I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18075

# Facing Inflated Rules – Experimental Evidence from Threshold Public Goods Games

Christian Grund
Philipp Monschau

AUGUST 2025

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18075

# Facing Inflated Rules – Experimental Evidence from Threshold Public Goods Games

**Christian Grund**
*RWTH Aachen University and IZA*

**Philipp Monschau**
*RWTH Aachen University*

AUGUST 2025



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## Facing Inflated Rules – Experimental Evidence from Threshold Public Goods Games

We study the role of purpose-based rules for behavior and outcomes in a threshold public good game. Rules can be sufficient or even inflated in terms of proposing a fulfilling behavior. We conduct a lab experiment to describe the implications caused by the inflation of a rule. Our study shows that inflated rules are obeyed less. Yet, rule-following occurs also with inflated rules which leads to lower efficiency regarding exactly providing the threshold. A fair share option can help to coordinate efficiently. We complement our analysis by the investigation of the role of the implemented rules for the ex-post optimal behavior, i.e. evaluating the individual contribution depending on the individual payoff.

**JEL Classification:** C9, H41, M5

**Keywords:** rule-following, coordination, cooperation, groups, thresholds, public goods

**Corresponding author:**
Christian Grund
RWTH Aachen University
School of Business and Economics
Chair of Human Resource Management and Personnel Economics
Templergraben 64
52056 Aachen
Germany

E-mail: christian.grund@hrm.rwth-aachen.de

# 1 Introduction

Most people will agree that living together demands behavioral rules. These rules are intended to guide actions of individuals to secure an outcome that benefit either the group or society as a whole. Rules become relevant or effective when they actually prescribe a behavior that an individual would not have chosen independently. In such instances, the individually desired behavior is in conflict to the socially desired one. Consequently, it can be costly for individuals to follow such rules (Fischbacher & Föllmi-Heusi, 2013; Kimbrough & Vostroknutov, 2016; 2018). Nonetheless, there are several potential reasons why people still follow such rules such as intrinsic respect for rules, extrinsic incentives, social expectations and consequences and accept those costs (Gross & Vostroknutov, 2022; Gächter et al., 2025). Rules may have a purpose to reach a distinct objective, although even arbitrary rules are followed (Gross & DeDreu, 2021) due to the presence of social norms that encourage compliance (Bicchieri, 2005). Sometimes they shall prevent people from generating negative externalities for others (e.g. environmental pollution). In other situations, they force beneficial actions (e.g. flowing traffic). Thus, the individual cost on the one hand stands in opposition to a social benefit on the other hand.

Often, rules request a behavior that leads to a kind of overachievement of the objective to ensure its fulfilment. This overachievement then may not increase benefits, but leads to inefficient levels of individuals' costs. We refer to such rules as "inflated rules". Rule inflation can be seen as a validation against accidental inattention or conscious rule violation by individuals within a group. Therefore, one reason for inflated rules could be a possible trade-off between a larger probability for a successful achievement of the objective and potentially higher costs leading to inefficiency.

This general framework is highly relevant in the business context. For instance, consider a supervisor in a firm: A certain order by an important customer of the firm has to be finished until the end of the week. To ensure timely completion, she requests additional working effort of all workers of her team until the end of the week without having the chance to monitor or sanction individual behavior in detail. If this rule is followed by all team members, she can be sure that the order will be finished on time. However, this objective could be reached also by slightly less employee effort. The rule is thus inflated insofar as it allocates an inefficient amount of effort toward this specific task if all employees follow the rule. As a consequence, other relevant tasks may be neglected or workers may need to work overtime.

Our work contributes to the existing literature by focussing on non-arbitrary, potentially inflated rules. The main objective of this paper is to explore the behavior of individuals and consecutive outcomes if inflated rules are present. We experimentally apply a Threshold Public Goods Game (TPGG) (Marwell & Ames, 1979; van de Kragt et al., 1983) and examine whether thresholds are reached with a higher likelihood if rules are inflated (success rate) and to what extent contribution behavior is efficient in the sense that they do not or do exceed the threshold and, therefore, "waste" contributions. In doing so, we connect general rule-following with coordination tasks (Cooter, 1998; Bohnet & Cooter, 2003) and also complement the existing literature on different interventions in (Threshold) Public Goods Games (Croson & Marks, 2001; Galbiati & Vertova, 2008; Chaudhuri & Paichayontvijit, 2017; Johnson & Kecinski, 2025). We examine this possible success-efficiency trade-off by investigating differences among our different treatments which incorporate varying rules and thresholds. Reaching a common objective represents a coordination game. Thus, it is likely relevant whether a situation has focal points (Schelling, 1958) which create salience (Mehta, Starmer & Sugden, 1994; Crawford, Gneezy & Rottenstreich, 2008) for successful coordination. On the individual level, inflation may also cause incentive problems: The additional demand in behavior opens up opportunities for beneficial deviation from following the rule. Even if one does not obey the rule, the objective can still be achieved.

Our experiment yields meaningful results: The presence of rules leads to reduced variance in contribution behavior among participants. Variations in outcomes primarily emerge in the efficiency dimension between inflated rules and their opposite, the sufficient rules. Surprisingly, the threshold is reached with similar frequency across all treatments. Rule-following behavior is less pronounced, if rules are inflated. Nonetheless, efficient contribution behavior is rarer, when rules are inflated. Next to these results, we investigate the effect of the specific situation where the threshold amount can be equally distributed among all group members (so-called fair share option). We find that the threshold is met without wasting additional contributions more often when costs can be split equally. Finally, we assess optimal ex-post contribution behavior at the individual level. Despite concerns about potential beneficial deviations from following the rule, saving tokens is not always the payoff-maximizing strategy in the presence of inflated rules. We find evidence suggesting that also inflated rules can create situations where cooperative behavior (in terms of at least contributing one's fair share) is optimal.

The remainder of the paper is structured as follows: First, we outline the theoretical framework including the model and key definitions in chapter 2 before detailing our experimental design in chapter 3, explaining the procedure, the different treatments and deducting hypotheses from those. Chapter 4 presents our results in relation to these hypotheses before briefly discussing findings alongside limitations inherent in our design within chapter 5.

## 2 A simple Threshold Public Good framework

Following Cadsby et al. (2008) we consider a set $\mathcal{N}$ of $n$ players. Each player $i = 1, \dots, n$ has an endowment of $e_i > 0$ tokens from which they can contribute tokens to the group account. For simplicity, we do not multiply contributions by a cost factor but instead consider the costs as 1; thus, we denote the individual contribution choice as $c_i$. The threshold for public good provision is $T > 0$. Therefore, if $C = \sum_{i=1}^{n} c_i \geq T$, the public good is provided. Reaching the threshold yields additional payoffs $v$ for each player, which can be expressed as:

$$\pi_i = \begin{cases} e_i - c_i & if\ C = \sum_{i=1}^{n} c_i < T \\ e_i - c_i + v & if\ C = \sum_{i=1}^{n} c_i \geq T \end{cases} \tag{1}$$

It is important to note that contributions exceeding the threshold do not give additional utility (no refund), while contributions falling short of the threshold result in a complete loss (no rebate).

When a rule is announced, each individual receives a message containing a specific rule, referred to as $r_i$. This specific rule indicates how much player $i$ should contribute in the TPGG. The message also contains information about the other players' rules. Since the main component of the messages is the integer value, we can summarize the requested contributions of a group (group rule) as a sum of the individual rules: $R = \sum_{i=1}^{n} r_i$ . There are numerous combinations of individual rules adding up to the requested sum $R$. Hence, we represent the set of individual rules within a group as a vector $\bar{R} = (r_1, \dots, r_n)$.

The following definitions aim to clarify the setup, arguments and interdependencies among terms and constructs:

Rules

In our work, we define a *rule* as a prescription of a distinct behavior. This prescription constitutes cheap talk in the sense that it does not come along with a (probabilistic) penalty in case of a rule violation. We have previously referred to the concept of purpose. We are mainly interested in rules that, if obeyed, lead to an achievement of the purpose, i.e. rules that are *effective* ($R \geq T$). We distinguish between *sufficient* rules ($R = T$) and *inflated* rules ($R > T$). This categorization happens on the group level, i.e. an individual rule $r_i$ cannot be sufficient nor inflated; only the sum of individual rules $R = \sum_{i=1}^{n} r_i$ may fall into either category. The magnitude of the individual rule does not determine its classification regarding sufficiency or inflation.

