.012	0.417	0.396	0.049
PT-AT	0.375	0.426	0.268	0.973
SMT-AT	0.032	1.000	0.471	0.049

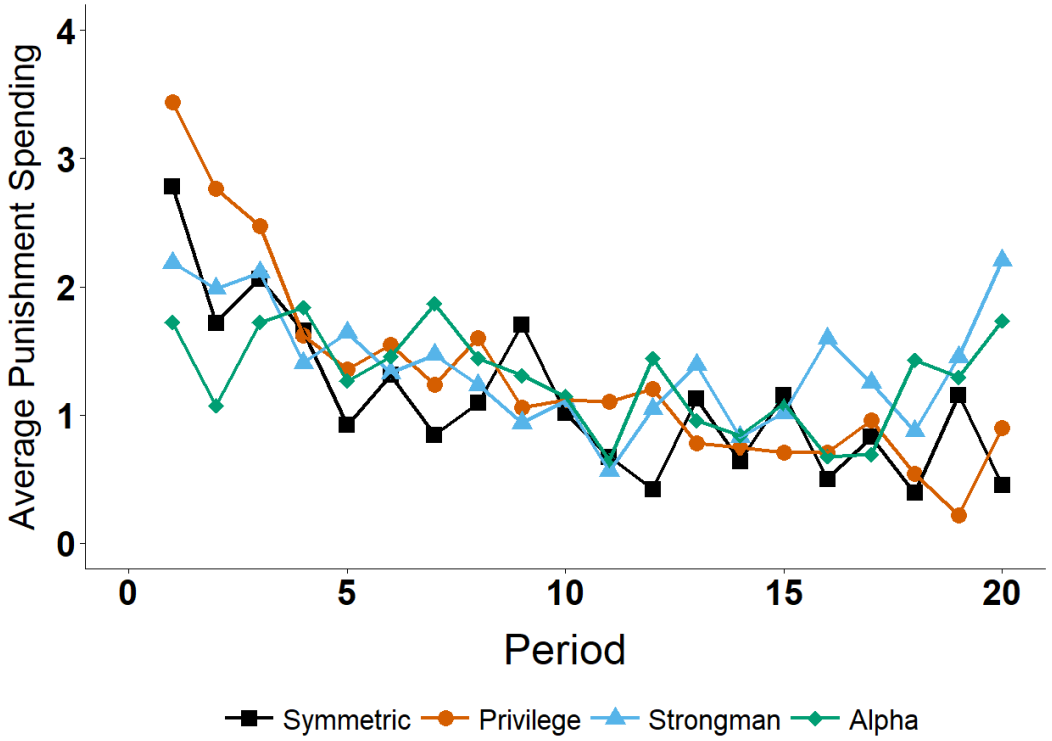


Figure 2.2: Evolution of average punishment spending by treatment.

2.4.3 Punishment

Result 3. *Limited evidence on differences in average punishment spending were detected.*

Figure 2.2 and the third column in Panel A of Table 2.2 provide a brief overview of average spending on punishment. Both suggest that average punishment spending did not vary considerably. This conclusion is supported by two-tailed MW tests on groups' averages (for p values see the third column of Panel B in Table 2.2).

To verify the above results we regressed punishment spending on treatment dummies, time variables, and interactions between them. The coefficients can be found in column two of Table B2. The regression analysis demonstrates that the punishment

spending in final periods was lower in the *Symmetric* treatment relative to the *Alpha* treatment (Wald test on "AT × Periods" + "AT × Periods3": $p = 0.043$). There is also weak evidence on punishment spending being lower during the last 3 periods in the *Privilege* treatment relative to the *Alpha* treatment (Wald test on "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.063$) and in the *Symmetric* treatment relative to the *Strongman* treatment (Wald test on "SMT × Periods" + "SMT × Periods3": $p = 0.085$). Moreover, there is weak evidence on decreasing punishment in the *Privilege* treatment relative to the *Strongman* treatment (Wald test on "PT × Periods" = "SMT × Periods": $p = 0.085$).

All other differences are not statistically significant even at 10% level.

Result 4. *(i) There is week evidence on the lower overall reduction of earnings due to punishment opportunities in the Strongman treatment in comparison with the Privilege treatment.*
(ii) The reduction in the final 3 periods was lower in the Symmetric treatment in contrast to both treatments with the non-uniform distribution of MPCR and in the Privilege treatment relative to the Alpha treatment.

Descriptive statistics (column four in the top panel of Table 2.1) suggest that asymmetries in punishment efficiency played a major role in determining the total loss of earnings due to the availability of punishment opportunities. On average losses were higher in the *Strongman* and the *Alpha* treatments, but none of the differences are significant according to two-sided MW tests applied to group averages (fourth column in the bottom panel, Table 2.1).

Nevertheless, the regression analysis indicates that some of them are significant (for coefficients see column 3 of Table B2 in Appendix B). Contrary to conclusions

derived from the descriptive statistics and the non-parametric analysis, the regression suggests that on average participants in the *Strongman* treatment experienced more modest reduction relative to the *Privilege* treatment, as seen from the weakly significant difference in intercepts (Wald test on "Privilege Treatment (PT)" = "Strongman Treatment (SMT)": $p = 0.064$) and despite the differences in trends (Wald test on "PT × Periods" = "SMT × Periods": $p = 0.013$), which persisted during the final periods of the experiment (Wald test on "PT × Periods" + "PT × Periods3" = "SMT × Periods" + "SMT × "Periods3": $p = 0.111$).

Other statistically significant differences include the lower average reduction in the *Symmetric* treatment versus the *Privilege* treatment (Wald test on "Privilege Treatment (PT)": $p = 0.050$; "PT × Periods": $p = 0.019$), in addition to the differences between the *Symmetric* and the *Strongman* treatments, and the *Alpha* treatment in the final periods of the game (Wald test on "SMT × Periods" + "SMT × Periods3": $p = 0.026$; "AT × Periods" + "AT × Periods3": $p = 0.004$). Furthermore, the the gap between the *Privilege* treatment and the *Alpha* treatment during the last 3 periods was statistically significant (Wald test on "PT × Periods" + "PT × Periods3" = "AT × Periods" + "AT × "Periods3": $p = 0.024$).

The analysis of punishment behaviour is finalised by looking at the tendencies to feud. We distinguish between intra-period and inter-period feuds. According to our definition, a within period feud occurred if in a particular period there were more than two punishment stages, in which at least one individual punished and was punished.

The second type of feuds should account for all punishment spillovers between periods. However, since motivations for punishment might be conflated, we define them as periods with equal contributions by the rival parties in the first stage and retaliation in the second stage for punishment meted out in some earlier period.

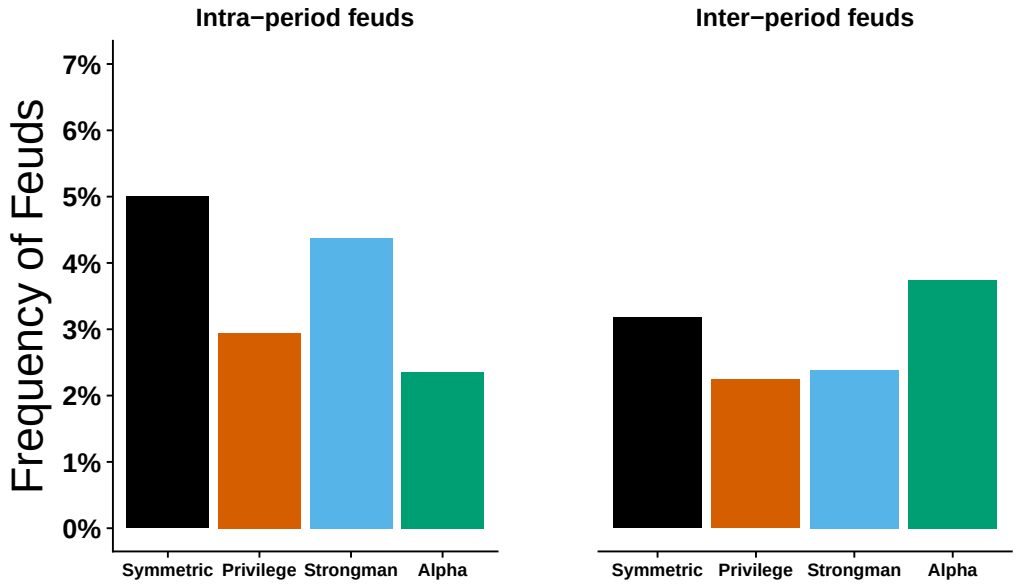


Figure 2.3: Frequency of feuds (out of the total number of periods).

Figure 2.3 shows that the *Symmetric* treatment was characterised by highest frequency of intra-period feuds, whereas inter-period feuds were most likely in the *Alpha* treatment. To test these conjectures we used logistic regressions with group-level random effects. The dependent variables get 1 for every period with a feud (0 otherwise), and independent variables include treatment dummies, a time variable (periods 11-20) and a time variable squared. The results (see Table B1) suggest that the differences in propensity to feud do not differ markedly between treatments, as both models are statistically insignificant. Such results should not surprise us, as feuds are rare in all treatments.

2.4.4 Earnings

Result 5. *Earnings in the Symmetric treatment were higher than in other treatments.*

Contribution and punishment patterns shaped participants' earnings. The evolution of earnings over time is shown in Figure 2.4, which suggests that treatments with symmetric MPCR had higher average earnings throughout the game. This impression is supported by the figures shown in Table 2.1 (fifth column, Panel A). However, according to MW tests (fifth column, Panel B of the same table), only the differences between the *Strongman* treatment and the treatments with heterogeneous MPCR are significant (MW test $p = 0.049$ for both).

The regression analysis (see column 4, Table B2 in the Appendix B) indicates that earnings in the *Strongman* treatment were higher than in the *Privilege* treatment (Wald test on "Privilege Treatment (PT)" = "Strongman Treatment (SMT)": $p = 0.025$; "PT × Periods" = "SMT × Periods": $p = 0.083$; "PT × Periods" + "PT × Periods3" = "SMT × Periods" + "SMT × Periods3": $p = 0.028$), but not than in the *Alpha* treatment (Wald test on "Strongman Treatment (SMT)" = "Alpha Treatment (AT)": $p = 0.121$; "SMT × Periods" = "Alpha Treatment (AT)": $p = 0.287$; "SMT × Periods" + "SMT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.883$).

According to the estimated model earnings in the last periods of the game were higher in the *Symmetric* treatment relative to the *Strongman* treatment (Wald test on "SMT × Periods" + "SMT × Periods3": $p = 0.006$) and relative to the *Alpha* treatment (Wald test on "AT × Periods" + "AT × Periods3": $p = 0.004$). Besides, the difference in intercepts of the *Symmetric* treatment and the *Privilege* treatment is weakly significant (Wald test on "Privilege Treatment (PT)": $p = 0.084$), suggesting the former had the highest level of average earnings, at least at 10% level.

Finally, the differences between the *Privilege* and the *Alpha* treatments are significant in the last 3 periods (Wald test on "Privilege Treatment (PT)" = "Alpha Treatment (AT)": $p = 0.522$; "PT × Periods" = "AT × Periods": $p = 0.553$; "PT ×

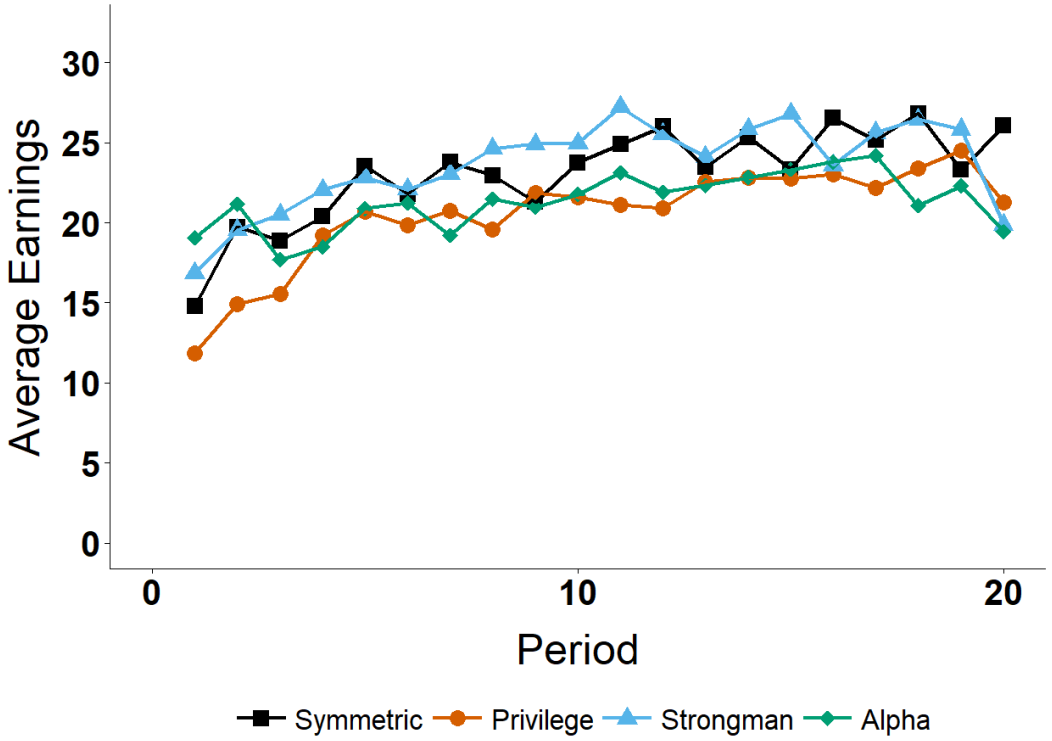


Figure 2.4: Evolution of average earnings by treatment.

Periods" + "PT × Periods3" = "AT × Periods" + "AT × Periods3": $p = 0.020$).

2.4.5 Privilege in contrast to Alpha

In the above sections, we established that at the aggregate levels the differences between the *Privilege* treatment and the *Alpha* treatment are either statistically insignificant or confined to the last periods of the game. Therefore, it is clear that "alpha" individuals had either no effect or even a slightly negative effect on the overall performance in the presence of tensions due to unequal MPCR. Nevertheless, in the previous chapter, we demonstrated that group level analysis unveiled more variation.

Figure 2.5 visualises the contribution, punishment and earnings data. From visual inspection of the figure both treatment seem to be similar, as clustering that we

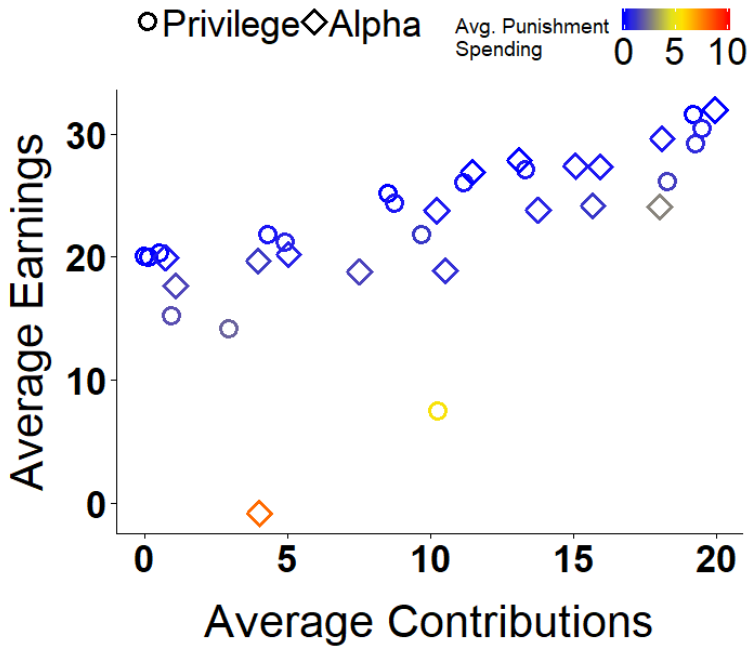


Figure 2.5: Groups' average contributions, earnings and punishment spending in ECU.

observed in the *Alpha* treatment in the previous study cannot be observed here. Furthermore, from the figure, it is seen that only one group in the *Alpha* treatment managed to cross 30 ECU threshold and another one had negative average earnings in the 10 periods of the analysis.

More detailed examination of the data shows that full cooperation was achieved only by 4 groups (one for 9 periods, two for 5 periods and another one for 4 periods) in the *Privilege* treatment and by 4 groups (one for 8 periods, one group for 7 periods and two others for a single period) in the *Alpha* treatment. Nine groups in the *Privilege* treatment and 7 in the *Alpha* treatment on average contributed less than half of the endowment.

Low transfers to the public account, accompanied by conflicts resulted in 6 groups in each treatment having average earnings lower than the initial endowment. One participant, in the group with negative average earnings, lost all earnings accrued

throughout the game, along with the initial endowment.

2.4.6 Bankruptcies

Three bankruptcies took place in the experiment described in Chapter 1. The earliest took place in the 27th period of the *Alpha* treatment. Because the current experiment was shortened to 20 periods and the initial capital remained unchanged, we expected no bankruptcies. Nonetheless, 2 subjects did go bankrupt. Here we provide some anecdotal evidence about their behaviour over 20 periods.

