

# Norm Compliance, Enforcement, and the Survival of Redistributive Institutions\*

Mehmet Y. Gürdal<sup>†</sup>  
*Boğaziçi University*

Orhan Torul<sup>‡</sup>  
*Boğaziçi University*

Alexander Vostroknutov<sup>§</sup>  
*Maastricht University*

March, 2019

## Abstract

We study the incentives that drive behavior in redistributive institutions with various levels of enforcement. We are interested in how the opportunistic incentive to use a redistributive institution for personal gain and the desire to follow the rules of a regulated community, populated by similarly obedient individuals, interact and determine the success or failure of an institution. In the experiment, subjects can repeatedly join one of three groups, which are defined by explicitly stated injunctive norms that require to put all, half, or any amount of income to a common pool for redistribution. The treatments differ in the level of enforcement of these norms. We find that contributions are sustained only in the case of full enforcement. However, a sizeable number of subjects persist in following the norms of redistribution even after experiencing many periods of losses due to free riding. We find that subjects with strong propensity to follow norms perceive the same level of income inequality as fairer, when it was achieved without breaking the norm, and favor redistributive mechanisms with more stringent rules. This suggests that well-defined redistributive norms can create a powerful incentive for cooperation as many individuals seem to prefer stable regulated egalitarian institutions to unregulated libertarian ones. Some form of enforcement is, nevertheless, required to protect egalitarian institutions from exploitation by free riders.

*JEL classifications: C91, C92*

*Keywords: social norms, taxation, redistribution, egalitarianism, libertarianism, limited enforcement.*

---

\*We would like to thank the audiences at Max Plank Institute for Research on Collective Goods (Bonn), Vienna Center for Experimental Economics, and Florence-Konstanz Workshop on Behavioral Social Sciences for insightful comments. All errors are ours.

<sup>†</sup>Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey.  
email: [mehmet.gurdal@boun.edu.tr](mailto:mehmet.gurdal@boun.edu.tr)

<sup>‡</sup>Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey.  
email: [orhan.torul@boun.edu.tr](mailto:orhan.torul@boun.edu.tr)

<sup>§</sup>Address: Department of Economics (AE1) Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands email: [a.vostroknutov@maastrichtuniversity.nl](mailto:a.vostroknutov@maastrichtuniversity.nl)

# 1 Introduction

Redistribution is ubiquitous in human communities and can take various forms from food-sharing among hunter-gatherers to complex taxation and social insurance systems in developed economies. In this respect, the prevalence and sustainability of redistribution can be considered crucial for the successful functioning of any society. It is well documented that in indigenous communities redistribution is maintained by systems of social norms that involve punishment of free riders, while in the modern world these norms have become institutionalized as enforceable laws (Henrich, 2015). It might seem that the only thing that can prevent the collapse of redistributive mechanisms is the fear of retribution, as would be suggested by standard economic theory. However, there are many examples of institutions that are potentially threatened by free riding but, nevertheless, maintain various forms of redistribution with little involvement of explicit punishment options. Ostrom (1990) documents multiple communities that have managed to sustain common pool resources (CPR) for quite a long time without constant monitoring of the resource usage and with no cases of serious infringements. Alm and Torgler (2011) review empirical studies on tax compliance and point out that, given the general infrequency of tax returns audits, the rate of tax evasion is strikingly low.

The examples above share one feature in common. It seems that the enduring success of redistribution emanates from *voluntary* adherence to social norms that guide the community-conscious behavior and, in addition, from the preference for the *stability* of the institution in the future over the immediate gains from free riding (Proto et al., 2018). Studies in experimental economics support this view. Kimbrough and Vostroknutov (2016) demonstrate that some individuals—rule-followers—are willing to abide by arbitrary rules at the cost to themselves and that these same individuals show higher degree of cooperation in various social dilemmas than others who are not willing to follow costly rules (rule-breakers). For example, groups of rule-followers are able to sustain contributions in a repeated Public Goods game, while groups of rule-breakers follow the well-known pattern of contribution decay (Fehr and Gächter, 2000). This is not to say that rule-followers are a kind of people who are ready to forgo any personal gain in order to stick to rules or norms: when put in a group with rule-breakers they act as conditional cooperators and stop contributing as a response to free riding of others. Moreover, Kimbrough and Vostroknutov (2015) show that in a dynamic CPR game propensity to follow rules is a necessary, but not at all sufficient condition for the sustainability of the common resource: low growth rate of CPR prevents groups of rule-followers from successful management of the resource. Overall, experimental research on social norms suggests a view according to which heterogeneity in social behavior within given context is explained by the individual differences in propensity to follow rules (Kimbrough and Vostroknutov, 2018), and, thus, by the trade-offs between monetary gains and following norms, whereas the difference in behavior between contexts is a result of variation in social norms (Krupka and Weber, 2013).

The picture presented above provides a somewhat conflicting evidence on the functioning of redistributive mechanisms and the role of social norms and enforcement. On the one hand, the

evolution of complicated cheater detectors (Cosmides and Tooby, 1992, 2005) with the associated desire to punish violators (Fehr and Fischbacher, 2004) and the ubiquity of sanctions for tax evasion (OECD, 2017) strongly suggest that redistribution cannot be implemented without restricting the incentives to free ride. On the other hand, the prevalence of social norms that commend the acts for greater good and the abundance of examples of norm following at personal cost (Bicchieri, 2005) point towards a possibility that enforcement, at least under some circumstances, is not necessary to sustain redistribution. Therefore, it remains unclear how all these ingredients—propensity to follow social norms, incentives to free ride, and degree of enforcement—mix with each other and how success or failure of a redistributive institution is determined by their interaction.

The goal of this paper is to shed some light on the intrinsic incentives that drive behavior in various redistributive mechanisms; how these incentives shape the outcome of redistribution for the group and the satisfaction of its members with this outcome; and the role of enforcement in it. We are interested in the interaction of two opposing forces: the opportunistic incentive to use a redistributive institution for personal gain and the desire to live in and follow the rules of a regulated community populated by similarly obedient individuals. The first question we ask is whether or not a redistributive institution can be successful under a threat of free riding. Or, in other words, if the incentive to follow the rules of an institution alone is enough to maintain some level of redistribution. To find this out we allow subjects in the experiment to repeatedly choose among three institutions, which are defined by “redistribution rules.” A rule of an institution is simply a statement that requires those who have joined it to contribute a certain percentage of their income to a tax pool for redistribution. The rules are non-binding and bear no direct cost of violating them. Thus, we are interested in whether the desire to follow the rules of an institution is overcome by free riding. Next, we compare the fairness ratings in order to understand how the satisfaction with overall income inequality depends on redistribution rules and experienced opportunistic behavior. Finally, by varying enforcement levels we investigate how it helps with the sustainability of the redistribution and, in addition, how preferences for rule-following and free riding are associated with the preference for a particular redistribution mechanism.

Our experiment consists of two tasks: the Rule-Following task (Kimbrough and Vostroknutov, 2018) and the Institution Choice and Redistribution task. In the Rule-Following (RF) task, subjects choose how much to follow an arbitrary costly rule set by the experimenter. The performance in this task allows us to quantify the individual degree of “norm compliance” or the “rule-following propensity.” In the Institution Choice and Redistribution task, prior to learning their randomly generated income, subjects choose among three different redistributive institutions. These institutions differ in terms of the announced tax rate, which can be 100%, 50%, or 0%. In the Enforcement treatment, subjects have to abide by the rule of the institution that they choose to join and should contribute all, half, or any chosen amount of their income to the tax pool depending on the institution. In the No Enforcement treatment subjects are free to choose any contribution between zero and their entire initial income in all three institutions. Finally, the Exclusion treatment is identical

to the No Enforcement treatment except that random checks of the contributions are performed and subjects who did not follow the prescribed rule are excluded from future participation in an institution. When all contributions are collected, the tax pool is equally divided among all subjects who decided to join an institution. This task is repeated 20 times, which allows us to observe the dynamics of institution choices and respective contributions.

