Institution design in social dilemmas: How to design if you must?

Bettina Rockenbach and Irenaeus Wolff

University of Erfurt, CEREB

17. July 2009

Institution design in social dilemmas: How to design if you must?

Bettina Rockenbach and Irenaeus Wolff
University of Erfurt

Abstract:
Considerable experimental evidence has been collected on how to solve the public-good dilemma. In a 'first generation' of experiments, this was done by presenting subjects with a pre-specified game out of a huge variety of rules. A 'second generation' of experiments introduced subjects to two different environments and had subjects choose between those. The present study is part of a 'third generation', asking subjects not only to choose between two environments but to design their own rule sets for the public-good problem. Whereas preceding 'third-generation' experiments had subjects design and improve their strategies for a specified game, this study is the first to make an attempt at answering the question of how people would shape their environment to solve the public-good dilemma were they given full discretion over the rules of the game. We explore this question of endogenous institution design in an iterated design-and-play procedure. We observe a strong usage of punishment and redistribution components, which diminishes over time. Instead, subjects successfully contextualize the situation. Interestingly, feedback on fellow-players' individual behavior tends to be rendered opaque. On average, rules do improve with respect to the welfare they elicit, albeit only to a limited degree.

Keywords: Public good; strategy method; experiment; public choice

JEL-Classification: C72, C92, D71, D72

Address: University of Erfurt, Nordhäuser Str. 63, 99089 Erfurt, Germany.
Tel: +49-361-7374521 (BR), +49-361-7374524 (IW);
e-mail: bettina.rockenbach@uni-erfurt.de (BR), irenaeus.wolff@uni-erfurt.de (IW)

Acknowledgements: First and foremost, we would like to thank our seminar participants for their enthusiasm while taking part in this study. We are also deeply indebted to Reinhard Selten and participants at seminars in Bonn and Erfurt for their inspiring comments, as well as our colleagues at the chair of microeconomics for their tireless help in testing the experimental software programs, as well as for the fruitful discussions along the way.

Policies based on the assumptions that individuals can learn how to devise well-tailored rules and cooperate conditionally when they participate in the design of institutions affecting them are more successful in the field. (Ostrom, 1998, p.3)

I. Introduction

Despite an impressive body of work on the issue, the high degree of cooperation amongst unrelated humans still remains a puzzle for those concerned with the study of human interactions (cf. e.g., Ledyard, 1995, Fehr and Gächter, 2000a, 2000b, or Fehr and Fischbacher, 2004). [1] Why would an agent spend resources on promoting a public good in a one-shot situation without signaling opportunities when she can do better by free-riding on others' cooperative efforts? On the other hand, while substantial cooperation is going on amongst mankind, there are many instances in which one would hope for more, the efforts on the preservation of our very planet being an eminent example. Therefore the study of institutional settings that overcome the dilemma is advanced in various disciplines and with high intensity.

In much of this research, public-good games have been used as an experimental paradigm to study cooperation in a specific type of social-dilemma situation: subjects may either contribute (parts of) their resources to a common project that benefits society as a whole or they may keep them for private consumption. While it is socially beneficial to contribute fully, it is in the material interest of each subject to keep everything for private consumption. In the laboratory, it is commonly observed that contributions start off far below the socially optimal level and approach individually 'optimal' free-riding with increasing subject experience.[2]

In order to overcome socially disadvantageous free-riding, experimental subjects have been presented with a huge variety of institutional regulations. Among the most prominent rule features examined are punishment opportunities,[3] communication,[4] leadership,[5] reputation opportunities,[6] and ostracism.[7] In this "first generation" of experiments, subjects were exogenously exposed to the experimenter-determined institutional rules. This provides us with valuable information on the performance of these rules. However, these settings cannot answer the question whether subjects

1 For an overview of the theoretical approaches concerned with explaining cooperation, cf. e.g. Fehr and Schmidt (2002).
2 Cf. Ledyard (1995) or Ostrom (2000).
3 E.g. Yamagishi (1986), Fehr and Gächter (2000a), Masclet, Noussair, Tucker, and Villeval (2003), Nikiforakis (2008), Nikiforakis and Normann (2008), or Falk, Fehr, and Fischbacher (2005).
4 E.g. Isaac and Walker (1988), Ostrom, Gardner and Walker (1994), Cason and Khan (1999), Brosig, Weimann and Ockenfels (2003), or Bochet, Page, and Putterman (2006).
5 E.g. Vesterlund (2003), Potters, Sefton and Vesterlund (2005), Arbak and Villeval (2007), or Güth, et al. (2007).
6 E.g. Milinski, et al. (2006), or Sommerfeld, et al. (2007).
7 E.g. Cinyabuguma, Page and Putterman (2005), Maier-Rigaud, Martinsson and Staffiero (2005), or Güth, et al. (2007).

would actually develop or choose a specific rule. Consequently, subsequent studies have addressed the question of collective rule choices, both by ballot voting and by 'voting-with-one's-feet'.[8] These studies provide the important insight that the initial acceptance of punishment mechanisms is quite low but grows over time up to the extent that the mechanism may ultimately be accepted completely. The limitation of these "second-generation" studies is that they shed light on the acceptance of experimenter-given mechanisms, but cannot answer the question which mechanisms would actually be developed.

A "third generation" experiment

In this paper, we proceed by introducing an experiment of a "third generation" in which we endogenize the design of the institutional regulations. This approach bears a close resemblance to the studies of Axelrod (1984), Selten, Mitzkewitz and Uhlich (1997), Keser and Gardner (1999), and Keser (2000). While Axelrod asked scholars of game theory to specify complete strategies for a prisoners' dilemma, which could be refined in a second round, Selten, Mitzkewitz, and Uhlich had student subjects do the same for an asymmetric Cournot duopoly over several rounds. Keser and Gardner (1999) and Keser (2000) applied the method to a common-pool resource and a public-good problem, respectively. While, in the studies mentioned, subjects were completely free in their *design of a strategy* for a given situation, to the best of our knowledge, the present study is the first tackling the question of *institution design* for social-dilemma situations in this way. In other words, in our experiment, subjects are not only free to choose and adapt their behavior, but they actually act as lawmakers empowered to shape the institutional environment of the game.

At the same time as providing us with valuable insights about real-world institution formation, the chosen approach comes with a high level of complexity, which renders the gathering of a number of observations sufficient for sensible statistical analysis virtually impossible. For this reason, we do not intend to single out statistically significant effects of subtle changes in the environment but see our contribution as an explorative advance into the white-waters of how experimental subjects approach the problem of designing adequate institutional frameworks for social-dilemma situations -- and how well they do in this task.

The experiment was conducted over 3 months in the framework of two student seminars in which

[8] For studies building on ballot votes, cf. e.g. Potters, Sefton and Vesterlund (2005), Sutter, Haigner and Kocher (2005), or Guillen, Schwieren and Staffiero (2007), 'voting-with-one's-feet' was used by Gürerk, Irlenbusch and Rockenbach (2006) or Rockenbach and Milinski (2006). Kosfeld, Okada and Riedl (forthc.) also use a 'voting-with-one's-feet' mechanism, but in contrast to Gürerk, Irlenbusch, and Rockenbach, subjects choose to form and become subject to a sanctioning institution, rather than choosing between two separate worlds.

the participants' task was to design institutional regulations to overcome a social dilemma. Before designing the institutional regulations subjects gathered experience in playing the basic public-good game in an anonymous laboratory setting. Following that, they were given a week to develop a set of rules of play for this game. There was no predefined set of rules so that subjects could freely choose whatever rules they wanted to implement, as long as the incentive structure still exhibited the social-dilemma characteristics. To achieve a certain degree of external validity we attached a certain cost to each of the proposed rules according to the true (relative) costs such an institution would give rise to in common real-world settings. After this first design phase, a different set of subjects played the public-good game under these rules and the designers were rewarded according to the efficiency, i.e. sum of players' profits minus rule costs, the players achieved under their rule. The design-and-play process was repeated three times. This quite elaborate design (described in detail in section 2) enables us to study the development process of the institutional mechanism in an iterated-improvement procedure.

Research questions

Even though by taking a step further than most studies we forfeit the advantage of examining the effect of changing a single treatment variable holding all other things constant, the research questions that inspired our design remain firmly grounded in the existing literature:

1. Will subjects make use of the solutions researched in the literature, or are they going to attempt at finding their own ways out of the dilemma? Will their rule sets contain elements of a) punishment, b) communication, c) leadership opportunities and d) reputational mechanisms, and e) ostracism?

2. Are rule sets going to be based on a single rule characteristic or are they going to be intricate combinations of several such characteristics?

3. Will rule sets "converge" to a single, possibly successful set that is the same across all groups, and possibly across seminars?

What we find is (1) that they make extensive but decreasing use of punishment, they hardly provide opportunities to communicate, and no opportunities to increase contributions by assigning leaders or ostracizing others. Also reputation-building elements are present only in one third of all rule sets. Instead, they often make use of framing and moral appeals, and they try to create positive incentives by redistribution. (2) For the most part, subjects try to combine two or even more rule components instead of relying on a single-component rule set. Finally, (3) there is no clear "winner" rule component or combination, and thus, rule sets retain and even increase the diversity of earlier tournaments also in the final tournament instead of "converging" to a single combination of rule components.

The paper is organized into six parts. In section II of the paper, we present the design of our experiment in due detail, as well as specifying the game-theoretic model of the underlying basic public-good game. In section III, we analyze the observed rule sets, classifying them along four basic component categories. We evaluate rule performance and the rule-adaptation process in section IV and perform a typicity analysis to determine what may have been 'typical mistakes' and 'typical improvements' in section V. Also in this section, we explore the relationship between rule-component typicity and rule-set performance, finally arriving at a tentative 'rule of choice'. Finally, in section VI, we discuss our findings and derive some possible implications for real-life situations.

II. Model and experimental design

In order to find out what kinds of rules people would give themselves were they free to do so, we recruited 24 students of economics at the University of Erfurt for two separate seminar-type courses (12 in each seminar). The seminars ran over 70 days from April to July 2007. Upon arrival, students were allocated to rule-development groups of 4 who stayed together for the entire seminar. We asked the participants of both seminars not to interact with the participants of the other seminar, however, we could not control that this was actually complied with. The course of each seminar is summarized in Table 1.

During a preliminary meeting, potential seminar participants were introduced to the schedule. However, they were not informed on what game they would be going to face during the seminar.[9] The seminar was not accompanied by any lectures on theoretic or experimental investigations into social dilemmas, neither in this term nor in previous terms. On the first day of the seminar (Day 1 in Table 1), the participants experienced the basic game by playing it in a laboratory setting over 25 rounds in a partner design. The basic game was a standard public-good game with four players. Each player i would make a contribution of x_i from an endowment of 20 tokens to a common project and keep the remainder. The total contributions were multiplied by 1.6 and divided evenly amongst the players, so that the public good exhibited a constant marginal per-capita return of 0.4.

The resulting payoff function of subject i is therefore:

$$\Pi_i = 20 - x_i + 0.4 \Sigma x_j.$$

Subjects were informed only about the sum of contributions, not about individual contributions and there was neither punishment nor any other additional non-standard rule feature.

[9] See Appendix A for the information passed to those present at the preliminary meeting, as well as to the instructions for the basic game.

