DFG Research Group 2104

Need-Based Justice and Distribution Procedures

Information-sensitive Leviathans

Andreas Nicklisch, Kristoffel Grechenig and Christian Thöni

Working Paper Nr. 2016-02

http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2016-02.pdf

Date: 2016-08-16

DFG Research Group 2104 at Helmut Schmidt University Hamburg
http://needs-based-justice.hsu-hh.de

Information-sensitive Leviathans[*]

Andreas Nicklisch,[‡] Kristoffel Grechenig,[§]
Christian Thöni[¶]

August 16, 2016

Abstract

We study information conditions under which individuals are willing to delegate their sanctioning power to a central authority. We design a public goods game in which players can move between institutional environments, and we vary the observability of others' contributions. We find that the relative popularity of centralized sanctioning crucially depends on the interaction between the observability of the cooperation of others and the absence of punishment targeted at cooperative individuals. While central institutions do not outperform decentralized sanctions under perfect information, large parts of the population are attracted by central institutions that rarely punish cooperative individuals in environments with limited observability.

Keywords: centralized sanctions, cooperation, experiment, endogenous institutions
JEL codes: C92, D02, H41

[*] For helpful comments and discussion, we thank the editor, two anonymous referees, Berno Büchel, Ernst Fehr, Urs Fischbacher, Guillaume Frechette, Simon Gächter, Sonja Köke, Manfred Milinski, Nikos Nikiforakis, Louis Putterman, Arno Riedl, Bettina Rockenbach, Roberto Weber, and the seminar participants at the University of Pennsylvania, the Max Planck Institute (Collective Goods, Bonn), the University of Nottingham, the University of Lausanne, and the University of Cologne. We would also like to thank the Max Planck Society and the German Research Foundation (Ni 1610/1-1) for financial support. An earlier version of this paper circulated under the title "Information-sensitive Leviathans: The Emergence of Centralized Punishment" (Nicklisch, Grechenig, & Thöni, 2015).

[‡] University of Hamburg and German Research Foundation, Research Unit "Needs Based Justice and Distributive Procedures"

[§] Max Planck Institute for Research on Collective Goods, Bonn

[¶] *Corresponding author:* University of Lausanne, email: christian.thoeni@unil.ch.

1 Introduction

Human life in Thomas Hobbes' natural state is lonely, short, and brutal, "a time of war where every man is enemy to every man" (Hobbes, 1651). To redress this grim fate of violence and distrust people appoint a central authority—a *Leviathan*—to enforce cooperative behavior. People voluntarily delegate their sanctioning power to the Leviathan, in the hope for a more efficient outcome.

In contrast to Hobbes' bleak view contemporary research suggests that people successfully use *decentralized sanctions* (peer-to-peer punishment) to enforce cooperation (Ostrom, Walker, & Gardner, 1992; Fehr & Gächter, 2000) and reach efficient outcomes in the long run (Gächter, Renner, & Sefton, 2008). If human societies are able to organize themselves in a decentralized fashion, one would expect to find many self-governed societies. However, the opposite is the case: We live in a world where centralized sanctions play are very important role, on the national and even on the supranational level.[1] Why did modern societies develop centralized institutions to enforce norms? Under which conditions are people willing to renounce their sanctioning power in favor of a central authority?

We use an experimental approach to these questions, and we analyze a voting by feet mechanism in favor of or against central authorities. We introduce an environment where players ('citizens') participate in a social dilemma. Prior to this, they can vote by feet for one of three institutions: centralized punishment (*CenPun*), decentralized punishment (*DecPun*), and a sanction-free institution (*NoPun*). In *CenPun*, an additional (randomly drawn) subject (the 'authority') can punish the citizens in his institution, while citizens are not allowed to punish each other. The authority's payoff is increasing in the citizens' contributions, and the authority does not have to bear the costs of punishment. In *DecPun* citizens can punish other citizens in the same institution, at their own expenses.

Our analysis builds on three major challenges for self governance which have been identified in the literature: antisocial punishment, revenge, and incomplete information. Antisocial punishment (or perverse punishment) refers to the observation that some subjects target their punishment at cooperative subjects. There is ample evidence that the strength and frequency of antisocial punishment negatively relates to contributions.[2] Related is the problem of retaliation for received punishment. Some studies find that re-

[1] Examples are institutions like the European Union, the International Military Tribunal in Nuremberg in 1945/46, or the United Nations Security Council.

[2] See e.g. Gächter, Herrmann, and Thöni (2005); Bochet, Page, and Putterman (2006); Herrmann, Thöni, and Gächter (2008).

taliation weakens decentralized punishment institutions because cooperative individuals are less willing to punish free riders (Denant-Boemont, Masclet, & Noussair, 2007; Nikiforakis, 2008; Nikiforakis, Noussair, & Wilkening, 2012), while others do not find such a general effect (Kamei & Putterman, 2015). Finally, decentralized punishment can become inefficient in increasing contributions when subjects receive only imperfect information about the contributions of others. Contrary to intuition (but in accordance to the theoretical analysis we develop below) subjects tend to punish more when information becomes more noisy (Grechenig, Nicklisch, & Thöni, 2010; Ambrus & Greiner, 2012).[3]

All three problems are closely related. For instance, less information leads potentially to more punishment of cooperative subjects, which might in turn trigger retaliatory punishment. While in principle one could exogenously vary multiple dimensions of this complex interaction we restrict our design to a manipulation of the informational quality. In terms of the underlying phenomenon (establishing cooperation in groups) we think that it makes sense to see informational conditions as an exogenous characteristic of the environment, while the individual propensity to engage in antisocial punishment or revenge seems endogenous in its nature. Thus, our approach is to vary the quality of information exogenously and study its impact on the relative popularity of the three institutions. More specifically, we introduce three environments differing with respect to the accuracy of information citizens and the authority receive about the contributions of others. In treatment condition ONE, they receive accurate signals about the contributions; in POINT-NINE, they receive signals which are correct in 90 percent of the cases, while in POINT-FIVE, the signals are correct in 50 percent of the cases. We measure the popularity of an institution by the fraction of citizens it attracts.

We find that the treatment variation significantly influences institutional choices. In particular, imperfect information lowers the popularity of *DecPun*. We show that the punishment of cooperative citizens significantly influences institutional choices. Finally, with regard to our main research question we find that *CenPun* becomes the most popular institution only when there is imperfect information and at the same time the central authority (the Leviathan) applies a punishment strategy which minimizes the punishment of cooperative citizens. At the same time, revenge motives seem to be less important in our design and cannot explain differences across treatments.

[3] On the other hand, Leibbrandt, Ramalingam, Sääksvuori, and Walker (2015) provide evidence that antisocial punishment increases when *more* information is provided, i.e., when subjects can identify individual punishers in the group.

Our study complements and expands recent discussions on the formation of centralized institutions. Dal Bó, Foster, and Putterman (2010) compare the effect of a democratically chosen and an exogenously imposed policy intervention aimed at eliminating the attractiveness of free-riding. They find that democratically installed interventions increase cooperation significantly compared to exogenously imposed interventions. Moreover, endogenously introduced regimes with centralized sanctions perform well, even when sanctions are non-deterrent (Tyran & Feld, 2006), or, in some cases, outperform decentralized sanctions (O'Gorman, Henrich, & Van Vugt, 2009).

Another important aspect of centralized institutions is the way how sanctions are implemented. In contrast to our approach, the majority of articles focus on sanctions that are executed automatically. If both, decentralized and automatically executed centralized punishment are available, the latter seems to crowd out the use of the former (Kube & Traxler, 2011). Markussen, Putterman, and Tyran (2014) investigate the choice of centralized sanctions through voting, when centralization is costly (and executed automatically). They find that people are particularly responsive to the fixed costs of having a centralized sanctioning scheme in place, more so than they respond to whether or not the sanctioning scheme is fully deterrent. Putterman, Tyran, and Kamei (2011) allow participants to vote on the rules of an automatically executed sanctioning scheme. The results show that many groups quickly implement sanctions that induce efficient outcomes.

Kosfeld, Okada, and Riedl (2009) analyze the choice for automatically executed punishment mechanism which may govern only a subset of players. They show that participants are unwilling to implement equilibrium punishment which allows some players to free-ride. Andreoni and Gee (2012) investigate the formation of centralized sanctions through voting for a sanctioning scheme that punishes only the lowest contributor and find that full contributions are quickly achieved at very low punishment costs. Importantly, these articles focus on sanctions that are executed automatically; that is, once an implemented rule is violated, players are punished with a certain probability while contribution decisions are perfectly observable.[4]

Closer to our approach is Fehr and Williams (2013). They offer citizens the choice between uncoordinated decentralized, coordinated decentralized, or centralized punishment, which is executed by a democratically elected leader. They show that centralization of sanctions leads to high cooperation along with the selection of pro-social leaders who refrain from punishing high contributors. Similarly, Gross, Méder, Okamoto-Barth, and Riedl

[4] See also Sutter, Haigner, and Kocher (2010), who study endogenous choices between positive and negative sanctioning systems.

(2016) explore the emergence of central punishment authorities under perfect information. They demonstrate that if individuals can transfer their punishment power to others, cooperators empower subjects who have previously indicated their willingness to sanction free-riders. As a consequence, groups with centralized punishment and high cooperation emerge.

Summarizing the previous literature, both centralized as well as decentralized sanctioning sustain cooperation if chosen endogenously. If available, evidence suggests that citizens choose very selectively centralized institutions. That is, effective centralized sanctioning is demanded, but citizens are unwilling to accept centralized punishment that violates their fairness considerations (e.g., allowing some players to free-ride, or punishment targeted at contributors).

In our setting, it is up to the authorities to deliver effective sanctioning. Like in Fehr and Williams (2013), we introduce the authority as a player, who may use punishment in a similar, potentially erroneous or malevolent fashion as his citizens.[5] We do so as we believe that the feature is of particular importance to explain the choice of authorities in earlier societies. That is, we compare centralized and decentralized sanctioning when authorities are not equipped with better mechanisms to guide behavior than citizens (e.g., our authorities are not better informed than citizens, nor do they have more efficient punishment technologies than citizens). Rather, our authorities are autocratic leaders, holding absolute punishment power. Furthermore, like in a feudal society the authority is not appointed by a competitive procedure, but he is merely born into his position.[6]

Following previous works showing that decentralized sanctions prevail over a sanction-free environment (Gürerk, Irlenbusch, & Rockenbach, 2006) and over a pure reputation-building environment (Rockenbach & Milinski, 2006), we let our players choose their institution by leaving societies (exit), but not by vote (voice).[7] Consequently, each citizen is free to migrate to his most preferred institution. In addition, due to the third alternative *NoPun*, our setting requires citizens to choose actively in favor of one punishment institution, which allows us to interpret citizens' institutional choice predominantly as a choice in favor of centralized or decentralized punishment

[5] This is similar to Carpenter and Matthews (2012), who analyze the effect of third-party punishment for contributions in public good games.

[6] For the effect of democratically appointed leaders see also Hamman, Weber, and Woon (2011), Corazzini, Kube, Maréchal, and Nicolò (2014).

[7] Historically, the importance of exit mechanisms for the organization of tribes, or even the fall of entire nations (e.g., East Germany), is well documented (Hirschman, 1970, 1978). Contemporary exit mechanisms capture competition between jurisdictions for corporations or tax payers.

rather than a decision against the alternative sanctioning institution which is not chosen.

Our article is structured as follows. Section 2 describes our basic game and derives an expression for deterrent punishment; in Section 3 we introduce the experimental setting and discuss behavioral predictions. Section 4 presents the results, and Section 5 concludes the paper with a discussion.

2 Model

2.1 The game

We set up a game which embeds competition between centralized punishment, decentralized punishment, and a punishment-free institution in a public goods game. We combine a *voting by feet* mechanism between different sanctioning regimes (Gürerk et al., 2006) with *imperfect information* about individual contributions (Grechenig et al., 2010). There are ten citizens and one authority. The game consists of three stages. In stage one, each citizen i independently chooses an institution. There are three institutions, each associated with a specific punishment rule: centralized punishment (*CenPun*), decentralized punishment (*DecPun*), and no punishment (*NoPun*). We denote by $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{N}$ the set of citizens in the three institutions. Citizens in a given institution play a public goods game as long as at least two citizens are present.

In stage two, each citizen receives an endowment of 20 experimental currency units (ECU). Citizens simultaneously choose a contribution $g_i \in \{0, 2, 4, ..., 20\}$ to the public good. Each ECU contributed to the public good is multiplied by 1.6 and the resulting amount is divided equally among the citizens in the respective institution. This payoff function keeps the marginal social return from the public good constant for different group sizes, so that there are no productivity advantages for larger groups. Consequently the marginal per capita return decreases in group size.[8] At the end of stage two a citizen i in the institution *CenPun* earns a profit of

$$\hat{\pi}_i = 20 - g_i + \frac{1.6}{c} \sum_{k \in \boldsymbol{C}} g_k, \tag{1}$$

[8] We designed the game to be neutral with regard to the optimal group size. Our payoff function ensures that any given average contribution results in the same average profit for all group sizes. It is of course still possible that the change in the marginal per capita return introduces group size effects, as suggested in the literature for public goods games without punishment (Nosenzo, Quercia, & Sefton, 2015).

where $c \equiv |\boldsymbol{C}|$ denotes the number of citizens in *CenPun*. For citizens in the other two institutions the same payoff function holds with respect to the sets $\boldsymbol{D}$ and $\boldsymbol{N}$.