We find, unsurprisingly, that in the No Enforcement and Exclusion treatments, where subjects can contribute any amount of their income, the contributions to the tax pool in all three redistributive institutions fall to zero as time unfolds. Nevertheless, in early periods there is a strong effect of the announced rules of the institutions. In particular, a sizeable number of subjects choose to follow institution rules even at personal cost due to free riding. Moreover, we find that subjects, who are classified as “rule-followers” by the Rule-Following task, contribute significantly more than “rule-breakers” even in the late periods. This demonstrates that people with a strong propensity to follow rules are willing to maintain their contributions (albeit, only at 10% level) even when facing a severe free riding problem created by the rule-breakers.

Next, we find that rule-followers rate the final income distribution in the Enforcement treatment as significantly fairer than that in the No Enforcement treatment. This effect is especially pronounced for subjects who were mostly joining the “egalitarian” institution, which prescribes to contribute all income to the tax pool. Conversely, rule-breakers consider both the Enforcement and the No Enforcement treatments equally fair (at the same level as rule-followers perceive the No Enforcement treatment). From this, we conclude that differential fairness judgments of income distributions are directly linked to the propensity to follow rules and reflect the degree to which norms are followed by others. That is to say, we show that perceived fairness of an income distribution depends on whether others followed the rules and not on the income inequality per se (Starmans et al., 2017).

Finally, we analyze the choices of institutions that our subjects make. We find that, in the Enforcement treatment, where joining an institution automatically implies abiding by its rules, rule-followers prefer the “egalitarian” institution, where all income should be contributed to the tax pool, while rule-breakers prefer the “anarchic” institution in which any amount of contribution can be made. In the No Enforcement treatment the preference of rule-followers and rule-breakers switches to the opposite due to free riding. We show that this differential separation of types can only be explained by a *preference over redistribution rules* and the disutility that rule-followers receive from being in the environment where norms are violated, and not by risk or social preferences. Thus, we demonstrate that not only rule-followers persist in following the rules in the egalitarian institution (to a degree in case of free riders), but also that they prefer the institution with “strong” rules (contribute everything to the tax pool) to the institution without rules (Richerson et al., 2016). This type of preference is not captured by the standard norm-dependent utility proposed by Kessler and Leider (2012), which only accounts for the disutility of deviation from the norm.

In order to show that preference over rules is necessary to explain our experimental results, we develop a model of institution choice based on [Acemoglu and Jackson \(2017\)](#) which assumes an extended norm-dependent utility function. From the model, it follows that the patterns found in our data can only be reconciled with utility maximization when, in addition to the disutility from deviations from the norm, agents have disutility from others not following the norm. This shows that preferences for following rules are coupled with the desire to be in an institution with well-defined rules and populated by others with similar preferences.

## 2 Literature Review

This paper is related to four different strands of literature. The first one focuses on the relation between norm-compliance and pro-social behavior in various settings. Here, studies that involve economic experiments commonly find that tendency to comply with social norms and/or declared rules is a strong determinant of pro-social behavior in stylized games. [Kimbrough and Vostroknutov \(2016\)](#) show that people who are more likely to follow arbitrary rules differ from the rest of the population as they set higher acceptance thresholds in ultimatum games, sustain higher contributions in public good games, and exhibit greater reciprocity in the trust game. [Krupka and Weber \(2013\)](#) introduce an incentive-compatible way to measure prevalent social norms by eliciting beliefs. They show that a model that combines the propensity for norm-compliance and concern for money can accurately predict behavior in dictator games. On the other hand, [Gächter et al. \(2017\)](#) show that perceived social norms are affected by the behavior of peers in the environment, but they fail to find a moderating effect of these norms on individual behavior. We add to this literature by giving subjects a choice to join groups with *explicitly* stated rules of conduct that require different degrees of redistribution. This allows us to study how preferences for opportunism and the desire to abide by the well-defined norms determine the success of redistributive institutions under different levels of enforcement.

The second strand of literature is concerned with the prevailing fairness ideals in experimental settings. [Cappelen et al. \(2007\)](#) consider an environment where income depends on both luck and effort. They estimate the relative frequency of subjects who prefer strict egalitarianism, libertarianism, and liberal egalitarianism. The first two of these principles correspond to group 100 and group 0 in our experiments, whereas the last one is a modified form of egalitarianism where subjects are held accountable for the choices they make but not the luck factor. This is a principle constructed along the same lines as that of the accountability principle of [Konow \(2000\)](#). In a recent study, [Starmans et al. \(2017\)](#) show that people are mainly disturbed by the unfair processes generating unequal incomes rather than the final inequality itself. [Klor and Shayo \(2010\)](#) experimentally show that the knowledge about the mean income of a group influences subjects' preferences over redistributive arrangements. [Durante et al. \(2014\)](#) elicit demands for redistribution under various conditions and find non-negligible effects of self-interest, risk avoidance, and social concerns. In our experiment we

elicit subjects' rule-following propensity and make a novel connection between fairness perceptions and income inequality.

In some experimental studies, subjects are allowed to choose between different institutions throughout the experiment, similar to what we do in the current paper. In [Gürrer et al. \(2006\)](#), subjects choose to play a public goods game in an environment where they can sanction other group members or in an environment where this option is not available. The results reveal that a sanctioning institution is overwhelmingly preferred over the alternative after a learning period. [Sutter et al. \(2010\)](#) study similar environment where subjects unanimously decide to introduce punishments, rewards, or neither. The authors report higher contributions when the institutions are chosen endogenously rather than when they are exogenously implemented. In a recent paper, [Dal Bó et al. \(2017\)](#) allow players to vote to play a Prisoner's Dilemma or a Harmony game, where cooperation is the Nash equilibrium in the latter but not in the former. Even though subjects earn more in the Harmony game a substantial fraction votes for Prisoner's Dilemma. [Kosfeld et al. \(2009\)](#) study an  $n$ -person public goods game and let players decide on joining an "organization" prior to playing the game. The members of the organization are expected to contribute all of their endowment and impose punishment on those who do not follow the rule. The authors find that institutions are commonly formed and induce higher contributions. [Putterman et al. \(2011\)](#), study a public goods game where they allow endogenous choice of sanctioning parameters via voting. In this game, subjects are observed to converge to the efficiency enhancing parameters for the sanctioning mechanism.

Finally, our work is related to the models of tax compliance that use augmented utility function in order to reconcile the empirically observed low rates of tax evasion. [Gordon \(1989\)](#) introduces non-pecuniary costs, related to the act of filing a false tax declaration, that can be interpreted as guilt, anxiety, or damage to self-image. [Myles and Naylor \(1996\)](#) assume that agents derive additional utility from acting honestly, which leads to higher tax compliance. The model that we use to explain the behavior in our experiment has similar features, but applies more generally to any environment where the rules of conduct are well-defined.

### 3 Experimental Design

The experiment consists of two tasks: the Rule-Following task and the Institution Choice and Redistribution task with the former always preceding the latter (see Appendix J for instructions). These two main tasks were followed by a questionnaire with items related to demographics and attitudes towards fairness of hypothetical income distributions (see Appendix K).

### 3.1 The Rule-Following Task

In the Rule-Following task (Kimbrough and Vostroknutov, 2018) subjects have 100 balls that they can put one-by-one into two buckets: yellow or blue. For each ball in the yellow bucket they receive 0.10 TL (Turkish Liras) and for each ball in the blue bucket they receive 0.05 TL (see Figure 1). The current earnings from the two buckets are shown above them. The total earnings are the sum of earnings from the two buckets. The position of the buckets on the screen is randomized across subjects.