Fair share option

The main differences between regular Public Good Games and Threshold Public Good Games consists in the relevance of a threshold. We distinguish between two types of thresholds depending on whether they can be divided evenly among group members without remainder. We call this a *fair share option*. In such cases, there is a discrete value that every player can contribute so that their total exactly sums up to meet the threshold. If no fair share option exists, then achieving precisely this threshold requires some heterogeneity in contributions across players.

Categories of group outcomes

For an evaluation of the group outcome for a single TPGG, we can identify two general states:

*Failure*: The group contributions are less than the threshold which means that the public good is not provided and nobody receives the extra payoff $v$.

*Success* is the opposite state. In that case, the threshold was reached by the group contributions and the public good is provided, so everybody receives the extra payoff $v$.

In case of success we distinguish between two situations regarding the question, whether the threshold is exactly met.

*Efficiency*: The group contributions are equal to the threshold which implies that both the public good is provided and no contributions are wasted. Therefore, the payoff for the whole group is maximized.

*Excess*: The group contributions are higher than the threshold which indicates that although the public good is provided but the contributions that were paid additionally are wasted ($C - T > 0$). Therefore, the payoff for the whole group is not maximized.

Individual decisions

*Rule-following*: For each individual, we will be able to determine whether or not they have followed the rule it was presented to them. Even though the rule specifies a distinct number of tokens that shall be contributed, we consider the specified amount more like a lower boundary for contributions, i.e. with all contributions equal or above the value specified in the rule, the behavior is considered as a rule-following behavior. Consequently, all contributions below the requested value imply a rule violation.

## 3 Experimental Design

We apply the TPGG setup described above to a lab experiment and specify the design including the experimental process and parametrizations as well as different treatments. Because our analysis is built mainly on treatment differences, we derive our hypotheses only after their introduction at the end of this chapter.

### 3.1 Procedure

Figure 1 illustrates the main procedure of our experiment, which we programmed in o-Tree (Chen et al., 2016). After subjects are randomly assigned to seats in the experimental laboratory, explanations are given and participation is conditioned on correctly answered control questions.

After that, the subjects were randomly assigned to a distinct treatment in which they were assigned to groups of four ($n = 4$). They faced a threshold public good game and played a total of 20 rounds, divided in two phases of ten rounds each. The constant threshold that they encountered was common knowledge. Every subject received an endowment of ten tokens per round ($e_i = 10 \ \forall i$). Since we apply a constant step return of 2, we defined the valuation of the public good as $v = T/2$. Subjects only were told how it is calculated and it was not stated directly in the decision situation. However, the subjects had to answer control questions in the beginning which were pointing out the valuation mechanism and which were adjusted for the respective threshold. After each round, subjects got feedback on how much was contributed by the whole group and how much they earned in that round.

The first phase of the experiment (rounds 1 to 10) does not include a rule. In the second phase, subjects continue to play the same TPGG, but initially received a message that contained the rule for the entire group with the individually highlighted personal rule. Rules differ across treatments that we are going to explain in chapter 4.2.

For example, the slide including the rule, as Player 3 saw it, was presented like this (translation from the German original):

"In the following 10 rounds (phase 2), the following rules are presented to all players as kind of a guideline. If all players contribute the amount which is specified in the rules, a total of 22 tokens will be collected, thus reaching the threshold.

The rules for all players are as follows:

Player 1: Contribute 5 tokens.

Player 2: Contribute 5 tokens.

**Player 3 (You): Contribute 6 tokens.**

Player 4: Contribute 6 tokens."

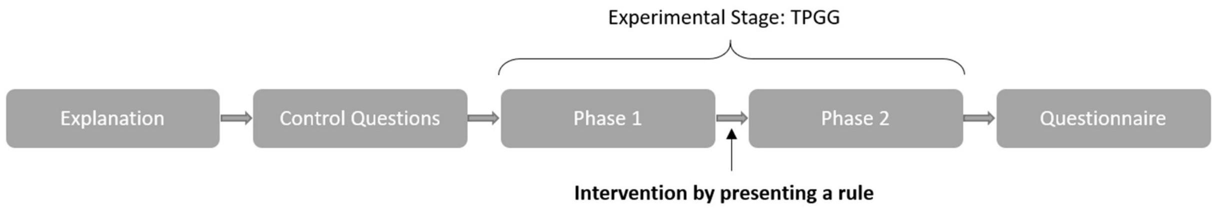
In each round 11 to 20, they now faced the decision situation as before plus their personal rule again that does not differ across rounds. For our example, for a group that had encountered a threshold of $T = 20$ and now receives an inflated rule we refer to it with the following parametrization: $R = 22 > T$; $\bar{R} = (5, 5, 6, 6)$. It is noteworthy that subjects face the identical individual and group rule over all rounds 11 to 20.

After the last round, subjects learned which phase was randomly chosen as being payoff relevant. For that phase, they saw how much they earned throughout the entire phase and how much this earning is worth in terms of money (0.05 € per token). Subjects received a fix payment of 6 € plus their tokens from the selected phase converted into money.

After everybody finished the decision stage, the experiment proceeds to a questionnaire that elicited various demographics, perceptions of the rule and their social appropriateness[1], beliefs about the behavior and the perception of others and the big five personality traits (Soto & John, 2017; Rammstedt et al., 2018).

[1] The questions concerning social appropriateness followed the framing of Krupka & Weber (2013) but excluding the underlying incentive scheme.

Figure 1: Experimental Procedure




We conducted data of 220 subjects from the XXX laboratory subject pool in XXX. Due to the restrictions to have groups of four (caused by the design), we had to conduct 19 sessions between August and October 2024. Most of our sample were students (211), 58.18 % of all were men. 44 subjects (= 11 groups) were allocated to each treatment, respectively. The allocation was adequately randomized across treatments concerning the demographics (gender: $\chi^2 - \text{test}, p = 0.638$; age: Kruskal-Wallis test, $p = 0.777$).

### 3.2 Treatments

We designed four different treatment groups which vary in the threshold and the rule in a 2x2-design. During phase 1 treatments only differ in the threshold. We implement two threshold levels $T = 20$ or $T = 22$, implying the incidence of a fair share option in the first case, and the absence of it in the second. In phase 2 we introduced rules that are either sufficient or inflated. In detail, in treatments with $T = 20$, we present either a sufficient rule with $R = 20$ and $\bar{R} = (5,5,5,5)$ or inflated rule with $R = 22$ and $\bar{R} = (5,5,6,6)$. $R = 22$ and $\bar{R} = (5,5,6,6)$ are also used as parameters regarding a sufficient rule in the presence of $T = 22$. The corresponding inflated rule for the threshold value of T=22 is $R = 24$ with $\bar{R} = (6,6,6,6)$. Additionally, we added a control treatment with T=22 that did not encounter any rule. We assign names to our treatments as described below and additionally summarize parameters in Table 1.

(1) FAIRSUFF: The first treatment has a threshold with a fair share option. In the second phase, the subjects encounter a for this threshold sufficient rule.
$T = 20; R = 20$ with $\bar{R} = (5,5,5,5)$

(2) FAIRINF: The second treatment has a threshold with a fair share option. In the second phase, the subjects encounter a for this threshold inflated rule.
$T = 20; R = 22$ with $\bar{R} = (5,5,6,6)$

(3) UNFAIRSUFF: The third treatment has a threshold without a fair share option. In the second phase. Subjects encounter the identical rules as in FAIRINF, which sufficient in this treatment, though.

$T = 22; R = 22$ with $\bar{R} = (5,5,6,6)$

(4) UNFAIRINF: The fourth treatment has a threshold without a fair share option. In the second phase, the subjects encounter a for this threshold inflated rule.

$T = 22; R = 24$ with $\bar{R} = (6,6,6,6)$

(5) CONTROL: The control group has a threshold without a fair share option. In the second phase, the subjects encounter no rule at all.