In stage three, players receive signals about the contribution of the other citizens in their institution. For each citizen i a signal is produced, such that

$$s_i = \begin{cases} g_i & \text{with } prob = \lambda, \\ \widetilde{g}_i & \text{with } prob = 1 - \lambda, \end{cases} \tag{2}$$

where $\widetilde{g}_i$ is randomly drawn from the set $\{0, 2, 4, \ldots, 20\} \setminus \{g_i\}$ with uniform probabilities. Thus, for each citizen, there is an independent random draw determining whether the signal corresponds to the true contribution or not. If not, another independent draw selects a different contribution. The signal s_i is communicated to all other citizens in i's institution, and, in case of *CenPun*, also to the authority. Citizen i does not know whether the other participants receive a true or false signal about his contribution.

In addition, all citizens receive an extra endowment of three units. Depending on their institution, players assign punishment points (that is, citizens in *DecPun* and the authority in *CenPun*), and the final payoffs are realized. The three institutions differ only in stage three. For a citizen in *NoPun* the payoff equals the profit after stage two plus the extra endowment:

$$\pi_i = \hat{\pi}_i + 3 \qquad \forall\, i \in \boldsymbol{N}. \tag{3}$$

In *DecPun* all citizens decide simultaneously over punishment $p_{i \to k}$ with $k \in \boldsymbol{D} \setminus \{i\}$. Each punishment point assigned to another citizen leads to a deduction of three units from the punished citizen's payoff and reduces the punisher's payoff by one unit. Each citizen can spend up to her extra endowment for punishment, that is, $\sum_k p_{i \to k} \leq 3$.[9] Units not spent on punishment are credited to the citizens' payoff. For a citizen i in *DecPun*, the payoff equals

$$\pi_i = \hat{\pi}_i + (3 - \sum_{k \in \boldsymbol{D} \setminus \{i\}} p_{i \to k}) - 3 \sum_{k \in \boldsymbol{D} \setminus \{i\}} p_{k \to i} \qquad \forall\, i \in \boldsymbol{D}. \tag{4}$$

In *CenPun* all punishment decisions are delegated to the authority. The authority decides over punishment $p_{\to k}$ with $k \in \boldsymbol{C}$. Like in *DecPun* each

[9] This design of the punishment stage has the property that larger groups have more resources for punishment. It is, however, unclear whether this means that punishment is more severe in larger groups, because larger groups might be faced with more deviators, or subjects might be more likely to act as a bystander on the punishment stage. For a discussion of the adaptation of the punishment mechanism to various group sizes see also Roux and Thöni (2015).

punishment point assigned to a citizen leads to a deduction of three units from the punished citizen's payoff and costs one unit. In *CenPun* these costs have to be borne equally by all other citizens in the institution. In sum, the authority can spend up to the extra endowment of all its citizens for punishment, i.e., $\sum_k p_{\to k} \leq 3c$. In addition, maximum punishment targeted at a single citizen is restricted to $3(c-1)$. Units not spent on punishment are credited to the particular citizen's account. Hence, *DecPun* and *CenPun* are identical with regard to the feasible set as well as the financial consequences of punishment. The only difference is that punishment *decisions* are taken by the authority. For citizen i in *CenPun*, the payoff equals

$$\pi_i = \hat{\pi}_i + \left(3 - \frac{\sum_{k \in \boldsymbol{C} \setminus \{i\}} p_{\to k}}{c-1}\right) - 3p_{\to i} \qquad \forall\, i \in \boldsymbol{C}. \tag{5}$$

The authority's payoff equals the average profit after stage two of all citizens in institution *CenPun*

$$\pi_A = \frac{\sum_{i \in \boldsymbol{C}} \hat{\pi}_i}{c} \qquad \text{if } c \geqslant 2. \tag{6}$$

If there is only one citizen in an institution, there is no public good and no punishment. In this case, the citizen receives a payoff of 20. If there are less than two citizens in *CenPun* the authority receives a payoff of 20.[10] All parameters, the signal technology (λ), and payoff functions are public information.

We vary the information environment λ across treatment conditions. In treatment ONE citizens and the authority receive perfect information regarding the contributions of members of their institution ($\lambda = 1$). In treatment POINT-NINE we set $\lambda = .9$, such that citizens and the authority receive a signal about the others' contributions that displays the accurate information in nine out of ten cases (and a different contribution in the remaining case). Finally, in POINT-FIVE players receive accurate information in five out of ten cases ($\lambda = .5$).

2.2 Deterrent punishment

In the main text we restrict our theoretical analysis of the game to the punishment stage. In particular, we derive an expression for *deterrent* punishment, that is, the strength of punishment required to render unilateral deviation from a situation with mutual cooperation unprofitable. At the end of this

[10] These payoffs ensure that the authority has an incentive attract at least two citizens, and citizens have an incentive to form groups.

section, we sketch the equilibria of the entire game, but characterize it fully in the online appendix.

Assume a central punishment institution seeks to enforce a contribution of $\bar{g} > 0$ in a group of c citizens. Let us assume that a citizen i is risk neutral and has selfish preferences, and all other citizens contribute $g_{-i} = \bar{g}$. As we show in the appendix, minimal punishment required to make i indifferent between contributing $g_i = \bar{g}$ and $g_i = 0$ is

$$p^*_{\to i}(s_i, \lambda, c, \bar{g}) = \frac{(10c - 16)}{3c(11\lambda - 1)} \max\{\bar{g} - s_i, 0\} \quad \text{for } c \geqslant 2, \lambda > \tfrac{1}{11}. \tag{7}$$

Signals equal to $\bar{g}$ or above trigger no punishment. For signals lower than $\bar{g}$ punishment is linearly decreasing in the signal. The absolute slope of the punishment function in s_i ($|\frac{\partial p}{\partial s_i}|$) is increasing in λ and approaching infinity as $\lambda \to \frac{1}{11}$, which refers to the case of an uninformative signal. Thus, the lower the accuracy of the signals the more punishment is required at a given signal to achieve deterrence. Enforcing higher contributions ($\bar{g}$), as well as increasing the group size (c) requires more punishment for a given signal.

The punishment necessary to deter free-riding is independent of the punishment institution. In case of *CenPun* the authority uses Equation (7) to punish all the $i \in \boldsymbol{C}$ citizens in the group. In case of *DecPun* the expression would be the same with the exception that we have to replace c by d, the number of players in *DecPun*. However, it lies in the very nature of this institution that players face a coordination problem in the punishment stage. Equation (7) just specifies the total punishment that should be assigned to player i, but does not specify the allocation to punishers. A natural benchmark would be that each citizen bears the same share of the punishment costs for citizen i, i.e., all players j punish player i by $p_{j\to i} = \frac{p^*_{\to i}}{d-1}$, $\forall\, j \in \boldsymbol{D} \setminus \{i\}$ punishment points. If all players follow this punishment strategy, then *DecPun* and *CenPun* would be equivalent in terms of payoffs for the citizens.

In the introduction we stressed the role of antisocial punishment as a determinant for contributions. We understand this term in the sense of describing an act with the *intent* to punish cooperative subjects. Under imperfect information there is a potential misalignment between the intended action and the realized action, either due to false signals, or due to signals suspected to be false. This makes it difficult to qualify a punishment act as antisocial punishment. To account for this we introduce two terms, both relating to antisocial punishment: For punishment which—independent of the signal—hits cooperative citizens (with $g_i > \bar{g}$) we use the term *contributor punishment*; for punishment acts targeted at citizens for which the punisher receives a high signal ($s_i \geqslant \bar{g}$) we will use the term *misguided punishment* (independent of g_i). In both cases we will refer to the opposite punishment (either $g_i < \bar{g}$, or $s_i < \bar{g}$) as free-rider punishment. Punishment according to Equation (7) therefore rules out misguided punishment, while contributor punishment becomes stronger the lower the quality of the signals.[11]

Given the payoff functions it is in the interest of all players in *DecPun* and *CenPun* to enforce maximum contributions ($\bar{g} = 20$). However, depending on λ this might not be feasible. In particular, high levels of noise in the signals require amounts of punishment which are outside of the feasible set of the punisher(s). In Appendix A.1 we show that enforcing maximum contributions is feasible in ONE and POINT-NINE, but typically not in POINT-FIVE. Since the punishment endowment and technology are identical in *DecPun* and *CenPun* this holds equally for both institutions. Under standard assumptions (selfish preferences and subgame perfection) punishment in *DecPun* is a non-credible threat and should not occur. Consequently, the central authority should be able to attract all citizens in ONE and POINT-NINE. Countless experiments suggest, however, that this is not an accurate description of punishment behavior in decentralized sanctioning institutions. In the next section we specify the experimental setup and use theoretical arguments and stylized facts on punishment to formulate behavioral predictions.

3 Implementation

3.1 Experimental setup

The experiment is played in matching groups of eleven subjects. Prior to the start of the game we randomly allocate one subject in each matching group to the role of the authority and ten subjects to the role of the citizen.[12] Roles remain the same throughout the experiment.

Because the game is fairly complicated, and because we think that interesting things might unfold with time we implement a repeated game of 32 periods. Participants know that they play the game for the finite number of periods.[13] Since we want to provide the three institutions with some time to

[11] When analyzing the data we cannot observe $\bar{g}$ and we will use the mean contribution (or the mean signal) instead.

[12] Grosse, Putterman, and Rockenbach (2011) use the same technique to introduce a observer in their public good game.

[13] English translations of the instructions are reported in Appendix A.2. Before the experiment starts, subjects have to solve a set of control questions on the computer screen.

establish cooperation before they are put into competition with other institutions, the citizens in our experiment choose their institution every fourth period only. Thus we implement a game with eight phases consisting of four periods each. At the beginning of each phase all subjects allocated to the role of citizens choose one of the three institutions and remain there during the phase.

Each period consists of three steps, a contribution step, a punishment step, and an information step. Appendix A.3 shows the information provided on the screens during the experiment. In the punishment step, all citizens and the authority receive the signals from the citizens in their institution. If applicable, citizens or the authority choose their punishment points. The identification number of citizens are randomly reassigned between periods. In the information step, citizens learn their period payoff including the total amount of punishment received. Citizens do not receive information about their own signal, that is, they do not know whether other subjects were correctly informed about their contribution or not. Citizens learn only the total amount of punishment received, and not the number nor the identifier of the citizens who punished them.

At the beginning of each phase all citizens are informed about the outcome in all institutions (see screenshot in Figure A2). In particular, when choosing an institution citizens know (i) the number of citizens (ii) the average contribution, and (iii) the average profit in all three institutions and for all previous periods. At this point all information is undistorted. In the light of this information citizens choose their institution for the next phase. There is no cost attached to switching an institution.[14]

3.2 Behavioral Predictions

In this section we develop predictions about the effect of the treatment variation on the main outcome variable, the institutional choice. Our null hypothesis is that the amount of noise in the signals does not affect the number of citizens attracted by the three institutions. The alternative hypothesis is that the popularity of the three institutions systematically varies with the amount of noise.

While our theoretical analysis of the institutional choice briefly mentioned at the end of Section 2.2 does not offer a compelling prediction, we still want to be more specific as to what we expect from the three treatments. In the

[14] It is certainly an extreme assumption that moving from one institution to another is costless. However, we decided against introducing an arbitrary switching cost because we want to measure preferences for institutions unaffected by other considerations such as the sunk cost fallacy.

following we use theoretical arguments as well as stylized facts from previous experimental research to develop three conjectures about the direction of the effect. We start with ONE, the treatment with perfect information.

A large body of evidence on public goods games with punishment shows that the majority of individuals is willing to use costly punishment to sanction free-riders in games with decentralized punishment institutions (Chaudhuri, 2011). While it is difficult to explain costly punishment under standard assumptions, theories of social preferences explain punishment either by assuming inequality averse preferences or reciprocal preferences.[15] In the former cooperative citizens are willing to punish free-riders to eradicate their payoff differences; in the latter free-riding is perceived as an unkind act, which motivates retaliatory punishment.

While we do not offer a theoretical treatment of our game with social preferences, it seems intuitive that both flavors of social preferences can rationalize that citizens contribute if (and only if) others contribute as well, and that unequal contributions trigger punishment to deter free-riding.

In line with this perspective, earlier experimental studies find that under perfect information a majority of the subjects end up in the punishment institution when given the choice between decentralized punishment and no punishment (Gürerk et al., 2006). Subjects' willingness to punish free-riders creates a credible threat and coordinates behavior on high contributions, and only little actual punishment is required to enforce this outcome. Based on this stylized fact, we expect that citizens manage to reach and maintain high contributions in *DecPun* in ONE.[16] Furthermore, previous experimental evidence suggests that subjects have a preference for retaining authority (Fehr, Herz, & Wilkening, 2013), and they might gain satisfaction from punishing defectors themselves (De Quervain et al., 2004). In addition, citizens may fear that the central authorities punish excessively due to the fact that they do not bear the marginal cost of punishment. For treatment ONE we thus expect *DecPun* to be the prevailing institution:

[15] For inequality aversion there are fairly specific predictions for free-rider punishment (Fehr & Schmidt, 1999), as well as antisocial punishment (Thöni, 2014). We are not aware of an application of contemporary models of reciprocity (such as Dufwenberg and Kirchsteiger, 2004 or Falk and Fischbacher, 2006) to the punishment decision in public goods games. Ambrus and Pathak (2011) analyze public good games with reciprocal preferences.

[16] Herrmann et al. (2008) show that there is large cross-societal variation in the public goods game with punishment. We conducted the experiments in Bonn, Germany, where previous evidence points towards high contributions and little antisocial punishment in *DecPun*. Consequently, the conjectures we develop here are sensitive to the societal background in which the experiment is conducted.