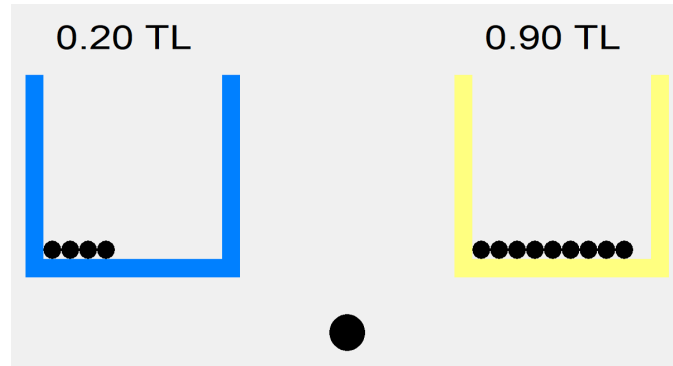


Figure 1: The Rule-Following task.

The instructions explicitly state that “the rule is to put the balls into the blue bucket” (see Appendix J). Subjects have 100 balls to allocate, thus, their earnings can vary from 5 TL, if they follow the rule to the letter, to 10 TL, if they break the rule and put all the balls into the more profitable yellow bucket.<sup>1</sup>

### 3.2 The Institution Choice and Redistribution Task

This task consists of 20 identical periods. In each period, subjects first choose a redistributive institution that they would like to join (in what follows we will use the terms *institution* and *group* interchangeably). There are three institutions to choose from which have different redistribution rules. The rules are described with the following normative statements:

- Group 100: We should all put our whole income (100%) to the common pool;
- Group 50: We should all put half of our income (50%) to the common pool;
- Group 0: We can contribute any amount to the common pool at our own discretion.<sup>2</sup>

<sup>1</sup>When subjects asked for clarification about the statement that “the rule is...,” the experimenters always answered that “this is the rule of the experiment,” and when they asked whether anything will happen to them if they don’t follow the rule, the experimenters responded that the information about all possible contingencies that can occur is contained in the instructions.

<sup>2</sup>In the experiment the three groups were called A, B, and C.

After this choice, subjects are assigned to their respective institutions and observe their initial (pre-tax) income in the current period. The income (in tokens) is a randomly chosen integer from the interval  $[0, 50]$  (uniform distribution). One token was exchanged for 1 TL. Next, subjects choose how much income to contribute to the common pool. For each institution, the contributions to the common pool are summed up and divided equally among all subjects who have chosen to join that institution.<sup>3</sup> We decided to give subjects random income instead of a fixed known one to make sure that the choice of an institution is not influenced by their wealth. Another reason was to create a perception among subjects that groups 100 and 50 are beneficial for the “society” since following the rules of redistribution works as a hedge against random income shocks.

At the end of each round subjects were informed about the performance of the entire group. In particular, they observed their initial income and the average initial income of other group members; their contribution to the common pool and the average contribution of other group members; the amount they received from the common pool and the average amount distributed to other group members from the common pool; and their final income and the average final income of other group members.

The experiment has three treatments which differ in the degree of enforcement of the rules of the institutions. In the Enforcement treatment subjects are constrained by the computer to abide by the rule of the institution that they choose to join. In particular, subjects who chose group 100 can only contribute all their initial income to the common pool; subjects in group 50 can contribute at least 50% of their initial income (any amount at or above 50%); and subjects in group 0 can choose any number between 0 and their initial income as their contribution to the common pool. In the No Enforcement treatment subjects are free to choose any contribution between zero and their entire initial income in all three groups. In the Exclusion treatment subjects can be banned from participating in an institution under certain conditions. In particular, their contributions were subject to random and idiosyncratic outside checks, which occurred with probability 20%. In case of groups 100 and 50, a subject is not allowed to join either of these groups ever again if the outside check reveals that her contribution was lower than 80% of the respective norm (100% or 50% requirement). By introducing a cost of free riding in this way we intended to test if compliance with the rules of the institutions can be sustained.<sup>4</sup>

Subjects’ payment is calculated as follows. The income at the end of each period (post-tax income) is equal to the initial, pre-tax, income minus the contribution plus the return from the

---

<sup>3</sup>In case only one subject has chosen an institution the contribution choice was not provided.

<sup>4</sup>In choosing the parameters of the Exclusion treatment we followed several guidelines. First, we favored exclusion over fines because making subjects pay for non-compliance would entail a complex cost-benefit analysis of the decision to cheat involving the tax rate and the fine size. Second, the probability 20% of an outside check was chosen so that it is small enough to not disincentivize all cheating, but at the same time large enough so that in 20 periods all cheaters are excluded: with 20% checks, the probability of undetected cheating for 20 periods is 1%, whereas it is 12% (too high) with 10% checks. Finally, we set the threshold of cheating detection at 80% of the required tax rate to separate subjects who genuinely follow the norm from those who want to comply just because of the threat of punishment.



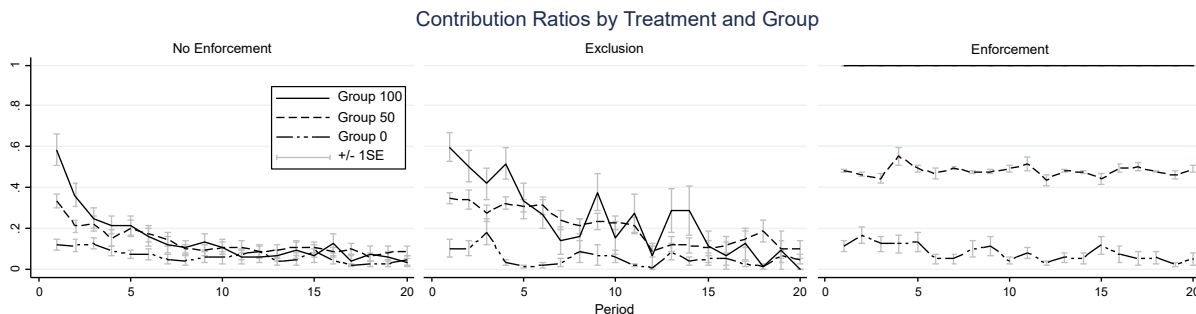
“public good” (for each group in each period, the sum of all contributions in the group divided by the number of subjects in the group). One period out of 20 was chosen randomly for payment.

The experiment was run at Boğaziçi University, Istanbul, Turkey in February-April 2017. 158 subjects participated in the No Enforcement treatment (9 sessions), 72 in the Exclusion treatment (4 sessions), and 106 in the Enforcement treatment (6 sessions). The number of subjects in a given session was between 16 and 19. There were no other sessions or pilots. No data were discarded. Experiment was programmed in z-Tree (Fischbacher, 2007).

## 4 Results

### 4.1 Descriptive Statistics

We begin with presenting some aggregate statistics of the choices of our subjects in the redistribution task. We define the variable *contribution* which, in each period, is equal to the ratio of the amount contributed to the income received. This is the only meaningful way to analyze contributions, since in our setup income is random in each period. Thus, looking at absolute contributions can be misleading.