We already mentioned that one bankruptcy transpired in the presence of an "alpha" individual (after 14 periods) and the second one took place in a group with a "strongman" (after 15 periods). In both groups, average contributions were not high, 3.80 ECU and 8.81 ECU respectively, but were not the lowest in their treatments.

What distinguishes these groups from others is the abnormal levels of punishment spending. The average punishment spending levels in other groups were 0.75 ECU and 0.65 ECU in the *Strongman* and the *Alpha* treatments, compared to 7.74 ECU and 7.33 ECU in the respective groups with bankruptcies. The bankrupt subjects spent on average 15.45 ECU and 14.20 ECU per period in the *Strongman* and the *Alpha* treatments respectively.

In both cases, the conflicts involved groups' contest winners and participants whose score was the second best. The group dynamics differed, however, making it difficult to tell how important the contest ranking was for the abnormal behaviour.

2.4.7 Pivotal participants in contrast to others

The analysis is continued with a comparison of behavioural trends of contest winners relative to the other group members. For this purpose, we estimated models similar to the ones we used earlier, but without the individual-level random effects. Explanatory variables included interaction terms of "Participant 1" dummy with the independent variables used in the models we described earlier (for coefficients check Table B3). The baseline category is the *Symmetric* treatment's rank and file.

Result 6. *Contest winners on average contributed more and spent more on punishment than others only in the Privilege treatment. However, the overall loss of earnings was similar for both categories in all treatments.*

The regression analysis reveals significantly higher contributions (Wald test on "PT - Participant 1" = "PT - Rank and File": $p = 0.029$) and punishment spending (Wald test on "PT - Participant 1" = "PT - Rank and File": $p = 0.005$) by the contest winners relative to others in the *Privilege* treatment. The latter is true despite the fact that punishment spending of the pivotal individuals was falling at a higher pace (Wald test on "PT - Part.1 $\times$ Periods" = "PT - R&F $\times$ Periods" $p = 0.034$).

No significant differences in contributions or punishment spending are detected in the other three treatments. Furthermore, the differences in overall income reduction due to the presence of punishment opportunities are not substantial.

Result 7. *(i) "Alphas" and "privileged" on average earned more than the other members of their groups.*
(ii) Average earnings of neither "alphas", nor "privileged" exceeded that of the participants in the Symmetric treatment.

According to the estimated models, pivotal participants in *Privilege* and *Alpha* earned more than their peers (the respective Wald test p values for the differences in intercepts are 0.014 and 0.033). Unlike their counterparts, contest winners in treatments with symmetric marginal per capita returns from cooperation on average did not enjoy higher levels of earnings compared to others. Nonetheless, neither "alphas", nor the "privileged" earned markedly more than the participants in the *Symmetric* treatment. The former even earned significantly less than "Participant 1" players in the *Symmetric* treatment during the final periods of the game (Wald test on "ST - Part.1 × Periods" + "ST - Part.1 × Periods3" = "AT - Part.1 × Periods" + "AT - Part.1 × Periods3": $p = 0.042$).

2.4.8 Comparison with Chapter 1

We mentioned earlier that the experimental design of the current study differs from the one in the previous chapter in two dimensions, namely the number of periods per session and the process of designation to the advantageous positions. Therefore, we are not comparing them quantitatively, but rather focusing on the changes in treatments' relative performance.

That being said the cooperation levels for asymmetric treatments are roughly the same in both treatments. Therefore, the weakening of the differences, relative to Chapter 1, between the *Symmetric* treatment and the treatments with asymmetric MPCR must be attributed to a decrease in cooperation levels of the former. Moreover, after the introduction of the contest *Symmetric* lost its position as the treatment with the highest average contributions to *Strongman*.

The *Symmetric* treatment still had the highest average earnings, as in the previous

chapter. However, the gap between *Symmetric* and *Privilege* turned out to be weakly significant, and the difference with the *Alpha* treatment is statistically significant only for the last 3 periods.

Another negative effect is an increase in the frequency of intra-period feuds in the treatments with asymmetric punishment mechanisms and the *Symmetric* treatment. In Chapter 1 we report on the higher likelihood of feuds in the *Privilege* treatment, but this is no longer the case in the current study, as feuds in other treatments were as common as in *Privilege*.

Lastly, the contest leads to prompt bankruptcies. The earliest bankruptcy in the previous study occurred by the end of the 25th period, whereas in the current study one of the participants lost everything after 14 periods, in spite of identical initial capitals in both studies. This observation can be interpreted as evidence for increased tensions, resulting from the participation in the contest at the beginning of the game.

2.5 Discussion and conclusions

In this chapter, we continued to explore the effect of asymmetries in marginal per capita returns from a public good and punishment efficacy on cooperation, punishment behaviour, and welfare. The core of the experimental design remained unchanged with respect to Chapter 1, namely two by two design with each period being comprised of the contribution stage followed by a punishment phase with an endogenous number of punishment stages (Nikiforakis and Engelmann, 2011).

The main difference is in the assignment process to the advantageous positions. The random allocation of Chapter 1 was replaced by a pre-game contest, so that

group members had to compete to obtain the "Participant 1" ID and the associated benefits. After the contest winners were determined, 20-period sessions began, which were shorter by 10 periods than in the previous chapter.

The purpose of the change was to look into the possibility that designation to the pivotal roles can instill a greater sense of entitlement of some advantages, compared to the random assignment. Specifically, we were interested in seeing whether the pre-game tournament that mimics the real world competition for the top position in a hierarchy could improve the performance of the *Alpha* treatment and other asymmetric treatments.

The results suggest that the contest had the most profound (negative) impact on the *Symmetric* treatment, despite having no real effect on its main parameters, MPCR, and punishment efficacy. Some other treatments also experienced adverse effects, such as increased tendency to feud or prompt bankruptcies, but the cooperation levels both relatively and in absolute terms remain quite stable.

These observations point out the fragility of the symmetric institutions, whose success in maintaining high levels of cooperation seems to be the result of stylised laboratory environments. The competitive mood that is reflected by the relatively high interest in winning the contest, even in the absence of benefits in the form augmented capacities, could have played a role in lowering cooperation in the *Symmetric* treatment. Therefore, it might be interesting to check its significance by adding a cooling off period after the contest. Bolle et al. (2014) demonstrate that a cooling off period reduced aggression in a different game.

On the individual level, pivotal participants in treatments with asymmetric returns on average kept higher earnings than other co-players. Nonetheless, contrary to the results in the previous chapter, the level of earnings of"alpha" individuals not only

did not exceed that obtained by the contest winners in the *Symmetric* treatment, but even was lower during the last three periods of the game. Therefore, the individual level data suggest that when tensions are high, even the dominant individuals might be better off by maintaining egalitarian institutions.

The similarity in the cooperation rates between the *Privilege* treatments in both studies hints that the task used by Nikiforakis et al. (2012) could have had a minor effect on the behaviour in their asymmetric treatments unless different tasks are perceived differently and lead to varying results. Therefore, future research may not only look into precise effects of the contest by comparing equally long treatments but also check how well various real effort tasks affect the willingness of others to accept inequalities in public goods games.

2.6 Bibliography

Bolle, F., Tan, J. H., and Zizzo, D. J. (2014). Vendettas. *American Economic Journal: Microeconomics*, 6(2):93–130.

Cinyabuguma, M., Page, T., and Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9(3):265–279.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Gächter, S., Huang, L., and Sefton, M. (2016). Combining "real effort" with induced effort costs: the ball-catching task. *Experimental Economics*, 19(4):687–712.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Güth, W. and Tietz, R. (1986). Auctioning ultimatum bargaining positions. *Issues in West German Decision Research, Frankfurt: Lang*, pages 60–73.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games.

Hoffman, E. and Spitzer, M. L. (1982). The Coase Theorem: Some Experimental Tests. *The Journal of Law & Economics*, 25(1):73–98.

Hoffman, E. and Spitzer, M. L. (1985). Entitlements, Rights, and Fairness: An

Experimental Examination of Subjects' Concepts of Distributive Justice. *The Journal of Legal Studies*, 14(2):259–297.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Nikiforakis, N. and Engelmann, D. (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization*, 78(3):319–332.

Nikiforakis, N., Normann, H.-T., and Wallace, B. (2010). Asymmetric enforcement of cooperation in a social dilemma. *Southern Economic Journal*, pages 638–659.

Nikiforakis, N., Noussair, C. N., and Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, 96(9):797–807.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a Sword: Self-Governance Is Possible. *American Political Science Review*, 86(02):404–417.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1):110–116.

Appendix A: Instructions

Please note that in the instructions below, we mark treatment-specific text as follows: [Symmetric] {Privilege}, |Strongman| ,<Alpha>. The actual instructions only contained the relevant text with no markings.

You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions.

The instructions were distributed to all participants. **Please do not communicate with the other participants during the experiment.** Should you have any questions please raise your hand.

During the experiment we shall not speak of Pounds, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECU. At the end of the experiment the total amount of ECU you have earned will be converted to £ at the rate of **1 ECU = 2 p** and will be immediately paid to you in cash. Your **£6** show up fee consists of two parts. **£3** will be paid by the end of the experiment regardless of your results. Another **£3** will be used to cover potential negative earnings that might arise during the experiment.

At the beginning of the experiment the participants will be randomly divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of each group will remain the same throughout the experiment.**

The experiment will last 20 periods. In the beginning of the experiment, after you are randomly divided into groups, **each participant in your group will be given a number for identification** (i.e. Participant 1, Participant 2, Participant 3, and Participant 4). **Each participant will keep this number throughout the experiment.** No one will learn the participants' identities after the experiment.

Each of the 20 periods is divided into a number of stages.

The first stage

At the beginning of each of the 20 periods each participant will receive 20 ECU. In the following, we shall refer to this amount as the "endowment". Your task in the first stage is to decide how to use your endowment. **You have to decide how many of the 20 ECU you want to contribute to a project** (from 0 to 20) **and how many of them to keep for yourself**. The consequences of your decision are explained in detail below.

Once all the participants have chosen their contribution to the project you will be informed about the group's total contribution, your earnings from the project and your earnings at the end of the first stage. Your earnings at the end of the first stage are calculated using the following formula. If you have any difficulties do not hesitate to ask us.

Earnings at the end of stage 1 = Endowment of ECU - Your contribution to the Project + α*Total contribution to the Project

---

This formula shows that your earnings at the end of the first stage consist of two parts:

1. The ECU which you have kept for yourself (endowment – contribution)

2. The earnings from the project, which equal to the α share of the group's total contribution.**The α [\|is 0.4 for all four members of the group\|] <{for Participant 1 is 0.7 and 0.3 for the other three members of the group}>.**

The earnings of each group member from the project are calculated in the same way. Suppose the sum of the contributions of all group members is 60 ECU. In this case, [\|earnings of each group member from the project are 0.4*60=24 ECU\|] <{Participant 1 earns 0.7*60=42 ECU from the project, while the other members of the group receive 0.3*60=18 ECU}>. If the total contribution to the project is 9 ECU, then earnings are [\|0.4*9=3.6 ECU\|] <{0.7*9=5.6 ECU and 0.3*9=2.7 ECU respectively}>.

You always have the option of keeping the ECU for yourself or contributing them to the project. Each ECU that you keep raises your end of period earnings by 1 ECU. Supposing you contributed this ECU to the project instead, then the total contribution to the project would rise by 1 ECU. Your earnings from the project would thus rise by [\|**0.4***1 ECU\|] <{α*1 ECU}>. The earnings of the other group members would also rise [\|by 0.4 ECU\|] <{ in accordance with their α }>, so that the total earnings of the group from the project would be increased by 1.6 ECU. Your contribution to the project therefore also raises the earnings of the other group members. On the other hand you also gain earnings for each ECU contributed by the other members to the project. In particular, for each ECU contributed by any member, you earn [\|0.4 ECU\|] <{0.7 ECU in case you are Participant 1 and 0.3 otherwise}>.

The second stage

At the second stage you will be informed about how much each group member contributed individually to the project at the first stage. At this stage you can **reduce the earnings of each member of your group by distributing points or leave unaltered by assigning 0 points**. The other group members can also reduce your earnings if they wish to.

To reduce another participant's earnings you will have to distribute points. Each point you distribute will cost you 1 ECU and will reduce the earnings of the person you assign it to by β ECU. **[{The β is 3 for all four members of the group (each assigned point reduces the earnings of a targeted participant by 3 ECU).}] <|The β for Participant 1 is 9 (each point that Participant 1 assigns, reduces the earnings of a targeted participant by 9 ECU) and $\beta = 1$ for the other three members of the group (each point that other participants assign, reduces the earnings of a targeted participant by 1 ECU).|>** By choosing 0 points you will leave the earnings of a participant unchanged.

Example: [{Suppose that you are Participant 1 and you assign 2 points to Participant 2. This costs you 2 ECU and reduces Participant 2's earnings by 2*3 = 6 ECU.}]

<|Suppose that you are Participant 1 and you assign 2 points to Participant 2. This costs you 2 ECU and reduces Participant 2's earnings by 9*2 = 18 ECU. In case you are not Participant 1, 2 points still cost you 2 EUCs, but reduce the earnings of a targeted individual by 1*2 = 2 EUC.|>

Your total earnings from the two stages are therefore calculated as follows:

Total earnings (in ECU) at the end of stage 2 =

Earnings from the 1^{st} stage – β*Points you receive – Points you assign

---

Please note that your payoff for a round can be reduced to 0 in case the total amount of points to be deducted, due to the received points, is equal to or exceeds your earnings in ECU. Only if your earnings are insufficient to cover the cost of the points you assigned to others, your earnings become negative. Once your earnings are negative or zero, you no longer can assign points to others and no further points can be assigned to you. This is true for any stage of any round.

For example assume that your earnings from previous stage are 30 ECU. In the current stage you received [{4 points from Participant 2 and additional 8 points from Participant 1}] <|9 points from Participant 2 and additional 3 points from Participant 1|>. Therefore, your earnings should be reduced by [{3*4+3*8=36 ECU}] <|9*1+3*9=36 ECU|>. According to the rules, because 36>30 your earnings will be reduced to zero, provided you allocated no points in this round. If, in addition, you allocated 5 points in total, your earnings will fall to -5 by the end of this stage, as a result you will not be able to reduce the earnings of others and others will not be allowed to reduce yours till the end of the round.

If your earnings are negative at the end of the period, this will be covered by the points you accumulated in the previous stages and by the 150 ECU that we gave you in the beginning. It is very unlikely, but possible, that due to negative payoffs in multiple periods you will run out of ECU. Under such circumstances you will have to wait till

the end of the session and be paid the second half of your participation fee.

If none of the members of your group distributes points then the period finishes and the next period begins again with stage one. The period also finishes if none of the members of your group are allowed to assign further points. This will be explained below. Otherwise, a third stage will follow.

The third stage

In the third stage, you will be informed of the points that each person in your group assigned to you and the other members in your group. Similarly, the other members of your group will be informed about how many points you assigned to each of them. Then you can again **reduce or leave unaltered the earnings of each member of your group**. The other group members can also reduce your earnings if they wish to.

Note that in order for someone to be able to assign points at this stage he or she must have at least 1 ECU (i.e. the cost of one point). Additionally, if one member has non-positive earnings, the others will not be able to reduce her earnings further.

The costs of assigning points, as well as the earnings reduction caused by each point remain the same as before.

Since a group member can only assign points if her current earnings are at least 1 ECU, the period ends when the earnings of all members of a group has been reduced to less than 1 ECU. Furthermore, since it is not allowed to assign points to a member who already has zero or negative earnings, the period ends if three members have non-positive earnings, because the fourth member of the group cannot reduce their

earnings further; hence the third stage will only be entered if after the second stage at least one group member has 1 ECU left; at least another one has retained earnings greater than 0 and points were assigned in the second stage.

Fourth stage and beyond

Next stages are the **same as stage 3**. **After being informed of the points distributed in your group you will be able to assign further points.** The costs of assigning points, as well as the earnings reduction caused by each point remain the same as before. The fourth, fifth, or any further stage will only be entered if at least one member of the group assigned points in the previous stage.

When does a period end?

A period ends and a new one begins when one of the following occurs.

- No points are distributed in a given stage.
- Points are distributed, but no participants would be allowed to assign any more points if another stage followed. This can happen if nobody in a group has at least 1 ECU left or if three of the individuals have non-positive earnings.