Conjecture 1 *Under perfect information (treatment* ONE*) the majority of citizens choose DecPun.*

What changes under imperfect information? We begin with treatment POINT-NINE, where signals are accurate in 90 percent of the cases. The results of Grechenig et al. (2010) suggest that a small amount of noise does not hamper the enforcement of high contributions in decentralized punishment. While they find more punishment in POINT-NINE than under perfect information, profits in later periods are still higher than in the treatments without punishment. Consequently, as in treatment ONE, institutions with punishment should have a competitive advantage.

For three reasons we think that in this treatment citizens might find it attractive to delegate the sanctioning decisions to the central authority. First, unlike in ONE, there is the risk that punishment acts do not hit the right citizen, and thereby do not serve as retaliation (in case the punishment was motivated by reciprocal preferences), or increase instead of decrease the inequality in the group (for inequality averse agents). Since the authority does not have more information than the citizens, erroneous punishment acts are just as likely in *CenPun*. While an inequality averse player does not care about who pulled the trigger, reciprocally motivated agents do. Thus, the latter type of agent might prefer to shift the responsibility for punishment to the authority.

A second reason is that punishment which mistakenly hits high contributors might motivate revengeful reactions in the form of misguided punishment (see Herrmann et al., 2008; Leibbrandt et al., 2015). While retaliatory punishment may also play a role in ONE, we think that the levels of misguided punishment in our subject pool (Univ. of Bonn) are too low to trigger such vicious cycles. The treatment POINT-NINE, on the other hand, might introduce the right amount of ambiguity in order to mess things up in *DecPun*. *CenPun* could then be an attractive alternative, because this institution delegates all responsibility for punishment to the authority and rules out retaliatory punishments among the citizens.

Third, the results of Ertan, Page, and Putterman (2009) show that subjects prefer institutional environments which do not allow for punishment of high contributors. While these results stem from experiments with perfect information, we interpret this as evidence that the more punishment of cooperative citizens is perceived a problem, the more citizens are willing to tolerate restrictions in their punishment authority. For the treatment with low noise levels we therefore expect:

Conjecture 2 *Under low noise levels (treatment* POINT-NINE*) the majority of citizens choose CenPun.*

In treatment POINT-FIVE signals are accurate only half of the time. In experiments with exogenous institutional environments, Grechenig et al. (2010) find that contributions in the treatment with decentralized punishment are similar as in the treatment without punishment, despite the fact that the subjects use punishment no less than in the treatments with accurate signals. Because punishment is costly, profits are lower in the treatment with punishment. In a similar setting, Ambrus and Greiner (2012) also find lower profits when punishment is available. This should give *NoPun* a competitive advantage over *DecPun* in POINT-FIVE. Furthermore, our theoretical analysis in Appendix A.1 shows that enforcing full contributions under high noise is often impossible, because the deterrent punishment levels are outside the feasible set. Unlike under perfect information, where the threat of punishment suffices, enforcing contributions in POINT-FIVE requires high levels of punishment even if the group fully cooperates. For this reason we expect that in such a noisy environment *CenPun* cannot offer substantial advantages over *DecPun*, and thus neither of the punishment institutions will prevail:

Conjecture 3 *Under high noise (treatment* POINT-FIVE*) the majority of citizens choose NoPun.*

4 Results

We ran 15 experimental sessions with 30 independent populations (330 participants, 110 per treatment). Each subject participated in only one population. The experiments were conducted at the laboratory for economic experiments (EconLab) at the University of Bonn with mostly undergraduate students from various fields. Six percent of participants were non-students, 56 percent of participants were females, and age ranged between 18 and 64 (median 22). The experiment was programmed in z-Tree (Fischbacher, 2007); we used ORSEE (Greiner, 2015) for recruiting. A session lasted for about 120 minutes. Payoffs were converted at an exchange rate of 1 Euro per 75 ECUs; payoffs accrue over all periods. Subjects earned on average 15.64 Euros, including a show-up fee of 4 Euros.

The results section is structured as follows: First we show that noise influences institutional choices in a systematic way: Citizens opt predominantly for *DecPun* in ONE, for *NoPun* in POINT-FIVE, while all three institutions attract similar shares in POINT-NINE (Result 1). Then we relate institutional choices to punishment behavior and show that punishment towards cooperative citizens predicts exit in *DecPun* and *CenPun* (Result 2). In the next step, we analyze contributions and profits and show that both decrease when signals become noisy (Result 3). We then demonstrate for the final stage of the game that central authorities choose punishments close to the deterrent levels, while punishment in *DecPun* is typically stronger (Result 4). Finally, we show that, under imperfect information, authorities who avoid misguided punishment gain a competitive advantage and are able to attract the majority of citizens (Result 5).

4.1 Choice of institution

For the choice of institution in the first phase, *NoPun* attracts the majority of the population in all treatments. About two thirds of the subjects choose this institution in POINT-NINE and even more so in the other two treatments. This is in line with the results of Gürerk et al. (2006), who also find that their punishment institution is not popular early in the game. Centralized punishment initially attracts 21 percent of the citizens in POINT-NINE, compared to 13 and 7 percent in ONE and POINT-FIVE, respectively. These differences in the initial choice of institutions are significant across treatments ($p = .027$, Fisher's exact test). Over time, most citizens move to the two punishment institutions. Comparing the three institutional choices across treatments we can reject our null hypothesis: The allocation of citizens into *DecPun*, *CenPun*, and *NoPun* is significantly different across treatments ($F(2.49, 72.3) = 3.41$, $p = .029$ for all phases; $F(2.85, 82.7) = 3.14$, $p = .032$ for the final phase, Pearson χ^2 statistic with correction for dependence within matching group, see Rao and Scott, 1984).

The top panels of Figure 1 show the average choice of institutions for each treatment. Across all phases we find evidence for our three conjectures: In ONE, the modal choice is clearly *DecPun*, while in POINT-FIVE the modal choice is *NoPun*. In POINT-NINE the modal choice is *CenPun*, although only by a small margin over the two other institutions.[17]

The bottom panels of Figure 1 show the relative share of the institutions over time. In all treatments, *NoPun* loses a lot of citizens during the first three phases. Most of the adjustments happen through the first half of the 32 periods and we observe relatively stable shares of institutions in the second half of the experiment in ONE and POINT-FIVE. In POINT-NINE, the share of *CenPun* is stable, but *NoPun* loses in favor of *DecPun* throughout the 32 periods. Thus, while the evidence supports our Conjectures 1 and 3, the results are less clear with regard to Conjecture 2, which postulated the

[17] One-sample Pearson χ^2 tests for the null hypothesis of equal probabilities for all three institutions (corrected for dependence within matching group) are insignificant for ONE ($p = .196$) and POINT-NINE ($p = .892$), and significant for POINT-FIVE ($p = .003$). In the final phase we have ONE: $p = .051$, POINT-NINE: $p = .064$, and POINT-FIVE: $p = .538$.

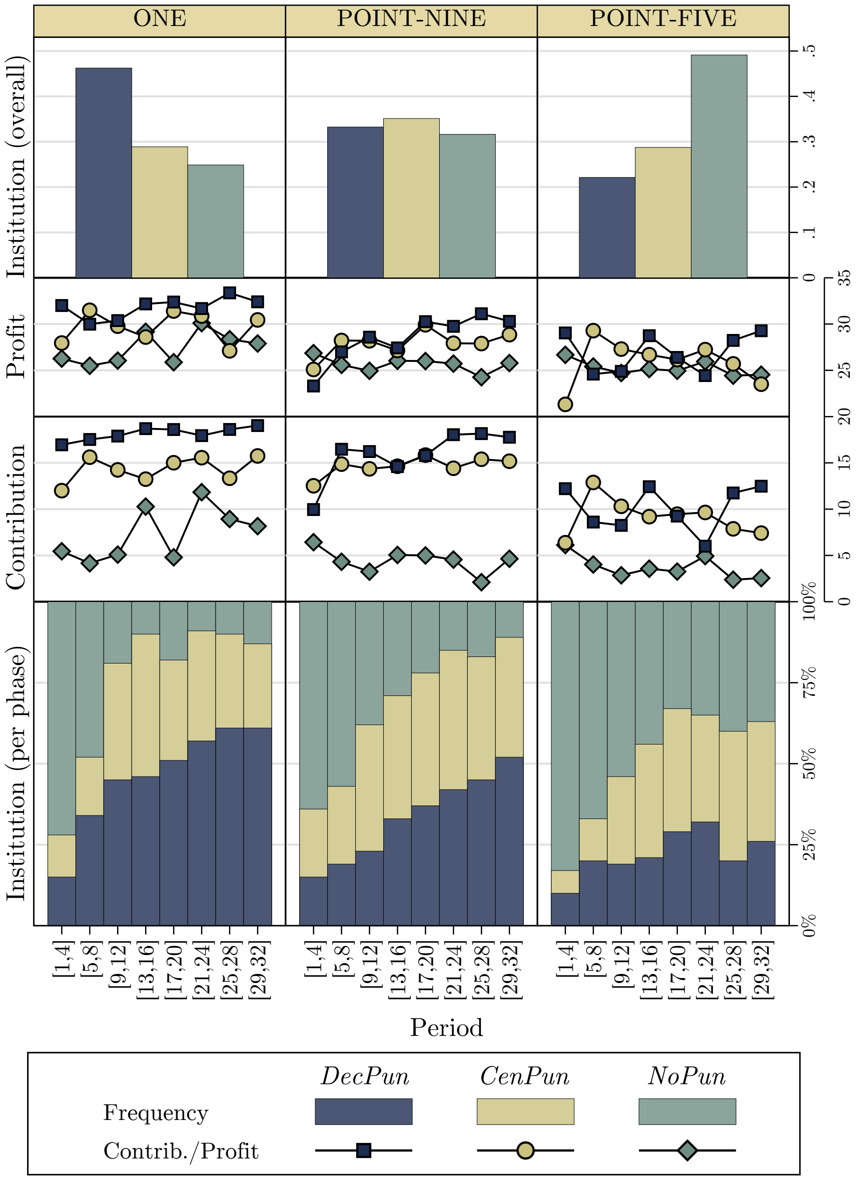


Figure 1: Top panel: Average choice of institution over all periods and by treatment. Middle panels: Average profits and contributions in *NoPun*, *DecPun*, and *CenPun* across time. Dots show averages in a phase of four periods. Bottom panel: Choice of institution during the eight phases.

dominance of *CenPun* in treatment POINT-NINE. Summarizing our results on the choice of institutions we find:

Result 1 *Institutional choices are significantly affected by the level of noise in the signals. After some early adjustments, citizens choose predominantly DecPun in* ONE*, while NoPun retains highest shares in* POINT-FIVE*. In* POINT-NINE *all three institutions attract similar shares of the population.*

In the next step we want to take a closer look at the determinants for the choice of an institution. Recall that when citizens can move between institutions, they are informed about the outcomes in the three institutions. In particular, citizens learn (i) the number of citizens, (ii) the average contribution, and (iii) the average profits earned in each of the three institutions in all previous periods. We use multinomial probit models to explain the choice of institution between phases. For each citizen we observe seven institution choices with information about the outcome of the prior phase. In Model (1) of Table 1 we explain the choice of institution by the average profit of the citizens in each institution in the previous phase.[18] We use two dummies for the treatments ONE and POINT-NINE, with POINT-FIVE being the omitted case. We also add two dummies for the institution in which the subject is currently in, with *NoPun* as the omitted case, and we add a linear time trend (variable *Phase*). The treatment dummies indicate that citizens are less likely to choose *NoPun* over *DecPun* in the two treatments with relatively accurate or perfect information.

We find evidence for inertia in the choice of institution. Having been in *NoPun* before significantly increases the chance of choosing *NoPun* relative to *DecPun*, as shown by the significant negative effects of both institution dummies. The coefficients of the three profit variables show that this information is indeed a strong determinant for the institutional choice. Observing high profits in *NoPun* significantly increases the probability of choosing *NoPun* over *DecPun* for the next phase, while the opposite is true for high profits in *DecPun*. The profits in *CenPun* do not seem to affect the choice between *NoPun* and *DecPun*. The estimates for choosing *CenPun* (the second set of covariates in Table 1) show a very similar pattern. High profits in *CenPun* increase the probability of choosing that institution for the next phase over *DecPun*, while the opposite is true for high profits in *DecPun*.

Although the relation between relative profits and institution choice is strong, it is not informative with regard to the ultimate causes of the relative attractiveness of the institutions, because profits are merely a result of

[18] In case there were no citizens in a given institution we cannot observe a profit. In the estimates we use the same profit as in the case when there is only one citizen in a given institution.

Table 1: Choice of institution.