**Figure 2:** The dynamics of contribution ratios in the three treatments divided by group. The spikes are  $\pm 1SE$ .

Figure 2 shows the dynamics of contributions in the No Enforcement, Exclusion, and Enforcement treatments. In the Enforcement treatment the picture is clear: subjects cannot contribute less than a fixed percentage of their income.<sup>5</sup> In the No Enforcement treatment the contributions decay fast in all three groups, which is not surprising in the light of similar results with repeated Public Goods games (e.g., Fehr and Gächter, 2000). However, it is important to notice that the average contributions in the first four periods are ranked according to the percentage of income that is required to be shared by the rule of the group. This demonstrates that in early periods at least some subjects do try to follow the rule of the group, but are later overwhelmed by free riders, and, consequently, decrease their contributions as a conditional response. In Appendix B we support

<sup>5</sup>The averages sometimes fall below the required contribution of 50% in group 50. This happens because subjects were allowed to only contribute integer amounts, so they could contribute the amount just below 50% if their income was an odd number (e.g., if income is 3 they could contribute 1, which is 33%).

this conclusion with an additional analysis, which shows that around half of subjects start with abiding by the announced rules in early periods.

In the Exclusion treatment average contributions also go to zero with time as in the No Enforcement treatment. However, this happens much slower: while in the No Enforcement treatment the average contributions fall below 20% after period 5, in the Exclusion treatment the averages are above 20% for groups 100 and 50 until around period 12. This demonstrates that exclusion does have an effect on average contributions.

**Result 1.** *In the early periods of the No Enforcement and the Exclusion treatments contributions are ranked according to the announced tax rate of the institutions: around half of subjects start with following the norm of an institution. Contributions decay faster in the No Enforcement treatment than in the Exclusion treatment.*

## 4.2 Rule Following and Contributions

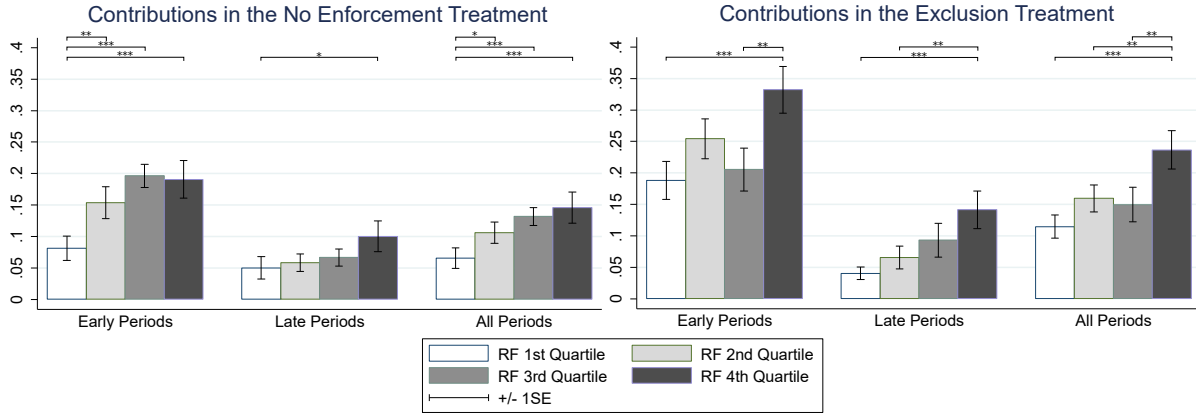
To substantiate the idea that the announced rules of the institutions are perceived as expected norms of behavior, we look at the contributions as dependent on subjects' rule-following propensity.<sup>6</sup> Figure 3 shows contributions by quartiles of rule-following and early/late periods in the No Enforcement and Exclusion treatments. The contributions fall from early to late periods due to free riding. However, it is clear that in both treatments rule-followers (4th quartile) contribute significantly more than rule-breakers (1st quartile) even in the late periods.<sup>7</sup> Overall, contributions increase with rule-following propensity on average. This demonstrates that rule-followers try to adhere to the rules of the groups, to some extent, even when they drop their contributions as a conditional response to low cooperation from others. The same conclusion is supported by the regression analysis of the contribution choices reported in Table 5 in Appendix D: the coefficient on rule-following propensity is positive and significant. From this analysis we can conclude that, even though conditional response to free riding eventually overcomes the desire to follow the rules of the groups, rule-followers are still trying to maintain higher contribution levels than rule-breakers. This shows how strongly they are willing to adhere to the rules of institutions, even if these rules are essentially just non-binding statements about how institutions ought to work.

The next observation is that contributions are significantly higher in early periods of the Exclusion treatment than in early periods of the No Enforcement treatment and overall: the coefficients on the Exclusion treatment dummy in the regressions in Table 6 (Appendix D) are positive and significant (except for late periods). Interestingly, there are no significant effects of interactions between Exclusion treatment dummy and dummies for quartiles of rule-following. Therefore, in early

---

<sup>6</sup>The descriptive statistics for the behavior in the rule-following task are provided in Appendix C.

<sup>7</sup>The significance levels reported in Figure 3 correspond to the  $t$ -tests of the differences in coefficients in the random effects regressions in Table 4 (Appendix D). We chose this approach over non-parametric tests since there are several observations per subject averaged together. Tests like rank-sum assume independence of observations inside comparison groups, thus making them invalid for such data.



**Figure 3:** Contributions in the No Enforcement and Exclusion treatments by quartiles of rule-following and early/late periods. Significance levels are obtained from the random effects regressions reported in Table 4 in Appendix D. \* -  $p < 0.1$ ; \*\* -  $p < 0.05$ ; \*\*\* -  $p < 0.01$ .

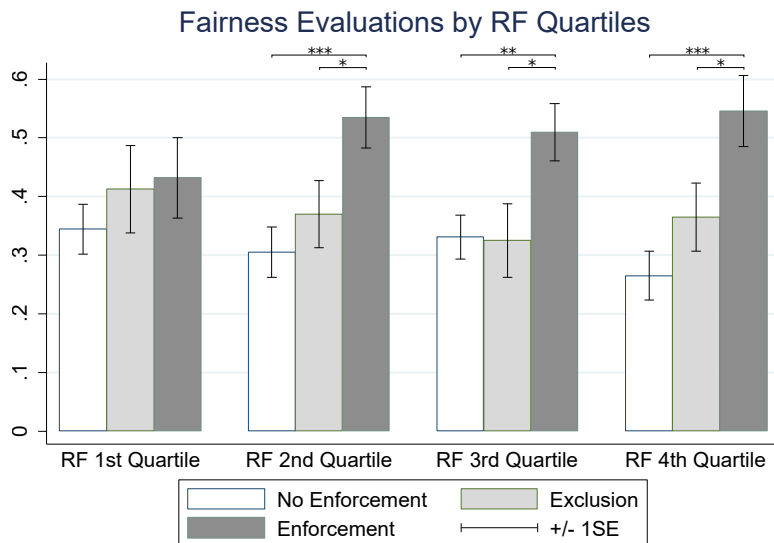
periods the possibility of exclusion increases contributions of both rule-followers and rule-breakers, which suggests that the latter increase contributions because of the threat of exclusion.

**Result 2.** *The announced rules of the institutions have a strong positive effect on the contributions of rule-followers, even when free riding is rampant. This shows that the announced rules are perceived as norms. The possibility of exclusion increases the contributions of both rule-followers and rule-breakers in early periods.*

### 4.3 Perceived Fairness of Redistribution

Given the previous result, that rule-followers are trying to abide by the announced rules of the institutions even when being constantly exploited by free riders, it is interesting to investigate their perceptions of fairness of redistribution. Thus, we look at how fair our subjects deem the income distributions after the main task. In the questionnaire subjects were asked a question “*How fair do you find the income distribution that resulted in the experiment?*” on a 0 to 10 Likert-scale. The average answers are 3.11 in the No Enforcement treatment, 3.68 in the Exclusion treatment, and 5.08 in the Enforcement treatment (Figure 9 in Appendix I shows the distributions of answers in the three treatments). Rank-sum tests show a significant difference in the distributions in the No Enforcement and Enforcement treatments ( $p < 0.0001$ ), and the Exclusion and Enforcement treatments ( $p = 0.0010$ ).