[{**Example:** Assume that after the third stage, the earnings are as follows:

Participant 1: 2 ECU

Participant 2: 10 ECU

Participant 3: 3 ECU

Participant 4: 2 ECU

Assume that in the fourth stage Participant 4 assigns 2 points to Participant 3 and no more points assigned. Hence the earnings are

Participant 1: 2 ECU

Participant 2: 10 ECU

Participant 3: 0 ECU (= max (3 ECU – 3*2 ECU, 0)

Participant 4: 0 ECU (= 2 ECU – 2 ECU)

Since points were assigned in stage four and two participants have positive earnings, a fifth stage follows. If in the fifth stage Participant 2 assigns 2 points to Participant 1 and Participant 1 assigns 1 point to Participant 2, the earnings are

Participant 1: -1 ECU (= max (2 ECU – 3*2 ECU, 0) – 1 ECU)

Participant 2: 5 ECU (= max (10 ECU - 3*1 ECU, 0) – 2 ECU)

Participant 3: 0 ECU

Participant 4: 0 ECU}]

<|**Example:** Assume that after the third stage, the earnings are as follows:

Participant 1: 2 ECU

Participant 2: 12 ECU

Participant 3: 3 ECU

Participant 4: 2 ECU

Assume that in the fourth stage Participant 3 assigns 3 points to Participant 2 and no more points assigned. Hence the earnings are

Participant 1: 2 ECU

Participant 2: 12 ECU

Participant 3: 0 ECU (= 3 ECU – 3 ECU)

Participant 4: 0 ECU (= max (2 ECU – 3*1 ECU, 0)

Since points were assigned in stage four and two participants have positive earnings, a fifth stage follows. If in the fifth stage Participant 2 assigns 2 points to Participant

1 and Participant 1 assigns 1 point to Participant 2, the earnings are
Participant 1: -1 ECU (= max (2 ECU – 2*1 ECU, 0) – 1 ECU)
Participant 2: 1 ECU (= max (12 ECU - 1*9 ECU, 0) – 2 ECU)
Participant 3: 0 ECU
Participant 4: 0 ECU|>

The period now ends, because only Participant 2 can assign more points, but because all other participants have non-positive earnings, Participant 2 is not allowed to assign points to any of them.

Before the actual experiment begins you need to answer some questions to check your understanding of the rules. You can find these questions on the next page and the answers will be submitted on your computer.

After all participants have answered all questions correctly each group will have a short contest. In the contest you will see balls dropping on your screen and you can catch balls with a paddle which can be moved clicking the "Left" and "Right" buttons. You get 4 points for every ball caught, while each left or right click costs you 1 point. Your goal is to accumulate as many points as possible. **The person in your group accumulating the most points in one minute will be given the ID and <|{the advantages}|> of "Participant 1".** The other 3 IDs will be randomly allocated among the other three group members.

There will be one practice round before the actual contest. Please note that other than determining who is to be Participant 1, the points in the contest will not affect your actual earnings or the flow of the experiment as described in the instructions.

If you have any further questions please raise your hand and one of the supervisors will come to help you.

Control Questionnaire

1. Each group member has an endowment of 20 ECU. Nobody (including you) contributes any ECU to the project. How high are:

   (a) Your earnings after the first stage, if you are Participant 1?

   (b) Your earnings after the first stage, if you are Participant 3?

2. Each group member has an endowment of 20 ECU. You contribute 20 ECU to the project. All other group members contribute 20 ECU each to the project. What are:

   (a) Your earnings after the first stage, if you are Participant 1?

   (b) Your earnings after the first stage, if you are Participant 4?

4. Each group member has an endowment of 20 ECU. The other three group members contribute together a total of 30 ECU to the project. What are:

   (a) Your earnings after the first stage if you contribute 0 ECU to the project, if you are Participant 1?

   (b) Your earnings after the first stage if you contribute 0 ECU to the project, if you are Participant 2?

5. Your earnings from the first period 25 ECU. How much will your earnings be at the end of the second stage be if:

   (a) You receive 2 points from Participant 2, but do not assign any yourself?

   (b) You receive 2 points from Participant 1 and assign 3 points yourself?

6. *True or false:* Each participant will be given an identification number which will allow others in their group to follow their actions in the experiment.

7. *Yes or NO?* Assume you assign 2 points to another group member, no one else in your group assigns any points and all members in your group have positive earnings. Will another stage follow?

8. *Yes or NO?* Assume no member of your group assigns points including you. Will another stage follow?

Appendix B: Regression Tables

This section contains all regression tables mentioned in the paper.

Table B1 Propensity to feud.

	(1)	(2)
Privilege Treatment	−0.468	−0.917
	(0.940)	(1.956)
Strongman Treatment	−0.074	−1.818
	(0.913)	(1.964)
Alpha Treatment	−0.564	0.347
	(0.953)	(1.682)
Periods (11-20)	0.903	−5.357
	(1.022)	(1.252)
Periods^2	−0.027	0.016
	(0.033)	(0.041)
Constant	−11.326	−1.775
	(7.912)	(9.601)
Observations	660	660
Wald chi^2	1.62	0.91

Columns 1 - 2 report the coefficients of logistic regressions with clustered standard errors fitted to the incidence of within period and inter-period feuds respectively. The independent variables include treatment dummies with appropriate names; 'Periods' is a time variable for periods 11-30; 'Periods^2' is a time variable squared. The omitted category is the *Symmetric* treatment. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2 Main variables as a function of treatment and time.

	(1) Contrib.	(2) Pun. Spending	(3) Loss of Earnings	(4) Earnings
Privilege Treatment (PT)	0.485	1.351	6.607**	−6.316*
	(2.483)	(1.063)	(3.365)	(3.653)
Strongman Treatment (SMT)	5.536**	−0.250	0.025	2.149
	(2.594)	(1.240)	(3.536)	(3.815)
Alpha Treatment (AT)	1.108	0.794	4.328	−3.913
	(2.418)	(1.294)	(3.431)	(3.808)
Periods (11-20)	0.321***	0.015	0.092	0.100
	(0.065)	(0.047)	(0.133)	(0.139)
Periods3 (18-20)	−0.663***	−0.065	−0.280	−0.118
	(0.184)	(0.133)	(0.374)	(0.392)
PT × Periods	−0.318***	−0.089	−0.448**	0.258
	(0.104)	(0.063)	(0.191)	(0.194)
SMT × Periods	−0.205*	0.038	0.067	−0.096
	(0.114)	(0.077)	(0.204)	(0.207)
AT × Periods	−0.230**	−0.044	−0.248	0.136
	(0.096)	(0.081)	(0.197)	(0.208)
PT × Periods3	0.317	0.129	0.715	−0.524
	(0.292)	(0.177)	(0.538)	(0.547)
SMT × Periods3	−0.158	0.243	0.900	−1.123*
	(0.320)	(0.217)	(0.574)	(0.582)
AT × Periods3	0.182	0.395*	1.468***	−1.422**
	(0.270)	(0.229)	(0.553)	(0.586)
Constant	8.587***	0.547	1.554	23.599***
	(1.707)	(0.786)	(2.367)	(2.621)

Continued on the next page.

Table B2 - continued from the previous page.

	(1)	(2)	(3)	(4)
σ(group)	5.590***	1.385***	5.410***	5.886***
	(0.514)	(0.177)	(0.530)	(0.709)
σ(subject)	2.386***	1.655***	3.151***	6.816***
	(0.148)	(0.105)	(0.234)	(0.377)
σ(residual ST)	2.935***	2.118***	5.976***	6.268***
	(0.086)	(0.062)	(0.174)	(0.184)
σ(residual PT)	3.734***	1.940***	6.357***	6.275***
	(0.107)	(0.055)	(0.179)	(0.180)
σ(residual SMT)	4.181***	2.734***	6.966***	6.879***
	(0.124)	(0.082)	(0.211)	(0.202)
σ(residual AT)	2.935***	3.078***	6.718***	7.177***
	(0.086)	(0.088)	(0.194)	(0.207)
Observations	2640	2640	2640	2640
Wald chi2(11)	62.75***	20.12**	36.89***	48.13***

Columns 1 - 4 report the coefficients of multilevel mixed random effects models, which allow to control for individual and group differences. The respective dependent variables are contributions, punishment spending, overall loss of earnings due to punishment and earnings. The independent variables include treatment dummies with appropriate names; 'Periods' is a time variable for periods 11-20; 'Periods3' is a time variable for periods 18-20; interaction variables are also included. The Symmetric treatment is the omitted category. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3 Main experimental variables as function of treatment, time and role

	(1) Contrib.	(2) Pun. Spending	(3) Loss of Earnings	(4) Earnings
ST - Participant 1	1.172	−0.776	−3.857	2.686
	(2.212)	(1.732)	(4.503)	(4.632)
PT - Participant 1	5.567	4.256**	1.209	4.888
	(3.398)	(1.643)	(4.920)	(5.965)
PT - Rank and File	−0.818	0.124	7.120*	−9.155**
	(2.702)	(1.266)	(3.740)	(4.376)
SMT - Participant 1	7.177*	−0.132	−3.372	5.053
	(3.807)	(2.496)	(5.998)	(5.697)
SMT - Rank and File	5.380*	−0.548	−0.128	2.077
	(2.892)	(1.669)	(4.241)	(4.269)
AT - Participant 1	3.953	0.810	2.785	6.668
	(3.441)	(2.322)	(5.539)	(6.307)
AT - Rank and File	0.550	0.530	3.556	−6.545
	(2.720)	(0.581)	(4.018)	(4.533)
Periods (11-20)	0.335***	0.004	0.035	0.144
	(0.077)	(0.061)	(0.157)	(0.162)
Periods3 (18-20)	−0.620***	−0.085	−0.243	−0.199
	(0.218)	(0.170)	(0.443)	(0.455)
ST - Part.1 × Periods	−0.056	0.041	0.231	−0.175
	(0.154)	(0.121)	(0.314)	(0.323)
PT - Part.1 × Periods	−0.536***	−0.243**	−0.086	0.119
	(0.192)	(0.108)	(0.315)	(0.382)
PT - R&F × Periods	−0.263**	−0.024	−0.492**	0.246
	(0.128)	(0.080)	(0.223)	(0.257)
SMT - Part.1 × Periods	−0.411*	0.035	0.275	−0.192
	(0.225)	(0.170)	(0.395)	(0.361)
SMT - R&F × Periods	−0.155	0.053	0.075	−0.122
	(0.145)	(0.110)	(0.262)	(0.247)
AT - Part.1 × Periods	−0.352*	−0.035	−0.222	0.315
	(0.196)	(0.157)	(0.362)	(0.408)
AT - R&F × Periods	−0.208	−0.036	−0.180	0.018
	(0.060)	(0.103)	(0.093)	(0.270)

Continued on the next page.

Table B3 - continued from the previous page.

	(1)	(2)	(3)	(4)
ST - Part.1 × Periods3	−0.175	0.041	−0.151	0.326
	(0.435)	(0.121)	(0.886)	(0.911)
PT - Part.1 × Periods3	0.476	−0.243**	0.238	−0.622
	(0.542)	(0.108)	(0.888)	(1.077)
PT - R&F × Periods3	0.206	−0.024	0.823	−0.383
	(0.360)	(0.080)	(0.627)	(0.725)
SMT - Part.1 × Periods3	−0.272	0.035	0.607	−0.589
	(0.635)	(0.170)	(1.113)	(1.017)
SMT - R&F × Periods3	−0.179	0.053	0.947	−1.192*
	(0.407)	(0.110)	(0.737)	(0.609)
AT - Part.1 × Periods3	0.294	−0.035	1.603	−2.183*
	(0.542)	(0.157)	(1.019)	(1.150)
AT - R&F × Periods3	0.087	−0.034	1.372	−1.060
	(0.365)	(0.103)	(0.691)	(0.761)
Constant	8.829***	0.741	2.518	22.927***
	(1.805)	(0.952)	(2.652)	(2.870)
σ(group)	5.708***	1.587***	5.606***	6.777***
	(0.503)	(0.151)	(0.509)	(0.609)
σ(residual ST)	3.009***	2.356***	6.126***	6.301***
	(0.085)	(0.067)	(0.173)	(0.178)
σ(residual PT)	4.087***	2.077***	6.339***	8.038***
	(0.112)	(0.057)	(0.174)	(0.221)
σ(residual SMT)	4.764***	3.571***	8.159***	7.259***
	(0.135)	(0.101)	(0.231)	(0.206)
σ(residual AT)	4.179***	3.371***	7.559***	8.695***
	(0.115)	(0.093)	(0.133)	(0.239)
Observations	2640	2640	2640	2640
Wald chi2(23)	142.87***	46.38***	38.08**	828.68***

Columns 1 - 4 report the coefficients of random effects models, which allow to control for group differences. The respective dependent variables are contributions, punishment spending, overall loss of earnings due to punishment and earnings. The independent variables include the interaction terms of treatment dummies with the role dummy (1: contest winner/pivotal player; 0: otherwise); 'Periods' is a time variable for periods 11-20; 'Periods3' is a time variable for periods 18-20; interaction variables with the time variables are also included. 'Symmetric - Rank and File' is the omitted category. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Punishment: getting even or sending a message?

Abstract

The willingness to sacrifice resources to punish others in one-shot interactions has been explained with distributional concerns and intentions. In this paper, we take a different approach by looking into the messaging dimension of punishment. We conduct two different one-shot online studies. In one we check whether individuals are ready to replace own punishment with a costless penalty, inflicted by a random mechanism, or do they insist on making a statement by not reducing own punishment to zero. In the other, targeted individuals might remain unaware of the punishment and subjects are given the opportunity to affect the odds of this happening. Both studies provide evidence for the messaging being among primary motives to a few punishers. Moreover, in the second study, we find a variety of behaviours that include the willingness to decrease own earnings to hide punishment decisions (both punishment and lack of it) and readiness to pay to assure others know that no punishment was meted out by that individual.

3.1 Introduction

The experimental literature provides abundant evidence on the inclination of some individuals to take costly action to affect others, even if they yield no immediate or future material benefit. This robust phenomenon has been observed in various settings and has been replicated in numerous studies. For example, Güth et al. (1982) and others (Forsythe et al. (1994), Bolton and Zwick (1995) and Schotter et al. (1996) are among them) reported rejections of strictly positive offers in the Ultimatum game that resulted in both participants receiving nothing. In the public goods game participants punish others despite the absence of potential future gains due to the anonymity of interaction and zero probability for a rematch, as in Fehr and Gächter (2002), or the period being the last in a series of interplays, for instance, Noussair and Tucker (2005), Sefton et al. (2007), Engelmann and Nikiforakis (2014), and the previous two chapters. Furthermore, Andreoni et al. (2003), who allowed receivers in Dictator game to punish allocators, also find punishment in the final period of the game. In Power-to-Take game (Bosman and Van Winden, 2002) participants relinquished earnings just to decrease appropriators' payoffs, even though no further interactions were to take place and no reputational benefits could be reaped.

The inability of the rational choice paradigm under the assumption of complete selfishness (maximisation of earnings) to accommodate such behavior prompted the rise of new theoretical models and further experimental inquiry into the motives behind the non-strategic punishment. The new models, prescribe agents non-pecuniary concerns, such as others' intentions (Cox et al., 2007; Dufwenberg and Kirchsteiger, 2004; Rabin, 1993), distribution of outcomes (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) or both (Charness and Rabin, 2002; Falk and Fischbacher, 2006), to

explain the driving forces behind punishment.

Despite the potential differences in motives, functionality-wise, punishment in the many studies shares common characteristics. It fulfills two purposes: 1) redistribution of resources 2) informing a targeted party about this. Which of the two is valued by the penalizing individuals is an open question.

In case sanctioners find only the distributional concerns vital they should be indifferent to the state of knowledge of the punished, hence a stealth act of punishment, which keeps the targeted party oblivious to it, in the eyes of a punisher should be as valuable as one that makes the punished party aware of the retribution. Moreover, a transgressor's misfortune should be equivalent to punishment administered by the harmed individual. Nonetheless, it is possible that punishment serves as means of communication, as it allows to express disapproval of someone's action, thus even unlucky perpetrators would not be spared, and individuals would make an effort to assure targeted individuals get to know about the fact they were punished.

In the current paper, we probe into the importance of the messaging function of punishment, by conducting two online studies on Amazon Mechanical Turk (MTurk). One study is dedicated to exploring the messaging aspect of punishment in the presence of outcomes generated by an exogenous mechanism (fate) and in the other we take a new approach to check whether targets' awareness of sanctions is essential to those imposing them. Therefore the current research not only takes a new approach towards testing the importance of targets' awareness of retribution to the punishers but also provides new insights into subjects' consideration of punishment by fate.

Both studies suggest that for a considerable number of individuals who are willing to relinquish resources to affect the welfare of others, punishment is not just a tool for getting even, but also a means of communication. Results of the first study show that

while some individuals were less harsh with those affected by the random negative shock, others did not adjusting their punishment to account for targets' misfortune. We suggest that the latter were interesting in sending a message with their actions, as they kept punishment in place despite the fact that negative shock was equivalent in magnitude to the punishment they wished to assign if a targeted co-player was not affected by the shock.

In the second study about 20% of sanctioners paid to assure others' awareness of punishment. Additional insights include the readiness of a few to pay to make sure other group members get to know that no harm was done to them at the hands of their peers. Also, we report about individuals who showed interest in their hostile actions, or absence of them, being hidden from other group members, despite the one-shot nature of the interactions.

The remainder of the paper has the following structure. Section 2 presents the related literature and research questions. In Section 3 experimental designs are introduced. We show the results of both studies in two distinct subsections of Section 4. Finally, a general discussion in Section 5 is followed by the Appendix.