	Dependent variable: Institution in $t+1$			
	(1)		(2)	
Choose *NoPun*				
ONE	-0.417^{***}	(0.139)	-1.001^{***}	(0.240)
POINT-NINE	-0.298^{**}	(0.131)	-0.644^{***}	(0.205)
DecPun	-1.740^{***}	(0.181)	-2.137^{***}	(0.196)
CenPun	-0.824^{***}	(0.163)	-0.894^{***}	(0.188)
Phase	-0.003	(0.026)	-0.113^{***}	(0.031)
Profit *NoPun*	0.100^{***}	(0.018)		
Profit *DecPun*	-0.124^{***}	(0.014)		
Profit *CenPun*	-0.007	(0.011)		
Free-rider pun × *DecPun*			0.045^{**}	(0.020)
Contributor pun × *DecPun*			0.117^{***}	(0.045)
Free-rider pun × *CenPun*			0.074^{**}	(0.035)
Contributor pun × *CenPun*			0.040	(0.066)
Constant	1.753^{***}	(0.586)	1.607^{***}	(0.162)
Choose *CenPun*				
ONE	-0.208	(0.165)	-0.473^{*}	(0.287)
POINT-NINE	-0.072	(0.162)	-0.183	(0.285)
DecPun	-0.778^{***}	(0.190)	-1.286^{***}	(0.180)
CenPun	0.787^{***}	(0.162)	1.219^{***}	(0.168)
Phase	-0.006	(0.028)	-0.034	(0.032)
Profit *NoPun*	0.035^{**}	(0.016)		
Profit *DecPun*	-0.123^{***}	(0.015)		
Profit *CenPun*	0.133^{***}	(0.015)		
Free-rider pun × *DecPun*			0.028	(0.021)
Contributor pun × *DecPun*			0.114^{**}	(0.046)
Free-rider pun × *CenPun*			-0.010	(0.028)
Contributor pun × *CenPun*			-0.125^{**}	(0.057)
Constant	-1.005^{*}	(0.577)	0.313	(0.202)
Wald χ^2-test	1724.2		606.2	
p	0.000		0.000	
N	2100		2100	

Notes: Multinomial probit estimates. Dependent variable: Chosen institution for the next phase (*DecPun* is the omitted case). Independent variables are treatment dummies (POINT-FIVE as omitted case), dummies for the institution in the previous phase (*NoPun* as omitted case), Phase, average profits in the actual phase in the respective institution, and free-rider and contributor punishment in the respective institutions during the previous phase. Robust standard errors, clustered on matching group, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the activities in a given phase. The profits are mainly linked to contributions (for *NoPun* they are linearly dependent). If we replace the profits by contributions in Model (1) of Table 1 we get very similar results (not shown in the table), that is, high contributions in an institution increase the probability of choosing the respective institution. However, the main source of the relative popularity of the two punishment institutions should be determined in the way the citizens and the authority use the punishment option.

In Model (2) of Table 1 we investigate the use of the punishment option as a determinant of institution choice. Just adding the frequency or strength of punishment used in a given institution is, however, not an adequate measure of how well cooperation norms are enforced. Punishment can not only hit low contributors, but also at high contributors. We classify received punishment into free-rider punishment (if the punished citizen contributed less than the group average) or contributor punishment (otherwise). We replace the covariates for the profits by variables measuring free-rider, and contributor punishment, interacted with the dummy for the two institutions allowing for punishment. The results in the upper half of Table 1 show that punishment in *DecPun* increases the probability of leaving the institution in favor of *NoPun*. Interestingly this holds both for free-rider punishment and contributor punishment, although the latter effect seems to be stronger (the coefficients are not significantly different). In the lower half of the table we find clear evidence that the occurrence of contributor punishment is decisive for the choice between the two institutions allowing for punishment. High contributor punishment in *DecPun* significantly increases the probability of choosing *CenPun*, and vice versa. The strength of free-rider punishment, on the other hand, does not significantly affect the choice between these two institutions.

Result 2 *The choice of institutions is importantly influenced by the punishment behavior in the previous periods. In particular the amount of punishment towards cooperative citizens (contributor punishment) significantly predicts exit, both for DecPun and CenPun.*

When exiting an institution where punishment is possible citizens can either opt for *NoPun* or for the alternative punishment institution. In ONE and POINT-NINE the majority of citizens leaving *DecPun* opt for *CenPun* in the next phase (71.4 and 65.5 percent), while in POINT-FIVE we observe a majority of moves towards *NoPun* (57.7 percent of the cases). For citizens deciding to leave *CenPun* the results are less clear. In all treatments we observe a move to *DecPun* in slightly more than half of the cases (ONE: 57.4%, POINT-NINE: 56.4%, POINT-FIVE: 51.3%).

Given the crucial role of punishment of cooperative citizens in the choice of institution we will analyze the punishment behavior in response to the treatment variation in more detail in section 4.3. Before we do that, we focus on two other outcome variables of interest, contributions and profits.

Table 2: Contributions and profits

		ONE	POINT-NINE	POINT-FIVE
Contribution	Overall	14.2	12.3	6.6
	NoPun	5.8	4.7	4.0
	DecPun	18.4	16.6	9.8
	CenPun	14.4	14.8	9.0
Profit	Overall	29.2	27.3	25.4
	NoPun	26.1	25.5	25.4
	DecPun	31.6	28.8	25.9
	CenPun	29.2	27.9	25.7

Notes. Average contributions and profits for the three treatments, both overall and for each institution separately. Averages are calculated based on individual observations.

4.2 Contributions and profits

Varying the noise in the contribution signals does not only affect institutional choices, but also the degree to which citizens manage to mitigate free-rider problems within their population. Table 2 shows the average contributions in the three treatments. Averages over all institutions are highest in ONE, followed by POINT-NINE, and POINT-FIVE. These treatment differences are significant at $p = .000$ (Kruskal-Wallis test on matching group averages). The same holds for average profits across institutions ($p = .002$).

The averages per institutions show that both contributions and profits are typically highest in *DecPun*, followed by *CenPun* and *NoPun*.[19] The middle panels of Figure 1 show the average profits and contributions in the three institutions over time. In most of the phases profits are higher for the two institutions allowing for punishment, but overall differences are not pronounced. This is not surprising, given that there is free movement between institutions every fourth round.

The profits of *DecPun* are comparable to the results of Grechenig et al. (2010, p. 861) (GNT hereafter), who use the same signal technology and marginal per capita return, but randomly assign subjects to an institution and use a partner matching with groups of four subjects. In ONE (POINT-FIVE) we observe somewhat higher average profits of 31.6 relative to 29.7 in GNT (25.9 vs. 24.2 in GNT).[20] In POINT-NINE the profits are very similar

[19] We calculate the averages based on all individual decisions or outcomes. The overall average is therefore not equal to the average of the values for the three institutions.

[20] The values of GNT are corrected for the fact that their punishment was higher (10)

(28.8 vs. 28.5 in GNT). On the other hand, profits in *NoPun* are lower than in the corresponding treatments in GNT (26.1 vs. 28.5 in GNT for ONE, and 25.4 vs. 28.3 in GNT for POINT-FIVE, POINT-NINE not available).

Figure 1 suggests a positive correlation between the share of the population and average profits in an institution. This correlation is strong and significant for ONE ($\rho = .711$, $p = .000$, correlations based on matching group averages for the three institutions) and POINT-NINE ($\rho = .700$, $p = .000$), while the two are virtually uncorrelated in POINT-FIVE ($\rho = -.016$).

Table 3 shows the results from OLS estimates explaining contributions. In Model (1) we include treatment dummies and two controls for time effects. The first identifies the period within each phase of four periods, the second controls for the time trend over the course of the eight phases. The results confirm that the treatments POINT-NINE and ONE result in significantly higher contributions. Interestingly contributions significantly drop within a phase, but at the same time significantly increase over the eight phases.

In Model (2) we add dummies for the institutions to the model (with *NoPun* as omitted case), and we control for the number of citizens in the institution. We allow for non-linear effects of group size by adding a quadratic term as well (n and n^2). While the controls in Model (1) are clearly exogenous, this is no longer the case for Model (2). The results should therefore not be interpreted in a causal way. Both institutions with a punishment are related to higher contributions compared to *NoPun*. Furthermore, the coefficient for *DecPun* is significantly larger than for *CenPun* ($p = .014$). These results are similar when using profits instead of contributions as dependent variable (models not shown in the table): Profits are highest in *DecPun*, followed by *CenPun*, and *NoPun*. The differences are significant in the overall sample. The result that *DecPun* leads to higher profits that *CenPun* is, however, mainly driven by treatment ONE. Both in contributions and in profits the differences become insignificant if we estimate the models based on the treatments POINT-NINE and POINT-FIVE only ($p > .2$).[21]

With respect to group size it seems that either large or small groups are conducive to high contributions, whereas groups of around five citizens tend to have lowest contributions.[22] In the remaining models $(3) - (5)$ we repeat

than in the present experiment (3).

[21] Alternatively we estimated the model with all observations, but allowed for an interaction between *DecPun* and ONE. For both profit and contributions the interaction is positive and significant, while the difference between the dummies for *CenPun* and *DecPun* becomes insignificant ($p > .27$).

[22] Thus, unlike what could have been expected from the literature discussed in footnote 8, the fact that the marginal per capita return decreases in group size does not produce a similar monotonic pattern in the contributions. In case of decentralized punishment this

Table 3: Contributions

	Dependen variable: Contribution				
	All observations		ONE	POINT-NINE	POINT-FIVE
	(1)	(2)	(3)	(4)	(5)
POINT-NINE	5.719^{***}	3.745^{***}			
	(1.262)	(0.908)			
ONE	7.591^{***}	4.587^{***}			
	(1.216)	(0.799)			
Period in phase	-0.592^{***}	-0.592^{***}	-0.395^{*}	-0.462^{**}	-0.921^{***}
	(0.097)	(0.097)	(0.203)	(0.158)	(0.077)
Phase	0.816^{***}	0.092	0.239^{**}	0.172	-0.260^{***}
	(0.138)	(0.067)	(0.078)	(0.135)	(0.071)
DecPun		9.572^{***}	11.707^{***}	11.066^{***}	5.417^{***}
		(0.685)	(0.557)	(0.630)	(1.293)
CenPun		7.352^{***}	7.860^{***}	9.528^{***}	5.197^{***}
		(0.789)	(1.798)	(1.199)	(0.677)
n		-1.616^{***}	-1.680^{*}	-0.172	-2.976^{***}
		(0.459)	(0.772)	(0.576)	(0.630)
n^2		0.152^{***}	0.136^{**}	0.042	0.233^{***}
		(0.035)	(0.059)	(0.041)	(0.049)
Constant	4.363^{***}	7.130^{***}	10.625^{***}	4.840^{**}	15.564^{***}
	(0.979)	(1.731)	(2.701)	(1.519)	(1.812)
F-test	84.6	237.9	184.8	350.1	49.0
Prob > F	0.000	0.000	0.000	0.000	0.000
R^2	0.215	0.445	0.451	0.446	0.217
N	9160	9160	3064	3052	3044

Notes: OLS estimates. Dependent variable: contribution. Independent variables: treatment dummies (with POINT-FIVE as omitted case), period within a phase $(1 - 4)$, and phase $(1 - 8)$, dummies for the institution (*NoPun* as omitted case), and two measures for the number of citicens in the institution, n, n^2. Robust standard errors, clustered on matching group, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the estimates for the three treatments separately. In all three treatments contributions are highest in *DecPun*, followed by *CenPun*. The differences are, however, only significant in ONE. It seems that higher levels of noise close the gap in contributions between *DecPun* and *CenPun*. For the group size we observe a significant u-shaped effect with a minimum around five to six participants for ONE and POINT-FIVE, while the coefficients are insignificant for POINT-NINE.

Result 3 *Contributions and profits decrease as the information about other players' behavior becomes noisy. The two institutions with punishment result in higher contributions than NoPun. Contributions in DecPun tend to be higher than in CenPun, but the difference becomes small and insignificant under noise.*

is in line with previous research (Carpenter, 2007; Roux & Thöni, 2015).

4.3 Punishment strategies

Our results above suggest that the use of the punishment option importantly influences the choice of institutions. In Section 2.2 we derived an expression for minimal deterrent punishment (see Equation 7). In the following we compare the punishment decisions observed in the experiment to this theoretical benchmark. Recall that the expression requires to specify a contribution level $\bar{g}$ to be enforced. To calculate the benchmark we set $\bar{g}$ to a 'typical' contribution level. More precisely, we set $\bar{g}$ equal to the median of the signals an authority in *CenPun*, or a citizen in *DecPun* receives in a given period.[23]

The left panel of Figure 2 shows the results for the punishment of authorities in the three treatments. On the horizontal axis we depict the difference between the signal and the median signal in the group $(s_i - \bar{g})$. For example, a value of -20 refers to the case where the signals indicate that the citizen in question contributed zero and the majority of the other citizens contributed fully. The bars indicate the average number of punishment points meted out for the respective deviation. The horizontal lines show the minimal deterrent punishment, which is decreasing in the signal for free riders (negative deviations) and zero thereafter.[24] The top left panel of Figure 2 shows the results for ONE. Punishment clearly follows the theoretical pattern, but tends to be somewhat lower than predicted, with the exception of moderate negative deviations of four to two units. Zero or positive deviations trigger almost no punishment. For POINT-NINE we observe that punishment for negative deviations tends to be very close to the predicted level, while again misguided punishment seems to be negligible. We will, however, argue below that misguided punishment still plays an important role for the popularity of the central authority. In POINT-FIVE we observe a different pattern. For negative deviations punishment is much lower than deterrent punishment. In addition, punishment seems almost invariant across the deviation classes.[25]

In a next step we want to contrast the punishment in *CenPun* to the

[23] For even numbers of signals we slightly deviate from the usual calculation of the median and take the higher of the two middle values, that is, in case of the signals $\{20, 18, 12, 0\}$ we set $\bar{g} = 18$. Alternatively one could argue that punishers should always try to enforce full contributions and thus set $\bar{g} = 20$. This would result in higher predicted levels of punishment.

[24] The expression in Equation 7 is linearly decreasing in s_i for $\bar{g} > s_i$, while the horizontal lines in Figure 2 are not. The reason for this is that we combine all cases with various values for $s_i - \bar{g}$ and c into a single average per bar.