$T = 22; R = \emptyset$

Table 1: Treatments

| | FAIRSUFF (1) | FAIRINF (2) | UNFAIRSUFF (3) | UNFAIRINF (4) | CONTROL (5) |
|---|---|---|---|---|---|
| Phase 1 | $T = 20$ No rule | $T = 20$ No rule | $T = 22$ No rule | $T = 22$ No rule | $T = 22$ No rule |
| Phase 2 | $T = 20$ $R = 20$ $\bar{R} = (5,5,5,5)$ | $T = 20$ $R = 22$ $\bar{R} = (5,5,6,6)$ | $T = 22$ $R = 22$ $\bar{R} = (5,5,6,6)$ | $T = 22$ $R = 24$ $\bar{R} = (6,6,6,6)$ | $T = 22$ No rule |

### 3.3 Deriving Hypotheses

Although we formulate our hypotheses regarding the specific treatments of our experiment, we still aim to preserve some generalizability in these statements. We argue regarding the relevance of (i) the incidence and type of rules and (ii) the incidence of a fair share option.

*Incidence and type of rules:*

The introduction of a rule is the core feature presented in our work. Since all rules are effective, we assume that individuals are more likely to contribute an adequate number of tokens when facing a rule. Simultaneously, unnecessarily high contributions are largely ruled out. The rules are intended to fulfill both normative and coordinative roles. Many people have preferences for behaving socially appropriately (Bicchieri, 2005); by following a rule, one conforms to a social norm. Thus, for a substantial portion of subjects the rule can ensure a behavior that leads to successful outcomes. At the same time, participants face a coordination problem: The threshold value can be distributed in terms of contributions (and accompanying costs) among the group in numerous ways. A successful (and ideally efficient) outcome on the group level can only be achieved if every player selects distinct actions that match the actions of others players as well as possible. Hence, the rule is supposed to propose a specific distribution that enhances its salience compared to alternative distributions. This proposed distribution leads either to an efficient or excessive outcome since the rule is either sufficient or inflated. We assume that groups not reaching the threshold at all (failure) or not in an efficient way (excess) in the first phase, will shift their strategy. This leads to a higher fraction of successful and efficient outcomes in the second phase. Simultaneously, subjects with already efficient outcomes will likely stick to their strategy choice avoiding waste. Compared to situations without a rule, we thus hypothesize that on average, subjects will reach the threshold more frequently when faced with a behavioral prescription and the outcomes will improve in efficiency terms.

**H1: The incidence of rules leads on average to (a) higher success rates and (b) higher efficiency rates.**

If we distinguish between sufficient and inflated rules, we consider the two goals success and efficiency as kind of a trade-off. It is possible to attempt ensuring success by aiming for higher contributions while disregarding efficiency concerns. This rationale might explain the existence of inflated rules. Therefore, we hypothesize that inflated rules lead to higher success rates than sufficient rules since inflation serves as a safety net against rule-breakers.

**H2: Inflated rules lead to higher success rates than sufficient ones.**

Conversely, if efficiency is prioritized as the main goal, then a sufficient rule may prove more useful as an instrument. Since no extra tokens are solicited, we anticipate that efficiency rates will be higher in treatments with sufficient rules than they are with inflated ones.

**H3: Sufficient rules lead to higher efficiency rates than inflated ones.**

*The incidence of a fair share option:*

Success occurs when the sum of contributions within a group meets or exceeds the threshold. Each player has an incentive to reach the threshold because of the provision of a corresponding additional payoff $v$ for each player. Concurrently, every individual wants to keep as much of their own endowment, since every contribution lowers their own payoff regardless of the group outcome. This is the general (social) dilemma in our setting. We argue that a successful provision of the public good becomes easier if thresholds incorporate fair share options. In cases where efficient contribution behavior prevails, it prevents players from inequality costs (as described by Fehr & Schmidt, 1999). Additionally, this option grants higher salience regarding an effective *and* efficient allocation costs of contribution among the players by payoff symmetry (see Crawford, Gneezy & Rottenstreich, 2008)[2]. The height of the threshold is, in our setting, not crucial for the contribution behavior since the step return (see again Croson & Marks, 2000) is constant among all of our treatments. This implies that the higher threshold value and associated higher costs are compensated with a higher bonus payoff.

The same considerations hold for arguments on efficiency: Absence of fair share options implies there exists no salient option to split the contribution amount among the players efficiently. To be successful, subjects could coordinate on another equal distribution of costs of contribution which then has to exceed the threshold and by that lose efficiency.

**H4: The incidence of a fair share option leads to (a) higher success and (b) higher efficiency rates.**

*Individual behavior:*

A more detailed examination leads us to individual level considerations: We have asserted that people follow the rules within our setting due primarily either normative or coordinative reasons. The latter point warrants closer scrutiny when differentiating between sufficient and inflated rules. Since the threshold value is always common knowledge, subjects might evaluate the rule and its "coordinative value" in reference to the aimed contribution level. A sufficient rule states a distribution that represents a Nash equilibrium, since for everybody a deviation would lead to an income loss (if one deviates downwards, the income loss occurs for all other players, too). An inflated rule, however, does not have this property because the inflation implies the option for every player by deviating to save own contributions but still reach the threshold. Thus, the own payoff can be increased by deviation from the rule as long as the majority of team mates follows their individual rules. In such situations, the rule is explicitly costly (see our definition before). Although experimental literature has shown that even such rules are followed by some people (Fischbacher & Föllmi-Heusi, 2013; Kimbrough &

[2] The salience we refer to here is the payoff symmetry in their paper that makes the label salience quite less effective.

Vostroknutov, 2016; 2018), we hypothesize that the rule-following rate in treatments with inflated rules is lower compared with sufficient rules.

**H5: Sufficient rules lead to a higher rule-following rate than inflated ones.**

Given our presumption that inflated rules are being followed less than sufficient ones, we can also check whether this holds true for specifications related to a fair share option. This option functions as a somewhat "natural" coordination device that suggests a certain allocation of costs similarly to a rule. When this option exists, there is less need for additional interventions like rules. Conversely, if no fair share option is present, there arises an increased necessity for a coordination device. Then, even an inflated rule can provide a benefit by suggesting (and with that pointing it out saliently) a distinct distribution. By that, an inflated rule should be followed more frequently, if there is no other salient distribution like the fair share option. Therefore, we hypothesize that the fair share option has a moderating effect on inflated rule-following behavior.

**H6: The negative effect of inflated rules for rule-following is emphasized, if a fair-share option is present.**

## 4 Results

This chapter is structured as follows: First, we describe the effect of the bare incidence of a rule, including also the effects of differentiation between sufficient and inflated rules. Subsequently, we explore the role of the fair share option. We will conclude with a consideration of the individual perspective regarding rule following and payoffs.

### 4.1 Incidence of rules

Our first hypothesis indicates effectivity and efficiency increasing effects of a rule's incidence. We therefore, start our analysis with a within-subject comparison between phase 1 without and phase 2 with a rule employing one-sided Wilcoxon Signed-Rank Tests. In a first step, we aggregate all treatments with incidence of a rule in phase 2. Beyond that, a comparison between the treated groups (who received a rule) and the control group can provide independency of any possible restart effects. The results from previous rounds are likely to influence the contribution choices. To tackle potential path dependencies within a phase, we calculated means over the 10 rounds of each phase.

Figure 2 and 3 illustrate the results for the success rate and efficiency rate, respectively. In the first phase without a rule, the groups across all treatments (without the control group) reach the threshold in 83 % of cases. In one of three cases, they exactly meet the threshold value. After rules were introduced at the beginning of the second phase, both outcome measures raised in a meaningful way. The success rate is over 10 percentage points higher, which determines a significant increase ($p = 0.003$). The efficiency rate nearly even doubled up to 60 % ($p = 0.001$). It might look like a strong rule effect, but up to here it might stem also from a simple restart. Thus, we have to check if we find similar effects in the control group without any rules. By coincidence, the level of successful rounds in the control group lies below the treated groups during phase 1. The increase by 10 percentage points from phase 1 to 2 is almost the same as in the treatment groups, though. Consequently, the difference between phases is significant, too ($p = 0.031$). Also, the efficiency rate starts on a different level and increases significantly ($p = 0.004$). However, this increase in efficiency is considerably lower compared to the treated groups resulting in an efficiency rate of 32 % which is only half of the treated groups (60 %; Mann-Whitney U Test: $p = 0.023$).

A difference-in-difference analysis reveals that in both measures the increases are not significant from that of the treatments with rule incidence although the increase in efficiency is more than twice as high in treated groups than it is for the control group (28 vs. 13%; Mann-Whitney U Test: success rate: $p = 0.893$; efficiency rate: $p = 0.211$). Hence, these first results are only partly in line with H1 (a) and H1 (b) since increases in success rates appear primarily driven by a restart effect.