3.2 Related literature and research questions

The abundant experimental results on individuals' readiness to punish and reward in non-strategic settings not only led to the development of other-regarding theoretical models but also to further experimental enquiry into punishment preferences. To this end, strategic considerations were curtailed through the implementation of the stranger (Andreoni, 1988) and the perfect stranger (Fehr and Gächter, 2000) protocols or by conducting one-shot anonymous sessions.

Two of such project, undertaken independently, showed that though altruistic punishment defies some of the neoclassical assumptions, economic logic can still be applied to it. Both Carpenter (2007) and Anderson and Putterman (2006) used public goods games to demonstrate that demand for punishment is sensitive to the price of punishment. Moreover, Carpenter demonstrates that people behave as if punishment was an inferior and inelastic good.

De Quervain et al. (2004) used a trust game in which player A by transferring his endowment to player B could quadruple the received amount. After the transaction was complete, B could either keep it all, the augmented transfer and her initial endowment or opt for the equal split. In the following stage, A was allowed to mete out punishment points, while his brain was scanned. The authors found that punishment of free riders activates brain areas associated with rewards. Also, they report a correlation between the strength of activity in the area and punishment spending. Their results, therefore, support the augmented utility functions of social preference models that award utility for punishing ill-intentioned or distributionally unfair actions.

While the improved utility functions are supported by the above study, the experimental evidence on motivation is inconclusive. Dawes et al. (2007) used a game with random allocation of different- sized earnings, under perfect stranger protocol, to separate between norm enforcement and equality as motivators of punitive behaviour. Low-earning participants were willing to pay a cost to reduce the income of high earners. Therefore, authors conclude that inequality is the primary factor in motivating reduction (or augmentation) of others' incomes.

A similar conclusion was reached by Raihani and McAuliffe (2012), after using a simplified version of the moonlighting game (Abbink et al., 2000). In each of their three treatments, the endowment of player 1 and the amount confiscatable by player 2 were

fixed, while the latter's endowment varied. They show that the punishment frequency increased markedly only when the confiscation resulted in an adverse inequality.

It can be argued that so long as confiscations of windfall gains resulted in greater equality, such acts were not regarded as unfair since people perceive windfall and earned gains differently (Oxoby and Spraggon, 2008). Only when the second player confiscated to get ahead, this behaviour was deemed harmful and negative emotions arouse. This kind of argument is made by Charness and Rabin (2002) and Falk et al. (2005), as well as by Falk et al. (2008), who stress the importance of intentions as a motive for penalizing others.

Charness and Rabin (2002) used Dictator, two-player and three player response games with a variety of payoff distributions to differentiate between reciprocity and distributional concerns. The authors find strong evidence for efficiency concerns (moderated by will to help those who are less well off). They also show that reciprocity has some relevance in explaining the behaviour in laboratory experiments, but not the inequality aversion, as they argue that "apparent adequacy of recent difference-aversion models has likely been an artifact of powerful and decisive confounds".

The inquiry into the other regarding preferences was continued by Falk et al. (2005), who used one-shot, three-player prisoner's dilemma game, with punishment opportunities. According to the results only spiteful punishment, executed by defectors against cooperators, was driven by payoff differences. The cooperators, however, whose punishment was more frequent and stronger, seemed to be moved by the desire to harm defectors, even if inequality could not be affected.

Economists are not the only ones interested in punishment. The psychology literature offers a different perspective on motives behind punishment. According to Frijda (1994) and Heider (2013) primary motivation of punishers is "pain equalization"

(relative suffering hypothesis). French (2001) hypothesise that without the awareness of the punished, a penalty has no value to a punisher.

Gollwitzer and Denzler (2009) used a lexicon decision task to measure the accessibility of aggression concepts before and after the punishment[1]. The authors allege that punishment is only satisfactory if conducted by the harmed, as opposed to a penalty by fate, and just if a transgressor understands the reasons for being punished. Therefore, their findings corroborate the notion that punishment is used to convey a message to the punished. Later Gollwitzer et al. (2011) found some weak evidence for relative suffering hypothesis and much stronger evidence for the understanding hypothesis, according to which punishment is only satisfactory if a transgressor understands the reasons for being punished.

In a revised version of my M.A. thesis a more common approach for economists, based on the principle of revealed preferences, was taken. The focus was concentrated on the messaging component of punishment, without addressing the understanding hypothesis. The awareness of the sanctioned individuals was manipulated, while disclosing to potential punishers the future state of knowledge of the former, in original versions of the dictator, public goods, and power-to-take games. The data demonstrate that the overall incidence of punishment was higher when an informative message about the punishment was delivered to perpetrators.

In the current paper, we continue to look into messaging role of punishment.

[1] Individuals were paired with an alleged partner. In reality, this was a preprogrammed machine, whose actions at the end of the stock market game led to an unfavourable outcome for the human participants. After participants were informed of the result, they carried out a "lexicon decision task", in which strings of letters were displayed one by one on the participants' screens, and players were asked to identify actual words they saw among the strings. According to a motivational priming model by Förster et al. (2005) after the display of the stock market game results, concepts related to aggressions became more accessible. The model suggests that the heightened accessibility was to disappear only after satisfactory punishment took place.

To this end, we use techniques popular with economists and conduct two different studies to answer two separate but related research questions: a) "is punishment by fate a perfect substitute for own punishment?" b) "is awareness of a penalty by targeted individuals valuable to punishers?". If the answer to the first question is positive, a targeted individual will face no punishment when the consequences of her bad luck match the outcome of the conceived sanctions. However, if messaging is valued, punishers will insist on meting out some punishment, even if the desired level of suffering by the target was reached. The 'yes' answer to the second question would mean that punishing individuals should be willing to sacrifice resources to affect punished awareness of the retribution.

3.3 Experimental designs

Experimental methods are useful for studying punishment preferences as they allow us to reveal individuals' true tastes by eliminating any concerns about future repercussions, such as counter-punishment, with one-shot anonymous interactions. In the current paper, we use a public goods game with punishment opportunities, which relates to the well-established literature on cooperation and its enforcement initiated by Fehr and Gächter (2000), as the basis of all treatments.

What makes this framework useful is the tension between personal and collective well-being. Incidents of punishment in such settings are frequent and well documented (for literature reviews see Fehr and Fischbacher (2003) and Chaudhuri (2011)). Two types of punishment are differentiated in the literature. A penalty assigned to an individual with a lower contribution to the public good than the punisher is called "pro-social". Punishment in the opposite direction is referred to as "anti-social".

3.3.1 Baseline treatment

The purpose of the *Baseline* treatment (henceforth BT) is somewhat uncommon. By serving as a benchmark for the aggregate levels of cooperation and punishment spending, it allows checking to what extent introduction of novel features in other treatments affects behaviour.

The BT consists of two stages. Stage one is the contribution stage, where three group members simultaneously decide how many tokens (c_i) of their 20 token endowments to contribute to the public account. For each token contributed to the account, every group member receives 0.5 tokens, regardless of own contribution. In this setting, collective earnings are maximised if all members contribute their whole endowments, but one can increase own earnings by contributing less, ceteris paribus.

Earnings of any participant at the end of the stage are given by the following payoff function

$$\pi_i^1 = (20 - c_i) + (0.5 \cdot \sum_{j=1}^{3} c_j).$$

After all members of a group finish submitting stage 1 decisions, they are informed of own contribution, group's average contribution, and their final earnings along the values in the parenthesis in the formula above.

In the next stage, a participant i learns of others' stage 1 contributions and can assign up to 7 deduction points to each group member (p_{ij}) at the cost of 1 token per point. Every such point reduces the earnings of a recipient by 2 tokens. To make the maximum punishment feasible under all circumstances (including the scenario of a single contributor, who contributes the total endowment) participants receive an extra 10 tokens at the beginning of the stage.

We already mentioned that one-shot anonymous interactions reveal individuals' actual punishment preferences by eliminating strategic considerations. However, their choices still might be moulded by expectations of others' punishment choices. Subjects might take into account the possibility of being at the receiving end of the punishment, or they might hope for the third group member to penalize the transgressor. To deal with this issue, stage 2 decisions of only one group member, picked randomly (R), are implemented. All other decisions are discarded and affect no one.

The following equations give the earnings of aparticipant i by the end of this stage

$$\begin{cases} \pi_i^2 = \pi_i^1 + 10 - \sum_{j=1, j\neq i}^{3} p_{ij} & \text{for } i = R \\ \pi_i^2 = \pi_i^1 + 10 - 2 \cdot p_{ji} & \text{for } i \neq R. \end{cases}$$

3.3.2 Study 1 - Fate treatment

The contribution stage in the FT is followed by an extra stage in which the earnings of the participants are affected by a discrete random shock (α). The shock can either increase the earnings by 5 tokens ('+5 outcome'), decrease them by the same amount ('−5 outcome') or leave them unchanged ('0 outcome'), with equal probability ($\alpha \in \{-5, 0, 5\}; P(-5) = P(0) = P(5) = \frac{1}{3}$). Individual shock outcomes are independent and are shown in the final report at the end of the game. Therefore no output is displayed, or input is required at this stage and participants continue directly to stage three.

The third stage, like the second stage of the BT, provides participants with an extra 10 tokens, information about others' contributions and the opportunity to assign deduction points at the same cost. A major difference between the two treatments

is the use of the strategy method in the FT. Instead of entering a single number, participants state their punishment preferences for each of the three outcomes of the shock ($p_{ij}^{\alpha} \in [\![0, 7]\!]$). This feature makes it possible to adapt punishment decisions to the (mis)fortune of a targeted individual.

The coordination problem is solved in the same manner, i.e., by picking and implementing decisions of one individual in each group. However, the introduction of the shock raises concerns over punishment decisions being shaped by participants' risk preferences. To relieve this concern the chosen individuals are not affected by the shock.

The earnings of a participant i after the punishment stage can be expressed by

$$\begin{cases} \pi_i^2 = \pi_i^1 + 10 - \sum_{j=1, j\neq i}^{3} p_{ij}^{\alpha_j} & \text{for } i = R \\ \pi_i^2 = \pi_i^1 + 10 - 2 \cdot p_{ji}^{\alpha_i} + \alpha & \text{for } i \neq R. \end{cases}$$

3.3.3 Study 2 - Awareness treatment

In the AT, after submitting contribution decisions, participants proceed to stage 2, where their earnings are changed by a continuous random shock (β). From the instructions, they know that small changes, positive or negative, are more likely than the big ones, but any outcome between doubling the earnings and nullifying them is possible[2]. The exact magnitude of the shock is not revealed at this point, and the punishment stage follows.

[2] The magnitude of the shock is determined by a draw from an adapted Irwin-Hall distribution, which is a symmetric distribution, generated by summing a finite number of identical random variables uniformly distributed. Our distribution was characterised by the following parameters n = 6, a support of [-1,1], an expected value of 0 and variance of $\frac{1}{6}$. The distribution is such that extreme outcomes are improbable and in our sessions, none of the participants even experienced a shock of 50%.

The punishment stages of the AT and the BT share most of the settings, namely participants receive an extra 10 tokens upon entering that stage, they learn about others' contributions and are allowed to assign deduction points to other group members ($p_{ij} \in [\![0, 7]\!]$). Also, decisions of only one member in every group are implemented, and for the same reasons as in the FT, the random shock does not affect her earnings.

By the end of the punishment stage an individual whose punishment choices are not implemented can expect a change in earnings from two sources: a) random shock and b) punishment. If only showed her final earnings, with no information about the origins of a change, a punished individual is not able to know whether she was penalised or just unlucky. This information is displayed with the probability of $\frac{1}{2}$, but it can be changed to 1 or 0 by paying 1 token (payable only if decisions are implemented) in stage 4 of the game (hereafter the options to reveal and hide punishment decisions referred to as "Info" and "No-Info")[3]. A costless option (denoted by "No-Pref" in the analysis section) that kept the probability unchanged was available for those who were not concerned with punished awareness.

The earnings of any participant at the end of the game can be expressed by[4]

$$\begin{cases} \pi_i^2 = \pi_i^1 + 10 - \sum_{j=1, j\neq i}^{3} p_{ij} - \gamma & \text{for } i = R \\ \pi_i^2 = \pi_i^1 \cdot (1+\beta) + 10 - 2 \cdot p_{ji} & \text{for } i \neq R, \end{cases}$$

where $\gamma \in 0, 1$ is the cost incurred in stage 4.

[3] Theoretically, if no punishment is assigned a potentially punished individual might become aware of not being penalised. This might occur if $\pi_i^1 \cdot (1+\beta) > \pi_i^1 \cdot 2 - 2$, but the probability is negligible.

[4] Negative earnings were allowed in theory, but to be consistent with the description of the HIT they could not affect the flat fee. Furthermore, no instances of negative earning took place.

3.4 Recruitment, procedures and sample

3.4.1 Recruitment and procedures

The experiment was conducted online. All subjects were recruited via Amazon Mechanical Turk (MTurk), a successful crowdsourcing platform that is suitable for recruiting participants for laboratory-style experiments[5]. The platform offers a pool of over 500,000 users, who complete over 40,000 Human intelligence tasks (HITs), posted by "requesters", every day (Difallah et al., 2015; Ipeirotis, 2010).

All HITs include a short description of the task. Our experiment was published under the description of a study on social decision making. We informed potential participants that all interactions are anonymous and that the saved data is used for scientific purposes only. MTurkers could take part in our study only once and as compensation for the expected 15 minutes of their time they were promised \$0.5 flat completion fee and up to \$2 as a bonus payment[6].

Those who accepted the HIT proceeded to the first screen of the task, which was implemented in LIONESS (Arechar et al., 2017). After reading the instruction and successfully answering all comprehension questions, they entered the "Lobby" stage, to be matched with another 2 participants from the same experiment[7]. Once a group was formed, the experiment began[8].

Upon completion of all stages, participants proceeded to answer a short ques-

[5] For evaluations of MTurk for conducting experiments in behavioral sciences, please see Arechar et al. (2017); Buhrmester et al. (2011); Horton et al. (2011); Paolacci and Chandler (2014); Rand (2012).

[6] Due to random allocation to various experiments in most sessions and since the expected value of the shocks is 0, the maximum bonus did not reflect potential increases that could occur because of the random shocks.

[7] Participants waited up to 5 minutes the "Lobby" stage. In case a group was not formed, they could either claim the participation fee or choose to wait for another 2 minutes.

[8] Instructions and screenshots are in Appendix.

tionnaire, comprised of demographic questions (age, gender, and education), a brief measure of impulsivity and sensation seeking (Webster and Crysel, 2012), along with a couple of questions about beliefs in supernatural entities, namely God/s and Karma.

The brief measure of impulsivity and sensation seeking comprises 8 questions and has been developed to save time, which is particularly scarce in online studies. The authors have demonstrated that its success in predicting impulsive behaviour is comparable to the full version of the questionnaire. These questions are included since impulsive individuals might find self-control more difficult, as a result, their punitive preferences might stand out, and we wish to control for that.

The extra two questions about individuals' beliefs in supernatural were included to account for a potential difference in perception of the game by believers. For instance, Purzycki et al. (2016) show that degree of belief in gods might be correlated with cooperation levels. In our setting, belief in God/s or Karma might be associated with lower punishment levels, if believers rely on the supernatural forces to penalise transgressors.

After the final batch of answers was submitted, the validation code and the final payment in USD were displayed.

3.4.2 Attrition and attention

Although the settings of our tasks might resemble those used in laboratories, online studies give rise to entirely new pitfalls. For example, only on rare occasions, participants in a laboratory experiment leave before a session ends. However, in online studies, there is no way of preventing participants from closing their web browsers or from failing to submit answers due to a technical issue. Moreover, the number of

potential distractions is likely to be greater outside of a laboratory.

We addressed the latter by asking participants to make decisions within the time limit displayed on their screens. A failure to comply resulted in removal from the HIT.

Such settings could make the problem of attrition (participant dropout) more severe. However, we argue that while attrition can be a major problem for studies with repeated interactions[9], it is of limited importance for the current study due to the shortness of the interactive part of our experiments. Once a group had been formed, subjects were just a few steps away from claiming their participation fees and bonuses. Considering the relative simplicity and the short span of the task, dropouts initiated by participants seem to be unreasonable.

Our data confirm the above assertion. We registered only 8, 10 and 12 cases (3%, 3.5% and 3.8% of observations with records of contribution decisions) in the BT, the FT and the AT respectively, of participants dropping out after making their decisions in stage 1. For some, we even got emails pointing at technical problems as the reason for the interruption.

In addition to data loss after stage 1, some participants did not complete the questionnaire or parts of it, as their browsers crushed (6, 10, 2 subjects in the BT, the FT and the AT respectively). These observations are included both in descriptive statistics and in analyses if the available data allows, and excluded otherwise.

[9] For instance, Arechar et al. (2017) report that 47% of their groups experienced at least one dropout.