[25] We use exact Wilcoxon signed rank tests for the difference between prediction and data (matching group averages). In case of ONE none of the differences are significant ($p > .125$). For POINT-NINE we observe a significant difference for the bars $[-8, -6]$ ($p = .039$), and $[-4, -2]$ ($p = .006$), all others are insignificant ($p > .4$). In case of POINT-FIVE all but one differences are highly significant.

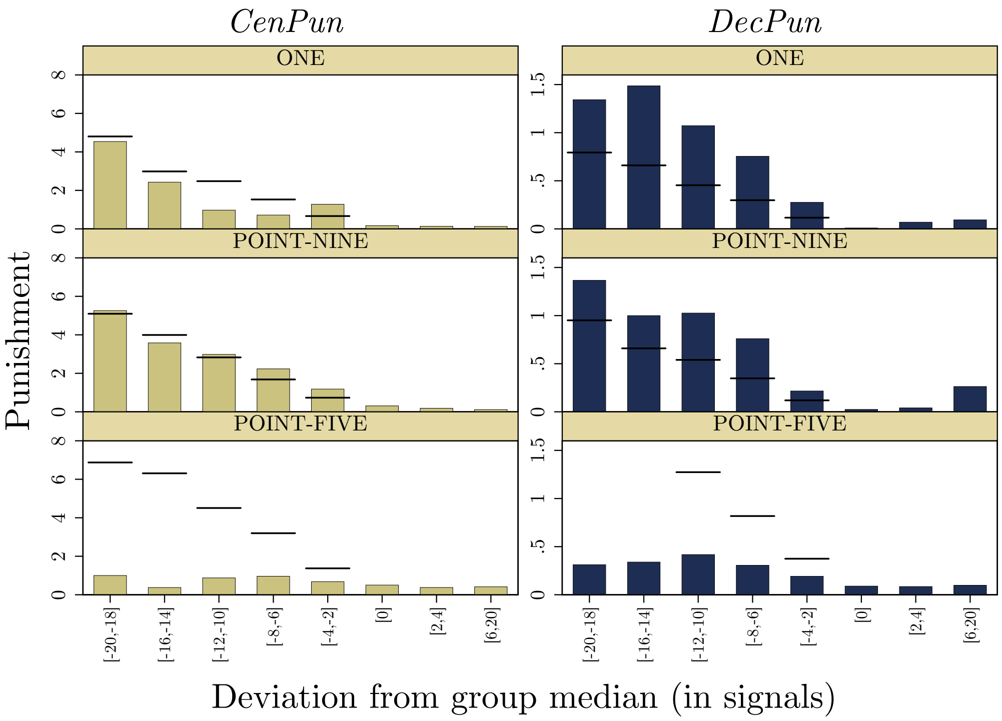


Figure 2: Predicted and actual punishment of authorities in *CenPun* (left panel) and citizens in *DecPun* (right panel) for the three treatments. Bars show average punishment targeted at a citizen dependent on the difference between the citizen's signal and the median signal in the group. Horizontal lines show the average of all deterrent punishments according to the theoretical prediction (values for the two first bars in the bottom right panel are outside the plotted range).

punishment of the citizens in *DecPun*. For the benchmark we assume that each citizen calculates $\bar{g}$ on the basis of the signals she receives, including her own contribution. In addition, for groups with more than two citizens we assume that each citizen punishes other citizens by $\frac{1}{d-1}$ (with d denoting the number of citizens in *DecPun*) of the minimal deterrent punishment according to Equation (7). The right panel of Figure 2 shows the results for *DecPun*. In ONE we observe that punishment for negative deviations is substantially higher than predicted. This holds also for POINT-NINE, where in addition we observe a clear increase in misguided punishment relative to *CenPun*. Finally, POINT-FIVE leads again to punishments far from deterrent and largely invariant in the deviation.[26]

Result 4 *In the treatments* ONE *and* POINT-NINE *average central authorities' punishments match the predicted minimal deterrent punishment patterns surprisingly well. In contrast, decentralized punishment for low signals is substantially higher than the theoretical benchmark. For* POINT-FIVE *punishments are largely invariant to the signal and lower than the deterrent level.*

Despite the fact that citizens face a second order free-rider problem in the punishment stage and have to bear the marginal cost of punishment, it seems that decentralized norm enforcement is *stronger* than centralized norm enforcement. Consequentially, as shown in Figure 1 decentralized punishment institutions tend to achieve higher contributions but lose some of the efficiency gains for stronger punishment. From the estimates shown in Model (2) of Table 1 we learned that free-rider punishment increases the likelihood of choosing *NoPun*, but does not seem to affect the choice between the two punishment institutions. On the other hand, contributor punishment significantly affects the relative popularity of the two punishment institutions. Misguided punishment is (highly) likely to result in contributor punishment in POINT-FIVE (POINT-NINE). Furthermore, when the signal is wrong, then a fraction of the punishments for signals below $\bar{g}$ result in contributor punishment. Taken together these observations suggest a rationale for the shift of the competitive advantage from *DecPun* towards *CenPun* once we move from perfect information to low noise. Central authorities tend to be more moderate in their punishment relative to the citizens in *DecPun*, both for free-rider and misguided punishment. Under noise, this leads to substantial differences of punishment for high contributors. For example, in POINT-NINE average punishment received by a full contributor is 0.48 units in *DecPun* compared to 0.27 *CenPun* ($p = .050$ exact Wilcoxon signed rank test based on matching group averages).[27]

When formulating Conjecture 2 we stressed the role of retaliatory motives for misguided punishment in *DecPun.* Note that our experimental design makes targeted revenge difficult, because citizens do not know who is responsible for the punishment (unless there are only two citizens in *DecPun*). Nevertheless citizens might punish others in response to punishment received

[26] All differences between prediction and data are at least weakly significant in ONE. In POINT-NINE the bars $[-16, -14]$ and $[-12, -10]$ are slightly above 10 percent, while all other comparisons are significant at $p < .031$ (same test as in Footnote 25).

[27] In ONE, we observe also a significant difference ($p = .040$), but in both cases average punishment is very small (0.05 and 0.03) such that the difference presumably does not matter anymore.

in the previous period. To investigate the role of retaliation we ran regressions explaining punishment decisions. In a model controlling for the deviation in signals, time, and group size effects we find a small but significant positive effect of received punishment in the previous round on misguided punishment ($\beta = 0.0052$, $p = .006$, clustered standard errors). If we interact the term with the treatment POINT-NINE we find the main coefficient unchanged and the interaction small and insignificant ($\beta = -0.0007$, $p = .813$). From this we conclude that revenge is a motive for misguided punishment in *DecPun*, but we do not see a strong indication that the effect varies across treatment.

Figure 2 suggests that average levels of misguided punishment are low, in particular in *CenPun*. This conclusion might, however, be premature. Because institutions are endogenous, popular authorities are strongly overrepresented in this analysis. In the next subsection we show that the authorities and the populations of citizens differ in their use of misguided punishment, which is a crucial determinant for the choice of institution under imperfect information.

4.4 Misguided punishment and the popularity of an institution

In the previous sections we provided evidence that contributor punishment is a crucial determinant of entry into and exit out of an institution. At the same time, misguided punishment occurs in *DecPun* and—to a lesser extent—also in *CenPun*. We now investigate the role of misguided punishment for the relative popularity of a punishment institution.

In the following we derive a measure for the relative strength of misguided punishment in *DecPun* and *CenPun*. In particular, we calculate for each population the frequency of punishment targeted at citizens with above average signals. For *CenPun* we use the punishment data from the subject in the role of the authority, while for *DecPun* we calculate the average over all the punishment decisions of citizens in *DecPun*. Populations in which we observe less frequent misguided punishment in *CenPun* than in *DecPun* are classified as populations with a 'good' authority. Conversely, if the authority metes out more misguided punishment than the citizens we speak of a population with a 'bad' authority.[28] Our classification is based on the data of

[28] The results in this analysis are robust with respect to a number of alternative criteria for good and bad authorities. In particular, the result that *CenPun* is clearly the modal institution in POINT-NINE does also hold if we (i) define good authorities by a median split according to misguided punishment in *CenPun* (ignoring punishment in *DecPun*), or (ii) if we define good authority by the average strength of contributor punishment in *CenPun* (instead of the frequency).

phases 1–7 and we explain the institutional choices in the final phase.[29] In ONE and POINT-FIVE this criterion leads to an equal split of the matching groups, while in POINT-NINE we classify 40 percent of the matching groups as populations with a good authority.

Panel A of Figure 3 shows that authorities attract only a small fraction of the citizens in ONE and POINT-NINE when they mete out a lot of misguided punishment relative to the citizens in *DecPun*. Panel B shows that good authorities manage to attract a larger share than bad authorities in all treatments. However, only under imperfect information *CenPun* is clearly the modal choice. Under perfect information not even good authorities are able to gain the support of the majority of the population.[30]

Panels C and D of Figure 3 provide information about the stability of the population in *CenPun* over time. Bars show the fraction of citizens in this institution, divided into incumbents (darker part) and immigrants (lighter part). Incumbents are citizens who were already in *CenPun* in the preceding phase; immigrants are citizens who were previously in *DecPun* or *NoPun*. The graph shows that bad authorities have a high turnover. Most of the time, more than half of their population are immigrants. Populations of good authorities are much more stable, with a large fraction of the citizens remaining in the institution.

Panel D also shows that, unlike in POINT-NINE, good authorities continuously lose support in the second half of the experiment in ONE. Presumably the low differences in misguided punishment between *CenPun* and *DecPun* are not sufficiently important for many citizens to be willing to subordinate themselves to the Leviathan.

Instead of dividing the observations in two groups we can also use the difference between the frequency of misguided punishment in *DecPun* and *CenPun* as a continuous measure of an authority's relative performance in punishing. If we use OLS to regress the share of the population in *CenPun* in the final phase (the middle bars in Figure 3) on this measure we observe a highly significant positive effect for POINT-NINE ($\beta = .721, p = .001$, robust standard errors, group averages as observations), but not for the other two

[29] In ONE we have three matching groups for which the punishment data is missing because the citizens never chose the respective institution. In these cases we use the average of the corresponding figures in the other matching groups in the same treatment as an estimate for misguided punishment. Qualitatively the results are the same if we drop the three matching groups from the sample: *DecPun* remains clearly the modal institution for good (67.5 percent) as well as bad authorities (46.7 percent).

[30] Testing for differences in the institutional choices between good and bad authorities (Panel A and B of Figure 3) results in $F(1.66, 48.1) = 3.07, p = .064$ (Pearson χ^2 statistic with correction for dependence within matching group).

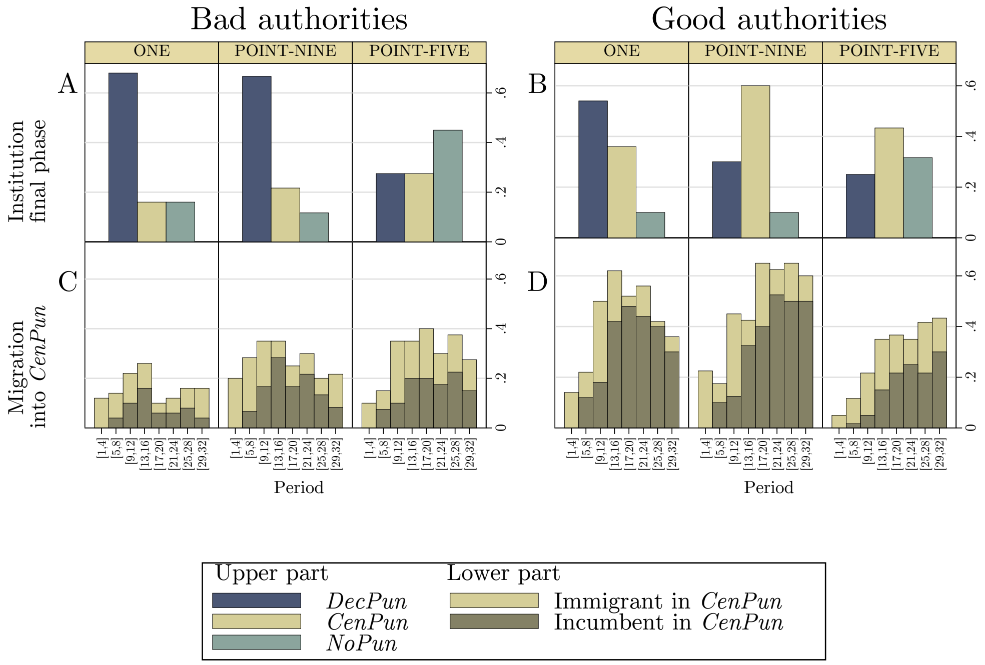


Figure 3: Choice of institution in the final phase of the game for matching groups with bad (panel A) and good authorities (panel B). Bars show the fraction of participants choosing *DecPun, CenPun, NoPun*, separated by the treatments with perfect information ONE and imperfect information POINT-NINE, POINT-FIVE. Panels C and D: Migration patterns in *CenPun*. The dark part of the bars shows the citizens who were in *CenPun* already in the previous phase (incumbents); the light part shows the immigrants.

treatments.