Time	Playing stages	Rule development
	Preliminary meeting	
Day 1:	play of the basic public-good game	*development groups formed*
Days 2-6:		development of 1st set of rules
Day 7:		handing-in of rules
Days 8-13:	software implementation (experimenter)	
Day 14:	1st rule tournament	
Days 15-20:		development of 2nd set of rules
Day 21:		handing-in of rules
Day 22-27:	software implementation (experimenter)	
Day 28:	2nd rule tournament	
Days 29-34:		development of 3rd set of rules
Day 35:		handing-in of rules
Days 36-41	software implementation (experimenter)	
Day 42:	3rd rule tournament (double weighting)	
Days 43-48:		4th set of rules
Day 49:		handing-in of rules
Days 50-55:	software implementation (experimenter)	
Day 56:	4th rule tournament (triple weighting)	
Day 70:	*Final meeting; discussion*	

Table 1: Schedule of the seminars

Rule design

At the end of the first meeting participants were randomly allocated to *design groups* of four students each who stayed together until the end of the seminar. The participants had one week to develop their own set of rules for the public-good game within these design groups. There was no pre-defined “menu” of rules and the subjects were free to develop whatever rules they wanted (in the boundaries of standard rules of ethics). Each rule (component) is potentially attached to costs. The costs were meant to reflect the expenditures such a regulation would imply in a real-world setting. Thus, we tried to approximate the rules' costs by estimating what the implementation of such a rule would entail in real-life situations. For some rule sets, costs were split into a fixed and a variable part: e.g. to set up the infrastructure to make public announcements as a fixed cost and the variable costs of actually making announcements. The most severe and therefore most expensive interventions are restrictions of the players’ action space, e.g. to exclude complete free-riding or to even enforce full contribution. These kinds of coercion would not only require a lot of enforcement

power, but severely change the nature of the game and therefore, the cost of full-contribution enforcement was set to 1200 points. This amount equals the maximum gains to be achieved by such an intervention and was chosen to make such a change of the nature of the game prohibitively costly. For minimum-contributions that were lower than 20, the attached cost was approximated by a linear function and thus equaled 60 times the minimum set. On the other end of the cost range are simple announcements or advertisements to the players which gave rise to costs of 50 points. A detailed listing of the introduced rules and the attached costs is provided in Table B1 in appendix B.

When an institutional rule was proposed, the experimenters quantified the attached costs. The cost scheme was hidden from the groups to avoid an "anchoring effect", but any subject was given the possibility to ask for the costs of a specific rule set at all times. We are well-aware that our cost estimates are arbitrary and that the chosen rule cost system may influence outcomes. We are all the more surprised that the groups maintained a large variety in rule sets in both seminars, yet achieving a high standard of rule-cost minimization in the final round. This clearly shows that our system of rule costs allowed for a variety of minimal-cost rules, rather than predetermining one particular set of rules as a natural winner of the efficiency contest. This is also reflected in the fact that our results are not very sensitive to whether we compare rule sets by the average contributions or the efficiency they elicited: as we shall see in section IV, the only substantial difference is the standard finding that *punishment or redistribution* as a rule component induces higher contributions but lowers overall efficiency. Even so, the ordering of rule sets by contributions or efficiency are similar: for example, the best three rule sets in either dimension coincide, and out of the best six sets in terms of one of the dimensions, five are among the best six in terms of the other. Therefore, we think that our results are relatively robust to changes in the system of rule costs.

Implementation, play, and feedback

At the end of the week, each design-group had to hand in a verbal description of their rule set which was subsequently implemented in the experimental software z-tree[10] and translated into neutrally-worded instructions by the experimenter.[11] After another week we met again for the first tournament. The subjects of a seminar were randomly allocated to *play groups* of 4. Each play group played under a different rule set developed by the rule design groups. To avoid rule designs tailored to a specific subject population, the play groups differed from the design groups. To guarantee there was enough incentive to create efficient rules and improve them as best as participants could, we had design groups compete in an efficiency tournament that would partly

[10] Fischbacher (2007).

[11] Subjects were given the possibility to have loaded instructions distributed, incurring the same rule costs as on-screen announcements during the experiment.

determine the seminar marks for the design group. More precisely, we measured efficiency as the sum of individual payoffs within the play group (who also had to bear the costs caused by the rule-set), as a fraction of the payoffs in the social optimum. To leave some room for initial experimentation, later tournaments were weighted higher than earlier tournaments.

By the end of each tournament, seminar participants were provided with detailed feedback on the performance of all rules within their seminar group by round and group, comprising (i) individual contributions, (ii) efficiency, (iii) returns from the public good abstracting from any costs, and (iv) variable costs, as well as the instructions for all rule sets.[12] Overall, there were five rounds of play, four of them – the tournaments – under rules developed by the design groups. The third (final) tournament was weighted two (three) times higher than the first two rule tournaments in terms of rule performance. In addition to their rule sets' performance, the profits gained in students' individual play also made up for the seminar mark, so that sabotage of alien rules would be costly to the saboteur. Note that individual performance in the different tournaments was weighted evenly.

As mentioned above, we had the students clearly separated into two distinct groups that met at different times, urging them not to communicate with students from the other group on the topic of the seminar. This was done to see whether the rule sets would "converge" to similar sets. Not only did rules not "converge" to the same set over the seminars; they did not even converge within a seminar group.

Data base

In our experiment we obtain a data base of 24 rule sets from two seminar groups composed of 12 participants each. The corresponding three rule groups in each seminar interacted in four tournaments. Out of these 24 rule sets, we excluded two for our analysis, leaving us with 22 data points.[13] In the following section, we report on what the rule sets looked like, classifying them in terms of basic rule components, before we proceed to analyze the development process in more detail in sections IV and V.

[12] Fixed rule costs were stated in the instructions and implied in the efficiency figures.

[13] To ensure that groups had the largest-possible freedom in their pursuit of avenues out of the social dilemma, we admitted one group’s idea to transform the game into a minimum-effort game, given the no-contribution equilibrium remained. The transformation was done by allowing the rule group to redistribute any contributed points surpassing the minimum contribution back to the contributing player before the sum of contributions was multiplied by the public-good factor. Nevertheless, this changes the game from being a social dilemma to a coordination game, a solution that is not generally applicable to social-dilemma situations and is thus off our research path. We excluded this rule set, which was applied by the group in tournament 2 and 3.

III. Rules used and disregarded rule features

The rules introduced by our subjects are summarized in table B.1 in appendix B. In that table, we briefly describe the rule sets, providing the contribution and efficiency level achieved as well as the costs they gave rise to, grouped by the seminar group and the tournament number. In order to have a better understanding of what the determinants of those rule sets were, we classified the rule components along different categories.

1. Punishment and redistribution

The role of punishment in social dilemma situations is widely and prominently discussed in the literature[14]. Although our subjects had at best a very incomplete knowledge of this literature, many of the submitted rule sets made use of punishment or redistribution features. Interestingly however, only two of them employed a peer-to-peer mechanism. Most often, rules specified the deduction of points from the lowest-contributor and, in case of redistributive mechanisms, the reallocation of these points in favor of the highest-contributing player. In some of the cases, deducted points were transferred to an account that was to be redistributed at the end to the player having contributed the most. One rule set had players contribute to the joint account as well as to an 'administration' that was constituted in a step-level fashion by those contributing to this additional account. In case of sufficient contributions to the 'administration', its members were allowed to punish non-members using a peer-to-peer mechanism. Another set incorporated a 'warning system': those contributing less than the mean were asked to increase their contributions in the following period. In case this advice was not followed, they were punished by an amount conditional on the earlier deviation from the mean. Most of the sets, however, had a direct, automatic punishment or redistribution mechanism based on contribution ranks, with differentiated punishment of the lower-contributors. These observations give rise to four (not mutually exclusive) components capturing different punishment and redistribution techniques. The components are formulated in such a way that the question of whether a rule set satisfies it can unambiguously be answered by yes or no.

1a. Punishment and redistribution (pun): the rule set provides for either destruction of a part of a player *i*'s points (punishment), or a reallocation thereof in favor of at least one other player *j* (redistribution), conditional on *i*'s behavior

This condition was fulfilled for 75% of all rule sets. Remarkably, in the first tournament, all rule sets include either punishment or redistribution, or both. However, only half the sets feature any of

[14] See e.g. Yamagishi (1986), Ostrom, Walker, Gardner (1992), Fehr and Gächter (2000a, 2002), Denant-Boemont et al. (2007), Nikiforakis (2008), Carpenter and Matthews (forthc.) for experimental studies, and Henrich and Boyd (2001), Boyd, Gintis, Bowles, and Richerson (2003), Hauert et al. (2007), Dreber et al. (2008) for theoretical approaches.

them in the final tournament, and those that do, use redistribution. In other words, 'pure' punishment is no longer observed in the final tournament, but “punishment” through redistribution or rule cost assignment conditional on the player’s contribution is. While this could be attributed to details in the cost scheme implemented, it cannot be ignored that the frequency of use of these mechanisms steadily declines over time.

To obtain a better understanding of the *pun* rules, we introduce three further classifying variables, the first of which relates to the frequency with which the deduction regimes played a role in the game:

1b. Punishment or redistribution in every period (punEP): punishment or redistribution (as described in 1a.) takes place in every single round

This category makes the distinction between rules with roundly punishment or redistribution and those implementing them only periodically or even only once. A rule set exhibiting *punEP* provides for roundly deduction of points, a characteristic that was displayed by 13 out of 24 rule sets. In terms of rule sets which exhibit characteristic *pun*, the fraction of rule sets with *punEP* decreases from six out of six (eight out of ten in tournaments one and two) to two out of three (five out of eight in tournaments three and four).

Another distinction can be made with regard to players’ influence on the points to be deducted:

1c. Redistribution endogeneity (redEnd): punishment or redistribution (as described in 1a.) is administered by players themselves rather than automatically

Most rule sets (16 out of 18 rule sets with *pun*) proposed by our rule groups deprived the players from any influence on the points to be deducted, once the contribution decisions had been taken. Only in one instance, punishment was in the form of peer-to-peer punishment as in the typical public-good experiments with punishment,[15] and in the other rule set, players contributing more than 14 tokens were allowed to jointly decide (through a voting procedure) on the allocation of rule costs among the remaining players.

Finally, we distinguish rules that involve a single, concentrated reward payment at the end of a session as a special case of a redistributive rule:

1d. 'Big bonus' (BB): redistribution takes on the form of a fund paid into and one distributive action (the allocation of a 'jackpot') at the end of round 25

[15] E.g., Fehr and Gächter (2000a, 2002).

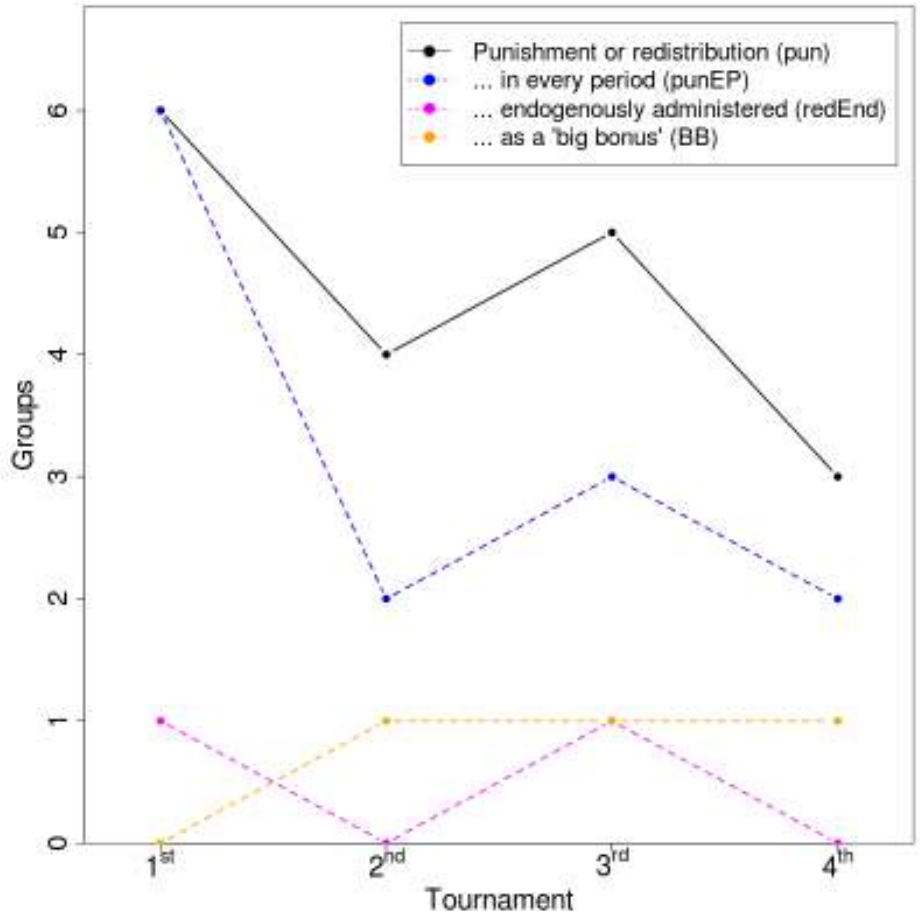

Figure 1: Frequency of punishment-or-redistribution characteristics by tournament.