3.4.3 Samples

All participants were recruited via MTurk. In total, excluding groups with dropouts, 789 participants[10] (263 groups) took part in all treatments. 219 participated in the BT, of them, 52.11% were women[11], their average age was 34.53 years, and 76.06% reported about having an academic degree. Additional 315 subjects took part in the FT. Almost half of them were women (44.92%), 33.98 years was their average age, and most of them had either an undergraduate or a postgraduate degree (70.49%). Remaining 255 subjects participated in the AT. The proportion of women in the treatment was similar to that in the FT, to be precise 46.64%. The mean age of participants in this treatment was 34.85 years, and 71.54% had a degree[12].

Average earnings in our study were around 1.9 USD (including the flat fee).

3.5 Experimental results

3.5.1 Study 1 - Fate

Fate versus Baseline

The purpose of comparison is to check whether the extra components of the FT affected subjects' behaviour in contrast to the BT. We begin the analysis by comparing the contribution levels in the two treatments.

[10] 18 participants were not able to complete a part or the whole the questionnaire due to technical issues. Their decisions are not omitted. Instead, we include them wherever possible.

[11] One participant in the BT, 2 in the FT and 3 in the AT chose not to disclose their gender.

[12] According to t-test none of the demographic differences is statistically significant. The p-values for the differences in the proportion of women are 0.774 for the BT versus FT, 0.377 for BT versus AT and 0.509 for the remaining pair; the respective p-values for the differences in age are 0.532, 0.744, 0.341 and for the proportion of individuals with academic degrees are 0.202, 0.510, 0.544.

Result 1. *The contribution levels in the BT and the FT do not differ significantly.*

According to the data on average subjects contributed 13.74 tokens in the BT and 13.05 tokens in the FT. The difference is not statistically significant according to the two-sided Mann-Whitney test ($p = 0.269$).

To control for demographic and other variables, we estimate a Tobit regression with right censoring[13], stage 1 contributions as the dependent variable and the BT as the base category. These are the independent variables included in the model: 'Fate' is the dummy for the FT; 'Age' is the age of the participants; 'Female' is a dummy variable for female participants, and 'Unidentified' is a dummy variable for individuals who chose not to declare their gender.

Another two variables are 'Belief' and 'Impulsivity'. Their values are determined by participants' answers to the post-experimental questionnaire. A higher value of 'Belief' stands for a greater agreement of a participant with the existence of supernatural forces, such as God/s and Karma. Higher 'Impulsivity' is associated with greater impulsivity and sensation seeking.

The regression analysis confirms that contributions are not markedly different (Wald test on 'Fate': $p = 0.453$)[14]. Also, the analysis shows that women contributed on average less than men (Wald test on 'Female': $p = 0.010$), and there is a positive correlation between participants' age and contributions to the public account (Wald test on 'Age': $p = 0.017$). However, neither belief in supernatural forces, nor impulsivity is statistically significant in explaining contributions (Wald test on 'Belief' and 'Impulsivity: p values are 0.199 and 0.438).

[13] Tobit model was chosen due to a high proportion (slightly more than 20%) of individuals that chose to contribute their full endowments.

[14] The coefficients of the regression model are reported in Table A1.

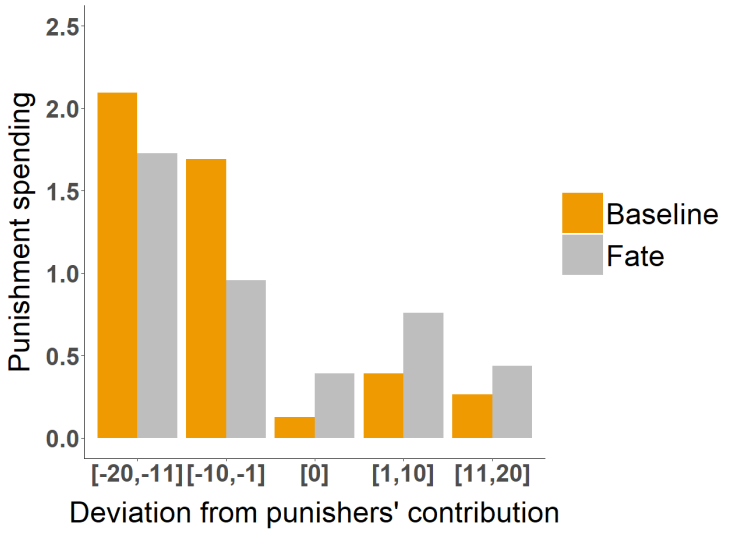


Figure 3.1: Punishment spending as a function of deviation from the contribution of punishers.

Result 2. *(i)Pro-social punishers in the BT on average assigned more deduction points than their counterparts in the FT.*

(ii)The level and likelihood of anti-social punishment were higher in the FT.

The statistical analysis suggests that contribution levels were similar in the two treatments, but not the levels of punishment spending. Figure 3.1 outlines lower pro-social and higher anti-social expected punishment spending in the BT in contrast to the FT (1.80 versus 1.15 tokens and 0.25 versus 0.56 tokens respectively).

Regression coefficients of the relevant models are reported in Table 3.1. The estimated models suggest that the variation in levels of pro-social and anti-social punishment are substantial (Wald test on "Fate": $p = 0.031$ and $p = 0.006$, columns 1 and 2 respectively). Furthermore, the likelihood of pro-social retribution is similar in the two treatments, unlike the anti-social punishment that was significantly more likely in the FT (Wald test on "Fate": $p = 0.113$ and $p = 0.004$, columns 2 and 4 respectively).

Table 3.1 Punishment in the Baseline and the Fate treatments

	Pro-Social		Anti-Social	
	(1)	(2)	(3)	(4)
Fate	−1.515**	−0.376	2.341***	0.903***
	(0.700)	(0.237)	(0.847)	(0.315)
Target's contribution	−0.139**	−0.025	−0.008	−0.007
	(0.071)	(0.022)	(0.052)	(0.019)
Other's contribution	0.089*	0.024	0.026	0.015
	(0.049)	(0.015)	(0.044)	(0.015)
Punisher's contribution	−0.029	−0.017	−0.057	−0.017
	(0.079)	(0.027)	(0.046)	(0.017)
Age	0.031	0.002	−0.032	−0.014
	(0.034)	(0.012)	(0.035)	(0.013)
Female	−0.288	−0.047	0.753	0.443*
	(0.733)	(0.260)	(0.660)	(0.251)
Unidentified	−0.040	−0.361		
	(3.995)	(1.202)		
Belief	0.394	0.206**	0.830***	0.307***
	(0.265)	(0.097)	(0.271)	(0.094)
Impulsivity	−0.038	0.053	1.048**	0.388**
	(0.503)	(0.172)	(0.459)	(0.159)
Constant	−2.502	−0.988	−11.200***	−4.032***
	(2.411)	(0.853)	(2.797)	(0.893)
N	826	826	1424	1424
F\Wald chi2	2.09**	17.73**	3.50***	37.51***

Columns 1 and 3 report the coefficients of Tobit regressions with left (0) and right (7) censoring and the number of assigned deduction points (expected value for the Fate treatment) as the dependent variable. Columns 2 and 4 reflect the estimates from logistic models fitted to the punishment decision (0: no punishment points were assigned; 1: at least one deduction point assigned (to any of the outcomes of the random shock in the FT)). The first two columns are for pro-social punishment and the last two for anti-social punishment. BT is the omitted category; 'Fate' is a dummy variable for the FT; 'Target's contribution' is the contribution of the punished participant; 'Other's contribution' is the contribution of the third player in the group; 'Punisher's contribution' is the contribution of the participant that assigned deduction points; 'Age' is the age of the participants; 'Female' and 'Unidentified'[15] are dummy variables for female participants and those who chose not to disclose their gender; 'Belief' is an average of the answers to questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for brief measure of impulsivity and sensation seeking. Robust standard errors clustered on individuals in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Result 3. *(i) The level of pro-social punishment is negatively correlated with contributions by the targeted individuals.*
(ii) The measure of impulsivity and sensation seeking, as well as the belief in higher powers, are positively correlated with levels and likelihood of anti-social punishment.

We find that a target's contribution played a substantial role in determining the level of pro-social punishment. On average, higher contributions by the punished individuals lead to fewer deduction points being assigned to them. Additionally, the short measure of impulsivity and sensation seeking, as well the measure of belief in Karma and God/s, are significant in explaining the levels of spending and the incidence of anti-social punishment. The latter is also positively correlated with the incidence of pro-social punishment.

Behaviour in the Fate treatment

Result 4. *On average both pro-social and anti-social punishers assigned most deduction points to the $+5$ outcome of the random shock, somewhat less to the -5 outcome and the least points were assigned to the 0 outcome of the shock.*

This part of the analysis is dedicated to punishment behavior in the FT. Because the focus is on the punishment levels for each of the shock outcomes, the sample is limited to individuals that assigned at least one deduction point to any of the shock results.

Visual inspection of Figure 3.2 implies that on average fewer deduction points were assigned to the 0 outcome of the random shock than to the other two results. To test the significance of above differences a Tobit model with right and left censoring, with

a number of deduction points assigned for each outcome as the dependent variable and a few independent variables as specified in Table 3.2, is estimated. The relevant coefficients are reported in columns 1 and 5.

The regression analysis supports the observation that for both pro-social (Wald test on "Random shock $+5$" and "Random shock -5": $p < 0.001$ for both coefficients) and anti-social punishment (Wald test on "Random shock $+5$" and "Random shock -5": $p < 0.001$ for both coefficients) the least deduction points were assigned to the 0 outcome of the shock. The differences between the amounts assigned to the other two results of the shock are significant for both pro-social (Wald test on "Random shock $+5$" = "Random shock -5": $p = 0.041$) and anti-social (Wald test on "Random shock $+5$" = "Random shock -5": $p = 0.004$) punishment.

A further examination indicates that on average punishment assigned to the $+5$ outcome of the random shock by the pro-social punishers was sufficient to eliminate the windfall benefits. However, the punishment did not suffice to bite into the excess earnings obtained through lower contributions in stage 1 (Wald test on "Random shock $+5$" $= 2.5$: $p = 0.606$, column 1). When it comes to the anti-social punishment, there is evidence of punishment being high enough to not only cancel-out the windfall gains, but also to reduce the earnings even further (Wald test on "Random shock $+5$" $= 2.5$: $p < 0.001$, column 5).

Result 5. *Pro-social and anti-social "zero shock punishers" on average assigned fewer deduction points to the -5 outcome of the random shock than to the neutral result.*

The overall proportion of individuals who showed the willingness to sacrifice resources to lower others' earnings in the FT was substantially higher than that in the

Table 3.2 Determinants of punishment - Fate treatment

	Pro-Social				Anti-Social			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Target's contribution	−0.172**	−0.133**	−0.082	−0.030	0.050	−0.043	0.327**	−0.097
	(0.072)	(0.065)	(0.086)	(0.043)	(0.053)	(0.115)	(0.158)	(0.084)
Other's contribution	0.064	0.050	0.009	0.009	0.056	−0.087	0.234*	−0.092
	(0.044)	(0.044)	(0.048)	(0.031)	(0.044)	(0.097)	(0.124)	(0.088)
Punisher's contribution	0.079	0.046	0.152	−0.079	−0.156***	0.056	0.013	0.108
	(0.068)	(0.078)	(0.130)	(0.054)	(0.048)	(0.079)	(0.095)	(0.067)
Random shock = +5	2.727***	0.254	0.635	0.156	3.353***	−0.745	−1.102	−0.435
	(0.439)	(0.262)	(0.645)	(0.245)	(0.534)	(0.470)	(0.751)	(0.479)
Random shock = -5	1.621***	−1.357***	−2.876***	−0.217	2.061***	−1.062**	−1.699**	−0.601
	(0.581)	(0.448)	(0.819)	(0.350)	(0.480)	(0.472)	(0.740)	(0.537)
Age	0.028	0.064***	0.051	0.051**	0.062**	0.066	−0.075	0.012
	(0.027)	(0.024)	(0.043)	(0.020)	(0.025)	(0.050)	(0.047)	(0.041)
Female	−0.885	−0.771	−0.660	−0.155	−0.560	−2.030**	−0.731	−0.828
	(0.593)	(0.584)	(1.031)	(0.263)	(0.555)	(0.963)	(1.106)	(0.995)
Belief	−0.031	−0.115	0.061	−0.133	0.263	0.374	1.513***	0.113
	(0.212)	(0.185)	(0.217)	(0.133)	(0.225)	(0.432)	(0.480)	(0.283)
Impulsivity	−0.401	−0.588	−0.531	−0.533**	0.288	−0.281	0.456	−0.954**
	(0.396)	(0.358)	(0.471)	(0.219)	(0.346)	(0.709)	(0.508)	(0.377)
Constant	−1.025	3.040	3.074	3.837**	−5.646***	2.899	−5.768	5.915***
	(1.924)	(2.175)	(2.334)	(1.598)	(2.003)	(4.737)	(4.142)	(2.115)
N	387	141	60	81	507	108	39	69
F	5.62***	5.94***	2.89***	3.98***	5.54***	1.44	370.00***	3.58***

Columns 1-8 report the coefficients of Tobit regressions with left (0) and right (7) censoring. The first four columns are for pro-social and the last four are for anti-social punishment. Columns 1 and 5 reflect the estimates based on the full sample of punishers; 2 and 6 sample restricted to decisions that include punishment of the 0 outcome of the shock; 3 and 7 sample restricted to decisions that include between 4 and 7 deduction points to 0 outcome of the shock; 4 and 8 sample restricted to decisions that include 1, 2 or 3 deduction points to the neutral outcome of the shock; The omitted category is punishment assigned to the 0 outcome of the shock. 'Target's contribution' is the contribution of the punished participant; 'Other's contribution' is the contribution of the third player in the group. 'Punisher's contribution' is the contribution of the participant that assigned deduction points; 'Random Shock = +5' and 'Random Shock = −5' are dummy variables for the punishment points assigned to the +5 and −5 outcomes; 'Age' is the age of the participants; 'Female' is a dummy variable for female participants; 'Belief' is an average of the answers to the questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for brief measure of impulsivity and sensation seeking. Robust standard errors clustered on individuals in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

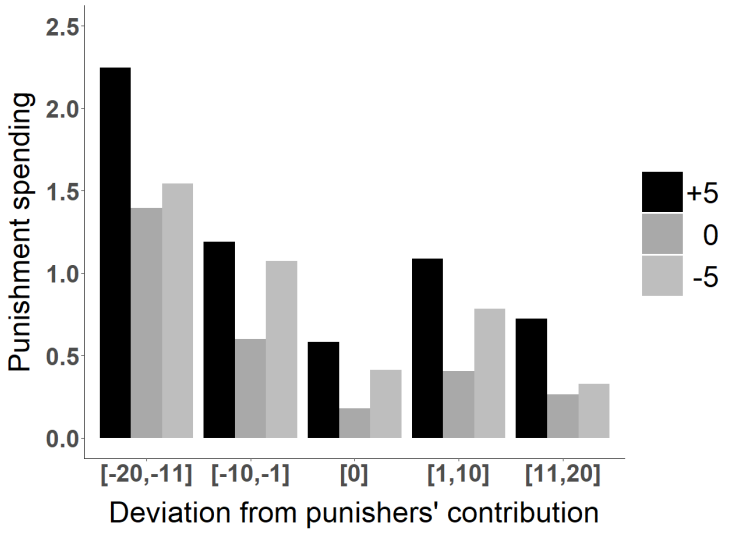


Figure 3.2: Average punishment spending (punishers only) in the Fate treatment, by outcomes of the random shock, as a function of deviation from contribution of punishers.

BT (49.84% and 32.42% respectively; Mann-Whitney test $p < 0.001$). Nevertheless, of 157 punishers in the FT, 96 assigned no deduction points in case a targeted individual was not affected by the shock (43 assigned deduction points to $+5$ outcome only; 15 to -5 outcome alone and 25 to both results of the shock; the rest were not consistent in their punishment decisions). The motivation for such acts of punishment can range from inequality aversion, in case deduction points are assigned to the positive outcome, to the nonlinearity of punishment preferences in case they add to the adverse consequences of the random process.

In the further analysis, we focus on decisions of 61 "zero shock punishers" that assigned deduction points to the 0 outcome of the shock, hence can be expected to mete out punishment in more standard settings that lack the shock and the strategy method. The average levels of punishment spending on various outcomes of the shock, as displayed in Figure 3.3, are compared to check participants' attitudes towards the stochastic changes in earnings of others.