Result 5 *Authorities who assign less misguided punishment than the citizens attract a larger share of the population in all three treatments. In the two treatments with imperfect information good authorities attract more citizens than the other two institutions, while in* ONE *decentralized punishment remains the modal institution.*

5 Discussion

Our study analyzes information conditions which lead subjects to voluntarily subordinate themselves to a central authority. We vary the accuracy of the information concerning the contributions of the other participants in the group. We show that an environment with perfect information tends to favor the decentralized punishment institution, while a high level of noise favors an institution with no punishment. Under low levels of noise we observe the highest support for centralized punishment.

In line with the literature on the importance of antisocial punishment we observe that the punishment of cooperative subjects plays a crucial role for the popularity of an institution. Institutions that punish cooperative citizens tend to lose citizens in favor of other institutions. In the treatments ONE and POINT-NINE we observe that punishment is stronger under decentralized punishment than under centralized punishment, which is remarkable given that centralized punishers do not have to bear the marginal cost of punishment, while decentralized punishers do. In ONE, citizens can easily avoid punishment by contributing (nearly) fully. In POINT-NINE, however, a fraction of the punishments targeted at citizens with low signals ends up with high contributors. While this holds for both punishment institutions equally, the fact that punishment is stronger in the decentralized institution leads to more contributor punishment in this institution. In addition we observe higher levels of punishment targeted at citizens with a high signal in decentralized punishment. These two observations explain why decentralized punishment loses support when we introduce imperfect information.

Taking into account differences between the central authorities we show that the interaction between imperfect monitoring and the availability of a central authority who prevents punishment of cooperative citizens boosts the choice of centralized punishment institutions. This is not the case for the treatment with perfect information. In this treatment there is very little punishment of cooperative citizens and citizens prefer not to delegate their punishment power.

According to the data reported in Herrmann et al. (2008) we conducted the experiments in a society with very low levels of antisocial punishment. In a subject pool with higher levels of antisocial punishment centralizing punishment might be more attractive, even under perfect information. Furthermore, the Leviathans in our experimental design have full discretion in the use of their power: There are no legal constraints, no noblesse oblige which limit the actions of authorities. Real world authorities typically face institutional and moral constraints. Moreover, there are arguably better selection mechanisms for authorities in place. Presumably, these societies come close to the outcome of good authorities in POINT-NINE, where *CenPun* is clearly the dominant institution. Therefore, we might underestimate the attractiveness of centralized punishment in our experiment.

We consider our treatment variations as prototypical for various epochs of the evolution of social structures in humans. Early societies allowing for nearly perfect observation of others tend to apply decentralized punishment regimes. In maturing societies with increasing agglomeration and complexity, it becomes difficult to monitor others' behavior. These are the circumstances, in which people are willing to sacrifice some of their autonomy and delegate the sanctioning power to a Leviathan. In times of social unrest and destabilized law enforcement systems, however, punishment by authorities becomes more erratic. Under these circumstances centralized sanctions lose their competitive advantage and, if possible, citizens migrate to other institutional arrangements.

Recently, the appearance of new media like social networks and mobile communication technologies give rise to another interesting development, as they increase transparency of actions among group members. As a consequence, we might expect a decentralization of the societal structures. The latest developments on the administration of mass protests during the Arab Spring via social networks are an example for this development (Hussain & Howard, 2013). Whether this is a first indication for a general shift towards more decentralized organizational structures is too early to tell.

References

Ambrus, A. & Greiner, B. (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review*, *102*(7), 3317–3332. doi:10.1257/aer.102.7.3317

Ambrus, A. & Pathak, P. A. (2011). Cooperation over finite horizons: A theory and experiments. *Journal of Public Economics*, *95*(7-8), 500–512. doi:10.1016/j.jpubeco.2010.11.016

Andreoni, J. & Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, *96*(11-12), 1036–1046. doi:10.1016/j.jpubeco.2012.08.003

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *60*(1), 11–26. doi:10.1016/j.jebo.2003.06.006

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*(1), 31–51. doi:10.1016/j.geb.2006.08.011

Carpenter, J. P. & Matthews, P. H. (2012). Norm enforcement: Anger, indignation, or reciprocity? *Journal of the European Economic Association*, *10*(3), 555–572. doi:10.1111/j.1542-4774.2011.01059.x

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. doi:10.1007/s10683-010-9257-1

Corazzini, L., Kube, S., Maréchal, M. A., & Nicolò, A. (2014). Elections and deceptions: An experimental study on the behavioral effects of democracy. *American Journal of Political Science*, *58*(3), 579–592. doi:10.1111/ajps.12078

Dal Bó, P., Foster, A., & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, *100*(5), 2205–2229. doi:10.1257/aer.100.5.2205

De Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258. doi:10.1126/science.1100735

Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, *33*(1), 145–167. doi:10.1007/s00199-007-0212-0

Dufwenberg, M. & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*, 268–298. doi:10.1016/j.geb.2003.06.003

Ertan, A., Page, T., & Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, *53*(5), 495–511. doi:10.1016/j.euroecorev.2008.09.007

Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*, 293–315. doi:10.1016/j.geb.2005.03.001

Fehr, E. & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. doi:10.1257/aer.90.4.980

Fehr, E., Herz, H., & Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, *103*(4), 1325–1359. doi:10.1257/aer.103.4.1325

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868. doi:10.1162/003355399556151

Fehr, E. & Williams, T. (2013). *Endogenous emergence of institutions to sustain cooperation.*

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. doi:10.1007/s10683-006-9159-4

Gächter, S., Herrmann, B., & Thöni, C. (2005). Cross-cultural differences in norm enforcement. *Behavioral and Brain Sciences*, *28*(6), 822–823. doi:10.1017/S0140525X05290143

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, *322*(5907), 2008. doi:10.1126/science.1164744

Grechenig, K., Nicklisch, A., & Thöni, C. (2010). Punishment despite reasonable doubt—a public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies*, *7*(4), 847–867. doi:10.1111/j.1740-1461.2010.01197.x

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. doi:10.1007/s40881-015-0004-4

Gross, J., Méder, Z. Z., Okamoto-Barth, S., & Riedl, A. (2016). Building the Leviathan – voluntary centralisation of punishment power sustains cooperation in humans. *Scientific Reports*, *6*, 20767. doi:10.1038/srep20767

Grosse, S., Putterman, L., & Rockenbach, B. (2011). Monitoring in teams: Using laboratory experiments to study a theory of the firm. *Journal of the European Economic Association*, *9*(4), 785–816. doi:10.1111/j.1542-4774.2011.01026.x

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*(5770), 108–111. doi:10.1126/science.1123633

Hamman, J. R., Weber, R. A., & Woon, J. (2011). An experimental investigation of electoral delegation and the provision of public goods. *American Journal of Political Science*, *55*(4), 738–752. doi:10.1111/j.1540-5907.2011.00531.x

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362–1367. doi:10.1126/science.1153808

Hirschman, A. O. (1970). *Exit, voice and loyalty. Responses to decline in firms, organizations and states.* Cambridge MA: Harvard University Press.

Hirschman, A. O. (1978). Exit, voice, and the state. *World Politics*, *31*(1), 90–107. doi:10.2307/2009968

Hobbes, T. (1651). *The Leviathan.* Oxford World's Classics Series. Oxford University Press, Incorporated, 1996.

Hussain, M. M. & Howard, P. N. (2013). What best explains successful protest cascades? ICTs and the fuzzy causes of the Arab Spring. *International Studies Review*, *15*(1), 48–66. doi:10.1111/misr.12020

Kamei, K. & Putterman, L. (2015). In broad daylight: Fuller information and higher-order punishment opportunities can promote cooperation. *Journal of Economic Behavior & Organization*, *120*, 145–159. doi:10.1016/j.jebo.2015.09.020

Kosfeld, M., Okada, A., & Riedl, A. (2009). Institution formation in public goods games. *American Economic Review*, *99*(4), 1335–1355. doi:10.1257/aer.99.4.1335

Kube, S. & Traxler, C. (2011). The interaction of legal and social norm enforcement. *Journal of Public Economic Theory*, *13*(2006), 639–660. doi:10.1111/j.1467-9779.2011.01515.x

Leibbrandt, A., Ramalingam, A., Sääksvuori, L., & Walker, J. M. (2015). Incomplete punishment networks in public goods games: Experimental evidence. *Experimental Economics*, *18*(1), 15–37. doi:10.1007/s10683-014-9402-3

Markussen, T., Putterman, L., & Tyran, J.-R. (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies*, *81*(1), 301–324. doi:10.1093/restud/rdt022

Nicklisch, A., Grechenig, K., & Thöni, C. (2015). *Information-sensitive Leviathans: The emergence of centralized punishment*, WiSo-HH Working Paper Series, Working Paper No. 24.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*, 91–112. doi:10.1016/j.jpubeco.2007.04.008

Nikiforakis, N., Noussair, C. N., & Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, *96*(9-10), 797–807. doi:10.1016/j.jpubeco.2012.05.014

Nosenzo, D., Quercia, S., & Sefton, M. (2015). Cooperation in small groups: The effect of group size. *Experimental Economics*, *18*(1), 4–14. doi:10.1007/s10683-013-9382-8

O'Gorman, R., Henrich, J., & Van Vugt, M. (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B*, *276*(1655), 323–329. doi:10.1098/rspb.2008.1082

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, *86*(2), 404–417. doi:10.2307/1964229

Putterman, L., Tyran, J.-R., & Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics*, *95*(9-10), 1213–1222. doi:10.1016/j.jpubeco.2011.05.001

Rao, J. N. K. & Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, *12*(1), 46–60. doi:10.1214/aos/1176346391

Rockenbach, B. & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature*, *444*(7120), 718–723. doi:10.1038/nature05229

Roux, C. & Thöni, C. (2015). Collusion among many firms: The disciplinary power of targeted punishment. *Journal of Economic Behavior & Organization*, *116*, 83–93. doi:10.1016/j.jebo.2015.03.018

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, *77*(4), 1540–1566. doi:10.1111/j.1467-937X.2010.00608.x

Thöni, C. (2014). Inequality aversion and antisocial punishment. *Theory and Decision*, *76*(4), 529–545. doi:10.1007/s11238-013-9382-3

Tyran, J.-R. & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, *108*(1), 135–156. doi:10.1111/j.1467-9442.2006.00444.x

Online Appendix

Information-sensitive Leviathans

Andreas Nicklisch, Kristoffel Grechenig, Christian Thöni

A.1 Mathematical Appendix

In the Appendix, we show for ONE and POINT-NINE that there exists an equilibrium where all citizens choose *CenPun* and cooperate fully. That is,

Proposition 1 *For $\lambda > .58$, all citizens play the same equilibrium strategies: (i) all choose CenPun in Stage 1, (ii) all contribute $g_i = 20$ in CenPun, and 0 otherwise in Stage 2, (iii) $p_{i \to k} = 0 \; \forall \; k \in \boldsymbol{D}$.*

In case of POINT-FIVE, the authorities lack the resources to enforce full contributions of all ten citizens. However, there exists an equilibrium in which only two citizens choose *CenPun* and contribute fully:

Proposition 2 *For $\lambda \leq .58$: (i) two random citizens choose CenPun, all other choose NoPun or DecPun in Stage 1, (ii) citizens contribute $g_i = 0$ if in DecPun or NoPun, in CenPun they contribute $g_i = 20$ if $c = 2$, $g_i = 18$ if $c \leq 4$, and $g_i = 16$ else, (iii) $p_{i \to k} = 0 \; \forall \; k \in \boldsymbol{D}$.*

To proof both propositions, we use backwards induction and start by analyzing the punishment stage.[31] The two simple cases are *NoPun* and *DecPun*, since deterrent punishment is not feasible in *NoPun* and *DecPun*:

Corollary 1 *There is no punishment in NoPun and DecPun under standard assumptions.*

Proof: In the first there is no punishment by design, in the second punishment is costly to the punisher, which means that all strategy profiles including punishment acts are not subgame perfect. This proofs the corollary.

In *CenPun* things are more interesting, since the authority does not bear the cost of punishment. Consequently, the entire set of possible punishment strategies can be part of a subgame-perfect Nash equilibrium in *CenPun.* We start by investigating punishment strategies for the authority in *CenPun*, and derive a function $p^*(s, \bar{g}, \lambda)$, indicating the amount of punishment necessary to deter a player from deviating. As the authority's payoff is increasing in

[31] Along the proofs for the proposition, we specify the equilibrium strategy of the authority below.

its citizens contributions, we are particularly interested in equilibria which yield the maximum contributions. That is, we look for punishment strategies which resolve the social dilemma character of the public goods game in stage two and make it individually rational for the citizens to contribute.