Many more subjects consider the distribution unfair in the No Enforcement and Exclusion treatments than in the Enforcement treatment. To illustrate this difference in more detail, Figure 4 shows the average fairness ratings for the four quartiles of rule-following propensity. For the rule-breakers in the 1st quartile the difference in fairness between all three treatments is insignificant. The difference between the Enforcement and the other two treatments becomes more pronounced as the rule-following propensity grows. Rank-sum tests show a significant difference in fairness estimates



**Figure 4:** Average fairness estimates for the four quartiles of rule-following propensity by treatment. The significance levels reported correspond to the rank-sum tests, Benjamini-Hochberg corrected for 12 comparisons: \* -  $p < 0.1$ ; \*\* -  $p < 0.05$ ; \*\*\* -  $p < 0.01$ .

between treatments.<sup>8</sup> Therefore, rule-followers, but not rule-breakers, find income distribution in the No Enforcement and Exclusion treatments less fair, as compared to the Enforcement treatment. The reason for this is that in both the No Enforcement and Exclusion treatments rule-followers, who try to follow the rules of the institutions, experience bad consequences of free riding, while in the Enforcement treatment they do not. We provide a more detailed regression analysis that supports this conclusion in Appendix E.

These findings are related to a recent discussion on the fairness perception of income inequality. Deaton (2017) claims that inequality is not mapped onto unfairness on the basis of its level, but by its origin: when inequality stems from deeply and obviously unjust foundations, only then it is regarded unfair. Starmans et al. (2017) also argue that people prefer fair inequality to unfair equality, and that they are concerned about economic inequality only when it is confounded with economic unfairness.<sup>9</sup> In Table 1 we illustrate the distributional properties of pre-tax and post-tax incomes. We find that in the No Enforcement treatment not only post-tax income inequality surpasses pre-tax one, but also the supposedly egalitarian group (group 100) generates the most unequal post-tax income distribution among the three groups. In other words, *de facto* redistribution across groups not only blurs, but also clashes with the announced norms, particularly in the case of the egalitarian

<sup>8</sup>2nd quartile:  $p = 0.0042$  between the No Enforcement and Enforcement treatments,  $p = 0.074$  between the Exclusion and Enforcement treatments; 3rd quartile: respective  $p = 0.0288$ ,  $p = 0.0854$ ; 4th quartile:  $p = 0.006$ ,  $p = 0.093$ . All  $p$ -values are Benjamini-Hochberg corrected for 12 comparisons.

<sup>9</sup>See also Rustichini and Vostroknutov (2014) for an experimental investigation.

group.<sup>10</sup> Rule-followers in the egalitarian group, who try hard to play by the rules and conform with the group norm by contributing sizeably, realize that they are the constant losers due to the free riding of those who contravene the announced group norm.<sup>11</sup> As a result of their legitimate resentment, they evaluate the unequal distribution resulting from bending the announced norm unfair.

		Income Gini		Fairness
		Pre-Tax	Post-Tax	
No Enforcement	All	0.340	0.343	0.311
	Group 100	0.340	0.352	0.261
	Group 50	0.344	0.336	0.324
	Group 0	0.338	0.339	0.320
Exclusion	All	0.322	0.324	0.368
	Group 100	0.332	0.350	0.300
	Group 50	0.322	0.307	0.333
	Group 0	0.316	0.319	0.410
Enforcement	All	0.336	0.242	0.508
	Group 100	0.325	0.121	0.556
	Group 50	0.339	0.215	0.519
	Group 0	0.345	0.344	0.504

**Table 1:** Inequality and Fairness Perception. Fairness perception of the groups refers to the average fairness perception scores of subjects who spent more than half of experimental rounds in that group.

In order to unveil subjects’ mapping of economic inequality onto fairness, in Table 1 we also report fairness scores of subjects who spent more than half of the experiment in one of the three groups. Our results are mostly in accordance with claims by Deaton (2017) and Starmans et al. (2017): comparing between-group variations by treatment, we report that the higher the post-tax income inequality, the lower the fairness perception of subjects for a given enforcement rule, albeit with moderate differences. Comparing across-treatment inequality mappings onto fairness, we observe that enforcement rule is decisive beyond inequality: while group 0 in the enforcement treatment generates moderately more unequal post-tax income inequality than its no-enforcement counterpart, its members evaluate the resulting distribution as significantly fairer (0.504) than when enforcement is absent (0.321), thereby providing support for the claim that the origin of inequality *is* pivotal.

To summarize, our results on fairness of redistribution demonstrate the importance of norms for the perception of inequality and the attitudes towards the redistribution mechanism. In the

<sup>10</sup>A careful reader could notice that the post-tax income Gini coefficient of group 100 in the Enforcement treatment in Table 1 is equal to 0.121, which is different from zero. The reason for this is that the reported post-tax income Gini coefficients are calculated with the subject-level post-tax incomes of group 100 participants over *all* periods. As such, depending on the period-specific income draws of group 100 subjects, average amount collected in the group 100 common pool varies over periods, especially during periods when the number of group 100 subjects is limited. As a result, post-tax incomes of group 100 subjects in the Enforcement treatment are not fully equalized, and the resulting Gini coefficient is small, but different from zero.

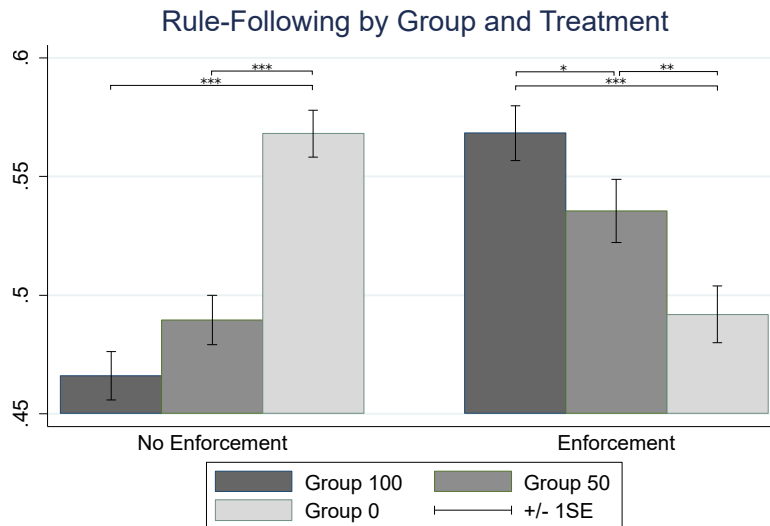
<sup>11</sup>We discuss this in more detail in Section 4.4. Figure 6 provides support for our point: it shows gains and losses due to contribution and redistribution by rule-following propensity.

No Enforcement treatment, a mere *announcement* of a non-binding rule that all income should be contributed to the tax pool creates a world with high inequality and many dissatisfied subjects who consider the redistribution unfair. Thus, a seemingly benevolent attempt at convincing people to contribute more achieves an opposite result than what might have been intended. This is a tangible consequence of the intrinsic rule-following propensity that people have. Our experiment, therefore, shows that established norms of redistribution can have a strong influence on inequality, its perception, and, as a result, general satisfaction with economic policies.