Figure 2: Success rate by treated vs. control group (difference in phases)

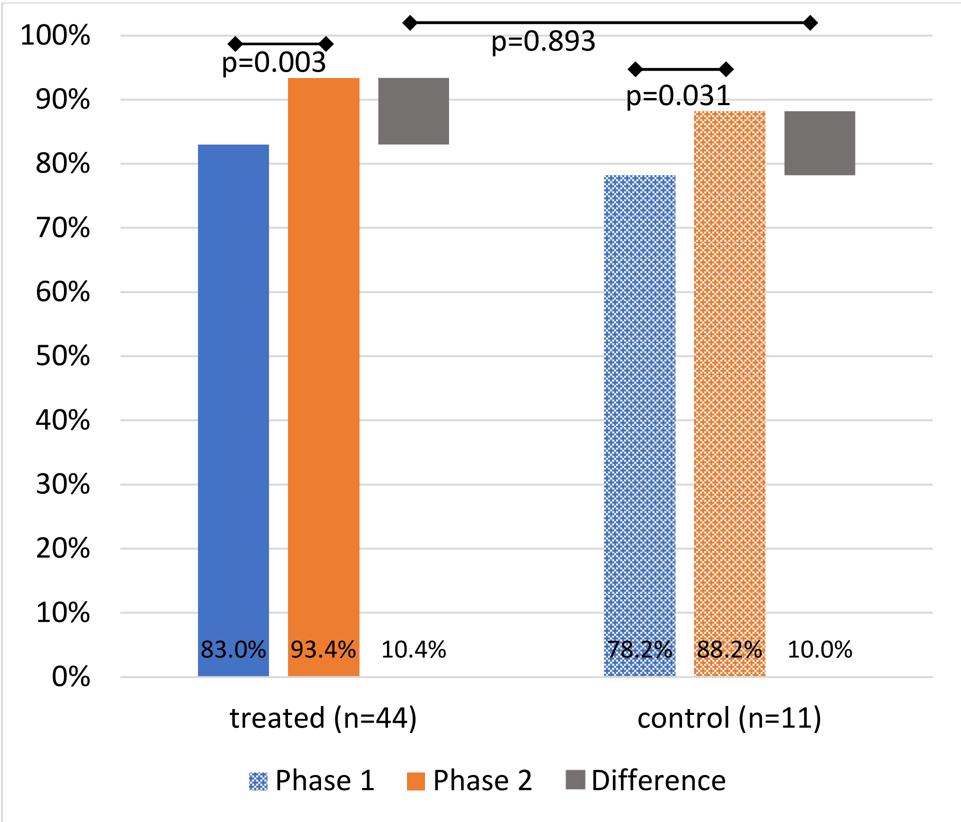
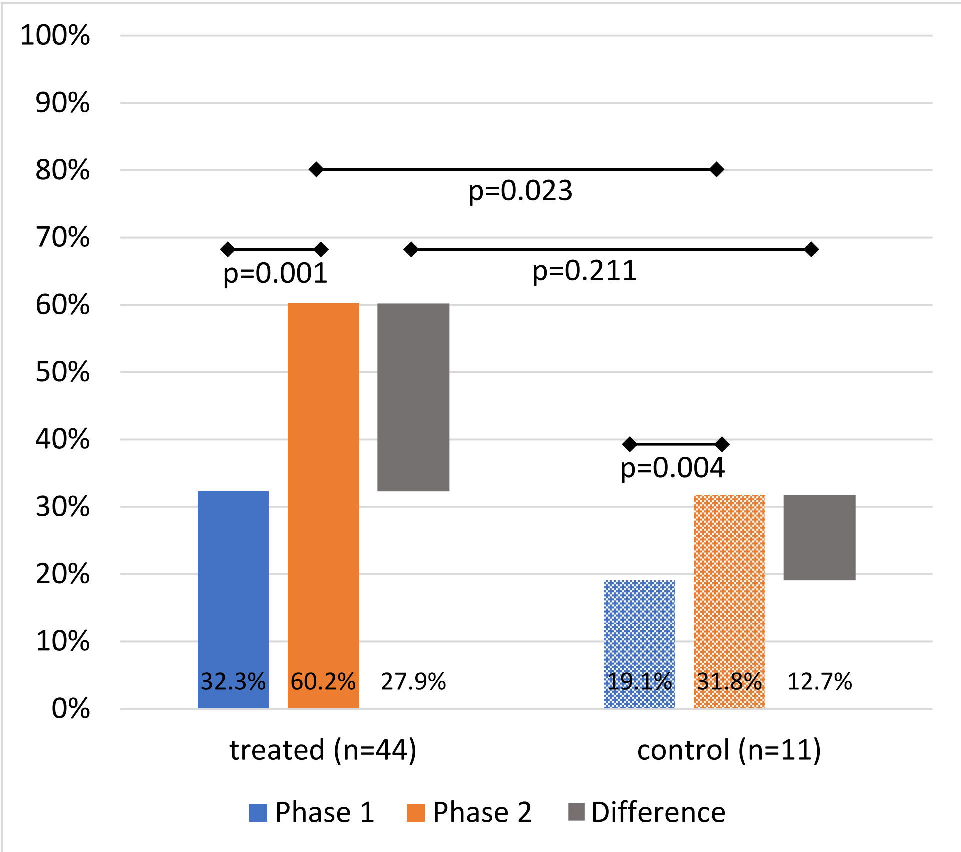
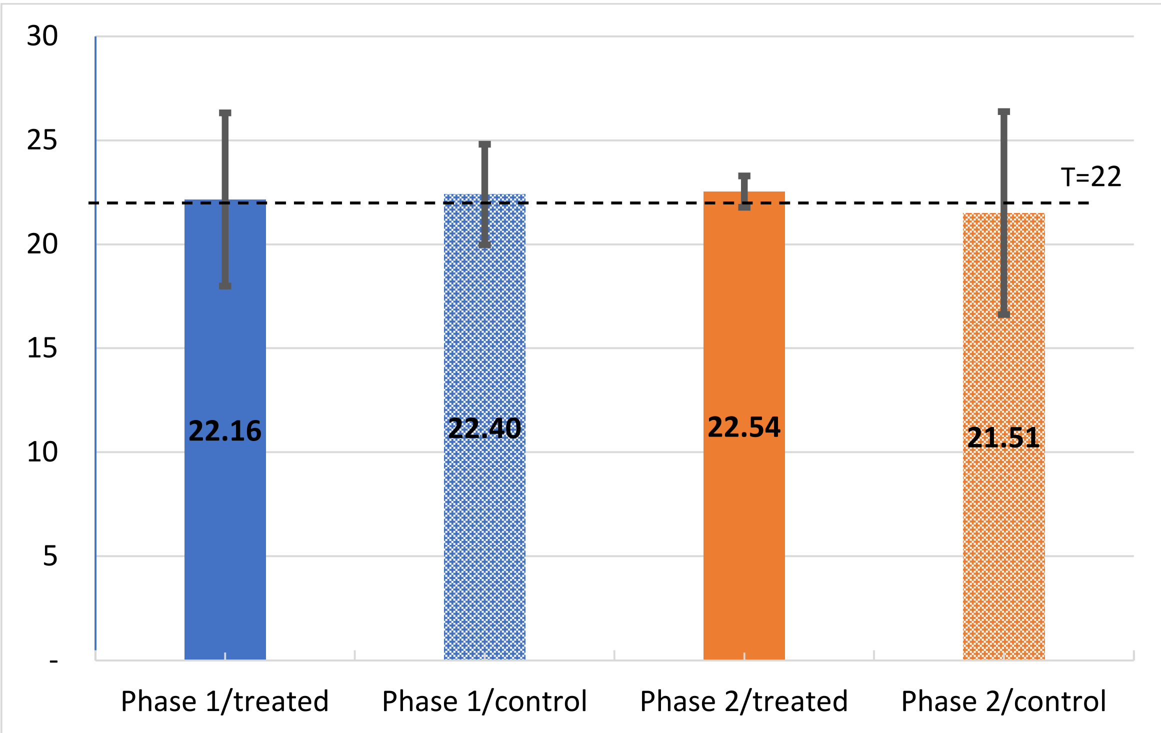



Figure 3: Efficiency rate by treated vs. control group (difference between phases)



The substantial difference in efficiency rates between treated groups and the control group in phase 2 hints for larger heterogeneity in contributions when no rule is present. To investigate this further, we compare the standard deviations of group contributions between treated groups and control group across both phases. Figure 4 illustrates mean group contributions alongside standard deviations by phase. Our analysis focuses on groups facing a threshold of 22, which applies equally to our control group. In phase 1, we observe no significant difference in the standard deviations of group contributions between treated groups (4.16) and the control group (2.42; $p = 0.081$), which is reasonable given that there are no design differences during the first phase. Upon entering phase 2, however, the standard deviation of group contributions decreases for the treated groups that received a rule down to 0.76, while it increases for the control group up to 4.88. Both changes between phases are significant (treated: $p < 0.001$; control $p = 0.037$). Also, the difference within the second phase between the groups is significant ($p < 0.001$) which underlines the coordinating effect of the rules: Contribution behavior varies less in the presence of rules.