If deterrent punishment is feasible and if it does not require too much contributor punishment it prevents unilateral deviation from contributing a certain level $\bar{g}$ $(20 \geqslant \bar{g} > 0)$. Then the game has an equilibrium in which all citizens choose *CenPun* and contribute $\bar{g}$. How could a deterrent punishment strategy look like? If the $c-1$ other citizens contribute $\bar{g}$, citizen i's profit before punishment is

$$\hat{\pi}_i(g_i) = 20 - g_i + \frac{1.6}{c}\Big[(c-1)\bar{g} + g_i\Big]. \tag{8}$$

Taking the derivative with regard to g_i leads to the marginal disutility of contributing, $\frac{1.6-c}{c}$, which is increasing (in absolute terms) in the number of citizens in *CenPun*. To be deterrent the authority must ensure that the payoff gains of $g_i < \bar{g}$ are set off by an equivalent or larger payoff reduction through punishment. In the following, we focus our attention to the least expensive punishment strategy which exactly matches the profit from every deviation $g_i < \bar{g}$ in expectation. Let $p(s)$ be the authority's punishment function, mapping signals into punishment for citizen i (we omit the subscript, because the punishment strategy is the same for all citizens). We derived $p^*(s)$ under the assumption that each citizen is punished only dependent on his own signal:[32]

Proposition 3

$$p^*(s) = \left[\frac{(10c-16)(\bar{g}-s)}{3c(11\lambda-1)}\right]_{[0}. \tag{9}$$

Proof: If there is no uncertainty $(\lambda = 1)$, then a simple linear punishment with the slope $p' = \frac{1.6-c}{3c}$ for all $s_i < 20$ and $p(20) = 0$ suffices to induce full cooperation. With imperfect signals punishment inevitably leads to punishment of cooperative subjects. The value of λ determines the informational value of the signal. For $\lambda = \frac{1}{11}$ the signal contains no information about the

[32] Alternatively, the authority could adopt even more complicated punishment strategies $p(\boldsymbol{s})$, where $\boldsymbol{s} = (s_i, s_j, \ldots)$ is a vector of all signals observed. We also derived punishment strategies for the rules (i) punish only the citizen(s) with the lowest signal(s) in $\boldsymbol{s}$, (ii) punish the lowest signal only if unique, and (iii) punish if and only if there is a single signal lower than $\bar{g}$. The expected expenditures for disciplining a fully cooperative group are identical to the case of $p^*(s)$ for all punishment strategies (i), (ii) and (iii). The notation $[a]_{[0}$ is equivalent to $\max\{0, a\}$.

contribution, which renders deterrent punishment impossible. In the following we restrict our attention to signals with a $\lambda \in \left(\frac{1}{11}, 1\right]$. With such signals the best guess about the true contribution is the signal. Signals of $\bar{g}$ are taken as indication of cooperative behavior and are not punished. Signals above $\bar{g}$ are also not punished. The condition for the least costly punishment function is

$$\begin{aligned} &\hat{\pi}_i(\bar{g}) - 3(1-\lambda)\tfrac{1}{10}\sum_{s_i \in \boldsymbol{S}} p(s_i) \\ &\qquad = \hat{\pi}_i(g_i) - 3\lambda p(g_i) - 3(1-\lambda)\tfrac{1}{10}\left[\sum_{s_i \in \boldsymbol{S}} p(s_i) - p(g_i)\right], \quad (10) \end{aligned}$$

where the left-hand side shows expected payoff of contributing $\bar{g}$, consisting of the stage two payoff minus three times the expected contributor punishment points. Contributor punishment occurs with probability $(1-\lambda)$ and consists of the expected punishment for all possible wrong signals, where $\boldsymbol{S} = \{0, 2, 4, \ldots, 20\}$ is the set of all signals. The right-hand side shows the expected utility for any contribution $g_i < \bar{g}$, consisting of the deviation payoff from stage two minus the 'correct' punishment, as well as the punishment triggered by false signals. Here, we have to subtract the punishment for the true contribution g_i from the sum (this term is zero on the left-hand side). Rearranging leads to

$$\hat{\pi}_i(g_i) - \hat{\pi}_i(\bar{g}) = 3\lambda p(g_i) - \frac{3}{10}(1-\lambda)p(g_i), \qquad (11)$$

where the left-hand side shows the increase in stage two profits from deviating and the right-hand side shows the increase in expected punishment from deviating. The latter consist of the punishment based on the true signal reduced by the decrease in punishment due to false signals. In case of perfect signals the latter would be zero, in case of uninformative signals ($\lambda = \frac{1}{11}$) the right-hand side equals zero, which confirms our statement above that deterrence is impossible under these circumstances. Using equation (8), we can solve equation (11) for the punishment function dependent on the signal and yield equation (9). This proofs the proposition.

Having shown that deterrent punishment is possible for $\lambda > \frac{1}{11}$ raises the question of its feasibility. Given our design of the punishment mechanism the authority faces two 'incentive compatibility constraints'. The first one (IC_t) is due to the restriction on total punishment, the second one (IC_i) by the restriction on individual punishment.

Lemma 1 *(i) Given the parameters of our experiment, for $\lambda > .58$ neither IC_t nor IC_i are binding; fully cooperative outcomes can be enforced by the authority. (ii) For $\lambda = .5$ as implemented in the experiment, IC_t or IC_i are binding for $c > 2$ and $\bar{g} = 20$; fully cooperative outcomes can be enforced for $c = 2$. (iii) For $\lambda = .5$ and $\bar{g} = 18$, IC_t or IC_i are binding for $c = 3$ and $c > 4$; $\bar{g} = 18$ can be enforced for $c = 2$ and $c = 4$. (iv) For $\lambda = .5$ and $\bar{g} = 16$ (or lower) neither IC_t nor IC_i are binding; $\bar{g} \leq 16$ can be enforced for all numbers of citizens.*

Proof: First, let us derive IC_t. For this, we calculate the expected punishment expenditures necessary to deter a group of citizens with one deviator. We take the case of the most expensive deviation, which is a contribution of zero. We relate the expected expenditures to the authorities budget constraint, which is $3c$:

$$\lambda p^*(0) + (1-\lambda)\tfrac{1}{10} \sum_{s \in \boldsymbol{S} \setminus \{0\}} p^*(s) + (c-1)\,(1-\lambda)\tfrac{1}{10} \sum_{s \in \boldsymbol{S}} p^*(s) \leqslant 3c. \qquad (12)$$

The first two elements of the left-hand side refer to the expected punishment for the free-rider, followed by the expected punishment for the remaining citizens who contribute $\bar{g}$. This expression allows us to find the enforceable contribution levels dependent on the number of citizens and the noise in the signals.

Second, to derive IC_i recall that maximum punishment imposed on a single citizen is $3(c-1)$. Depending on $\bar{g}$, λ, and c there are situations in which this constraint does not allow for the punishment necessary to deter free riding, that is, $p^*(0) > 3(c-1)$.

Figure A1 shows the numerical results for these two conditions. All lines indicate combinations of λ and c for which one of the conditions holds with equality and ruling out the cases to the left of the line. Solid (long dashed, short dashed) lines indicate the incentive compatibility constraints for $\bar{g} = 20, (18, 16)$. The monotonically increasing lines refer to the constraint on total punishment, whereas the mostly decreasing lines indicate the constraints on maximum punishment for a single citizen. Bold lines indicate the envelope of all constraints.[33]

[33] For the more complicated punishment strategies $p(\boldsymbol{s})$ discussed in footnote 32 the expected punishment costs as calculated in equation 12 tend to be smaller, thus relaxing IC_t. However, this does not open up more equilibria for $\lambda = .5$, because IC_i is violated for all cases with $c > 2$. Intuitively, these strategies use punishment less often, but require stronger punishment when applied.

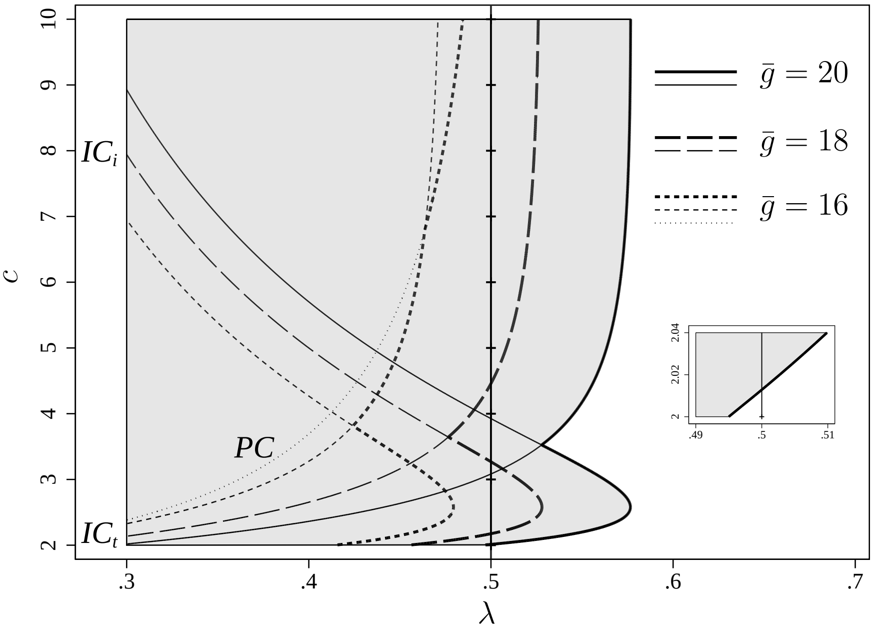


Figure A1: Feasibility of cooperative outcomes in *CenPun*, dependent on the quality of the signals λ and the number of citizens c. Positively sloped dashed or solid lines indicate the incentive compatibility with respect to total punishment (IC_t), mostly downwards sloped lines show the incentive compatibility for individual punishment (IC_i). In both cases the area to the right is feasible. The dotted line indicates the participation constraint for $\bar{g} = 16$ (not binding for higher $\bar{g}$). Bold lines indicate the envelope of all constraints. The small subfigure shows the region around $\lambda = .5$ for small groups enlarged.

For $\lambda > .58$ none of the constraints are binding and fully cooperative outcomes can be enforced by the authority. The grey area indicates constellations for which a fully cooperative outcome cannot be enforced. For $\lambda = .5$ (which we implemented in the experiment) the constraint on total punishment is binding in case of $c > 3$, while $c = 3$ is ruled out by the constraint IC_i, leaving $c = 2$ as the only possibility (see small subgraph enlarging the area). Going for $\bar{g} = 18$ relaxes the constraints (long dashed line) such that, in addition to $c = 2$, $c = 4$ becomes feasible. Enforcing $\bar{g} = 16$ (or lower) is feasible for all numbers of citizens. This completes the Lemma.

Note that in addition to the two incentive compatibility constraints (IC_t, IC_i), we also have to satisfy a participation constraint, ensuring that the punishment costs of a cooperative group do not surpass the efficiency gains created by contributing $\bar{g}$ instead of zero in another institution. Similar to the expression in equation (12) we calculate the expected punishment costs for a fully cooperative group of citizens. Different from before, we have to take into account that the income reduction is not only due to received punishment, but also due to the financing of the punishment of others in the group, that is, we have to ensure that

$$4(1-\lambda)\tfrac{1}{10}\sum_{s\in\boldsymbol{S}} p^*(s) \leqslant \tfrac{3}{5}\bar{g}. \tag{13}$$

Dotted lines indicate the participation constraints. It turns out that the participation constraint is only binding for $\lambda < .5$, $\bar{g} = 16$ and for $c \geqslant 7$, which is not considered in the experiment. We added the constrain nonetheless to complete the discussion.

The Lemma along the earlier Corollary allows us to formulate a set of Nash equilibrium strategies for the three levels of λ used in our experiment. For $\lambda = .9$ and 1 there exists an equilibrium in which all citizens choose *CenPun* and contribute fully. Equilibrium strategies are as follows: The authority punishes all citizens in $\boldsymbol{C}$ based on the signals according to $p^*(s)$ as defined in equation (7). All citizens play the same equilibrium strategies:

- Stage 3: No punishment in *DecPun*: $p_{i\to k} = 0\ \forall\ k \in \boldsymbol{D}$
- Stage 2: Contribute $g_i = 0$ if in *DecPun* or *NoPun*, contribute $g_i = 20$ if in *CenPun*
- Stage 1: Choose *CenPun*

This completes the proof of Proposition 1.

For $\lambda = .5$ this equilibrium is not feasible. The authority's preferred outcome would be to attract only two citizens, because then the fully cooperative outcome is enforceable. To do so the authority might play a punishment strategy with punishment according to $p^*(s)$ in case of $c = 2$ and punishment independent of the signal otherwise. The best response of the citizens would be two out of ten entering, while the others choose a different institution. The problem is that this punishment strategy is not subgame perfect. If, for example, four citizens choose *CenPun* the authority would prefer to implement a punishment inducing contributions of 18. Thus, for any punishment strategy it must hold that it enforces the highest level of contributions given c. Despite this additional condition the authority can reach maximum payoff in equilibrium by playing the following strategy in $\lambda = .5$:

- Punish according to $p^*(s)$ for the highest $\bar{g}$ feasible given c.

- In addition, use the remaining punishment points to punish all citizens in $\boldsymbol{C}$ except citizen i and j by equal amounts, independent of the signal.

This strategy ensures that the participation constraint is only met for citizen i and j if deterrent punishment leaves enough resources to punish citizens other than i and j. For our parameters this is the case. For instance, in case of $c = 4$ enforcing $\bar{g} = 18$ requires 8 out of 12 punishment points and the citizens earn an expected payoff of 25.8 when punished according to $p^*(s)$. Using the remaining 4 punishment points to reduce the income of two citizens reduces their expected income by 6 units each, making them worse off than the outside option of 23. Consequently, the citizens' strategies for $\lambda = .5$ are

- Stage 3: No punishment in *DecPun*: $p_{i \to k} = 0 \; \forall \; k \in \boldsymbol{D}$
- Stage 2: Contribute $g_i = 0$ if in *DecPun* or *NoPun*. In *CenPun* contribute the highest enforceable contribution given c, that is, $g_i = 20$ if $c = 2$, $g_i = 18$ if $c \leq 4$, and $g_i = 16$ else.
- Stage 1: Citizen i and j choose *CenPun*. All other choose *NoPun*.

This completes the proof of Proposition 2.

A.2 Experimental instructions

This section includes a translation of the instructions handed out on paper (original instructions were in German). The instructions are identical for all treatments and all roles (citizen, authority), with exception of the description of signal accuracy which we put in brackets.