**Result 3.** *The perception of fairness of redistribution is strongly associated with the rule-following propensity and the experience of behavior that violates the norms. Rule-followers find redistribution unfair if norms of contribution were violated. Rule-breakers’ fairness perceptions are not influenced by the norm violations.*

#### 4.4 Choice of an Institution

The results on the perceived fairness of redistribution suggest that rule-following subjects in their choice of institution may gravitate towards the one that, from their perspective, is fairer than others under given enforcement rule. To test this idea, we analyze how subjects choose the institution (or redistribution rules), depending on subjects’ characteristics and treatment. As we have hypothesized in the introduction, we expect to find a connection between the rule-following propensity and the preference over redistribution rules. Namely, we expect rule-followers to prefer more “structured” institution, which, in our design, is represented by the group where 100% of income is expected to be contributed (group 100). To the contrary, rule-breakers should prefer the least structured institution with 0% contribution requirement (group 0). However, if subjects also care about the fairness of redistribution—or, in other words, how often the redistribution rules are broken—then their choice can be influenced by the level of enforcement.



**Figure 5:** Average rule-following propensity by treatment and group.

Figure 5 shows average rule-following propensities by treatment and group. In the No Enforcement treatment rule-followers are congregated in group 0, where there are no rules of redistribution, whereas rule-breakers go to group 100, where the rules are to contribute 100% of income. In the Enforcement treatment, the picture is the opposite: rule-followers prefer group 100 and rule-breakers group 0. Permutation tests on differences of means show significant differences between group 0 and groups 100 and 50 in both treatments.<sup>12</sup> The same observation can be made if we look at the average rule-following propensities in groups 100 and 0 period by period. Figure 10 in Appendix I shows that in the No Enforcement treatment the average rule-following propensity is always higher in group 0, and in the Enforcement treatment it is always higher in group 100. In Appendix F we discuss several regression specifications that support these findings.

The question now is How can we rationalize these observations? In Appendix G we provide evidence that risk preferences and social preferences cannot account for the pattern observed in Figure 5. Thus, we propose an alternative explanation. It is easier to start with the Enforcement treatment where subjects only have a pure choice between institutions which is uncontaminated by the choice of the contribution. We postulate that the observed pattern of group choices shown in the right panel of Figure 5 can only be explained by assuming a separate *preference over redistribution rules*. In particular, rule-followers prefer institutions which are demanding in terms of rules (contribute 100% of income), whereas rule-breakers prefer institutions without rules (contribute any desired percentage of income). This can explain the higher average rule-following propensity in group 100 as compared to group 0 in the Enforcement treatment.

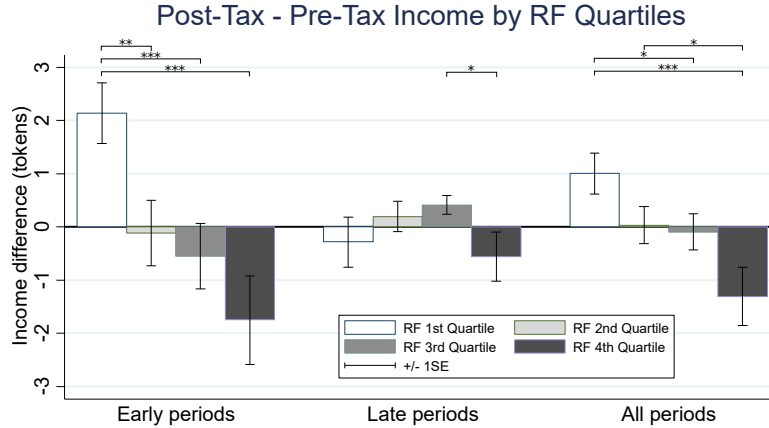
In the No Enforcement treatment subjects choose an institution and also the contribution amount. We see that the pattern of rule-following by group changes to the opposite. In the light of just postulated preference over rules, this means that rule-followers, as much as they care about being in group 100, nevertheless choose to switch to group 0. There are two reasons why this might happen. One is the decrease in payoffs due to the influx of free riders—rule-breakers who switch from group 0 to group 100. Another is the fairness considerations: everything else being equal rule-followers prefer the group where fewer people break the norm, whichever norm that is. The contributions are roughly the same and close to zero in all three groups in the No Enforcement treatment. Thus, rule-followers prefer group 0, which is “fairer” than others in the sense that no one at least is breaking the norm by contributing very little.

To see that rule-followers have a tangible incentive to avoid group 100 due to free riding we analyze the differences between the pre-tax income, the amount that subjects receive in the beginning of each period, and the post-tax income, equal to the pre-tax income minus tax contribution plus

---

<sup>12</sup>For the No Enforcement treatment:  $p = 0.0003$  for comparison between groups 0 and 100;  $p = 0.0005$  between groups 0 and 50. In the Enforcement treatment:  $p = 0.0002$  for comparison between groups 0 and 100;  $p = 0.0258$  between groups 0 and 50, and  $p = 0.0754$  between groups 100 and 50. All  $p$ -values are Benjamini-Hochberg corrected for 6 comparisons. We chose to use permutation tests since, as before in Figure 3, same subject appears in each group multiple times. However, now we cannot use random effects regression to correct for multiple observations since the dependent variable is rule-following propensity, which is fixed for each subject. Rank-sum tests are again invalid due to non-independence of observations.

the return from the public good. Figure 6 shows the difference between post-tax and pre-tax income for the four quartiles of rule-following propensity in group 100 of the No Enforcement treatment. In the early periods, rule-breakers earn more income than they receive in the beginning of the period, while rule-followers lose money, which may force them to leave group 100 and join group 0.<sup>13</sup> In the late periods, when subjects decrease their contributions (see Figure 2), the difference, expectedly, becomes zero for rule-breakers and a little negative for the subjects in the fourth quartile of rule-following (this is consistent with the pattern of contributions in Figure 3).



**Figure 6:** Differences between post-tax and pre-tax incomes in group 100 of the No Enforcement treatment in early and late periods. Permutation tests: significance levels Benjamini-Hochberg corrected for 18 comparisons: \* -  $p < 0.1$ ; \*\* -  $p < 0.05$ ; \*\*\* -  $p < 0.01$ .

To support the fairness reason for rule-followers to leave group 100 in the No Enforcement treatment we notice that groups 100 and 50 are close in terms of average rule-following propensity but both are different from group 0 (Figure 5). Since contributions are very low in all groups in late periods, this observation is consistent with the idea that rule-followers choose group 0 because no one is breaking the norms in it, which makes it different from groups 100 and 50.

To clarify how different incentives shape the preferences of rule-following and rule-breaking subjects in our treatments we propose a reduced form model based on [Acemoglu and Jackson \(2017\)](#) that we describe in Appendix H. The model assumes a specific norm-dependent utility function and explains the observed patterns of institutional choice by rule-followers and rule-breakers shown in Figure 5. Specifically, we show that the self-selection of rule-followers and rule-breakers into different groups in the No Enforcement and Enforcement treatments that we observe can only come about if

<sup>13</sup>We use permutation tests to compare the means of income differences on Figure 6. All  $p$ -values are Benjamini-Hochberg corrected for 18 comparisons. Early periods: quartiles 1 and 2,  $p = 0.0310$ ; quartiles 1 and 3,  $p = 0.0078$ ; quartiles 1 and 4,  $p = 0.0016$ . Late periods: quartiles 3 and 4,  $p = 0.0652$ . All periods: quartiles 1 and 3,  $p = 0.0891$ ; quartiles 1 and 4,  $p = 0.0045$ ; quartiles 2 and 4,  $p = 0.0776$ . No other comparisons are significant.



rule-followers receive utility from being in the group where others abide by the norm (or disutility from being in the group where the norm is violated).