There were three rule sets that involved roundly payments into a fund that was awarded to the highest-contributing player at the end of the respective session (with an equal-split rule in case of a tie). All of them were designed by the same rule group.

The incidence of the rule components summarized under punishment and redistribution over time is depicted in Figure 1. As mentioned before, we can see a clear trend away from *pun* rules, as well as, within that category, from rules that prescribe punishment or redistribution to be effective in every round. A 'big bonus' was employed by one rule group in all tournaments following the first one, and endogenous administration of *pun* was only introduced twice, by different groups.

2. Feedback on individual behavior

Several studies (e.g., Milinski, et al., 2006, Sommerfeld, et al., 2007) have shown that reputation-building opportunities may be very effective in promoting cooperation. A necessary prerequisite for being able to build a reputation is having some kind of identity. Bohnet and Frey (1999) show that identification alone suffices to increase the degree of pro-social choices in prisoner's dilemma and dictator games. In principle rule sets could allocate identification, e.g. in terms of a fixed player ID to subjects, but in our seminars, this happened in a minority of the cases.

2a. Feedback on individual behavior (ID): players are informed not only on the sum of contributions, but also on individual contributions

In fact, subjects were provided with identity-based feedback only in one third of all rule sets. The remaining two thirds of the rule sets did not provide feedback based on identity numbers, and not even on individual contributions. In two cases, players faced a rather special situation we also want to separate: players under these two rule-sets did not receive any feedback on contributions at all.

2b. No feedback on contributions (noFB): the sum of contributions is concealed from the players.

In both cases, this rule was coupled with a 'big malus', a severe punishment action of the lowest-overall-contributor at the end of the game, unless a pre-specified cumulative contribution level had been achieved by the group.[16] In the first instance of the rule-set, a considerable amount of contributions was achieved (1271 tokens, as compared to the 834 required to circumvent the 'big malus'). Not being satisfied by the efficiency achieved, the group modified their rule-set, asking for a group-contribution of 1667 tokens, which the play-group failed to comply with by a margin of 120. The application of the 'big malus' led to a disastrous result in terms of efficiency and the group abandoned this strategy.

3. Communication: subject-to-subject and "lawmaker"-to-"people" (framing)

Another tool that has proven to be very effective in inducing cooperation in social-dilemma situations is communication (e.g. Isaac and Walker, 1988, Ostrom, Gardner and Walker, 1994, Cason and Khan, 1999, or Brosig, Weimann and Ockenfels, 2003). Therefore, we would not have been surprised if our seminar participants had included elements of communication in their rule sets, such as a pre-play chat, or the selection and subsequent distribution of pre-specified messages.

3a. Communication (comm): there is some form of communication between players, e.g. in form of a signaling opportunity such as in cheap-talk agreements

To our surprise, rule sets rarely provided for signaling opportunities, and no group implemented an open-communication component in the form of a virtual chat-room. Only in the final tournament, one rule group in each seminar opted to allow players to engage in signaling behavior through a unanimity vote on a (non-binding) covenant, with roundly renewal (contingent on that there have not been more than two breaches in the past) and requiring subjects to contribute fully in one case, and periodical votes coupled with a proposed minimum of 15 tokens in the other.[17]

[16] We do not introduce an additional 'big malus' category, as the two rule-sets are already uniquely classified by the *noFB* category. No other rule-set employed a 'big malus', and thus, such a category would not add any information.

[17] The latter rule group envisaged the possibility of changing the minimum requirement in case of general adherence to the covenant; however, their covenant did not succeed in inducing the expected cooperation.

While it has been shown that framing can substantially influence behavior in public-good games (Ross and Ward (1996); Cookson (2000); for a different conclusion, see Rutte, Wilke and Messick (1987); Rege and Telle (2004)), we did not expect our subjects to make an attempt at contextualizing the situation. Even neglecting the studies not finding a framing effect, we had reason to believe so: all subjects were – by then – experienced players of a laboratory public good and this was common knowledge. Furthermore, they knew (as players) that the framing was artificially imposed and had no relevance to the game actually played (for seminar marks).

3b. Some frame (frame): the game is played after the issuing of an appeal (to moral sentiments, the advantageousness of the social optimum, or general social-welfare considerations), with a background story, or even with feedback conditional on behavior

Contrary to our expectations, less than half of all rule sets did without a framing (13 out of 24 had some framing). Rule components classified as falling into this category include priming attempts, such as the display of roundly changing statements similar to “One cannot live without trusting others”. Further, they include individual feedback conditional on behavior such as messages “To live means believing in something. In our case, in the community, thanks for that!” in case of a full contribution. Or framing in the narrower sense, embedding the contribution decision in contexts like the building of a school in Afghanistan or public goods arising in a local neighborhood.

To further distinguish rules making use of a *frame* component, we separate those containing a one-time appeal at the beginning only from those envisaging repeated messages.

3b. Frame in every period (frameEP): an appeal, a background story, or behavior-conditional feedback is displayed in every single round.

Out of the 13 rule-sets containing a *frame*, six also had *frameEP* as a characteristic.

4. Participative elements

Experimental studies like Ostrom, Walker and Gardner (1992) and Tyran and Feld (2006) have shown that increased participation opportunities lead to more cooperative outcomes. We were curious to find out whether our subjects would anticipate this and implement features of endogenous rule adaption or even rule change.

4. Endogeneity (end): the rule-set provides for institutional change as a consequence of either player behavior or a vote

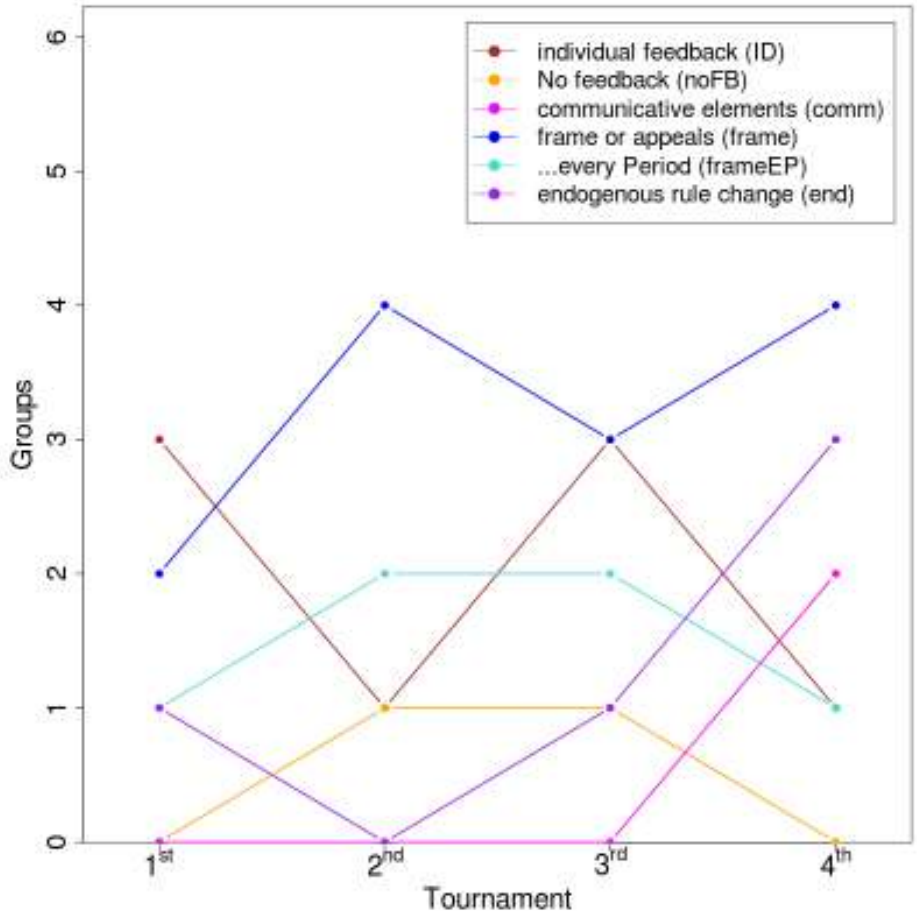

Figure 2: The use of communicative and participative elements over time.

Indeed, such features are used increasingly towards the end, however, only one fifth of all rule sets contain them and in one of the seminar groups, such elements are introduced only in the last tournament. Examples of endogenous rule features are a vote on removing any features that go beyond the basic game after the eighth round or the introduction of peer punishment contingent on players paying for a monitoring institution.

Figure 2 illustrates the occurrence of rule components 2. to 4. over the course of the seminar. As has been stated before, individual feedback was not a widespread rule component, exhibiting a downward trend over the four tournaments. Frames or appeals and endogenous rule changes, on the other hand, were used increasingly, while communicative elements where only introduced in the last tournament by two out of six groups.

In Table 2, we summarize the frequencies of usage of the different components and pair-wise combinations thereof. Note that the components are not mutually exclusive, and, in the case of the sub-categories of punishment or redistribution (i.e., *punEP*, *redEnd*, and *BB*), even contingent on the main category (*pun*). Therefore, the sum of border totals does not convey any sensibly interpretable information.[18] Furthermore, note that our color shading always refers to the frequency

[18] If it was not for the contingencies among the punishment and redistribution components, the sum of border totals

of the component combination as a percentage of the *row component*'s incidence. If, for example, the cell (BB, pun) wears an orange coloring, this is owed to the fact that all *BB* rules are also *pun* rules by definition. On the other hand, the turquoise shading of cell (pun, BB) marks that the three *BB* rules only account for roughly the sixth part of all *pun* rules.

What can be seen from Table 2 is that communication (*comm*) is never combined with a punishment-or-redistribution (*pun*) nor an individual-feedback (*ID*) component and that no feedback (*noFB*) is only combined with framing (in half of the cases), apart from the 'big malus' reported above. One finding that comes as a surprise is that feedback on individual behavior (*ID*) never goes together with a frame in every period (*frameEP*) and *ID* and some frame (*frame*) go together only once, even though neither category is extremely rare among the rule sets. Thus those engaging in the psychological techniques of framing, priming, and appealing explicitly chose to render others' individual behavior opaque.