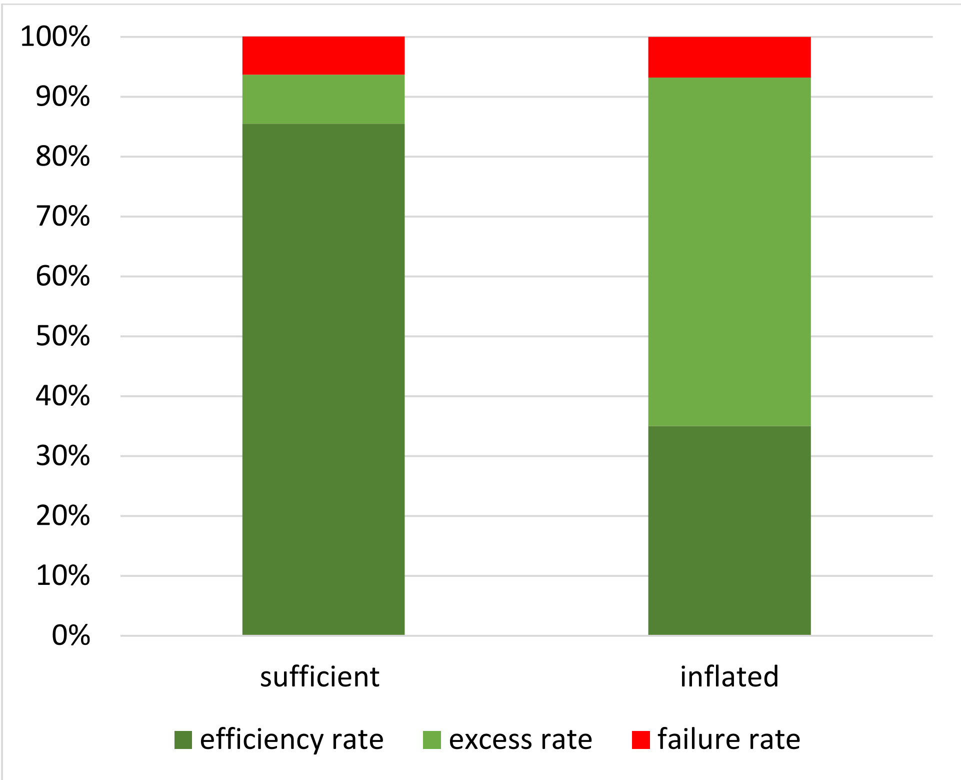
Figure 4: Mean and standard deviation of group contribution by phase and rule incidence (T=22)



Next, we aim to explore differences in outcomes between types of rules. As previously discussed, we discriminate between sufficient and inflated rules. We again aggregate the data of all ten rounds in the second phase (see Figure 5) and run Mann-Whitney U Tests. We do not observe differences in success rates between sufficient rules (93.6 %) and inflated rules (93.2 %; $p = 0.528$). Accordingly, our hypothesized positive effect of inflated rules on reaching the threshold cannot be confirmed and we do not have evidence in line with H2.

We observe clear differences regarding efficiency, though. Sufficient rules lead to an efficiency rate of 85.5 %, which implies that in only 8.1 % of cases subjects contribute too many tokens on aggregate. Facing an inflated rule instead, the efficiency rate is much lower (35.0 %; ($p < 0.001$), which lead to some excess of contributions in the majority of cases (58.2%). These results support H3. Bringing together the latter two results, we do not find convincing evidence for the suspected trade-off between effectiveness and efficiency.

Figure 5: Group outcomes in phase 2 for sufficient and inflated rules



### 4.2 The role of a fair share option

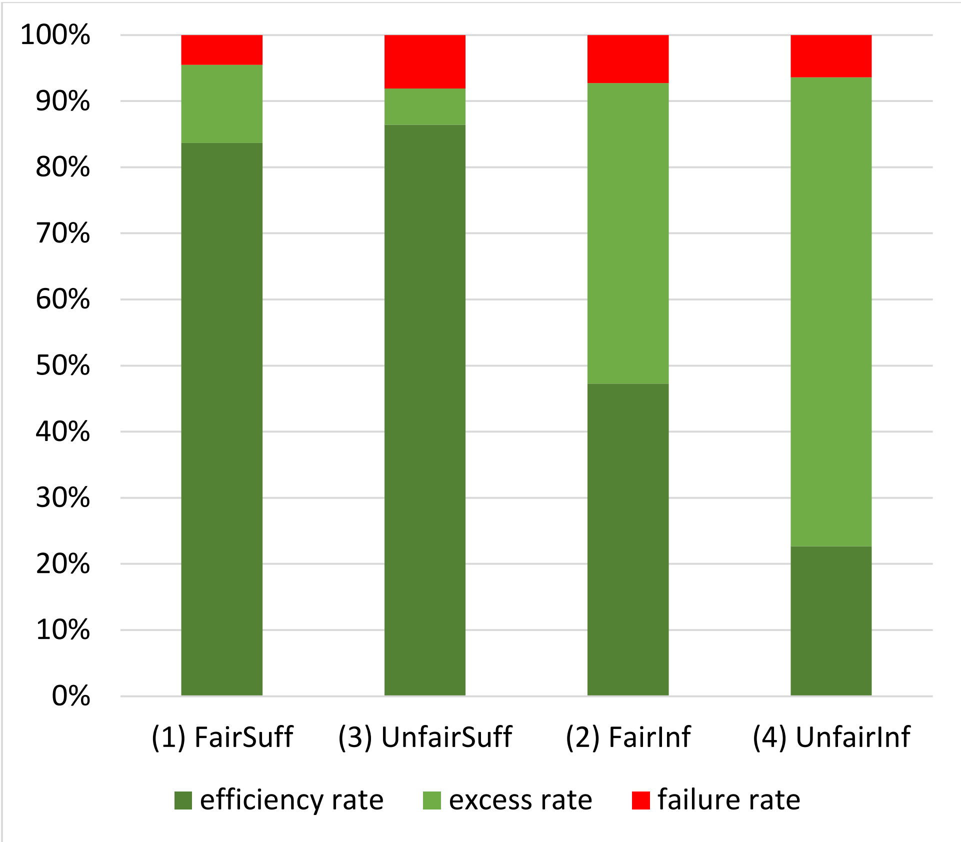
As described in Section 2 above, an additional important dimension in Threshold Public good games is the fair share option, i.e. the possibility to precisely hit the threshold with equal contributions across subjects. The fair share option is determined by the relationship between the threshold and group size. In our case, subjects encounter either a threshold of 20, which provides a fair share option (each of the four subjects provide 5), or 22, which does not offer the possibility.

First, we investigate the effect of the incidence of a fair share option on the success and efficiency rates. We focus on the relative frequencies in the first phase, where no rules were present yet, and compare treatments with a fair share option against those without a fair share option by applying Mann-Whitney U Tests. Groups that had a fair share option managed to reach the threshold on average in 83.6 % of the time. Groups without a fair share option have a slightly, though insignificantly, lower success rate (80.9 %; $p = 0.807$). However, the incidence of the fair share option has a clear effect on the efficiency of the public good provision. The threshold value of 20 was perfectly reached in 42.3 % of the cases, while T=22 without the fair share option was only hit in 21.2 % of cases ($p = 0.037$). Therefore, we can conclude that coordination in general is also possible without the fair share option, but at costs of efficiency. Since we hypothesized both higher success rates and efficiency rates, our result supports H4 (b).