**Result 4.** *The pattern of subjects' institutional choices in the No Enforcement and Enforcement treatments can only be explained by rule-followers' preference for being in more regulated institution and by their aversion to norm violations by others.*

## 5 Discussion

Herbert Simon, in an influential Science article (Simon, 1990), introduced the notion of a “docile individual” (disposed to be taught). Such individuals “tend to learn and believe what they perceive others in the society want them to learn and believe.” According to Simon’s bounded rationality argument, docile individuals learn to follow social norms and adopt preferences, opinions, and attitudes without checking if they contribute to their biological fitness (more progeny) since in a complex world it is either too hard or impossible. By means of a simple culture-gene co-evolution model (Boyd and Richerson, 1988), Simon showed that altruists, individuals who unconditionally forgo their fitness to benefit their community, can survive in the population of selfish individuals if they are docile and the altruistic trait is propagated through the learning of social norms. Our experiment provides a direct evidence to support this hypothesis. We find that rule-followers, defined as people who follow arbitrarily imposed rule at personal cost, are contributing more income to the tax pool than rule-breakers, people who are less inclined to tradeoff money for following an arbitrary rule. This is true even after rule-followers have experienced free riding on the part of rule-breakers (see Figure 3). Thus, we confirm that pro-sociality is indeed learned by docile individuals (in our terminology rule-followers) as socially-appropriate norm of behavior.

We are not the first to make the connection between rule-following and pro-sociality (Kimbrough and Vostroknutov, 2016; Krupka and Weber, 2013). One of the goals of our design was to go further and to explore the possibility that people with different propensities to follow rules might have different preferences over institutions with different degrees of redistribution. We conjectured that docile individuals, who are adept at social learning and are willing to accept the rules of social conduct without question, should prefer an institution which is governed by strict rules to the one with no rules at all. This idea is in line with the hypotheses expressed in Richerson et al. (2016). In this article, the authors argue that the evolution of contemporary complex social systems is only possible if docile individuals, on top of being good social learners, do also have preferences that push them towards more regulated societies. Our data and the model we propose suggest that this connection indeed exists (see Figure 5). This finding can have important implications for the policies related to the regulation of taxation and other public goods problems.

Another goal of our experiment was to see how much enforcement is necessary in order to make individuals abide by the rules of a redistribution mechanism. In the three treatments, ranked by

the degree of enforcement, we find that the contributions to the tax pool erode with time when there is no enforcement or when there is a medium level of enforcement (though, in the latter case the erosion is slower). This happens regardless of the specified redistribution “norm.” Thus, given the choice to join one of the groups, with 100% contributions, 50%, or none, many subjects, who would otherwise stay in the “libertarian” group, join the most regulated society in order to free ride. Rule-followers, given their aversion to norm violations by others, respond by moving to the libertarian group. This leads to an overall decrease of the tax contributions to almost zero in all groups reaching an equilibrium in which rule-breakers end up in the group with 100% contributions and rule-followers in the group with zero required contributions. When the rules of the groups are enforced we observe the preference for redistribution rules discussed above: rule-followers join the most regulated group, while rule-breakers go to the unregulated one.

Putting all these observations together, we can conclude the following. At least in our subject pool, there exists a large heterogeneity in the propensity to abide by rules and social norms. Subjects, who choose to follow norms at a personal cost, also possess an intrinsic desire to be in a regulated environment. However, this preference is not strong enough to overcome the conditional response to free riding, which leads to the decay of contributions and failure of the redistribution mechanism. Nevertheless, our results suggest that full enforcement, which in our design prevents free riding, might not be necessary as long as community consists of norm abiding individuals who are aware that other members are alike. Under these conditions, the preference for regulated society can solve the social dilemma and create an environment where people pay taxes voluntarily.

## 6 Conclusion

We conduct an experiment in order to understand how opportunistic incentives interact with the desire to abide by the rules of an institution. We find that a simple announcement of non-binding rules of redistribution makes many subjects try to follow them. However, in the absence of enforcement, free riding overcomes this tendency and leads to the decay of contributions. Experiencing rule violations changes rule-followers’ perception of fairness of redistribution: they find the same level of inequality fairer when it was achieved without anyone breaking the rules. This is coupled with the preference of rule-followers/breakers for institutions with strong/weak rules. Overall, well-defined rules of conduct provide a good incentive to maintain redistribution. However, some level of enforcement is necessary in order to protect institutions from free riding.

## References

- Acemoglu, D. and Jackson, M. O. (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association*, 15(2):245.
- Alm, J. and Torgler, B. (2011). Do ethics matter? tax compliance and morality. *Journal of Business Ethics*, 101(4):635–651.
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- Boyd, R. and Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.
- Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The Pluralism of Fairness Ideals: An Experimental Approach. *American Economic Review*, 97(3):818–827.
- Cosmides, L. and Tooby, J. (1992). *Cognitive Adaptations for Social Exchange*. Oxford University Press, Inc.
- Cosmides, L. and Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In Buss, D. M., editor, *The Handbook of Evolutionary Psychology*, chapter 20, pages 584–627. John Wiley & Sons, Inc.
- Dal Bó, E., Dal Bó, P., and Eyster, E. (2017). The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies*, 85(2):964–998.
- Deaton, A. (2017). How Inequality Works. *Project Syndicate*.
- Durante, R., Putterman, L., and van der Weele, J. (2014). Preferences for Redistribution and Perception of Fairness: An Experimental Study. *Journal of the European Economic Association*, 12(4):1059–1086.
- Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97:72–86.

- Gordon, J. P. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European economic review*, 33(4):797–805.
- Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770):108–111.
- Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Kessler, J. B. and Leider, S. (2012). Norms and contracting. *Management Science*, 58(1):62–77.
- Kimbrough, E. and Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, forthcoming.
- Kimbrough, E. O. and Vostroknutov, A. (2015). The social and ecological determinants of common pool resource sustainability. *Journal of Environmental Economics and Management*, 72:38–53.
- Kimbrough, E. O. and Vostroknutov, A. (2016). Norms Make Preferences Social. *The Journal of the European Economic Association*, 14(3):608–638.
- Klor, E. F. and Shayo, M. (2010). Social identity and preferences over redistribution. *Journal of Public Economics*, 94(3-4):269–278.
- Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, 90(4):1072–1091.
- Kosfeld, M., Okada, A., and Riedl, A. (2009). Institution formation in public goods games. *American Economic Review*, 99(4):1335–55.
- Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *The Journal of the European Economic Association*, 11(3):495–524.
- Myles, G. D. and Naylor, R. A. (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12(1):49–66.
- OECD (2017). Fighting tax crime: The ten global principles. Technical report, OECD.
- Ostrom, E. (1990). *Governing the Commons: the Evolution of Institutions for Collective Action*. Political economy of institutions and decisions. Cambridge University Press, Cambridge.
- Proto, E., Rustichini, A., and Sofianos, A. (2018). Intelligence, personality and gains from cooperation in repeated interactions. *The Journal of Political Economy*, forthcoming.

- Putterman, L., Tyran, J.-R., and Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics*, 95(9-10):1213–1222.
- Richerson, P., Baldini, R., Bell, A., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E., Narr, N., Newson, L., Ross, C., Smaldino, P., Waring, T., and Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, pages 1–68.
- Rustichini, A. and Vostroknutov, A. (2014). Merit and Justice: An Experimental Analysis of Attitude to Inequality. *PLoS ONE*, 9(12):e114512–25.
- Simon, H. (1990). A mechanism for social selection and successful altruism. *Science*, 250(4988):1665–1668.
- Starmans, C., Sheskin, M., and Bloom, P. (2017). Why People Prefer Unequal Societies. *Nature Human Behaviour*, 1:0082.
- Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *The Review of Economic Studies*, 77(4):1540–1566.