	Punishment and redistribution				Feedback on individual behavior		Communication			Particip. elements	TOTAL
	pun	punEP	redEnd	BB	ID	noFB	comm	frame	frameEP	end	
pun		12	2	3	8	2	0	6	3	3	17
punEP	12		1	3	5	0	0	5	3	1	12
redEnd	2	1		0	2	0	0	0	0	2	2
BB	3	3	0		0	0	0	3	2	0	3
ID	8	5	2	0		0	0	1	0	2	8
noFB	2	0	0	0	0		0	1	0	0	2
comm	0	0	0	0	0	0		2	1	2	2
frame	6	5	0	3	1	1	2		6	2	11
frameEP	3	3	0	2	0	0	1	6		1	6
end	3	1	2	0	2	0	2	2	1		5
TOTAL	17	12	2	3	8	2	2	11	6	5	22

Table 2: Usage of rule-set components (border totals) and pair-wise contingencies; colors highlight the frequency of the component combination as a percentage of the row component's incidence: 0-10% (blue), 10-40% (turquoise), 40-60% (green), 60-90% (yellow), 90-100% (orange).

A table classifying the different rule sets according to our characteristics can be found in Appendix B (see Table B2). The classification is also used for the typicity analysis reported in section V.[19] Before we do so, we explore the performance of rules exhibiting a certain component, in order to assess whether the rule-adjustment process went 'in the right direction' in section IV.

divided by the number of rule sets would be interpretable as the average number of rule components employed.

[19] To run a typicity analysis, characteristics displayed by less than half of all strategies have to be reverted (such that, e.g., *ID* becomes *noID*, i.e., "subjects are *not* informed on individual contributions but only on their sum"), for technical reasons outlined in Kuon (1993).

IV. Rule performance and rule adaptation

In the preceding section, we categorized subjects' rule sets and established the frequencies with which rule components were used in the different tournaments. In this section, we set out to evaluate those components' contribution to the performance of the rule set they are incorporated in. Figure 3 gives a first hint at an answer, displaying boxplots of the efficiency levels achieved by all rule-sets employing a certain component. Judging by Figure 3, one may argue that having characteristics like *BB*, *comm*, and *frameEP* in one's rule set seems to be a good idea, while employing *noFB* does not seem to lead to a good performance. To examine this conjecture more closely, we analyze our rule-sets using a simple distribution-free measure of success. We rank the rule sets by efficiency (with the highest rank corresponding to the highest efficiency) and calculate the average rank of all rule sets displaying a certain characteristic. Comparing this to the median rank of all rule sets provides us with an index for the discriminatory power of the component. Division by *(n-1)/2*, where *n* is the total number of rule sets, finally normalizes the index to lie within the interval [-1,1]. An index of 1 implies that only the best-performing rule set displays this component, while an index of -1 means that the characteristic is only made use of by the worst-performing rule set. A characteristic that is employed in every strategy would lead to an index of 0, as would any component for which the sum of rank deviations from the median above the median is equal to the sum of those below it. If R_i denotes the rank of rule set i, and Y_j is the set of all rule sets that exhibit characteristic j, our index ρ_j can be expressed as follows:

$$\rho_j = \begin{cases} \frac{2}{n-1}\left(\sum_{y \in Y_j} \frac{R_y}{|Y_j|} - \frac{n+1}{2}\right), & |Y_j| > 0 \\ 0, & \text{otherwise.} \end{cases}$$

This index is a rather simple approach assuming a linear separability of the performance of rule sets, neglecting any interaction effects between rule components. Note further that the indices are based on a relative ranking of marginal component effects, and can therefore only explain deviations from the median efficiency rank. Nevertheless, a virtual rank ordering of rule sets by simply adding up the components' ρ_j yields a surprisingly good prediction for the rank ordering by efficiency: the Spearman correlation coefficient between the two rankings is $r_s = 0.55$ ($p < 0.01$).[20] If we take into account that not all components contribute to efficiency to the same extend, this is a surprisingly high correlation.[21] In Figure 4, we depict the indices for our rule characteristics,

[20] Applying this analysis to the final tournament only yields a correlation coefficient of $r_s = 0.89$ ($p = 0.0333$).

[21] Running a mixed-effects estimate of normalised rank deviations from the median rank on the indices (with random effects for rounds, the interaction of rounds and seminar group, and rule groups), the rank correlation between the ensuing fitted ranks and the true ranks can be driven up further ($r_s = 0.90$, $p < 0.001$).

grouped by tournaments. The three components that seem to boast efficiency the most seem to be the use of framing, *not* providing subjects with feedback on individual contributions, and accentuating the supergame characteristic through a 'big bonus'. In contrast, a 'big malus' as in the case of the *noFB* rules seems to be counterproductive. While the indices for *pun* features other than a 'big bonus' do not seem to make up a clear picture across tournaments, this is notably different in the final tournament. It seems that our subjects have learnt to design efficient *pun* mechanisms by the end of the experiment. On the other hand, we see that the seemingly successful mechanisms of framing and communication (Figure 4c) do not perform as well compared to final-tournament *pun* rules (Figure 4b). When compared to rule-sets from other rounds, however, they still score above average.

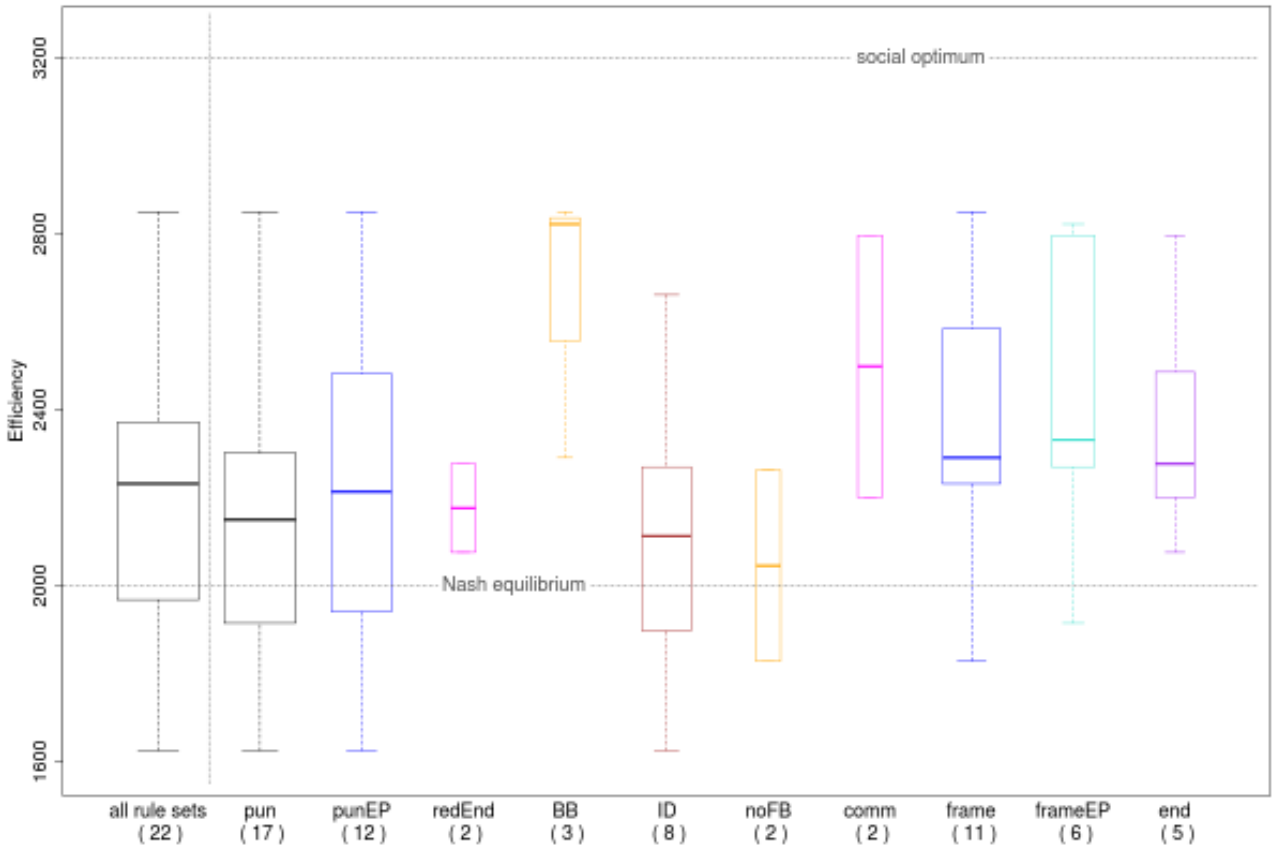

Figure 3: Boxplots of achieved efficiency levels, by rule components

Contrasting the efficiency-based indices in Figure 4 to their contribution-based counterparts depicted in Figure 5, we see that the main difference lies in that the *frameEP* components lose their positive influence, while *noFB* switches sides, which is at least in part owed to the realization of the 'big malus' in one of the two cases. This suggests that the *frameEP* characteristics are not the most cooperation-enhancing *per se*, but that they achieve the amount of cooperation they induce in the most cost-efficient manner for our given rule-cost scheme. What may come as a surprise is that the

increase in the *pun* mechanisms' performance is rather limited. While we are well-aware that the cost scheme we implemented may be prone to errors, the similarity of the indices based on efficiency and contributions may in our view be interpreted as a sign of robustness of our rule-cost scheme.

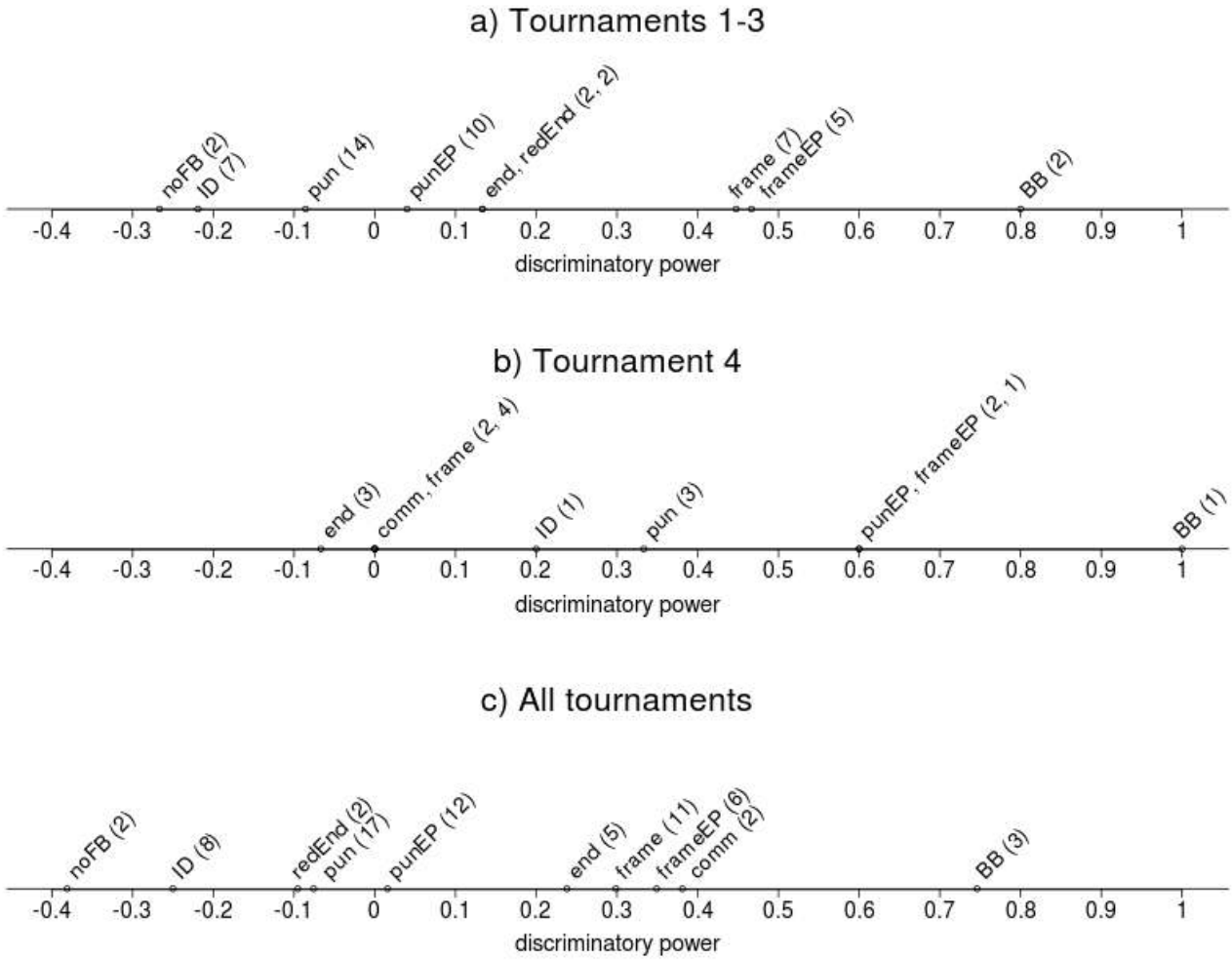

Figure 4: rule-component efficiency indices a) for tournaments 1-3 (upper panel), b) for tournament 4 (middle panel)), and c) for all tournaments (lower panel)). The numbers of rule-sets having a component are given in parentheses; if none, the index was omitted.