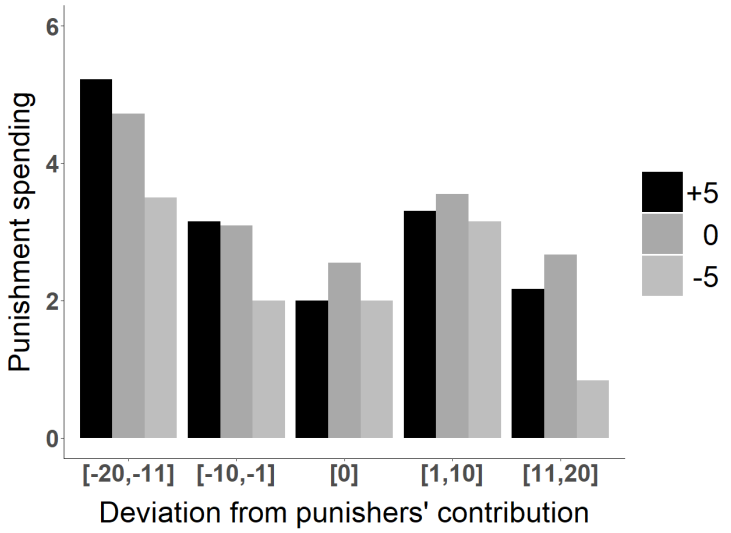


Figure 3.3: Average punishment spending, "zero shock punishers" only, by outcomes of the random shock, as a function of deviation from contribution of punishers.

For this purpose, we reestimate the previous Tobit model for the relevant subsample. The coefficients, presented in Table 3.2 (columns 2 and 6), suggest that "zero outcome punishers" took into account the negative consequences of the shock, but did not adapt their punishment behaviour to the positive outcome. That is true both for pro-social (Wald test on "Random shock $= +5$": $p = 0.334$; "Random shock $= -5$": $p = 0.003$, column 2) and anti-social punishers (Wald test on "Random shock $= +5$": $p = 0.116$; "Random shock $= -5$": $p = 0.027$, column 6).

Result 6. *"Zero shock punishers" that assigned up to 3 deduction points to the neutral outcome of the shock, did not reduce their punishment of targeted individuals who were about to be hit by the negative shock.*

Result 5 demonstrates that "zero shock punishers" on average adapted their penalties to the negative shock to others' earnings. Such a move allowed to achieve the desired distributional effect, which is inferred from the punishment assigned to the 0

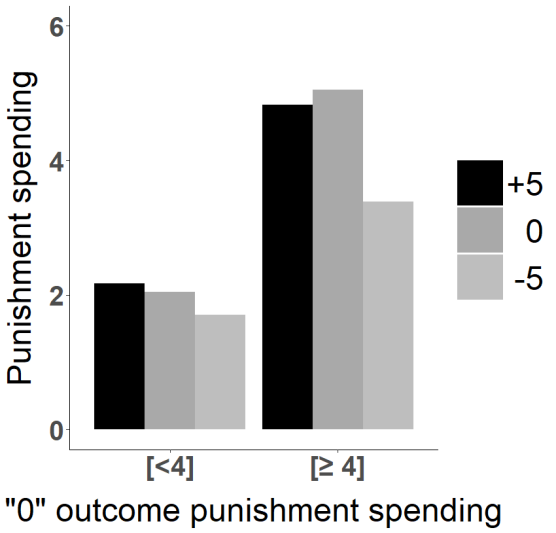


Figure 3.4: Average punishment spending, "zero outcome punishers" only, by outcomes of the random shock, as a function of the level of punishment assigned to the "0" outcome of the random shock.

outcome of the shock, at a lower cost. Because the negative effect of the shock was restricted to 5 tokens, replacing own punishment with a target's bad luck had different consequences for messaging for different levels of punishment.

Individuals who assigned up to 3 deduction points to the 0 outcome of the shock, could achieve a similar reduction of a target's earnings at no cost, but then no "message" was to be sent to the punished party. For high spenders, that assigned between 4 and 7 deduction points to the 0 outcome of the shock, the negative consequences of the stochastic process could not fully replace own punishment. To achieve the desired distributional effect, the negative shock had to be supplemented by own punishment, hence a "message" was communicated to the affected individual.

To account for the potential difference, we estimate two Tobit models for pro-social and another two for anti-social "zero shock punishers", with sub-samples being defined by punishment meted out to the 0 outcome of the shock. The coefficients for models that were estimated for those who spent at least 4 tokens, in case the earnings

of the punished were not changed by the shock, are presented in columns 3 (pro-social punishment) and 7 (anti-social punishment) of Table 3.2. The other two models, reported in columns 4 and 8 of Table 3.2, capture punishment choices of individuals who assigned no more than 3 deduction points to the 0 outcome of the shock.

The results indicate that pro-social high sanctioner adapted their punishment only to the negative outcomes of the shock, namely they did not assign more deduction points in case their targets were to experience a positive shock (Wald test on "Random shock $= +5$": $p = 0.330$), but less if the shock was negative (Wald test on "Random shock $= -5$": $p = 0.001$). The reduction was equivalent to the effect of negative shock (Wald test on "Random shock $= -5$" $= -2.5$: $p = 0.648$).

Individuals with the lower level of punishment spending (up to 3 deduction points to the 0 outcome of the shock) did not take the shock into to account at all. The average number of punishment points assigned to neutral outcome of the shock did not differ significantly from that assigned either to the positive (Wald test on "Random shock $= +5$": $p = 0.527$) or to the negative (Wald test on "Random shock $= -5$": $p = 0.538$) results of the shock.

The trends in the behavior of anti-social punishers were similar. Those who spent more than 3 tokens on deduction points to the 0 outcome decreased that amount if a targeted individual was to be affected by the negative shock (Wald test on "Random shock $= -5$": $p = 0.029$) and the reduction was equivalent in its magnitude to the negative shock (Wald test on "Random shock $= -5$" $= -2.5$: $p = 0.288$). Furthermore, the estimates suggest that the level of punishment assigned to the $+5$ outcome is similar to that of the 0 outcome (Wald test on "Random shock $= +5$": $p = 0.152$).

The conduct of less aggressive anti-social punishers that assigned up to 3 deduction points to the neutral outcome of the shock was also like that of their pro-social

counterparts. On average the numbers of deduction points assigned to 0 outcome of the shock and the other two potential results did not vary substantially (Wald test on "Random shock $= +5$": $p = 0.368$; "Random shock $= -5$": $p = 0.267$).

Discussion

In the first study, we investigated the importance of messaging in the presence of a stochastic mechanism that affects participants earnings. For the purpose, the one-shot BT was augmented through the introduction of a discrete random shock between the contribution and punishment stages. Implementation of the strategy method enabled participants to condition their punishment on three, equally likely outcomes of the random shock ($+5$, 0 or -5 tokens to the stage 1 earnings). The importance of punishment as a means of communication was determined by comparing punishment spending on the 0 and -5 outcomes of the random shock.

With the introduction of the novel components, the proportion of punishers increased (relative to the baseline treatment). Majority of punishers assigned no deduction points to the zero shock, so they were most likely moved by other goals rather than messaging. For example, some chose to punish only if their co-players were to enjoy the windfall gain, which can be explained by concern over their relative payoff. Others conditioned their punishment on the negative outcome of the shock. This kind of behaviour could be explained by either distributional concerns or reciprocity, but the utility function must account for non-linearity of preferences.

We found evidence for the messaging via punishment to be essential to a few subjects who assigned deduction points to the 0 outcome of the shock. They did not reduce the punishment of unlucky players to zero, even though the negative shock led to similar consequences as the punishment they assigned to the zero outcome of

the shock. This observation seems to in line with the claim that punishment by fate is not satisfactory in the eyes of punishers (Gollwitzer and Denzler, 2009). However, the more aggressive individuals, whose punishment of the zero shock exceeded the reduction caused by the negative shock, on average penalised the unlucky less severely. Even though this clearly shows that bad luck was perceived as a viable replacement of own punishment by a few, we cannot say much about the importance of messaging component to them without further research, which is beyond this paper.

Lastly, the behaviour of some "zero shock punishers" might have an alternative interpretation rather than care about sending a message through an act of punishment. It can be argued that those who did not adapt their punishment to the shock were driven by pure reciprocity, hence disregarded the distributional consequences of the shock.

3.5.2 Study 2 - Awareness

In the second study, the discrete random shock is replaced by a continuous one. By default, the information about the exact magnitude of the shock is withheld with probability $\frac{1}{2}$, hence punished individuals might not get to know if and how much they lost due to actions of another group member.

Awareness versus Baseline

This part of the analysis is dedicated to examining whether the extra components changed the contribution behavior in the AT in contrast to the BT.

Result 7. *On average participants in the AT and the BT contributed similar amounts.*

According to a Mann-Whitney test, average contribution levels in the BT and the AT (13.74 and 13.33 tokens respectively) do not differ significantly ($p = 0.729$). After controlling for gender, age, impulsivity, and belief in supernatural forces, by estimating a Tobit model (for coefficients see the first column of Table A2) a similar conclusion is reached, as both the treatment dummy (Wald test on "Awareness": $p = 0.934$) and the model as a whole are statistically insignificant (Prob $>$ chi2 $= 0.126$).

Result 8. *On average participants in the AT and BT spent similar amounts on deduction points. Moreover, there is no strong evidence for differences between the treatments in the probability of punishment to take place.*

The punishment stages in the AT and the BT are similar. In both participants are informed of others' contributions in stage 1 and are allowed to assigned deduction points. However, in the former, the punishment stage is preceded by the shock and followed by another stage, which enables participants to influence the odds of the punished to learn about the exact magnitude of the stochastic change and of the loss due to received deduction points.

According to the results reported in Table 3.3 the differences in punishment spending between the AT and the BT are statistically insignificant. This is true both for pro-social and anti-social punishment (Wald test on "Awareness": p equals 0.192 and 0.664, columns 1 and 3 respectively). We find weak evidence of higher likelihood of pro-social punishment to occur in the AT and no evidence for discrepancies in probabilities of anti-social punishment (Wald test on "Awareness": p equals 0.098 and 0.598, columns 2 and 4 respectively).

Result 9. *(i) The levels of pro-social and anti-social punishment are negatively*

correlated with contributions by the targeted individuals and by the punishers respectively.
(ii) The measure of impulsivity and sensation seeking is positively correlated with levels and likelihood of anti-social punishment.

The estimated coefficients suggest that pro-social punishers considered the amount contributed by the punished when assigning the deduction points (Wald test on 'Target's contribution': $p < 0.001$). There is also evidence on the negative correlation between the amount contributed by an anti-social punisher and the amount spent on reducing others' earnings (Wald test on 'Punisher's contribution': $p = 0.001$).

Among subjects that contributed less than their peers, higher impulsivity was associated with higher punishment spending and higher punishment frequency (Wald test on "Impulsivity": p equals 0.003 and less than 0.001, columns 3 and 4 respectively). In addition, there is weak evidence on the positive correlation between the belief in supernatural powers and spending, as well as frequency of anti-social punishment (Wald test on "Belief": p equals 0.082 and 0.062, columns 3 and 4 respectively).

Behaviour in the Awareness treatment

Result 10. *(i) Almost a third (28.24%) of all participants and punishers (30.06%) were willing to sacrifice some resources to affect the odds of others to learn their punishment decisions.*
(ii) Participants who chose to pay 1 token to conceal their actions contributed the least.

Previously we already mentioned that subjects could sacrifice 1 token to change the default odds (1:1) of information about the precise consequences of the shock and

Table 3.3 Punishment in the Baseline and the Awareness treatments

	Pro-Social		Anti-Social	
	(1)	(2)	(3)	(4)
Awareness	0.986	0.457*	−0.572	−0.221
	(0.753)	(0.277)	(1.315)	(0.419)
Target's contribution	−0.359***	−0.089***	0.013	0.006
	(0.080)	(0.025)	(0.106)	(0.033)
Other's contribution	0.134**	0.046**	0.208***	0.067***
	(0.057)	(0.019)	(0.070)	(0.026)
Punisher's contribution	−0.063	−0.078**	−0.317***	−0.094***
	(0.089)	(0.036)	(0.097)	(0.024)
Age	0.054	0.016	−0.014	−0.004
	(0.042)	(0.013)	(0.068)	(0.020)
Female	0.192	0.046	1.238	0.521
	(0.790)	(0.278)	(1.385)	(0.442)
Unidentified	5.135**	0.000		
	(2.132)	(.)		
Belief	0.500	0.190	1.038*	0.370*
	(0.330)	(0.120)	(0.595)	(0.199)
Impulsivity	−0.342	−0.211	3.508***	1.014***
	(0.530)	(0.198)	(1.159)	(0.296)
Constant	−1.834	0.325	−21.586***	−6.513***
	(2.873)	(1.006)	(5.080)	(1.500)
N	332	329	593	593
F\Wald chi2	4.89***	33.04***	3.74***	33.85***

Columns 1 and 3 report the coefficients of a Tobit regression with the left (0) and right (7) censoring, with the number of assigned deduction points as the dependent variable. Columns 2 and 4 reflect the estimates from logistic models fitted to the punishment decision (0: no punishment points were assigned; 1: at least one deduction point assigned). The first two columns are for pro-social punishment and the last two columns for anti-social punishment. BT serves as the omitted category; 'Awareness' is a dummy variable for the AT; 'Target's contribution' is the contribution of the punished participant; 'Other's contribution' is the contribution of the third player in the group; 'Punisher's contribution' is the contribution of the participant that assigned the deduction points; 'Age' is the age of the participants; 'Female' and 'Unidentified' are dummy variables for female participants and those who chose not to disclose their gender; 'Belief' is an average of the answers to two questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for brief measure of impulsivity and sensation seeking. Robust standard errors clustered on individuals in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the punishment stage decisions being revealed to other group members in the final reports. If they chose the free option, the odds remained intact.

According to our data, most participants (182) chose not to pay in stage 4. On average they contributed 13.41 tokens towards the public good. 52 subjects that decided to incur costs, so that other group members get the information about their choices in the punishment stage, on average contributed the most, namely 15.13 tokens, as shown in Figure 3.5. The remaining 20[16] participants paid to conceal their punishment choices and on average contributed 8.30 tokens.

The aforementioned differences in contributions are significant according to the non-parametric test we used earlier (Mann-Whitney test on "Info" = "No-Pref": $p = 0.019$; Mann-Whitney test on "No-Info" = "No-Pref": $p < 0.001$; Mann-Whitney test on "Info" = "No-Info": $p < 0.001$). After controlling for demographic variables, impulsivity, belief in supernatural powers, however, the gap between contributions by those who paid to unveil their punishment choices and those who went for the free option is statistically insignificant (Wald test on "Info": $p = 0.261$), as reported in Table 3.5. The differences in the other two dyads are substantial even after controlling for gender, age, impulsivity and participants' belief in supernatural (Wald test on "No-Info": $p = 0.006$; Wald test on "Info" = "No-Info": $p < 0.002$).

Result 11. *Punishers who preferred to reveal their actions on average spent more than others on deduction points, but the highest proportion of punishers was recorded among subjects that wished to veil their punishment decisions.*

A further probe shows that subjects care not only about others' awareness of punishment that took place but also of punishment that was not carried out. The distribution of stage 4 choices by contributions is displayed in Figure 3.6. Among the 52 individuals, who incurred costs to reveal their punishment stage decisions to

[16] The another participant dropped during the 4th stage for technical reasons.

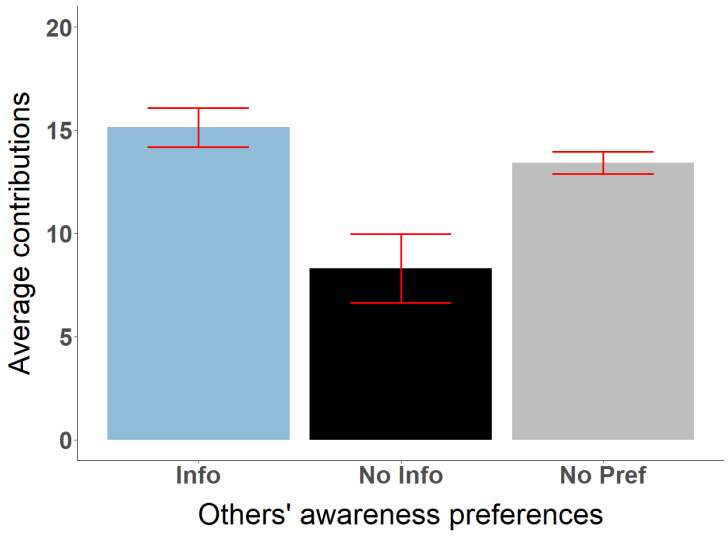


Figure 3.5: Average contributions by information preference of stage 4. Participants in the "Info" category paid to make the delivery of information about own punishment decisions certain; "No-info" paid to prevent the delivery of the information in the final report; "No pref" showed no strong preferences for any of the two information sets, as chose to pay nothing.