Next, we turn to the analysis of phase 2 and disentangle the relevance of specific rules. Figure 6 illustrates the results.[3] Among treatments with sufficient rules (FAIRSUFF and UNFAIRSUFF), we do not find significant differences in success and efficiency rates between the different thresholds (success rates: 95.5 % vs. 91.8 %, Mann-Whitney U test: $p = 0.587$; efficiency rates: 83.6 % vs. 87.3 %, $p = 0.927$). The difference in success rates at least aligns with our expectation since failure rates are almost twice as high in UNFAIRSUFF. The analogous comparison between FAIRINF and UNFAIRINF shows the expected differences in the efficiency rates, though. Groups, that received an inflated rule in a setting without a fair share option, only reached an efficient outcome in 22.7 % of cases, while groups that were able to split the threshold equally, coordinated efficiently more than twice as much (47.3 %; $p = 0.040$). Again, the results support only the second part of H4 concerning efficiency.

[3] To make the described comparison more visible, we rearranged the order of the treatments in this figure.

Figure 6: Group outcomes in phase 2 by treatments



### 4.3 Individual rule-following behavior

We now shift from the hitherto group-level analysis to individual decisions. Group outcomes are contingent upon individual contribution choices, which are influenced by the decision to adhere to or violate the received rule.

Overall, subjects follow their individual rule in 86.6% of observations. Overcontributions hardly apply (2.95%, in 52 out of 1760 cases) and are assigned to rule following. Applying a binary probit model with rule-following (1=yes) as the dependent variable, we want to investigate whether the design of our treatments affects the rule-following behavior. We include the round number as control variable and cluster standard errors on the group level to account for possible heteroscedasticity since we assume peer influenced behavior among group members. The results of this estimation are presented in Table 2. Model (1) contains observations from all treatments that received a rule, indicator variables for inflation, the existence of a fair share option and the interaction effect of those two variables. Our estimation delivers a significant negative effect for inflated rules on rule-following. Such rules are followed on average c.p. about 18 percentage points less than sufficient ones.

The decision to follow the rule is likely influenced by the attitude towards the rule (Krupka & Weber, 2013) and beliefs about other players' behavior (Gächter et al., 2025). Thus, in model (2) we control for answers from the questionnaire. We observe significant bivariate between-subject differences between sufficient and inflated rules regarding (i) whether one should follow the rule, (ii) whether the rule is an efficient tool to avoid overcontribution, (iii) and the estimated share of fellow rule followers (for all comparisons Mann-Whitney U Test: $p < 0.001$).

We find that the inflation effect reported in model (1) lowers by a third but stays still significant when we include answers from the questionnaire. The rather negative normative attitude towards an inflated rule partially explains the reduced rule following. Furthermore, the other control variables do not provide additional explanatory power. Overall, our evidence supports H5.

Interestingly, the fair share option does not have an effect on the rule-following behavior. Additionally, there is no interaction effect between fair share option and inflation.[4] Consequently, we do not observe a significant difference in rule-following numbers between the treatments FAIRINF and UNFAIRINF and we do not find evidence supporting H6.

Table 2: Binary Probit model on rule-following rates for the whole sample – marginal effects

| Marginal effects | (1) | (2) |
|---|---|---|
| inflated | -0.183 *** | -0.116 *** |
| (1=yes) | (0.046) | (0.039) |
| fair share option | -0.046 | -0.027 |
| (1=yes) | (0.057) | (0.041) |
| inflated # fair share option | -0.009 | -0.009 |
| | (0.064) | (0.047) |
| *Controls from questionnaire* | | |
| Attitude towards the rule | | 0.064 *** |
| | | (0.011) |
| Estimated share of rule followers | | -0.023 |
| | | (0.015) |
| Attitude towards rule efficiency | | -0.020 |
| | | (0.014) |
| Control for round | yes | yes |
| Number of observations | 1760 | 1760 |
| Pseudo-$R^2$ | 0.142 | 0.249 |

Note: Standard errors are clustered on group level; *** p<0.01, ** p<0.05, * p<0.1
Attitude towards the rule: "How should one behave regarding the presented rule?" 1: never follow the rule – 4: always follow the rule.
Estimated share of rule followers: "Please estimate the share of subjects in the whole experiment (in %) that has followed the rule."
Attitude towards rule efficiency: "I think of the rule that was introduced after the first phase as a useful tool to ensure low waste of contributions." 1: completely disagree – 5: completely agree.

Sufficient and inflated rules differ in the individual requests contained in the rules. This difference may explain the varying rule-following rates. Subjects who are asked to contribute

[4] We run additional binary probit estimations using sub samples (see Table A in the Appendix) where we differentiate between sufficient and inflated rules. For subjects facing inflated rules, we can observe a weakly significant negative effect of the fair share option on rule-following. But this effect disappears once controlling for the questionnaire answers.

5 tokens could be more willing to obey than subject who shall contribute 6 tokens. For our next analysis steps, depicted in Table 3, we will completely focus on the groups FAIRINF ($T = 20$, $R = 22$) and UNFAIRSUFF ($T = 22, R = 22$) which all received rules in the form of $\bar{R} = (5,5,6,6)$. Model (1) contains dummy variables for the treatment difference (1 = FAIRINF) and the individual rule (1 = *high* individual rule: $r_i = 6$). We see subjects receiving a "high" request followed the rule 14.5 percentage points less likely. Beyond that, we find a significant treatment effect which, following our previous result, in this case might stem completely from the inflated rule. With model (2), we check whether there is a further interaction effect between the high individual request and the treatment difference. We cannot find a significant difference apart from a simple overlay of effects. Again, the results remain after the integration of answers from the questionnaire (model (3)). Hence, we conclude that the individual request does not influence the reaction on inflated rules.

Table 3: Binary Probit model on rule-following rates for treatments FAIRINF and UNFAIRSUFF – marginal effects

| Marginal effects | (1) | (2) | (3) |
|---|---|---|---|
| Treatment<br>(1 = FairInf) | -0.214 ***<br>(0.032) | -0.201 ***<br>(0.048) | -0.184 ***<br>(0.045) |
| High individual request<br>(1 = "Contribute 6") | -0.145 ***<br>(0.038) | -0.128 **<br>(0.053) | -0.131 **<br>(0.054) |
| Treatment (FairInf) #<br>High individual request | | -0.026<br>(0.070) | 0.015<br>(0.058) |
| *Controls from questionnaire* | | | |
| Attitude towards the rule | | | 0.054 **<br>(0.025) |
| Estimated share of rule followers | | | -0.004<br>(0.022) |
| Attitude towards rule efficiency | | | -0.035 *<br>(0.019) |
| Control for round | yes | yes | yes |
| Number of observations | 880 | 880 | 880 |
| Pseudo-$R^2$ | 0.275 | 0.275 | 0.332 |

Note: Standard errors are clustered on group level; *** p<0.01, ** p<0.05, * p<0.1
Attitude towards the rule: "How should one behave regarding the presented rule?" 1: never follow the rule – 4: always follow the rule.
Estimated share of rule followers: "Please estimate the share of subjects in the whole experiment (in %) that has followed the rule."

| |
|---|
| Attitude towards rule efficiency: "I think of the rule that was introduced after the first phase as a useful tool to ensure low waste of contributions." 1: completely disagree – 5: completely agree. |

### 4.4 Individual payoff analysis

We complement our analysis by shifting the attention to individual payoffs as a combined result of the previously measured group outcomes and individual rule-following behavior.

To gain deeper insights into how the personal decisions influence the payoff, we will revisit the theoretical framework. We aim to ex-post evaluate the individual behavior ceteris paribus depending on the given setting (treatment and phase). Given that players in a treatment behave as they did on average, we seek to determine the individually optimal behavior which will correspond to the best response. Hence, we abstract from specific situations in which e.g. an individual could have contributed one additional token to reach the threshold, though. Instead, we are interested in a hypothetical (because not necessarily being an integer and thus not feasible) optimal contribution level that on average maximizes the expected outcome for a single subject by optimally balancing the trade-off of saving tokens and increasing the success probability. This hypothetical value represents the contribution incentives induced by a rule. Ultimately, we want to answer the question whether inflated rules create opportunities for beneficial rule violations.[5]

For each level of contribution, there are two possible outcome states: success or failure. The payoff varies between the cases of success and failure exactly by $v = T/2$ tokens. By observing an individual payoff for each individual contribution level, we are able to determine how often the threshold was reached. We interpret the individual payoff as a random outcome depending on a *success probability*. Therefore, we can state the individual payoff as a function of the success probability, influenced by the individual contribution. Marginal benefits of contributing an additional token in terms of an increased success probability is concave, whereas marginal costs are linear. Building on this reasoning, we can now separately estimate the effect of individual contributions and its square on individual payoffs for all treatments applying an OLS regression (see Table B.1 and B.2 in the Appendix). We receive a parabola for which we can calculate a maximum. This maximum describes the ex-post c.p. *optimal contribution level* $c^*$

[5] This behavior deviates from free-riding in its usual sense since rule violations do not hurt other group members as long as the threshold is met.

which is presented in Table 4. By performing this calculation for each setting (aggregating whenever threshold and rule are equal), we can compare the optimal contribution level across different treatments.