Based on our indices as a – however imperfect – measure of success, let us take a look at the rule changes over the tournaments as depicted in Figures 1 and 2. We observe that (i) punishment and redistribution systems lead to low efficiency (and only slightly higher contributions) unless they come with a 'big bonus', which corresponds with the trend away from such mechanisms. The fact that the decline in the use of *pun* is slow and only partial is surprising especially in seminar I, where there was one rule group that managed to achieve the highest degree of efficiency within their seminar in three out of four tournaments - using rules that did not contain either punishment or

redistribution. In fact, this was the only rule group to renounce the use of the *pun* component within this seminar group. In other words, participants in the remaining rule groups fail to recognize that not making use of punishment may actually be a competitive advantage: even in view of the superior results achieved by their competitor group, they fail to mirror that group's superior strategy.[22] In seminar II, rule sets abstaining from a 'monetary' sanctioning of behavior do not perform as well; in two out of three instances, they achieve only the second highest efficiency, the third set following one of the two just mentioned on third place.

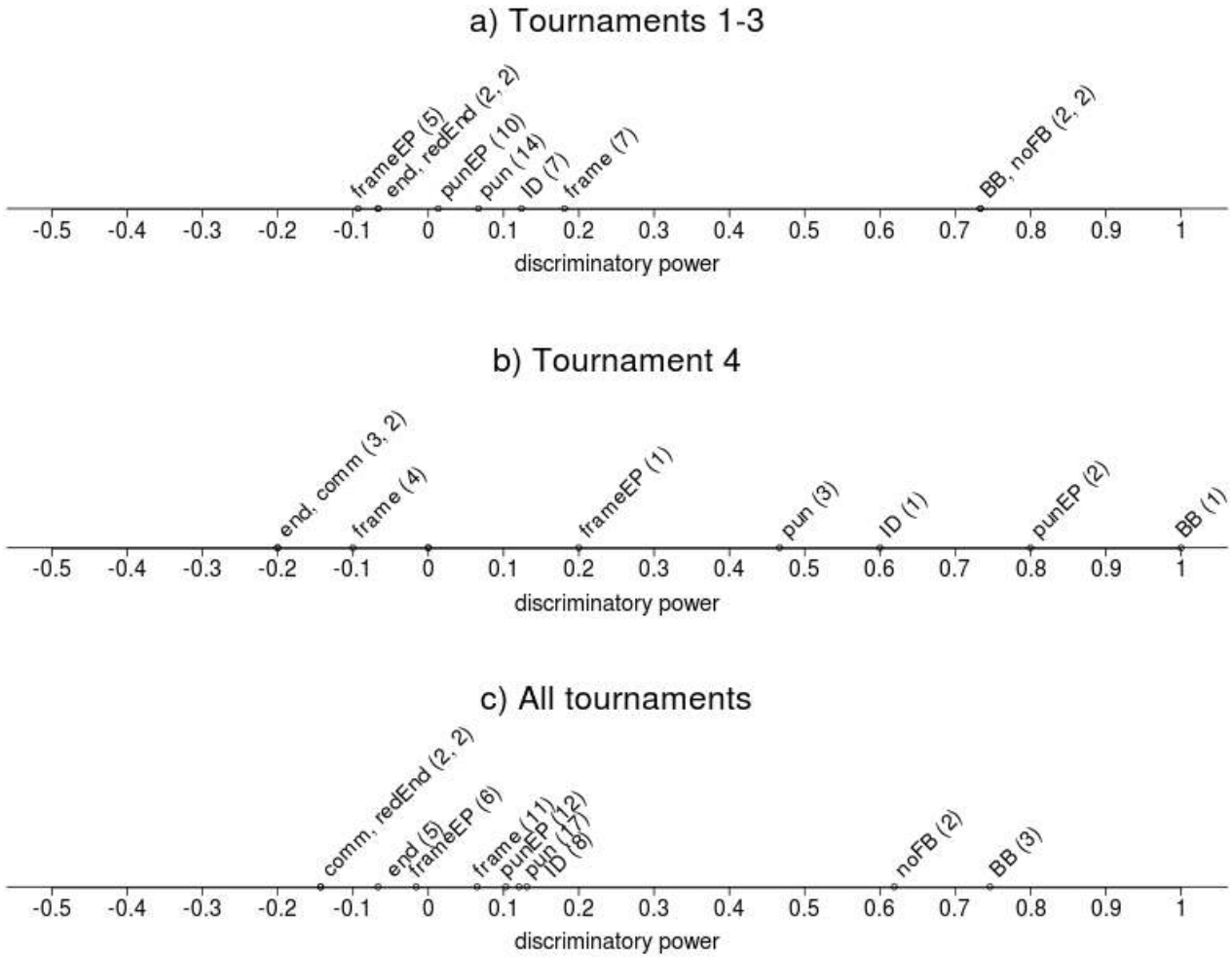

Figure 5: Rule-component contribution indices a) for tournaments 1-3 (upper panel)), b) for tournament 4 (middle panel)), and c) for all tournaments (lower panel)). The numbers of rule-sets having a component are given in parentheses; if none, the index was omitted.

[22] This reluctance to follow the good example of their competitor group could, of course, easily be explained by postulating a psychological bias to focus on contributions. However, from our discussions with the subjects we got the impression that they were well-aware of the fact that high(er) contribution averages do not necessarily lead to higher efficiency, but have to be measured against potentially increasing rule costs.

In the end, is the use of punishment a bad idea? Maybe not: *pun* mechanisms as designed in the last round (i.e., redistribution of rule costs as the only such mechanism) increase efficiency.[23] Furthermore, rules employing punishment or redistribution in every single round (*punEP*) consistently and increasingly outperform others ($\rho_{punEP,2} = \rho_{punEP,3} = 0.25$, $\rho_{punEP,4} = 0.6$, for tournaments two through four; for the first round, there is no such difference as all rule sets exhibit *punEP*). Adding the generally negative impact of *pun*, the findings suggest there may be a U-shaped productivity of punishment and redistribution in terms of application frequency, yielding best results for either roundly punishment or redistribution, or none at all. Nevertheless, the fraction of rule sets using *punEP* decreases over time. Finally, the use of a 'big bonus' at the end of the 25 rounds, making the supergame characteristic more salient, clearly seems to enhance performance, with an overall index of $\rho_{BB} = 0.75$. Nonetheless, only one of the six rule groups used this rule component. Looking at the general picture, what we can learn from our lawmaker's responses is that the effects of punishment are far from being obvious and understood.

From Figure 2, we know that the use of ID-based feedback tends to decline over time. Figure 4 shows that this is not as surprising as it may seem: the corresponding index for the first three rounds is $\rho_{ID,1\text{-}3} = -0.22$. Hence, ID-based feedback does not seem to have a positive effect on cooperation as long as there is no explicit award or commendation, or the possibility to be rewarded in an ancillary game as e.g. in Milinski, et al. (2006). Correspondingly, we observe the stated trend away from *ID*, with three rule sets having the component in the first and five sets in the final tournament. The abolition of the combination of *noFB* with a 'big malus' is easily explained by the fact that this attempt led to the lowest efficiency within that seminar group on both occasions it was made use of.

Turning to components with some communicative element, be it between subjects or between the 'lawmakers' and their 'people', we observe the following: the results of the two rule sets allowing for signaling do not allow for a clear conclusion, finding its expression in the corresponding index of $\rho_{comm,4} = 0$ despite an overall index of $\rho_{comm} = 0.38$. While one of them led to 87.4 % of the possible efficiency (70.1 % of possible contributions), the latter obtained only 68.8 % (20.9 %), compared to an overall average of 71.1 % (51.1 %). On the other hand, rule groups implementing a *frame* were correct in doing so, judging by their success: the implementation of some form of framing, be it a moral appeal or a story of building a school in Afghanistan and behavior-contingent feedback ("You did a good job on this one. The citizens of Kandahar are watching you with enthusiasm and getting you some fresh water."), led to indices of $\rho_{frame} = 0.30$. This is reflected in the slight trend towards framing, with two rule sets in the first tournament, and three or four sets ever after. However,

[23] Unfortunately, a meaningful statistical comparison using e.g. conditional average payoffs is not possible, given rules containing a feature and those not containing this feature within a seminar group are not independent.

subjects do not seem to have noticed that rule sets employing the framing in every period tended to perform even slightly better ($\rho_{frameEP}$ = 0.35): in the final round we only observe one such set. Finally, having the 'people' participate in the shaping of their world through endogenous features tends to foster efficiency (ρ_{end} = 0.24), while the figures for contributions are ambiguous: in the first round, $\rho^{cont}_{end,1}$ = 0.6, while in the last round, $\rho^{cont}_{end,4}$ = - 0.2. Note, however, that we only observe two rule sets containing the *end* characteristic.

Do 'lawmakers' change their rules for the better?

The final tournament was the most decisive for the students' success in the seminar and the intermediate tournaments were meant to improve the rule sets and their lower weights in the final grading enable experimentation. Participants have successfully used this opportunity. The median efficiency from the third to the fourth (final) tournament increases significantly (p_{34} = 0.0625, pair-wise two-tailed Wilcoxon signed-ranks test, where p_{xy} denotes the p-value for the test between tournament x and y), whereas it does not between the first and the second and between the second and the third, respectively (p_{12} = 0.3125 and p_{23} = 0.625). What did the rule groups do to achieve this improvement? To find an answer to this question and possibly shed some light on typical mistakes and typical improvements, we performed a typicity analysis based on the characteristics introduced in section III.