General Instructions for Participants

You are about to take part in an economic experiment. If you read the following instructions carefully, you can earn a substantial amount of money, depending on the decisions you make. It is therefore very important that you read these instructions carefully.

The instructions you have received from us serve your own private information only. **During the experiment, any communication whatsoever is forbidden**. If you have any questions, please ask us. Disobeying this rule will lead to exclusion from the experiment and from any payments.

During the experiment, we do not speak of Euro, but of Taler. Your entire income is hence initially calculated in Taler. The total number of Taler you earn during the experiment is converted into Euro at the end, at the rate of

75 Taler = 1 Euro.

At the end of the experiment, you will be paid **in cash** the amount of Taler you have earned during the experiment, in addition to 4 Euro for taking part in the experiment.

The experiment is divided into different rounds. In each round, you will be given an identification number, so that your decisions in the course of a round can be attributed to you. Please note that, after each round, the identification number allocated to you and the other members of your group changes randomly. Group members therefore cannot be identified beyond the rounds. All decisions are made **anonymously**, i.e., none of the other participants is told the identity of a person who made a particular decision. The payoff is also anonymous, i.e., no participant is told how high another participant's payoff is.

The exact procedure of the experiment is described on the following pages.

Information about the Exact Procedure of the Experiment

General Information

At the beginning of the experiment, you are randomly assigned to one of two halves, each of which has 11 participants. During the entire experiment, you interact only with participants from your half. At the beginning, one of the 11 participants is chosen at random for the entire duration of the experiment, receiving a different task from the one which the other participants are assigned to

Procedure

The experiment consists of 32 rounds. At the very beginning and, from then on, every four rounds (i.e., in rounds 1, 5, 9, 13,. . . , 29), you may choose a group. There are three different groups: A, B, and C. Each of the 32 rounds consists of 2 stages. In the first stage, you choose a contribution to the joint project. In the second stage, you can influence the income of the other participants in your group by means of deduction points. Groups A, B, and C differ in this regard.

Group Choice

Every participant chooses a group:

	Influencing the Income of the other Group Members:
Group	A: No points deducted B: Mutual point deduction C: Point deduction by the extra participant

You will find more details below about the way participants can influence the income.

Stage 1: Contribution to the Joint Project

In each round, you receive an endowment of 20 Taler. It is up to you to decide how many of the 20 Taler you wish to contribute to the joint project. All even numbers are possible contributions, i.e., $0, 2, 4, 6, \ldots, 18, 20$. All other participants in your group make the same decisions simultaneously. After this, the incomes from Stage 1 are calculated:

Your **income from Stage 1** is:
20 $-$ your contribution to the joint project
$+$ 1.6 $\times$ the average contribution

You therefore keep all Taler that you have not contributed to the project. In addition, you receive 1.6 times the average of the contributions from all group members (the average of the contributions is the sum of the contributions from all group members to the project, divided by the number of group members).

The income from the joint project is calculated by this formula for **all group members**.

Please note: Each group member receives the same income from the project, regardless of how much he or she has paid in, i.e., each group member profits from all contributions to the joint project.

Stage 2: Deduction points

(i) General Information

In Stage 2, all other participants in your group (A, B, or C) and the extra participant (if you are in group C) are told their contribution (henceforth referred to as the signal). [This signal is correct with a probability of 50% (90%). In other words, in 5 (9) out of 10 cases, the figure that the other participants see in your group corresponds to your actual contribution. In the remaining 5 (9) out of 10 cases, the other participants see a random other number that does not correspond to your contribution (here, all numbers can appear with equal probability).] You also receive a signal for each of the other members of your group, as well as for their contributions. [This information is also correct with a probability of 50% (90%).] In addition, you receive 3 extra Taler in Stage 2 of each round.

(ii) Groups

Group A: No Deduction Points

If you have chosen Group A, then you cannot take any action during this stage.

Your **income from Stage 2** is therefore:
3

Group B: Mutual Point Deduction

If you have chosen Group B, then you may **reduce** or **leave unchanged** the income of the other members of your group. You must decide how many of the **3 Taler** you wish to spend on **distributing** deduction points to other group members. Every deduction point that you give to another group member reduces this member's income by **3 Taler**. (Similarly, your own income is reduced by 3 Taler per deduction point distributed by another group member to you.) At the same time, every deduction point distributed by you to others costs you **1 Taler**. You **keep** the remaining Taler.

Your **income from Stage 2** is therefore:
3 − the sum of the deduction points you distribute to other group members in Group B
− 3 × the sum of the deduction points you receive from other participants in Group B

Group C: Point Deduction by the Extra Participant

In Group C, the **extra participant**, rather than the group members, decides on the distribution of deduction points (see the passage "Extra Participant (Group C)"). The extra participant also receives the information on the contribution decisions of the Group C participants. [This information, too, has a 50% (90%) likelihood of being correct.] If the extra participant gives you deduction points, then your income is reduced by **3 Taler**. The cost of deduction points that the extra participant gives to another Group C participant must be evenly divided among all other Group C participants. For instance, if 5 participants are in Group C and the extra participant gives one participant 2 deduction points, then the remaining four participants each have to shoulder the cost of 0.5 Taler.

Your **income from Stage 2** is therefore:
3 – (the sum of the deduction points from the extra participant to others) / (Number of participants in Group C - 1)
– 3 × the sum of the deduction points you receive

Extra Participant (Group C)

Should you have become the extra participant, the following refers to you. Unlike the other participants, you do not decide between the groups, and you cannot choose a contribution to the joint project either. However, like the other participants, you receive a signal about each Group C player's contribution. [This signal is correct with a probability of 50% (90%). In other words, in 5 (9) out of 10 cases, the figure corresponds to the actual contribution of the respective group members. In the remaining 5 (9) out of 10 cases, you see a random other number that does not correspond to your contribution (here, all numbers can appear with equal probability).]

Your task is to choose the deduction points for Group C. You may give each individual participant in Group C separate deduction points. In total, you can distribute a maximum number of deduction points that corresponds to three times the number of Group C participants.

Your income is determined by the mean income of Group C participants in Stage 1 (i.e., prior to the income reduction caused by deduction points). The higher the contributions in Group C are, the higher your income is as an extra participant.

Your **total income** in this round is therefore:
Average income of Group C participants in Stage 1

Special Case: Only Group Member

Should you be the **only member** in a group (in Group C, apart from the extra participant), you receive 20 Taler in Stage 1 and no Taler in Stage 2, i.e., your income for the round is 20 Taler. You have no possibility to take action, neither at the first nor at the second stage. If you are an extra participant or if there are zero participants or only one participant in Group C, you also receive **20 Taler** and have no possibility to take action.

Information at the End of the Round

At the end of the round, you receive a detailed overview of the results of your group. Each group member is told the own contribution to the project, the income from Stage 1, deduction points distributed (if possible), deduction points received (if possible), the income from Stage 2, and the income from the round. Every four rounds, you may choose whether you would like to be in Group A, B, or C for the next four rounds. For this decision, you are given an overview of the average round incomes of the last four rounds in Groups A, B, and C.

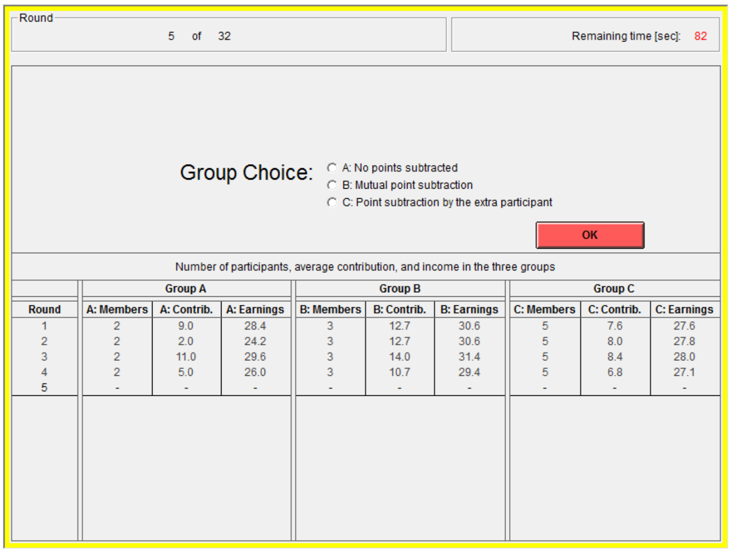
Round Income and Total Income

Your income from Stage 1 plus your income from Stage 2 taken together generate your income in each round. The total income from the experiment is calculated by adding the incomes from all 32 rounds.

Is anything unclear? Please contact someone in charge of heading the experiment!

A.3 On screen parts

After reading the instructions, the subjects had to solve six control questions on screen. The questions included hypothetical combinations of contribution and punishment decisions and the participants had to calculate the resulting payoffs. After all participants completed the control questions the experiment began. Only then did participants learn whether they were assigned the role of a citizen or an authority. Figure A2 shows the screen for the institution choice in period 5. Each subject is informed about the number of subjects, average contributions and payoffs in all three groups at the time of the decision about the institution for the next phase of four periods. In the punishment stage authorities were presented with the contributions of the participants in their group and had to choose deduction points. Figure A3 shows an example of a screen for stage 2. The screen for the citizens in *DecPun* looks alike with the exception that one of the rows contains the subjects own contribution and does not take an input on deduction points. Likewise, the screen for the citizens in *CenPun* as well as in *NoPun* contains the identical information about the contributions but does not take input.



Round 5 of 32

Remaining time [sec]: 82

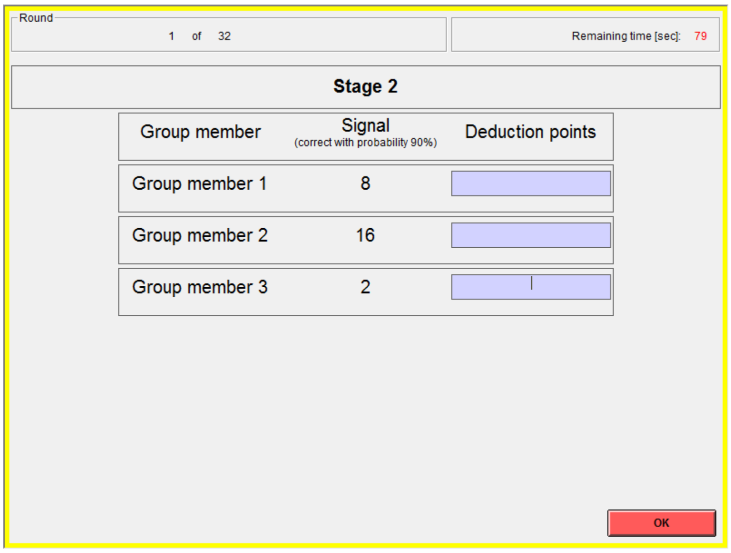
Group Choice:

- ○ A: No points subtracted
- ○ B: Mutual point subtraction
- ○ C: Point subtraction by the extra participant

Number of participants, average contribution, and income in the three groups

	Group A			Group B			Group C		
Round	A: Members	A: Contrib.	A: Earnings	B: Members	B: Contrib.	B: Earnings	C: Members	C: Contrib.	C: Earnings
1	2	9.0	28.4	3	12.7	30.6	5	7.6	27.6
2	2	2.0	24.2	3	12.7	30.6	5	8.0	27.8
3	2	11.0	29.6	3	14.0	31.4	5	8.4	28.0
4	2	5.0	26.0	3	10.7	29.4	5	6.8	27.1
5	-	-	-	-	-	-	-	-	-

Figure A2: Screen of institution choice stage after the first phase of four periods (only for citizens).



Round 1 of 32

Remaining time [sec]: 79

Stage 2

Group member	Signal (correct with probability 90%)	Deduction points
Group member 1	8	
Group member 2	16	
Group member 3	2	

Figure A3: Screen of punishment stage for a central authority with three citizens in POINT-NINE.

DFG Research Group 2104
– Latest Contributions

2016:

Paetzel, Fabian and Sausgruber, Rupert: Entitlements and loyalty in groups: An experimental study. Working Paper Nr. 2016-03. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2016-03.pdf

Nicklisch, Andreas, Grechenig, Kristoffel and Thöni, Christian: Information-sensitive Leviathans. Working Paper Nr. 2016-02. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2016-02.pdf

Greiff, Matthias and Paetzel, Fabian: Less sensitive reputation spurs cooperation: An experiment on noisy reputation systems. Working Paper Nr. 2016-01. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2016-01.pdf

2015:

Schramme, Thomas: The metric and the threshold problem for theories of health justice: A comment on Venkatapuram. Working Paper Nr. 2015-05. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2015-05.pdf

Nicklisch, Andreas, Grechenig, Kristoffel and Thöni, Christian: Information-sensitive Leviathans – the emergence of centralized punishment. Working Paper Nr. 2015-04. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2015-04.pdf

Schramme, Thomas: Setting limits to public health efforts and the healthisation of society. Working Paper Nr. 2015-03. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2015-03.pdf

Hinz, Jana and Nicklisch, Andreas: Reciprocity Models revisited: Intention factors and reference values. Working Paper Nr. 2015-02. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2015-02.pdf

Köke, Sonja, Lange, Andreas and Nicklisch, Andreas: Adversity is a school of wisdom: Experimental evidence on cooperative protection against stochastic losses. Working Paper Nr. 2015-01. http://bedarfsgerechtigkeit.hsu-hh.de/dropbox/wp/2015-01.pdf

FOR
2104

DFG Research Group 2104 at Helmut Schmidt University Hamburg
http://needs-based-justice.hsu-hh.de