We define a fixed value for every setting $c^{SP} = T/n$, implying an equal distribution of costs which we refer to as *social planner contribution level*. This benchmark represents *cooperative behavior* and helps us to evaluate $c^*$. The smaller the difference between $c^*$ and $c^{SP}$, the stronger cooperation incentives are in place. If $c^* \ll c^{SP}$, then there is space for a beneficial deviation of cooperation, i.e. it is not optimal to contribute the fair share, but one is better off c.p. by saving tokens.

Since we abstract from singular subjects in these considerations, we aggregate our treatments FAIRSUFF and FAIRINF as well as UNFAIRSUFF, UNFAIRINF, and CONTROL in the first phase (see the first row in Table 4) because their decision situations regarding the threshold do not differ.

Upon analyzing our data, several interesting implications emerge. Starting with the first phase, for individuals who face a threshold of 20, a contribution of 4.25 tokens would be optimal. For subjects in the FAIRSUFF treatment, we find a slightly higher $c^*$ of 4.58 tokens in the second phase. Compared to the social planner contribution level of 5, the c.p. optimal contribution level is lower in both phases meaning that deviating from the cooperative strategy can be somewhat beneficial. This difference between phases implies that the introduction of a rule marginally "improves" incentives for cooperation but does not significantly alter overall dynamics.

In contrast, the FAIRINF groups show a huge difference. After the introduction of the inflated rule, the optimal contribution is at 3.82 tokens, i.e. a randomly chosen person would have been best off by contributing this amount assuming everybody else behaved averagely. This seems to support our concern regarding a beneficial deviation from the inflated rule. By requesting an excessively high number of tokens on group level, the individual incentives for cooperative behavior are weakened.

The optimal contribution in the first phase for subjects facing a threshold of 22 lies at 4.74. Interestingly, the difference to $T = 20$ in phase 1 is nearly exactly as high as the difference between the social planner optima (= 0.5).

The results from UNFAIRSUFF warrant careful interpretation in the second phase due to insignificant estimated coefficients. This could stem from the low variance of contribution choices: In only 5 out of 440 choices, subjects picked a contribution different from 5 or 6. But based on this statistic, we can already see that the sufficient rule in a setting without fair share option does not seem to distort incentives towards lower contributions. The slightly higher optimal contribution level (4.93) supports assumption.

In the last treated group UNFAIRINF, we can observe a huge shift between phases: The optimal contribution raises up to 5.39 tokens in the second phase. To get the optimal payoff, one has to contribute over half an additional token in the second phase which is also really close to the social planner's choice of 5.5. The data shows clear evidence that the rule introduction shifts the incentives to a more social orientated contribution behavior. "Saving" tokens becomes less attractive, since it would now, in the second phase, give a smaller payoff than in the first one. This result strictly contradicts our hypothesis of beneficial deviation. The inflation of a rule in the absence of a fair share option does not distort but clearly improve the incentives to act cooperatively.

Interestingly, the optimal contribution level for subjects in the control group lowers in the second phase ($c^* = 4.58$). If no rule is introduced, the incentives for cooperative behavior get weakened. Recalling our result from before, rules make contribution behavior more consistent. The uncertainty about the other players' behavior probably makes saving tokens more attractive.

Summarizing our findings, we state that we can observe a shift in incentives only for inflated rules that occur in the presence of a fair share option (FAIRINF). Inefficiently high requests of contributions can create situations where it is best to contribute less than the average share when a simple coordination option exists. If there is no such option, such a beneficial rule violation is "punished" due to lower payoffs on average.

Table 4: Estimated ex-post optimal individual contributions ($c^*$) by threshold, phase, and treatment

| | T=20 ($c^{SP} = 5$) | | T=22 ($c^{SP} = 5.5$) | | |
|---|---|---|---|---|---|
| Phase 1 | 4.25 | | 4.74 | | |
| | FAIRSUFF | FAIRINF | UNFAIRSUFF | UNFAIRINF | CONTROL |
| Phase 2 | 4.58 | 3.82 | 4.93 | 5.39 | 4.58 |

## 5 Conclusion

We study the role of purpose-based rules for individuals' contributions and group outcomes (success and efficiency rates) in a threshold public good game. We explore the determinants of rule-following behavior, distinguishing between sufficient and inflated rules, which would lead to an over-fulfilment of the threshold if being followed. Additionally, we consider the role of a fair share option, i.e. whether it is possible to exactly meet the threshold through equal contributions by all group members. We test our hypotheses within an experimental setup.

Our main results include that rules promote consistency in contribution behavior as compared to the absence of a rule. Besides, sufficient rules are followed more frequently and result in efficient outcomes more often compared to inflated rules. Furthermore, the presence of a fair share option enhance efficiency, while it does not affect the number of successful outcomes.

We derive meaningful insights: First, the results reaffirm the coordinating power of interventions such as rules even without the possibility of penalties, which have been part of previous work such as in Bohnet & Cooter (2003). Decisions made under the influence of a rule result in fewer outliers, as the lower variance demonstrates. In general, the implementation of rules rather reduces the dispersion of behavior than it affects average contributions. Hence, rules make behavior of others more predictable. Accurate beliefs about other people's behavior are essential for successfully playing coordination games. This result underlines the general usefulness of rules in social dilemmas beyond a simple analysis of average behavior.

Second, our investigation does not find support for a trade-off between effectivity and efficiency as an explanation for the existence of inflated rules. In our setting, sufficient rules yield equally high success rates like inflated rules. Thus, in our setup inflation does not improve the outcomes but leads to additional costs in terms of wasted contributions. On the other side, since rule-following is reduced in the presence of inflated rules, subjects are aware of potential inefficiencies. Rather than blindly obeying the rule, some subjects adapt their behavior by conserving tokens. We want to point out yet that the level of inflation in our experiment is relatively small. The suspected trade-off between effectivity and efficiency may only be relevant for larger amounts of inflation.

Finally, we address concerns that inflated rules could create incentives to actually contribute less. Inflated rules in the presence of a fair share option can cause incentives to behave more selfishly and save tokens by contributing even less than the fair share. On the other hand, we also find circumstances (the absence of a fair share option) in which an inflated rule "improves" the incentives towards more cooperative behavior through higher optimal contributions ($c^*$). Thus, we can rule out the explanation that just higher requests lead to overcontribution which can be exploited by others. For an optimal incentive structure, these results suggest that the use of inflated rules is most promising in the absence of a fair share option.

Our rules suggest a specific allocation of individual contributions. This allocation can either be equal or unequal among the group members. The explicit request of an unequal allocation could trigger effects like inequity aversion (as described by Fehr & Schmidt, 1999). But since our design leads to an inverse relationship between rule inflation and the incidence of a fair share option, we cannot purely separate an inequity aversion effect from the inflation effect. To mitigate this concern, we note that subjects are informed only about their personal gain after each round, which shifts their focus away from others' outcomes.

The investigation of rule-following behavior and possible determinants is highly relevant in the business context as well as at societal level. The recent pandemic highlighted that many people questioned governmental rules that were too restrictive from their perspective. Our experiment gives evidence that people follow rules more often if the rules prescribe a behavior that leads to efficient goal achievement. Of course, in the case of a pandemic, uncertainty comes into play. In a business context however, the necessary effort for the output of a certain task is foreseeable more precisely for employees in rather small size groups. Thus, they can assess whether a rule is sufficient or inflated and therefore, base their choice to follow the rule on that categorization.