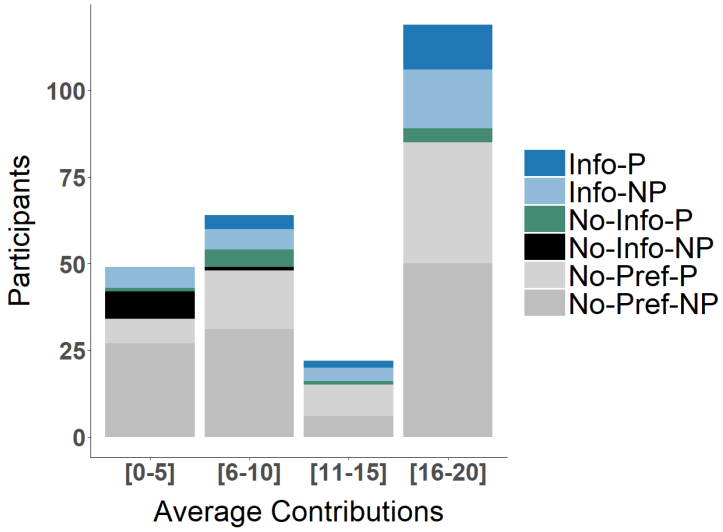


Figure 3.6: Stage 4 preferences and punishment decisions according to contributions. "Info-P", "No-Info-P", "No-Pref-P", denote individuals who paid to make the delivery of information about own punishment decisions certain, paid to hide them, chose the free option, respectively, and assigned some deduction points. "Info-NP", "No-Info-NP", "No-Pref-NP", made corresponding decisions in stage 4, but assigned no deduction points.

Table 3.4 Information preferences multinomial logistic regression

	Info	No-Info
Deduction points	0.328**	−0.010
	(0.135)	(0.182)
Target's contribution	0.010	−0.017
	(0.021)	(0.031)
Other's contribution	0.014	−0.050
	(0.018)	(0.033)
Punisher's contribution	0.025	−0.125***
	(0.025)	(0.040)
Age	0.019	0.041**
	(0.015)	(0.020)
Female	−0.320	0.487
	(0.345)	(0.568)
Unidentified	0.799	−11.084***
	(1.254)	(1.433)
Belief	0.095	0.040
	(0.129)	(0.201)
Impulsivity	0.207	0.308
	(0.255)	(0.378)
Punishment	−1.698***	1.294
	(0.628)	(0.817)
Constant	−3.248***	−2.796
	(1.232)	(1.922)
N	502	
Wald chi2(20)	336.78***	

The dependent variable is the stage 4 choice of information provided to the punished (1: Info option, information about punishment decisions is provided with certainty to others, at the cost of 1 token; 2: No-Info, information about punishment decisions is not presented in the final report, for 1 token; 3: No-Pref, is the reference group, information about punishment decisions is supplied with a probability of 0.5, free of charge). 'Target's contribution' is the contribution of the punished participant; 'Other's contribution' is the contribution of the third player in the group. 'Punisher's contribution' is the contribution of the participant that assigned the deduction points; 'Age' is the age of the participants; 'Female' is a dummy variable for female participants; 'Unidentified' is a dummy for individuals who chose not to declare their gender; 'Belief' is an average of the answers to questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for the brief measure of impulsivity and sensation seeking; 'Punishment' is the dummy variable that gets 1 if a positive number of deduction points was assigned.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5 Determinants of punishment - Awareness treatment

	Pro-Social		Anti-Social	
	(1)	(2)	(3)	(4)
Target's contribution	−0.222***	−0.120***	−0.060	0.032
	(0.077)	(0.036)	(0.127)	(0.053)
Other's contribution	0.021	0.049*	0.152	0.049
	(0.050)	(0.025)	(0.105)	(0.031)
Punisher's contribution	0.197***	−0.079	0.288	−0.079**
	(0.071)	(0.054)	(0.197)	(0.033)
Info	1.745**	0.147	4.188***	−0.982
	(0.791)	(0.405)	(1.135)	(1.151)
No-Info	−0.232	2.185**	−1.056	0.841
	(1.278)	(0.956)	(1.156)	(0.825)
Age	0.054	0.013	−0.023	−0.008
	(0.033)	(0.018)	(0.071)	(0.025)
Female	−0.346	−0.003	−1.292	0.568
	(0.662)	(0.372)	(0.934)	(0.711)
Unidentified	−2.195***	0.000		
	(0.629)			
Impulsivity	0.645	−0.256	0.738	1.107**
	(0.415)	(0.270)	(1.652)	(0.479)
Belief	0.166	0.078	0.734**	0.453
	(0.261)	(0.157)	(0.283)	(0.286)
Constant	−2.406	1.426	−3.383	−7.449***
	(2.531)	(1.479)	(5.312)	(2.114)
N	98	186	23	312
F\Wald chi2	8.71***	31.37***	37.73***	23.55***

Columns 1 and 3 report the coefficients of a Tobit regression with left (1) and right (7) censoring, with the number of assigned deduction points as the dependent variable, with the sample being limited to decisions that involved punishment. Columns 2 and 4 reflect the estimates from the logistic model fitted to the punishment decision (0: no punishment points were assigned; 1: at least one deduction point assigned. The odd columns are for pro-social punishment and the even for anti-social punishment. *No-Pref*, subjects that chose not to pay in stage 4 are the base category. 'Target's contribution' is the contribution of the punished participant; 'Other's contribution' is the contribution of the third player in the group; 'Punisher's contribution' is the contribution of the participant that assigned the deduction points; 'Age' is the age of the participants; 'Info' is a dummy variable for individuals that paid to conceal their punishment choices; 'No-Info' is a dummy variable for individuals that paid to reveal their punishment choices; 'Female' and 'Unidentified' are dummy variables for female participants and those who chose not to disclose their gender; 'Belief' is an average of the answers to questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for brief measure of impulsivity and sensation seeking. Robust standard errors clustered on individuals in parenthesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

others, only 19 meted out deduction points and the rest paid to let others know about the absence of punishment[17].

Interestingly, 9 individuals (out of 20) who paid to hide their punishment decisions assigned no punishment points. All of them were in the lower spectrum of contributions scale, with a mean transfer to the public account of only 3.63 tokens (SE = 1.21).

Table 3.4 reports the estimates of multinomial logistic regression with stage 4 choices as the dependent variable and the individuals who chose the free option as the base outcome. The estimated coefficients suggest that higher punishment spending, everything else constant, increases the likelihood of choosing to pay for the revelation of own punishment choices, relative to selecting the free option (Wald test on "Deduction points": $p = 0.015$). Higher odds of choosing to conceal own actions relative to choosing the free option are associated with lower own contributions (Wald test on "Punisher's contribution": $p = 0.002$) and higher age (Wald test on "Age": $p = 0.043$).

Models reported in Table 3.5 allow us to have another look at the same data from a different angle. The dependent variable in Tobit models is the number of deduction points assigned by a punisher, for a sample limited to individuals who assigned at least one deduction point. Logistic regressions were fitted to the decision whether to punish.

The results of the logistic regression for pro-social punishers point out the disproportionate number of punishers among individuals who preferred their punishment choices to be veiled (Wald test on "No-Info": $p = 0.022$; Wald test on "No-Info" = "Info": $p = 0.041$; column 2). The coefficients reported in the first column, however, inform

[17] The mode contribution gap between these individuals and other group members was 0 tokens, the mean -1.83 tokens and the S.D. 9.14 tokens

that on average they did not spend markedly different amounts than their peers with other stage 4 choices (Wald test on "No-Info": $p = 0.857$, column 1; "No-Info" = "Info": $p = 0.184$, column 1).

In contrast, punishers that relinquished 1 token to make the consequences of their choices reported to others spent substantially more than individuals that refrained from paying, both for pro-social and anti-social punishment (Wald test on "Info": p equals 0.030 and 0.002, columns 1 and 3 respectively). The difference in spending with the remaining third group is significant for the anti-social (Wald test on "Info" = "No-Info": $p < 0.001$, column 2), but not for the pro-social (Wald test on "Info" = "No-Info": $p = 0.184$, column 1) punishment.

Discussion

In the second study, the importance of the messaging function of punishment was explored in the context of punishment awareness. The awareness was undermined through the incorporation of a continuous shock to stage 1 earnings, which could be confounded with punishment unless full information about the shock and the received deduction points is received. Participants could pay 1 token fee to affect the probability of information delivery. In case the free option was selected the final report of a punished individual included full details about shock's magnitude and received punishment points with the probability of 0.5.

Differently, from the first study, the motives for sending a message in this one cannot be confounded with either reciprocity or relativity of some's payoff. The theoretical models we mentioned earlier ignore the awareness of punishers' actions, hence predict no spending in stage 4 of our experiment.

The majority of participants acted in line with the above prediction. Nevertheless,

a considerable number of subjects paid over the odds. Not all of them wished to inform others about the assigned punishment; some wanted to inform their co-players about the absence of punishment, others even paid to hide their punishment or lack of it. Building on above evidence we suggest that our findings might be related to a broader phenomenon of concerns about others knowledge/awareness, previously reported by Dana et al. (2006).

3.6 General Discussion

This paper is dedicated to the probe into the messaging aspect of the punishment, namely whether there is extra value in the message it sends to the punished. For this purpose, we conducted two online studies on Amazon Mechanical Turk. We used augmented versions of the basic public goods settings, with discrete and continuous random shocks in the first and the second studies respectively. Both studies provide evidence for messaging being essential to a few individuals.

Earlier in this paper, we suggested that the messaging via punishment might be linked to consideration of others' knowledge. The latter could be of importance for co-operation due to its impact on individuals' behavioural patterns. Masclet et al. (2003) demonstrate that a mere expression of disapproval might be sufficient to boost cooperation, more so when such a message comes with materialistic repercussions. Therefore, sanctions that lack the statement of disapproval will probably fail to achieve the same effect.

Future research might attempt to improve current settings to continue the inquiry into the importance of individuals' consideration of others' luck and awareness of retribution. For the former it is necessary to change the structure of the shock in study

1, to test whether high spenders are willing to completely replace own punishment with a negative shock to earnings of targeted individuals. Moreover, in the light of discrepancies between aggregate punishment spending in the *Baseline* and *Fate* treatments, it might be necessary to replace the strategy method with a direct-response method. In case of the *Awareness* treatment, one might consider enabling punishers to influence the awareness individually, rather than affecting a group as a whole.

The suggested improvements might complicate the interaction environment even further. Therefore, we believe that it would be reasonable to conduct future experiments in a laboratory.

3.7 Bibliography

Abbink, K., Irlenbusch, B., and Renner, E. (2000). The moonlighting game – An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, 42(2):265–277.

Anderson, C. M. and Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, 54(1):1–24.

Andreoni, J. (1988). Why free ride?. Strategies and learning in public goods experiments. *Journal of Public Economics*, 37(3):291–304.

Andreoni, J., Harbaugh, W., and Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, 93(3):893–902.

Arechar, A. A., Gächter, S., and Molleman, L. (2017). Conducting interactive experiments online. *Experimental Economics*.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Bolton, G. E. and Zwick, R. (1995). Anonymity versus Punishment in Ultimatum Bargaining.

Bosman, R. and Van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal*, 112(476):147–169.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5.

Carpenter, J. P. (2007). The demand for punishment. *Journal of Economic Behavior and Organization*, 62(4):522–542.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Cox, J. C., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., and Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446(7137):794–796.

De Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688):1254–1258.

Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., and Cudré-Mauroux, P. (2015). The dynamics of micro-task crowdsourcing: The case of amazon mturk.

In *Proceedings of the 24th International Conference on World Wide Web*, pages 238–247. International World Wide Web Conferences Steering Committee.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Engelmann, D. and Nikiforakis, N. (2014). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, pages 1–17.

Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73(6):2017–2030.

Falk, A., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness-Intentions matter. *Games and Economic Behavior*, 62(1):287–303.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137–140.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, pages 817–868.

Förster, J., Liberman, N., and Higgins, E. T. (2005). Accessibility from active and fulfilled goals. *Journal of Experimental Social Psychology*, 41(3):220–239.

Forsythe, R., Horowitz, J. L., Savin, N., and Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3):347–369.

French, P. A. (2001). *The virtues of vengeance*. University Press of Kansas.

Frijda, N. H. (1994). The lex talionis: On vengeance. *Emotions: Essays on emotion theory*, pages 263–289.

Gollwitzer, M. and Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, 45(4):840–844.

Gollwitzer, M., Meder, M., and Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European journal of social psychology*, 41(3):364–374.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4):367–388.

Heider, F. (2013). *The psychology of interpersonal relations*. Psychology Press.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads*, 17(2):16–21.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M. C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.

Noussair, C. and Tucker, S. (2005). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry*, 43(3):649–660.

Oxoby, R. J. and Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior and Organization*, 65(3-4):703–713.

Paolacci, G. and Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, 23(3):184–188.

Purzycki, B. G., Apicella, C., Atkinson, Q. D., Cohen, E., McNamara, R. A., Willard, A. K., Xygalatas, D., Norenzayan, A., and Henrich, J. (2016). Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, 530(7590):327–330.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83(5):1281–1302.

Raihani, N. J. and McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters*, 8(5):802–804.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299:172–179.

Schotter, A., Weiss, A., and Zapater, I. (1996). Fairness and survival in ultimatum and dictatorship games. *Journal of Economic Behavior and Organization*, 31(1):37–56.

Sefton, M., Shupp, R., and Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4):671–690.

Webster, G. D. and Crysel, L. C. (2012). "Hit Me, Maybe, One More Time": Brief measures of impulsivity and sensation seeking and their prediction of blackjack bets and sexual promiscuity. *Journal of Research in Personality*, 46(5):591–598.

Appendix

Tables

Below are two tables with coefficients of Tobit models fitted to contributions made in stage 1 of the experiment.

Table A1 Contributions to the public good - Fate treatment

Fate	−0.782
	(1.041)
Age	0.125**
	(0.052)
Female	−2.752**
	(1.067)
Unidentified	2.384
	(8.361)
Belief	0.545
	(0.423)
Impulsivity	−0.547
	(0.705)
Constant	13.625***
	(3.024)
N	517
LR chi2(9)	14.27**

The reported coefficients are of Tobit estimation with right-censoring (20). 'Fate' is the dummy for the Fate treatment; 'Age' is the age of the participants; 'Female' is a dummy variable for female participants; 'Unidentified' is a dummy for individuals who chose not to declare their gender; 'Belief' is an average of the answers to questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence); 'Impulsivity' is a variable for the brief measure of impulsivity and sensation seeking.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

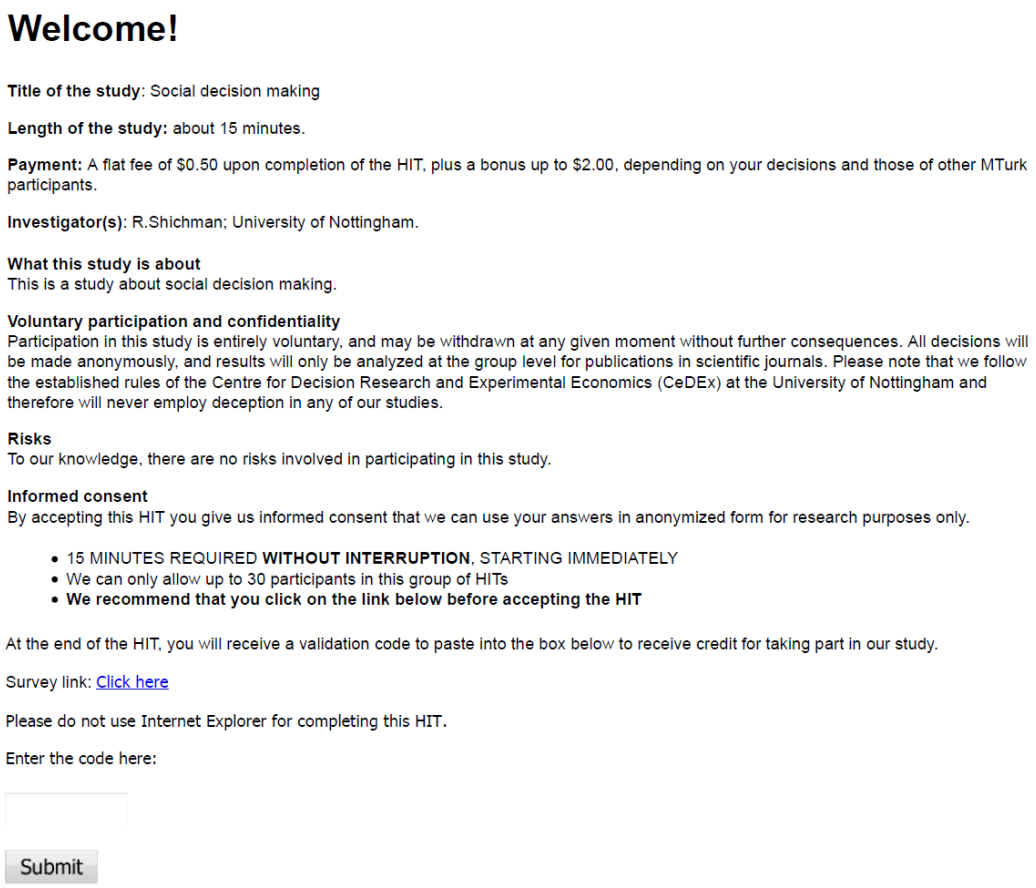
Table A2 Contributions to the public good - Awareness treatment

	1	2
Awareness	−0.096	
	(1.147)	
Info		2.969
		(2.006)
No-Info		−7.623***
		(2.698)
Age	0.075	0.163**
	(0.056)	(0.073)
Female	−2.250*	−2.412
	(1.199)	(1.594)
Unidentified	4.250	
	(7.901)	
Belief	0.935	0.440
	(0.487)	(0.636)
Impulsivity	1.085	0.440
	(1.054)	(1.423)
Constant	10.729**	8.030*
	(3.452)	(4.391)
N	464	299
LR chi2(6)	9.96	20.92**

The reported coefficients are of Tobit estimation with right-censoring (20). Column 1: the omitted category is the *Baseline* treatment; 'Awareness' is a dummy for the Awareness treatment. Column 2: the omitted category 'No-Pref', individuals that did not pay to affect others' awareness of their actions; 'Info' and 'No-Info' are dummy variables for individuals that paid to reveal and conceal their actions respectively. Other independent variables: 'Age' is the age of the participants; 'Impulsivity' is a variable for the brief measure of impulsivity and sensation seeking; 'Female' is a dummy variable for female participants; 'Unidentified' is a dummy for individuals who chose not to declare their gender; 'Belief' is an average of the answers to two questions about belief in God/s and Karma (ranging between 1 completely disagree and 5 completely agree with their existence).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Instructions

The description of the HIT of our MTurk experiment is below, followed by the instructions (both as seen by participants).