V. Typicity analysis

The preceding analyses considered the rule components in isolation and estimated their contribution to the rules' efficiency. In the following we analyze the interdependencies of the rule sets' characteristics with a typicity analysis. This type of analysis, introduced by Kuon (1993) and successfully implemented e.g. by Selten, Mitzkewitz and Uhlich (1997), is used to determine what a 'typical strategy' or, in our case, a 'typical rule set' is and what the typical characteristics of these rule sets are. The two measures are interdependently intertwined: the typical rule sets are the ones that carry the typical characteristics and the typical characteristics are the ones that occur in typical strategies. Technically the typicity analysis is the solution of an Eigen value problem, providing weights for the rule sets and the characteristics. Typicities of characteristics add up to one and are all equal to one divided by the number of characteristics if the number of strategies containing each of these features is the same, such that the reciprocal value of the number of characteristics forms the natural reference point against which typicities may be evaluated. On the other hand, the more typical a rule set's features are, the more typical is the rule set itself. More precisely, the typicity of a

rule set is the sum of its characteristics' typicities.[24] Once the typicities of rule sets and characteristics are established, we can look for whether typicity is in any way correlated with efficiency and thus assess how well-targeted our subjects were in terms of their search for better solutions. Given we have found rule sets in the final tournament to perform better than those in the preceding tournaments, we run two distinct analyses for these two sets and compare their typicities to find out how the typical rule sets differed. Table 3 lists the corresponding component typicities.

	pun	punEP	noRedEnd	noBB	noID	FB	noComm	frame	noFrameEP	noEnd
ρ [j,1-3]	-0.086	0.040	-0.019	-0.114	0.170	0.038	0.000	0.448	-0.212	-0.019
ρ [j,4]	0.333	0.600	0.000	-0.200	-0.040	0.000	0.000	0.000	-0.120	0.067
Typicity in trnmts 1-3	0.113	0.085	0.115	0.111	0.076	0.113	0.127	0.057	0.089	0.115
Typicity in trnmt 4	0.073	0.050	0.136	0.110	0.113	0.136	0.098	0.090	0.119	0.075

Table 3: Performance indices, as introduced in section IV (lines 1 and 2); rule component typicities over the first three and the final tournament (lines 3 and 4). Note that some characteristics have been reverted for the purpose of the typicity analysis, because – for technical reasons – the characteristics have to be formulated such that the majority of rule sets carry it. Therefore, for example *redEnd* has to be reverted into *noRedEnd*, and *BB* has to be reverted into *noBB*. The performance indices reported in lines 1 and 2 correspond to the changed component formulation and may hence be different from those reported above.

What we see in Table 3 is that the change in rule component typicities largely corresponds to the frequency changes we know from Figures 1 and 2. Also from Table 3, we observe that a positive component performance in tournaments 1 to 3 does not necessarily lead to an increase in the component's typicity. Furthermore, we find an increasing diversity in rule sets: average rule-set typicity declines from 0.888 in the first tournament to 0.773 in the final one even if we base these averages upon final-tournament typicities. This is a clear sign of a 'divergence' of rule sets. In light of this fact, a new question arises: is there a 'common denominator' in the sense of a typical rule set, as a paradigmatic institutional setting to employ?

Is the typical rule set a good choice for a (benevolent) lawmaker?

In Table 4, we present the correlations between rule typicity, on the one hand, and contributions and efficiency, on the other. What we see is that overall, there is no correlation of rule typicity with either contributions or efficiency. This statement has to be qualified slightly, as the typical rule set seems to improve over time, even if none of the correlations becomes significantly different from zero.

[24] For a thorough discussion of the mathematical properties of this method, cf. Kuon (1993).

Correlations between...	rule typicity and contributions (p-value)		rule typicity and efficiency (p-value)		
Overall	**0.010 (0.964)**		**-0.185 (0.410)**		
Tournament 1 - 3	**- 0.133 (0.623)**	- 0.235 (0.653)	**- 0.217 (0.419)**	- 0.059 (0.912)	1st tourn.
		- 0.200 (0.783)		- 0.500 (0.450)	2nd tourn.
		0.300 (0.683)		0.500 (0.450)	3rd tourn.
Tournament 4	**0.257 (0.658)**		**0.086 (0.919)**		4th tourn.

Table 4: Correlations of contributions and efficiency with rule typicity

The lack of a correlation between rule typicity and efficiency may be due to the presence of initial "typical mistakes" that are only hesitatingly done away with. Most prominently, and as already discussed in section IV, the use of *pun* seemed to generally decrease efficiency, as did the non-use of a 'big bonus' if a *pun* regime was introduced. What is most surprising, once again, is the reluctance of rule groups to follow the better example of other rule groups obtaining better results by not committing these mistakes.[25]

Notwithstanding, the redistribution rules of the final round are far more cost-effective, and thus, efficient than their predecessors. As "typical improvements" we would classify the increasing typicity of framing and the decline in the typicity of ID-based feedback. Finally, signaling opportunities *are* introduced in the final round, while open communication, leadership arrangements or opportunities for ostracism are not.

The (tentative) rule of choice

What would be the rule of choice were we to construct one solely based on our analysis? There would be a redistribution mechanism that is active in every round; its proceeds would (partially, if not completely) go into a 'jackpot' that is to be awarded to the player contributing most over time. There would be no identity-based feedback but an exchange of declarations of intent. Finally, we would try to activate social norms by some framing with individual, behavior-conditioned feedback messages. Such a rule would be very close to that used by rule group 4 in tournaments 2 and 3 and would have a typicity of about 77 %. The difference to the two observed rule-sets is the addition of an exchange of declarations of intent. Of course, such a "cocktail" of rule components may have to pay tribute to the possibility that combining rule features may lead to crowding-out effects, as e.g. Reeson and Tisdell (2008) have found for the combination of suasion (categorized as framing in our study) with a minimum contribution.

[25] As has been discussed before, in seminar I, we would have expected rule groups to copy the abstention of *pun* regimes, while in seminar II, it seemed to be most obvious after the second tournament that a 'big bonus' was a good way to go.

VI. Discussion and Implications

This paper reports on a novel approach in studying institutional frameworks to overcome social dilemmas. The novelty of the approach lies in completely transferring the institutional design process to the subjects. Previous studies have either confronted subjects with predefined institutional rules or given them the choice within an experimenter-defined rule set. The present approach allows us to study the emergence and development of institutional rules when subjects are unbound by any rule pre-specifications. We make three remarkable observations. First, punishment, in various disguises, was the initial focal point of all groups. This is noteworthy because punishment is an extensively studied mechanism in the literature and this result is a clear support for the importance of punishment rules in public goods settings. Remarkably, however, punishment mechanisms were not designed in the form of peer punishment, but rather in the form of pre-specified rules of deduction and/or redistribution contingent on complying with provision targets. These provision targets were either fixed levels (e.g. full provision) or contingent on the other group members (e.g. not being the lowest-contributing player).

A second important observation, namely the role of framing, is rather unexpected. This is noteworthy because subjects playing under these rule sets were experienced players, who, in their capacity as "lawmakers", have a relatively deep understanding of the logic of the game and were completely aware that the chosen frames have no connection to reality whatsoever, but are pure imagination. Nonetheless, framing was increasingly chosen and a successful means in achieving efficiency. While one of the groups opted to tell its subjects the public good was a school in Afghanistan in two of the tournaments, another group had its subjects play in a virtual neighborhood consisting of spouses and children, cats, dogs, and rat poison. Yet other rule sets displayed a "slogan of the round" such as: "Only within the community, people are beings conscious of their strength", or moral appeals directly asking subjects to contribute. What this tells us is that attempts at activating social norms through appeals or even general moral statements tend to be more than "just words": they seem to foster cooperation even amongst case-hardened addressees.

The third noteworthy finding is that subjects render information on a player’s fellow providers rather opaque - and that this tends to be a successful strategy. The implications are not straight forward and call for further research. Conditionally cooperative subjects are assumed to align their provision with what they believe others will contribute. Providing them with detailed information on past contributions of their peers may yield the most precise basis for the calculations. Yet, it seems that a certain degree of opaqueness by just providing the average contribution is more successful in enhancing cooperation. A possible explanation may be that the reduced information

precludes individual comparisons detecting advantageous as well as disadvantageous inequalities with respect to the other players. Even if these comparisons do not lead to equilibrium predictions different from the Nash equilibrium resting on the assumption of money maximizing actors, they may be important determinants of contribution dynamics. Engel and Rockenbach (2009) find individual payoff comparisons, and in particular experienced disadvantageous inequality to be a major source for the decline of contributions.

Interestingly, we also see prominent rules not used by subjects. Leadership opportunities and ostracism were not present in the rule sets at all. Additionally we find that communication is not used apart from scarce attempts to provide for cheap-talk signaling opportunities, such as non-binding 'contracts' specifying a contribution with the threat of abolishing the 'contracting' option after the third breach of agreement. Open communication, having proven to be a very effective mechanism in resolving social dilemma situations,[26] was not introduced at all. When we asked for the underlying reasons at the final meeting of the seminar, we were told communication was not introduced because subjects thought it would not work and it would cost too much. In other words, it would seem this very effective mechanism is not trusted enough to give it a try in a tense situation.

What are the implications of our study? Subjects design centralized punishment regimes. Interestingly, these regimes have so far been largely neglected in the experimental literature. Notable exceptions are Guillen, Schwieren, and Staffiero (2007), looking at a step-level public good with multiple (cooperative) equilibria, and, more relevant to the present study, Tyran and Feld (2006) who address the question of whether a "mild law", i.e., a punishment mechanism that is too weak to enforce cooperation, may nevertheless lead to a high level of contributions. They find two remarkable effects: (i) "mild law" does not lead to higher cooperation rates if exogenously imposed, but (ii) when endogenously introduced, cooperation is boosted. The first observation fits nicely with the limited success punishment regimes have in fostering contribution levels in our setting, the second may hint at the reasons for that. With two exceptions, rule groups introduce "mild law" centralized enforcement institutions *without* conveying their subjects an opportunity to change this institution. In other word, the introduction is endogenous with respect to the rule group only. On the other hand, the two punishment regimes involving endogenous features (the possibility of removing the institution by majority vote, for instance) seem to perform worse in our study. In light of this fact, further research on the issue seems necessary.

Our second main finding, the frequent and successful use of framing in an environment of case-

[26] E.g., cf. Ostrom, Walker and Gardner (1992), or Brosig, Weimann and Ockenfels (2003).

hardened players, is particularly strong and has several important implications. One is that the neutral frame commonly used in laboratory studies may lead to a biased estimate compared to cooperation in real-life settings. While most 'framing experiments' in the economics literature tend to examine presentation effects, such as the contrasting of a 'public good' game with an equivalent 'public bad' game in Sonnemans, Schram, and Offerman (1998) or the juxtaposition of a positive-frame investment in the public good and a negative-frame purchase of the private good in Andreoni (1995), there has also been research on contextualization. Cookson (2000), for example, shows that already having subjects calculate the effects of an action on group income rather than individual payoff can increase cooperation significantly. Liberman, Samuels, and Ross (2004), on the other hand, show just how important the context is. Building on older psychological research, they have subjects play identical prisoner's dilemma games, named either the "Wall Street Game" or the "Community Game". What makes their findings extraordinarily strong is that, while the context played a very substantial role, the players' behavior in real-life situations – as seen through others knowing them well – did not matter at all: people who were expected by others to default were no less likely to cooperate than those who were expected to do so. In other words, and relating it to our study, context seems to play so important a role that it works even when players know it is completely artificial. Bearing in mind the many things we see in commercials that are obviously and completely unrelated to the promoted product, this may not be such a big surprise, after all.

The low fraction of rule sets making use of detailed feedback on individual contributions, finally, points to yet another issue that has not been tackled by experimenters. While we know from Croson (2001) that individual feedback in the voluntary-contributions mechanism does not change average contributions but does increase their variance, in a step-level public good Croson and Marks (1998) find decreasing average contributions for detailed but anonymous information on others' contributions. When individual contributions are displayed alongside constant player-IDs, average contributions increase. In other words, feedback matters.[27] However, in a standard game with punishment, the effects of different types of feedback are still unknown. In public-good games with peer-to-peer punishment, detailed feedback is essential, but in a central-authority setting, this is no longer the case. Out of the studies using a centralized mechanism, Tyran and Feld (2006) do not give any feedback at all, while Guillen, Schwieren, and Staffiero (2007) do not give individual feedback. In other words, in terms of the feedback provided, neither allows for a comparison with existing peer-punishment studies. It is left to further research what role ID-based feedback may play in a public-good game with centralized punishment.