Future research in the area of rule-following may furtherly investigate the role of the purpose behind behavioral rules. Behavior can be affected by variations in the specific purpose of a rule, such as the cost-benefit ratio it entails, the individuals it affects, or the level of risk it involves. As usual in lab experiments, our subjects played for small monetary amounts. But behavioral rules often occur in circumstances where much higher values are at stake. Not rarely personal health is determined by rule obedience; e.g. during the Covid-19 pandemic, a lot of behavioral rules were introduced to prevent infections. It is likely that people adapt their behavior to the stakes of the game. Another facet in our experiment is that rules (and the players' decisions) affected only the players and their group members. But one could also think of other games than TPGGs in which players' actions affect the outcomes of non-active players (basically, an even stronger form of negative external effects). A rule in such a situation could have the purpose to protect people who cannot determine their own situation. For example, people in the global south are only marginally responsible for $CO^2$-emissions, but are affected even stronger.

Rules that force climate saving actions also have the purpose to protect people, who are mostly innocent of causing emissions. Additionally, one can think of rules that have a purpose but it is still unsure whether the objective will be reached in the case of total rule following. In many political discussions, the effectiveness of laws (i.e. governmental rules) is intensively debated. Uncertainty within a rule influences its coordinative power and by that most likely also people's behavior. Additionally, it becomes harder to categorize a rule as sufficient or inflated. Future additional investigations on rules with varying purposes could give insights in the reasoning for individual behavior.

## References


Bicchieri, C. (2005). The grammar of society: The nature and dynamics of social norms. Cambridge University Press.

Cadsby, C. B., Croson, R., Marks, M., & Maynes, E. (2008). Step return versus net reward in the voluntary provision of a threshold public good: An adversarial collaboration. Public Choice, 135, 277-289. https://doi.org/10.1007/s11127-007-9260-z

Chaudhuri, A., & Paichayontvijit, T. (2017). On the long-run efficacy of punishments and recommendations in a laboratory public goods game. Scientific reports, 7(1), 12286. https://doi.org/10.1038/s41598-017-12490-5

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88-97. https://doi.org/10.1016/j.jbef.2015.12.001

Cooter, R. D, & Bohnet, I. (2003). Expressive Law: Framing or Equilibrium Selection? UC Berkeley: Berkeley Law. Retrieved from https://escholarship.org/uc/item/65z2m533

Cooter, R. (1998). Expressive law and economics. The Journal of Legal Studies, 27(S2), 585-607. https://doi.org/10.1086/468036

Crawford, V. P., Gneezy, U., & Rottenstreich, Y. (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. American Economic Review, 98(4), 1443-1458. https://doi.org/10.1257/aer.98.4.1443

Croson, R. T., & Marks, M. B. (2000). Step returns in threshold public goods: A meta-and experimental analysis. Experimental Economics, 2, 239-259. https://doi.org/10.1023/A:1009918829192

Croson, R., & Marks, M. (2001). The effect of recommended contributions in the voluntary provision of public goods. Economic Inquiry, 39(2), 238-249. https://doi.org/10.1111/j.1465-7295.2001.tb00063.x

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. The quarterly journal of economics, 114(3), 817-868. https://doi.org/10.1162/003355399556151

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. Journal of the European Economic Association, 11(3), 525-547. https://doi.org/10.1111/jeea.12014

Gächter, S., Molleman, L., & Nosenzo, D. (2025). Why people follow rules. Nature Human Behavior, 1-13. https://doi.org/10.1038/s41562-025-02196-4

Galbiati, R., & Vertova, P. (2008). Obligations and cooperative behavior in public good games. Games and Economic Behavior, 64(1), 146-170. https://doi.org/10.1016/j.geb.2007.09.004

Gross, J., & De Dreu, C. K. (2021). Rule-following mitigates collaborative cheating and facilitates the spreading of honesty within groups. Personality and Social Psychology Bulletin, 47(3), 395-409. https://doi.org/10.1177/0146167220927195

Gross, J., & Vostroknutov, A. (2022). Why do people follow social norms?. Current Opinion in Psychology, 44, 1-6. https://doi.org/10.1016/j.copsyc.2021.08.016

Johnson, C. E., & Kecinski, M. (2025). Inequity aversion, mandates, and the provision of threshold public goods. Journal of Economic Behavior & Organization, 231, 106935. https://doi.org/10.1016/j.jebo.2025.106935

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. Journal of the European Economic Association, 14(3), 608-638. https://doi.org/10.1111/jeea.12152

Kimbrough, E. O., & Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. Economics Letters, 168, 147-150. https://doi.org/10.1016/j.econlet.2018.04.030

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. Journal of the European Economic Association, 11(3), 495-524. https://doi.org/10.1111/jeea.12006

Marwell, G., & Ames, R. E. (1979). Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem. American Journal of sociology, 84(6), 1335-1360. https://doi.org/10.1086/226937

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. The American Economic Review, 84(3), 658-673. https://www.jstor.org/stable/2118074

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2018). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. European Journal of Psychological Assessment. https://doi.org/10.1027/1015-5759/a000481

Schelling, T. C. (1958). The strategy of conflict. Prospectus for a reorientation of game theory. Journal of Conflict Resolution, 2(3), 203-264. https://doi.org/10.1177/002200275800200301

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. Journal of Research in Personality, 68, 69-81. https://doi.org/10.1016/j.jrp.2017.02.004

Van de Kragt, A. J., Orbell, J. M., & Dawes, R. M. (1983). The minimal contributing set as a solution to public goods problems. American Political Science Review, 77(1), 112-122. https://doi.org/10.2307/1956014

## Appendix

Table A: Binary Probit model on rule-following rates separately across sufficient and inflated treatments – marginal effects

| | sufficient | | inflated | |
|---|---|---|---|---|
| Marginal effects | (1) | (2) | (3) | (4) |
| fair share option | -0.019 | -0.008 | -0.096 * | -0.074 |
| (1 = yes) | (0.025) | (0.015) | (0.053) | (0.055) |
| *Controls from questionnaire* | | | | |
| Attitude towards the rule | | 0.019 ** | | 0.137 *** |
| | | (0.008) | | (0.038) |
| Estimated share of rule followers | | -0.008 | | -0.046 |
| | | (0.005) | | (0.042) |
| Attitude towards rule efficiency | | -0.003 | | -0.049 |
| | | (0.009) | | (0.034) |
| Control for round | yes | yes | yes | yes |
| Number of observations | 880 | 880 | 880 | 880 |
| Pseudo-$R^2$ | 0.0340 | 0.2176 | 0.0271 | 0.1315 |

Note: Standard errors are clustered on group level; *** p<0.01, ** p<0.05, * p<0.1
Attitude towards the rule: "How should one behave regarding the presented rule?" 1: never follow the rule – 4: always follow the rule.
Estimated share of rule followers: "Please estimate the share of subjects in the whole experiment (in %) that has followed the rule."
Attitude towards rule efficiency: "I think of the rule that was introduced after the first phase as a useful tool to ensure low waste of contributions." 1: completely disagree – 5: completely agree.

Table B.1: OLS regression on individual payoffs per round in phase 1 per thresholds

| | T = 20 | T = 22 |
|---|---|---|
| individual contribution | 1.363 *** | 1.619 *** |
| | (0.187) | (0.177) |
| individual contribution ^2 | -0.160 *** | -0.171 *** |
| | (0.019) | (0.019) |
| Intercept | 10.923 *** | 10.089 *** |
| | (0.509) | (0.473) |
| Number of observations | 880 | 1320 |
| Adjusted $R^2$ | 0.075 | 0.061 |

Note: *** p<0.01, ** p<0.05, * p<0.1

Table B.2: OLS regression on individual payoffs per round in phase 2 per treatments

| | FairSuff | FairInf | UnfairSuff | UnfairInf | Control |
|---|---|---|---|---|---|
| individual contribution | 2.225 *** | 2.377 *** | 2.146 * | 2.326 *** | 1.935 *** |
| | (0.341) | (0.372) | (1.104) | (0.836) | (0.178) |
| individual contribution ^2 | -0.243 *** | -0.312 *** | -0.217 ** | -0.216 *** | -0.211 *** |
| | (0.039) | (0.045) | (0.108) | (0.081) | (0.019) |
| Intercept | 9.605 *** | 10.421 *** | 9.463 *** | 8.376 *** | 10.794 *** |
| | (0.876) | (0.857) | (2.932) | (2.225) | (0.463) |
| Number of observations | 440 | 440 | 440 | 440 | 440 |
| Adjusted $R^2$ | 0.085 | 0.096 | 0.005 | 0.013 | 0.224 |

Note: *** p<0.01, ** p<0.05, * p<0.1