Welcome!

Title of the study: Social decision making

Length of the study: about 15 minutes.

Payment: A flat fee of $0.50 upon completion of the HIT, plus a bonus up to $2.00, depending on your decisions and those of other MTurk participants.

Investigator(s): R.Shichman; University of Nottingham.

What this study is about
This is a study about social decision making.

Voluntary participation and confidentiality
Participation in this study is entirely voluntary, and may be withdrawn at any given moment without further consequences. All decisions will be made anonymously, and results will only be analyzed at the group level for publications in scientific journals. Please note that we follow the established rules of the Centre for Decision Research and Experimental Economics (CeDEx) at the University of Nottingham and therefore will never employ deception in any of our studies.

Risks
To our knowledge, there are no risks involved in participating in this study.

Informed consent
By accepting this HIT you give us informed consent that we can use your answers in anonymized form for research purposes only.

- 15 MINUTES REQUIRED **WITHOUT INTERRUPTION**, STARTING IMMEDIATELY
- We can only allow up to 30 participants in this group of HITs
- **We recommend that you click on the link below before accepting the HIT**

At the end of the HIT, you will receive a validation code to paste into the box below to receive credit for taking part in our study.

Survey link: Click here

Please do not use Internet Explorer for completing this HIT.

Enter the code here:

Submit

Baseline treatment

Screen 1

Welcome

The setup of this HIT is **different** from HITs that you might be used to completing via MTurk.
You will be interacting in a group together with two other **real people** who also accepted this HIT, who are completing it **at the same time**.
It is therefore important that you complete this HIT **without interruptions**.
Including the time for reading these instructions, the HIT will take about 15 minutes to complete.

During the HIT, please **do not close this window** or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter the HIT and we will not be able to pay you!

In this HIT you can earn Tokens, which at the end of the HIT will be converted into real money (**25 Tokens = 1 Dollar**).
You will receive a code to collect your payment via MTurk upon completion.

Screen 2

Instructions

Your task

In your group of three, after receiving an ID (A, B or C) and you will make decisions.
The task consists of a number of Stages (each will take place just once).

Stage I

At the beginning of this stage, each participant receives 20 Tokens.
You have to decide how many of the 20 Tokens you want to **contribute to a group project**.
The other two members of your group make this decision at the same time.
The Tokens you do not contribute, you **keep for yourself**. These Tokens are added to your total.
After all group members have made their decision, all Tokens contributed to the **group project** are added up, and this number of Tokens is **multiplied by 1.5**.
The resulting number of Tokens is then divided equally among the group members (irrespective of how much they individually contributed to the group project).

Screen 3

In summary

Your earnings in Stage I =

The Tokens you keep for yourself

plus

The Tokens you receive from the group project

Two examples will follow.

Screen 4

Group project - Example 1

- All three players contribute 20 Tokens to the group project, hence sum of contributions is 60 Tokens (3 X 20).
- This amount is multiplied by 1.5, resulting in 60 X 1.5 = 90 Tokens.
- Each participant receives 90/3 = 30 Tokens from the group project.
- Therefore, in Stage I each participant earns 30 Tokens.

Screen 5

Group project - Example 2

- Participants A and B each contribute 20 Tokens to the group project.
- Participant C contributes 0 Tokens.
- Sum of contributions is 20 + 20 = 40 Tokens.
- This amount is multiplied by 1.5, resulting in 40 X 1.5 = 60 Tokens.
- Each participant receives 60/3 = 20 Tokens from the group project.
- Therefore, in Stage I Participants A and B each earn 20 Tokens.
- Participant C earns 40 Tokens, by keeping 20 Tokens and *getting* 20 Tokens from the group project.

Screen 6

At the end of Stage I, you will be informed of the average contributions to the group project, and your earnings so far. Then, Stage II begins.

Screen 7

Stage II

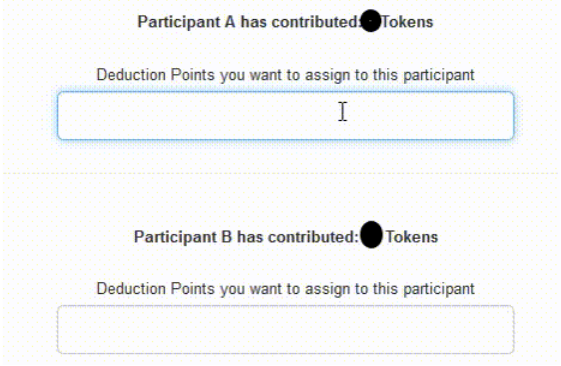
- At the beginning of Stage II, you are informed of the contributions to the group project of each of the other group members and receive an **extra 10 Tokens**.
- Then, each member of the group must indicate how many **Deduction Points** s/he wishes to assign to other group members.
- Click on the "Continue" button to see a screenshot of Stage II.

Instructions for this stage are continued on the next slide.

Screen 8

Stage II

- Here is an example of Stage II screen (you are Participant C in this example), where you will be able to assign Deduction Points.
- In the actual screen the black circles will be replaced by Stage I contributions of other members.
- After clicking on the boxes titled "Deduction Points you want to assign to this participant" you will be able to assign between 0 and 7 Deduction Points to each of the other group members.
- Each Deduction Point you assign reduces the earnings of the other participant by 2 Tokens and costs you 1 Token.



Screen 9

Stage II

- The other group members can assign Deduction Points to you if they wish to.
- However, decisions of **only one** group member picked **randomly**, will be implemented. Other decisions will be **discarded** and affect no one.

Screen 10

Control questions

Please answer all control questions. They serve as a test for your understanding of the HIT.

1. At the beginning of Stage I, each group member receives 20 Tokens.

Suppose nobody (including you) contributes any Tokens to the project.

How many Tokens do you earn in Stage I?

How many Tokens do the other members of your group earn in Stage I?

Screen 11

2. At the beginning of Stage I, each group member receives 20 Tokens.
Suppose the other two group members contribute a total of 30 Tokens to the group project.
Now suppose you contribute **0 Tokens** to the group project.

How many Tokens will each player receive *from the group project*?

How many Tokens will you earn in Stage I?

Screen 12

3. At the beginning of Stage I, each group member receives 20 Tokens.
Suppose that the other two group members contribute a total of 30 Tokens to the group project.
Now suppose you contribute another **10 Tokens.**

How many Tokens will each player receive from the group project?

How many Tokens will you earn in Stage I?

Screen 13

4. Suppose that in **Stage II** you assign the following Deduction Points to the other members of your group: 2 Deduction Points to Participant B and 4 Deduction Points to Participant C. Participant B assigns 3 Deduction Points to you. Participant C assigns no Deduction Points.

How many Tokens will be deducted from **your** total after assigning these Deduction Points if your decisions are implemented?

How many Tokens will be deducted from **your** total if decisions of Participant C are implemented?

How many Tokens will be deducted from the account of Participant B if your decisions are implemented?

Screen 14

All stages will take place just once.

True | False

In Stage II decisions of only one member in each group will be implemented.

True | False

Screen 15

You've finished answering the comprehension questions and are about to be matched with other participants.

Please make your decisions **within the time limit shown** on your screen. If you fail to do so, you will be removed from the HIT.

Additional and different instructions

Fate

Screen 7

Stage II

At this point, each member's earnings from Stage I are affected by a **random shock**. The outcomes of this shock, **+5, 0 or -5 Tokens**, are equally likely and might differ for members of the same group. You will be able to observe the result of this stage after completing the next stage.

Screen 8

Stage III

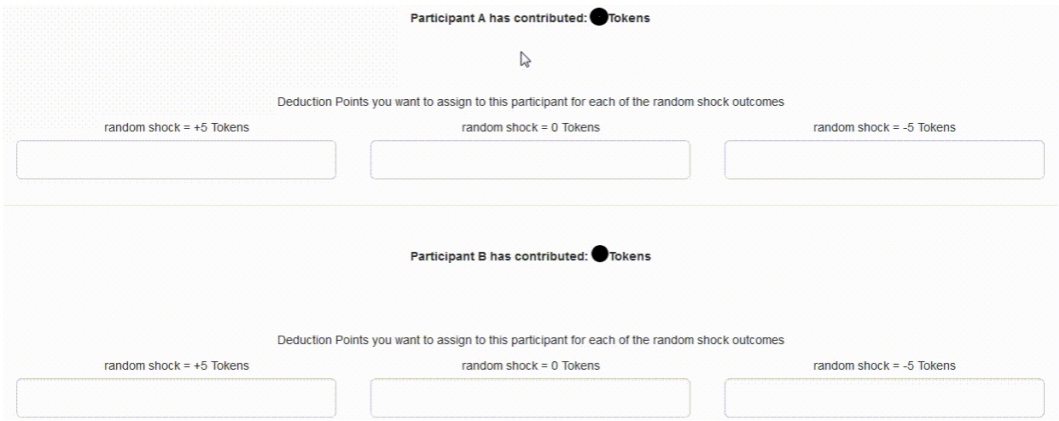
- At the beginning of Stage III, you are informed of the contributions to the group project of each of the other group members and receive an **extra 10 Tokens**.
- Then, each member of the group must indicate how many **Deduction Points** s/he wishes to assign to other group members, conditional on the outcomes of the previous stage.
- Click on the "Continue" button to see a screenshot of Stage III.

Instructions for this stage are continued on the next slide.

Screen 9

Stage III

- Here is an example of Stage III screen (you are Participant C in this example), where you will be able to assign Deduction Points.
- In the actual screen the black circles will be replaced by Stage I contributions of other members.
- After clicking on the boxes with titles of potential Stage II outcomes, you will be able to assign between 0 and 7 Deduction Points to each of the other group members for each of the three outcomes.
- Each Deduction Point you assign reduces the earnings of the other participant by 2 Tokens and costs you 1 Token.
- For example, the number you will enter to the top left box (labeled "random shock = +5), will be used if in Stage II earnings of Participant A were increased by 5 Tokens due to the random shock. In such case the final earnings of Participant A = Stage I earnings + 5 - (the number you entered X 2) + 10 extra Tokens received at the beginning of Stage III.



Screen 10

Stage III

- The other group members can assign Deduction Points to you if they wish to.
- However, decisions of **only one** group member picked **randomly**, will be implemented. Other decisions will be **discarded** and affect no one.
- *The chosen participant will **not** be affected by the shock of Stage II (it will be set to 0).*

Screen 15

5. Suppose that in **Stage III** you wish to reduce the earnings of some participant by 4 Tokens if his/her earnings were increased by 5 Tokens due to the random shock of Stage II. Moreover, you want to decrease his/her earnings by 6 Tokens in case in Stage II his/her earnings were decreased by 5 Tokens. Otherwise you assign no points. Please enter below the appropriate numbers of Deduction Points to achieve the above result.

Deduction Points you want to assign to this participant for each of the random shock outcomes

random shock = +5	random shock = 0	random shock = -5

Screen 16

6. Suppose that in **Stage III** you wish to reduce the earnings of some participant by 0 Tokens if his/her earnings were either increased by 5 Tokens or decreased by 5 Tokens due to the random shock of Stage II. Moreover, you want to decrease his/her earnings by 6 Tokens in case in Stage II his/her earnings were not changed by the random shock. Please enter below the appropriate numbers of Deduction Points to achieve the above result.

Deduction Points you want to assign to this participant for each of the random shock outcomes

random shock = +5	random shock = 0	random shock = -5

Screen 17

All stages will take place just once.

True | False

In Stage III decisions of only one member in each group will be implemented.

True | False

A participant whose decisions will be implemented will not be affected by the random shock.

True | False

Awareness

Screen 7

Stage II

At this point, each member's earnings from Stage I are affected by a **random shock**. This can either **increase** earnings (double in the best case scenario), **decrease** earnings (to zero in the worst-case scenario), or leave the earnings **unchanged**. Smaller changes are more likely than larger, but any outcome is possible.

The shock will probably vary for individuals within the same group. Participants may or may not be informed of the outcomes of this stage in their final report, as explained in Stage IV.

Screen 8

Stage III

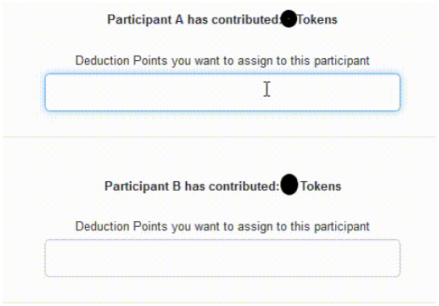
- At the beginning of Stage III, you are informed of the contributions to the group project of each of the other group members and receive an **extra 10 Tokens**.
- Then, each member of the group must indicate how many **Deduction Points** s/he wishes to assign to other group members.
- Click on the "Continue" button to see a screenshot of Stage III.

Instructions for this stage are continued on the next slide.

Screen 9

Stage III

- Here is an example of Stage III screen (you are Participant C in this example), where you will be able to assign Deduction Points.
- In the actual screen the black circles will be replaced by Stage I contributions of other members.
- After clicking on the boxes titled "Deduction Points you want to assign to this participant" you will be able to assign between 0 and 7 Deduction Points to each of the other group members.
- Each Deduction Point you assign reduces the earnings of the other participant by 2 Tokens and costs you 1 Token.



Screen 10

Stage III

- The other group members can assign Deduction Points to you if they wish to.
- However, decisions of **only one** group member picked **randomly**, will be implemented. Other decisions will be **discarded** and affect no one.
- *The chosen participant will **not** be affected by the shock of Stage II (it will be set to 0).*

Screen 11

Stage IV

In Stage IV you can:

- Pay **1 Token** (payable only if your Stage III decisions are implemented) to decide whether to inform or not inform other group members about your decisions from the previous stage, by providing/withholding information about the number of Deduction points you assigned to them.
- Pay **nothing**. In this case, with **probability of 1/2** (likelihood of getting heads/tails in a toss of a fair coin) other group members will be informed about your decisions in Stage III and with probability of 1/2 they will not.

If other individuals are not informed, their final report (displayed after Stage IV) will only include their earnings after Stage I, the extra 10 Tokens received in Stage III and their final earnings. Without information about the number of Deduction Points assigned to them, they will not be able to tell whether you affected their earnings or whether the change was solely due to the random shock.

Instructions for this stage are continued on the next slide.

Screen 12

Stage IV

For example, Participant A's earnings after Stage I are 20 Tokens. In Stage II the random shock **increases** Participant A's earnings by 5 Tokens. Stage III s/he receives the 10 extra Tokens and 2 Deduction Points from you, which **decrease** his/her earnings by 2 X 2 = 4 Tokens. Assuming that your decisions from Stage III are implemented, *final earnings of Participant A = 20 + 10 + 5 - 4 = 31*. If Participant A gets **no information about the received Deduction Points and the shock**, all s/he will see is the increase in earnings from 30 (Stage I earnings plus the extra 10 Tokens) to 31, hence will stay unaware of your actions in Stage III.

Reminder: if your Stage III decisions are implemented, you wont be affected by a Stage II random shock and have full information regarding your earnings.

Screen 17

All stages will take place just once.

True | False

In Stage III decisions of only one member in each group will be implemented.

True | False

A participant whose decisions will be implemented will not be affected by the random shock.

True | False

Participants will know how many points were lost because of received deduction points and how many due to a random shock.

True | False | Either True or False

Conclusion

Various versions of public goods games have been used by economists to study cooperation and punitive behaviour. In the current work, we used four different studies to add to this literature.

In the first two chapters, we systematically studied the effect of asymmetries in returns from a public good and punishment efficacy on cooperation. We have demonstrated that asymmetries are detrimental to cooperation in small groups and that self-selection to the positions of power did not play a great role in shaping the results. Furthermore, we also showed how devastating feuds could be for certain groups, stressing the importance of good institutions for groups' success in sustaining high levels of cooperation.

In the final chapter of this work, we looked into the possibility that punishment is used as a means of communication. Both studies in this chapter provide evidence that supports this idea. Moreover, the data suggest that some individuals seem to consider what others know about their decisions, even in one-shot, anonymous interactions.