[27] Another example fort he importance of feedback can be seen in a recent study by Nikiforakis (2009), who studies the effect of providing ID-based payoff information as opposed to information on individual contributions. He finds that in the payoff condition, contributions are significantly lower.

References

Andreoni, J. (1995). "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." Quarterly Journal of Economics **110**(1): 1--21.

Arbak, E. and Villeval, M.-C. (2007). Endogenous Leadership: Selection and Influence, Institute for the Study of Labor (IZA).

Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

Bochet, O., Page, T. and Putterman, L. (2006). "Communication and Punishment in Voluntary Contribution Experiments." Journal of Economic Behavior & Organization **60**(1):11-26.

Bohnet, I. and Frey, B. (1999). "The Sound of Silence in Prisoner's Dilemma and Dictator Games." Journal of Economic Behavior and Organization **38(1)**: 43--57.

Boyd, R., Gintis, H., Bowles, S. and Richerson, P. (2003). "The Evolution of Altruistic Punishment." Proceedings of the National Academy of Sciences of the United States of America **100**(6): 3531--3535.

Brosig, J., Weimann, J. and Ockenfels, A. (2003). "The Effect of Communication Media on Cooperation." German Economic Review **4(2)**: 217--243.

Carpenter, J. and Matthews, P. (forthcoming). "What Norms Trigger Punishment?" Experimental Economics.

Cason, T. N. and Khan, F. U. (1999). "A Laboratory Study of Voluntary Public Goods Provision with Imperfect Monitoring and Communication." Journal of Development Economics **58**(2): 533--552.

Cinyabuguma, M., Page, T. and Putterman, L. (2005). "Cooperation under the Threat of Expulsion in a Public Goods Experiment." Journal of Public Economics **89**(8): 1421--1435.

Cookson, R. (2000). "Framing Effects in Public Goods Experiments." Experimental Economics **3**(1): 55--79.

Croson, R. (2001). "Feedback in Voluntary Contribution Mechanisms: An Experiment in Team Production." Research in Experimental Economics **8**: 85--97.

Croson, R. and Marks, M. (1998). "Identifiability of Individual Contributions in a Threshold Public Goods Experiment." Journal of Mathematical Psychology **42**(2-3): 167--190.

Dreber, A., Rand, D. G., Fudenberg, D. and Nowak, M. A. (2008). "Winners Don't Punish." Nature **452**(7185): 348--351.

Denant-Boemont, L., Masclet, D., and Noussair, C. (2007). "Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment." Economic Theory **33**(1): 145--167.

Engel, C. and Rockenbach, B. (2009). We Are Not Alone: The Impact of Externalities on Public Good Provision.

Falk, A., Fehr, E., and Fischbacher, U. (2005). "Driving Forces behind Informal Sanctions." Econometrica, **73**(6): 2017--2030

Fehr, E. and Fischbacher, U. (2004). "Social norms and human cooperation." Trends in Cognitive Sciences **8**(4): 185--190.

Fehr, E. and Gächter, S. (2000a). "Cooperation and Punishment in Public Goods Experiments." American Economic Review **90(4)**: 980--994.

Fehr, E. and Gächter, S. (2000b). "Fairness and Retaliation: The Economics of Reciprocity." Journal of Economic Perspectives **14**(3): 159--181.

Fehr, E. and Gächter, S. (2002). "Altruistic Punishment in Humans." Nature **415**: 137--150.

Fehr, E. and Schmidt, U. (2002). Theories of Fairness and Reciprocity – Evidence and Economic Applications. Advances in Economics and Econometrics – 8th World Congress, Econometric Society Monographs. Dewatripont, M., Hansen, L. and Turnovsky, S., Cambridge: Cambridge University Press: 208--257.

Fischbacher, U. (2007). "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." Experimental Economics **10**(2): 171--178.

Guillen, P., Schwieren, C. and Staffiero, G. (2007). "Why Feed the Leviathan?" Public Choice **130(1-2)**: 115-128.

Gürerk, Ö., Irlenbusch, B. and Rockenbach, B. (2006). "The Competitive Advantage of Sanctioning Institutions." Science **312**: 108--111.

Güth, W., Levati, M. V., Sutter, M. and van der Heijden, E. (2007). "Leading by Example with and without Exclusion Power in Voluntary Contribution Experiments." Journal of Public Economics **91**(5-6): 1023--1042.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A. and Sigmund, K. (2007). "Via Freedom to Coercion: The Emergence of Costly Punishment." Science **316**(5833): 1905-1907.

Henrich, J. and Boyd, R. (2001). "Why People Punish Defectors: Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas." Journal of Theoretical Biology **208**: 79--89.

Isaac, M. R. and Walker, J. M. (1988). "Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism." Economic Inquiry **26**: 585--608.

Keser, C. (2000). Strategically Planned Behavior in Public Good Experiments, CIRANO, Montreal.

Keser, C. and Gardner, R. (1999). "Strategic Behavior of Experienced Subjects in a Common Pool Resource Game." International Journal of Game Theory **28**(2): 241--252.

Kosfeld, M., Okada, A. and Riedl, A. (forthcoming). "Institution Formation in Public Goods Games." American Economic Review.

Kuon, B. (1993). "Measuring the Typicalness of Behavior." Mathematical Social Sciences **26**(1): 35--49.

Ledyard, J. O. (1995). Public Goods: A Survey of Experimental Research. Handbook of Experimental Economics. Kagel, J. and Roth, A., Princeton University Press**:** 111--194.

Liberman, V., Samuels, S. and Ross, L. (2004): "The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner's Dilemma Game Moves." Personality and Social Psychology Bulletin **30**(9): 1175--1185.

Maier-Rigaud, F. P., Martinsson, P. and Staffiero, G. (2005). Ostracism and the Provision of a Public Good, Experimental Evidence, Max Planck Institute for Research on Collective Goods.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." American Economic Review **93**(1): 366--380.

Milinski, M., Semmann, D., Krambeck, H.-J. and Marotzke, J. (2006). "Stabilizing the Earth's Climate Is Not a Losing Game: Supporting Evidence from Public Goods Experiments." Proceedings of the National Academy of Sciences **103**(11): 3994-3998.

Nikiforakis, N. (2008). "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" Journal of Public Economics **92**(1-2): 91 -- 112.

Nikiforakis, N. (2009). Feedback, Punishment and Cooperation in Public Good Experiments.

University of Melbourne.

Nikiforakis, N. and Normann, H.-T. (2008). "A Comparative Statics Analysis of Punishment in Public-Good Experiments." Experimental Economics **11**(4): 358--369.

Ostrom, E. (1998). "A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997." The American Political Science Review **92**(1): 1-22.

Ostrom, E. (2000). "Collective Action and the Evolution of Social Norms." Journal of Economic Perspectives **14(3)**: 137--158.

Ostrom, E., Gardner, R. and Walker, J. A. (1994). Rules, Games, and Common-Pool Resources. Ann Arbor: The University of Michigan Press.

Ostrom, E., Walker, J. M. and Gardner, R. (1992). "Covenants with and without a Sword: Self-Governance Is Possible." The American Political Science Review **86**(2): 404--417.

Potters, J., Sefton, M. and Vesterlund, L. (2005). "After You--Endogenous Sequencing in Voluntary Contribution Games." Journal of Public Economics **89**(8): 1399-1419.

Reeson, A. F. and Tisdell, J. G. (2008). "Institutions, Motivations and Public Goods: An Experimental Test of Motivational Crowding." Journal of Economic Behavior \& Organization **68**(1): 273--281.

Rege, M. and Telle, K. (2004). "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." Journal of Public Economics **88**: 1625--1644.

Rockenbach, B. and Milinski, M. (2006). "The Efficient Interaction of Indirect Reciprocity and Costly Punishment." Nature **444**(7120): 718--723.

Ross, L. and Ward, A. (1996). Naïve Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. Values and Knowledge. Reed, E. S., E.Turiel and Brown, T., Lawrence Erlbaum (Mahwah, NJ)**:** 103--135.

Rutte, C. G., Wilke, H. A. M. and Messick, D. M. (1987). "The Effects of Framing Social Dilemmas as Give-Some or Take-Some Games." British Journal of Social Psychology **26**: 103--108.

Selten, R., Mitzkewitz, M. and Uhlich, G. R. (1997). "Duopoly Strategies Programmed by Experienced Players." Econometrica **65**(3): 517--555.

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D. and Milinski, M. (2007). "Gossip as an Alternative for Direct Observation in Games of Indirect Reciprocity." Proceedings of the National Academy of Sciences **104**(44): 17435-17440.

Sonnemans, J., Schram, A. and Offerman, T. (1998). "Public Good Provision and Public Bad Prevention: The Effect of Framing." Journal of Economic Behavior & Organization **34**(1): 143--161.

Sutter, M., Haigner, S. and Kocher, M. G. (2005). Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations.

Tyran, J.-R. and Feld, L. P. (2006). "Achieving Compliance When Legal Sanctions Are Non-Deterrent." Scandinavian Journal of Economics **108**(1): p135 - 156.

Vesterlund, L. (2003). "The Informational Value of Sequential Fundraising." Journal of Public Economics **87**(3-4): 627-657

Yamagishi, T. (1986). "The Provision of a Sanctioning System as a Public Good." Journal of Personality and Social Psychology Review **51**(1): 110--116.

Appendices

A1. Information on the course of the seminar

Welcome to our simulation-game seminar “social dilemma”!

During this semester, you will have the opportunity to participate in a business-game type of seminar. First, you will gain experience in interacting with other seminar participants in the particular “game situation”. Afterwards, you will develop “sets of rules” for the game situation within groups, gain experience with implementing these rules, and have the opportunity to refine these rules.

Schedule:

1st appointment, 19th april, preliminary meeting:

> Today we will clarify two things, above all: which tasks you are expected to comply with in this seminar, and how your grade is going to be composed based on this.

2nd appointment, 26th april, 1st round of play:

> After being handed out the instructions for the game, you will gain first experience with playing the game in randomly allotted groups. Furthermore, you will be randomly divided into groups of 4 in which you will have the opportunity to successively develop a set of rules for the game situation during the weeks to come. For simplicity, we will call these sets of rules “virtual rooms” (vRoom). Please appoint a contact person for your group by 3rd May, whose contact details are given to us for possible enquiry calls.
>
> Important note: You will be randomly allotted to a vRoom in each round of play, and you will not know with whom you are interacting in a vRoom in any of the rounds.

3rd appointment, 3rd May, 1st closing date:

> Handing in the rule set by midnight of the respective day; rules are to be handed in including a documentation in which you shortly (max. 2 pages) expose your deliberations on the set of rules handed in.

4th appointment, 10th May, 2nd round of play (single-weightet):

First-time play in the newly-designed vRooms; you will, again, be randomly assigned to your respective vRoom, and therefore, you cannot count on playing with the same people as the week before, nor with other members of your rule group.

5th appointment, 17th May, Ascension Day; handing-in of the 2nd rule set

6th appointment, 24th May, 3rd round of play (single-weighted)

7th appointment, 31st May, handing-in of the 3rd rule set

8th appointment, 7th June, 4th round of play (double-weighted)

9th appointment, 14th June, handing-in of the 4th rule set

10th appointment, 21st June, 5th round of play (final round) (triple-weighted)

11th appointment, 12th July, final discussion in the large group

12th appointment, 31st July, handing-in of the field reports

Your tasks:

- participating on all appointed days is an absolute necessity for this seminar, as otherwise there cannot be any play in at least one vRoom! In case of severe illnesses, we ask for the earliest possible notification!
- active participation in 5 rounds of play
- active participation in designing a rule set for the game situation. Additionally, a short motivation for your rule set. (Group task)
- preparation of a seminar paper in which you analyse your experience made in the seminar and in which you concludingly evaluate your rule set. (Individual task)

Composition of your grade:

- seminar paper and documentation of the rule sets: 40%
- individual play: 30%
- performance of your group's rule set in the efficiency tournament: 20%
- final discussion: 10%

The efficiency tournament of rule sets (vRooms):

Efficiency will be defined as follows: the sum of individual payoffs within the vRoom minus rule-set costs as a fraction of the sum of maximum possible payoffs without rule set. I.e., if players "do very well" under your rule set, you will earn points in the efficiency tournament of vRooms. In order not to give you any incentive to play "worse" in vRooms of others, your individual payoffs from the round of play are also taken into account for your grade.

Note, furthermore, that the efficiency that emerges in your vRoom during rounds of play 2 and 3

will contribute to your overall score in the efficiency tournament single-weightedly, while efficiency gained in rounds 4 and 5 will be weighted twice and three times, respectively. This is not the case for individual payoffs.

What is a “rule set”, what does its implementation cost, and which rule sets are feasible?

To start with, a “rule set” is everything that could be introduced to induce the players to perform actions that increase the total sum of contributions in the corresponding vRoom. Exceptions are technically unfeasible changes (or those that are feasible only under prohibitively large costs, as e.g. to transfer the laboratory to the attic, providing it with a glass cupola), as well as ethically questionable practices (shoot other players, cut their hands off). Additionally, rules are excluded, that make inferences about the true identity of players feasible.

As in “real life”, there is one thing to think about: “there ain’t no such thing as a free lunch” – every rule to be introduced gives rise to specific costs that will be taxed by us according to the difficulties the introduction/implementation of the rule would entail in “real life”.

In addition to the dates appointed above, each group should arrange with Mr Wolff for a consultation hour on the Tuesday or Wednesday preceding each handing-in of the rule sets to discuss the costs of any potential rule set and to find out whether specific rule changes are not possible due to technical, ethical, or other reasons, and to thus be able to change the rule set accordingly in this case. It is up to you to decide on how many group members go to these consultation meetings.

A2. Instructions for the basic game

Instructions for the experiment

General Information:

At the beginning you will be randomly assigned to **3 groups of 4 participants**. You will not be informed about the identity of the other group members.

Course of Action:

In every round, you will be given an endowment of 20 points you can invest in a common project. You have to decide how many of the 20 tokens you are going to contribute to the project. You will keep the remaining tokens.

Calculation of your payoff in stage 1:

Your period payoff consists of two components:

- **tokens you have kept** = endowment – your contribution to the project
- **earnings from the project** = 1.6 x sum of the contributions of all group members / number of group members

Thus, **your period payoff** amounts to:

20 – your contribution to the project
+ 1.6 x sum of the contributions of all group members / number of group members

The earnings from the project are calculated according to this formula for each group member. The total payoff from the experiment is composed of the sum of period payoffs from all 25 rounds. Payoff scores will remain anonymous, i.e. no participant will be informed of the payoff score of any other participant.

Please notice:

Communication is not allowed during the whole experiment. If you have a question please raise your hand out of the cabin. We will then come to you and answer your question privately.

Good luck!

Table B1: summary of all rule sets with elicited contribution levels, achieved efficiency and costs, according to seminar, group, and tournament round

Seminar	**Grp**	**Tournament 1**	**Contributions**	**Efficiency**	**Costs (variable costs)**	**Tournament 2**	**Contributions**	**Efficiency**	**Costs (variable costs)**
1	1	- endogenous state authority - punishment - endogenous redistribution - ID	825	2277	218 (158)	- (lagged) punishment after warning	668	1747	654 (454)
	2	- trial rounds - punishment - redistribution	550	1968	362 (62)	- redistribution of rule costs - no feedback (but own payoff)	641	2085	300 (0)
	3	- punishment - ID	789	1625	848 (648)	- roundly varying background stories - individual messages	533	2270	50 (0)
2	4	- punishment - ID - individual messages	1437	2302	560 (310)	- punishment - redistribution - individual messages - big bonus	1795	2822	255 (5)
	5	- punishment - redistribution	361	1915	302 (102)	- minimum-effort - appeal	1725	2583	452 (252)
	6	- punishment - appeal	597	1829	530 (280)	- no feedback - 'big malus' for minimum-contributor if welfare < 2500 (before rule-costs) - appeal	1271	2263	500 (0)

Seminar	Grp	Tournament 3	Contributions	Efficiency	Costs (variable costs)	Tournament 4	Contributions	Efficiency	Costs (variable costs)
1	1	- redistribution - ID	734	2150	290 (90)	- redistribution + redistribution of rule-costs - ID	1449	2661	208 (8)
	2	- (endogenous) redistribution of rule-costs administered by high-contributors - ID	627	2076	300 (0)	- redistribution of rule costs (every 4th round) - vote on institution	978	2486.0	101 (0)
	3	- individual messages (changing every round)	703	2372	50 (0)	- cheap-talk contract (full contribution; no longer available after 3rd breach) - roundly appeals - individual feedback	1411	2796.6	50 (0)
2	4	- punishment - redistribution - individual messages - big bonus	1091	2291	364 (114)	- redistribution - big bonus - example in instructions (msg)	2000	2850.0	350 (0)
	5	- minimum-effort - punishment - redistribution - appeal	2000	3000	200 (0)	- cheap-talk contract offered every 5th round: endogenous 'minimum-contribution' - individual messages	418	2200.8	50 (0)
	6	- no feedback - 'big malus' for minimum-contributor if welfare < 3000 (before rule-costs)	1547	1828	1100 (600)	- appeals every 5th round	390	2184.0	50 (0)

Legend (fixed rule costs):

appeal (50)	A statement is displayed to subjects before they play the game. The statement is possibly repeated every *k* rounds. Usually, the statement would appeal to moral values and/or a group sentiment.
big bonus (0)*	In every playing round, points are deducted from the players conditional on their contribution behavior and transfered into a 'jackpot account' applying the formula for redistribution. At the end of the 25 rounds, the 'jackpot' is awarded to the highest-contributing player, with an equal-split rule in case of a tie.
'big malus' (100)*	The lowest-contributing player is punished by 200 points if the group does not achieve a certain welfare benchmark. Costs are calculated as if the player is deducted 8 points each round (Total welfare cost in case of application: 200 + 400 = 600 points).
cheap-talk contract (50)	Players are given the opportunity to vote on a 'contract' requiring them to contribute a specified amount of tokens. The contract is 'enacted' only in case of unanimous consensus. The 'enactment' does not have any material consequences. In case of the endogenous 'minimum-contribution' levels, players were meant to vote on increasing the initially specified level of 15 tokens by one token in case of full compliance.
endogenous state authority (0)	Individual contributions are disclosed if the group invests 4 points into a secondary public good 'administration'; individual contributions to this 'administration' cannot surpass 2 points. If the 'administration' is 'constituted', its contributors may engage in peer-to-peer punishment (cost schedule as in *punishment*).
example in instructions	An example specifying the payoffs of a single free-rider amongst full-contributors is contrasted to the payoff of a full contributor amongst equals.
ID (200)	Players are given fixed ID numbers and informed of individual contributions of all players by ID number.
individual messages (50)*	Pre-defined messages, automatically displayed to players conditional on their behavior in the game.
minimum-effort (0)*	'Redistributing' back 'surplus contributions': any tokens contributed by a player *i*, which surpass the smallest contribution made by a group member are returned to *i*, diminished according to the redistribution formula, before the sum of contributions is multiplied by the public-good factor.
no feedback (but own payoff; 0)	Players do not receive any feedback (except on their own payoff), not even about the sum of contributions.
punishment (0)*	Deduction of *x* points, $x < 9$, from a player's income from the public-good according to the formula: $Cost(x) = x^2/4$. These costs are born by all players if punishment is administered automatically, and by the punishing player in case of peer-to-peer punishment. Higher deductions are possible, but no perfect punishment automata are available: in this case, players either have to administer punishment themselves, or non-contributions are punished only with a certain probability. The corresponding automata have additional fixed costs: High probability punishment (probability of deduction = 1/2): 600 points Low probability punishment (probability of deduction = 1/10): 200 points
(lagged) punishment after warning (0)*	A warning is issued to a player who does not comply with a specified condition. In case of repeated non-compliance, the player is

	punished automatically as described under *punishment*.
(endogenous) redistribution (0)*	Transfer of points from one player to another/others. The points available for allocation, *y*, as a function of points deducted, *x*, are determined according to the following formula: *y = sqrt(x - 1) + 1*. In the case of endogenous redistribution, individual players (determined in correspondence to their contributions) may decide on the reallocation of points.
redistribution of rule costs (0)*	In contrast to the default case, rule costs are allocated unequally among players. Under rule sets of this category, players are told they have to bear different shares of the rule costs depending on their contribution behavior.
roundly varying background stories (50)	The public-good is framed in a different way in each round. An example for a framing would read: "In your neighborhood, cats are poisoned more and more often. Residents of the area are worried that the rat poison laid-out may also be eaten by dogs or even by little children. They are forming a vigilante group in charge of combing through the streets for irregularities and to collect the poison. As you do not have the time to participate in these activities, your family is thinking about donating € 20 for support. Do you want that?", followed by a yes-or-no decision. In case of a "no", the story goes on: "It could be your little daughter who swallows such a toxic bait. And do you want to lock up your cat the whole day long? The vigilante group is a good thing, how much do you want to give them? It's for the safety of your family?"
trial rounds (100)	Players experience the rule set for a small number of rounds without payoff consequences before they play the game for 25 rounds.

*Rule features marked are feasible only in conjunction with monitoring; a monitor requires costs of 200 points, independent of whether she announces the observed contributions as in *ID* or not.

Table B2: Classification of rule sets according to the characteristics described in section 3; the table specifies for each rule set whether it displays (1) a certain characteristic or not (0).

Tournament	Seminar	Group	pun	punEP	noRedEnd	noBB	noID	FB	noComm	frame	noFrameEP	noEnd	minEff°
1	1	1	1	1	0	1	0	1	1	0	1	0	0
		2	1	1	1	1	0	1	1	0	1	1	0
		3	1	1	1	1	0	1	1	0	1	1	0
	2	4	1	1	1	1	1	1	1	1	1	1	0
		5	1	1	1	1	1	1	1	0	0	1	0
		6	1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	0	1	1	1	1	1	0	1	1	0
		2	1	1	1	1	1	1	1	0	1	1	0
		3	0	0	1	1	1	1	1	1	0	1	0
	2	4	1	1	1	0	1	1	1	1	0	1	0
		5	0	0	1	1	1	1	1	1	1	1	1
		6	1	0	1	1	0	0	1	1	1	1	0
3	1	1	1	1	1	1	0	1	1	0	1	1	0
		2	1	0	0	1	0	1	1	0	1	0	0
		3	0	0	1	1	1	1	1	1	0	1	0
	2	4	1	1	1	0	1	1	1	1	0	1	0
		5	1	1	1	1	1	1	1	1	1	1	1
		6	1	0	1	1	0	0	1	0	1	1	0
4	1	1	1	1	1	1	0	1	1	0	1	1	0
		2	1	0	1	1	1	1	1	0	1	0	0
		3	0	0	1	1	1	1	0	1	0	0	0
	2	4	1	1	1	0	1	1	1	1	1	1	0
		5	0	0	1	1	1	1	0	1	1	0	0
		6	0	0	1	1	1	1	1	1	1	1	0

° The category *minEff* is only listed for completeness. Those rule-sets displaying a “1” in this category were eliminated for our analysis, and the category was deleted.