# Appendix

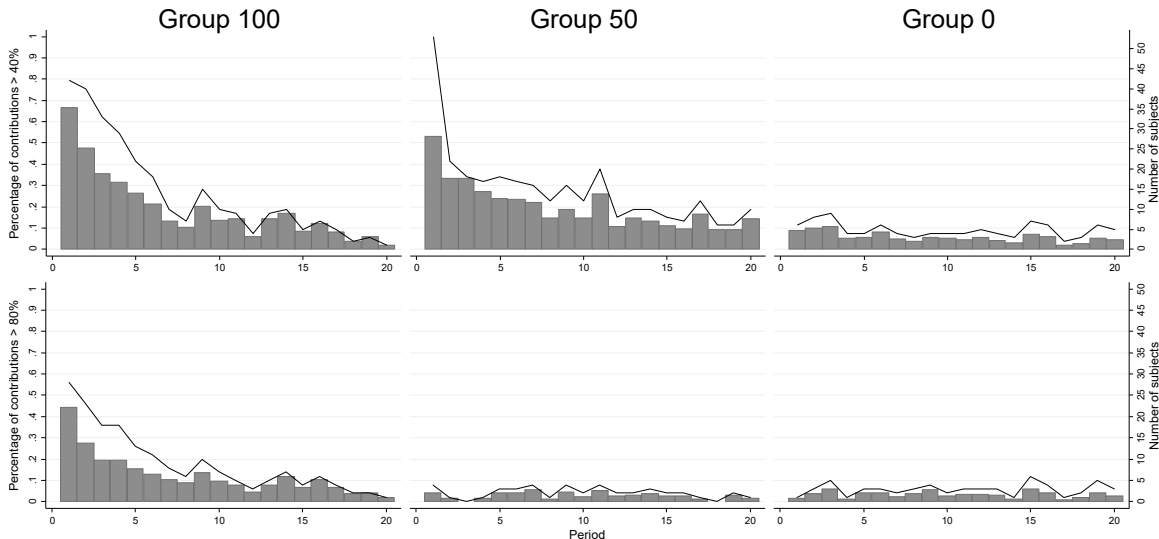
## A Variables

Variable	Range	Definition
No Enforcement Treatment	0/1	dummy for the No Enforcement treatment
Enforcement Treatment	0/1	dummy for the Enforcement treatment
Exclusion Treatment	0/1	dummy for the Enforcement treatment
Rule-Following	[0, 1]	<i>rescaled</i> number of balls in the blue bucket (1 if all in the blue bucket)
Age	[18, 29]	age of the subject
Gender	0/1	gender of the subject (1 if male)
Number of Siblings	[0, 7]	number of siblings of the subject
Number of Younger Siblings	[0, 7]	number of younger siblings of the subject
Number of Older Siblings	[0, 8]	number of older siblings of the subject
Economics or Business Major	0/1	major of the subject (1 if economics or business)
Number of Economics Courses	[0, 4]	number of economics courses taken by the subject
Organization Membership	0/1	membership (to an organization) status of the subject (1 if member)
Risk Appetite	[0, 1]	<i>rescaled</i> self-reported willingness to take risk (1 if risk-seeker)
Justification Tendency	[0, 1]	<i>rescaled</i> belief that cheating is justifiable (1 if justifiable)
Belief that the Poor are in Need due to Unfair Society	[0, 1]	<i>rescaled</i> belief about why the Poor are in need (1 if due to unfair society)
Belief that Incomes Should be More Equal	[0, 1]	<i>rescaled</i> preference for equality of incomes (1 if should be equal)
Belief that Hard Work Pays Off	[0, 1]	<i>rescaled</i> belief on the role of luck vs. effort (1 if hard work pays off)
Fairness Perception of the Income Distribution in the Country	[0, 1]	<i>rescaled</i> fairness evaluation of the subject (1 if fair)
Fairness Perception of the Income Distribution by the Experiment Country	[0, 1]	<i>rescaled</i> fairness evaluation of the subject (1 if fair)
Trust	0/1	subject belief on whether others are trustable (1 if trustable)
Contribution Ratio Group 100	[0, 1]	contribution ratio to the group common pool (1 if 100% contribution)
Group 100	0/1	dummy for the egalitarian Group 100 choice (1 if a subject joins Group 100)
Group 50	0/1	dummy for the semi-egalitarian Group 50 choice (1 if a subject joins Group 50)
Group 0	0/1	dummy for the libertarian Group 0 choice (1 if a subject joins Group 0)
Frequency in Group 100	[0, 1]	frequency of time spent in Group 100 over 20 periods (1 if only joined Group 100)
Frequency in Group 50	[0, 1]	frequency of time spent in Group 50 over 20 periods (1 if only joined Group 50)
Frequency in Group 0	[0, 1]	frequency of time spent in Group 0 over 20 periods (1 if only joined Group 0)
Switch	0/1	dummy for switching group over last period (1 if switched group)
Pre-Tax Income (in TL)	[0, 50]	income before contribution and redistribution
Post-Tax Income (in TL)	[0, 50]	income after contribution and redistribution
Income Change (in TL)	[-42, 33.37]	income change after contribution and redistribution (+ if better-off)

**Table 2:** Variables used in the analyses and regressions.

## B Effect of Rules on Contributions

An important difference between our design and other similar experiments is that we explicitly state the rules of conduct for each group, which subjects can join. In particular, in group 100 subjects know that in this group “we should all put our whole income (100%) to the common pool”, and in group 50 “we should all put half of our income (50%) to the common pool.” Thus, the norm of each group is well-defined unlike in typical public goods or redistribution experiments where nothing is said about what the contributions should be. While in our experiment free riding is perceived as a clear violation of the rules of the group, in standard experiments subjects might be not sure about what the appropriate level of contributions is and “discover” it as they choose repeatedly.



**Figure 7:** The percentages of contributions above 40% and 80% of income by group in the No Enforcement and Exclusion treatments (gray bars, left  $y$ -axis). Number of subjects who contributed above 40% and 80% of income (black lines, right  $y$ -axis).

We introduced group rules in order to study how opportunism and the desire to follow rules interact and influence redistribution choices. Even though free riding and conditional cooperation eventually decrease contributions in the No Enforcement and Exclusion treatments, we still observe a sizeable effect of the pre-specified rules on the contribution choices. Figure 7 shows the percentage and number of subjects in groups 100, 50, and 0 who contributed above 40% of income and above 80% of income (in the No Enforcement and Exclusion treatments together).<sup>1</sup> The differences are very noticeable: in group 0 virtually no one makes contributions above 40%; in group 50 many subjects make contributions above 40%, but almost no one above 80%; and in group 100 a significant number of subjects contribute above 80%.

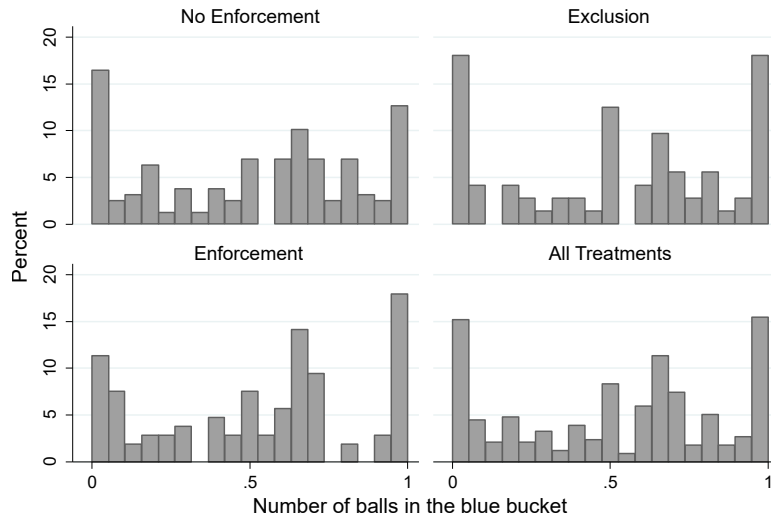
The rules of the three groups are simple announcements that are non-binding and bear no cost of violating them.<sup>2</sup> Nevertheless, they have a strong effect on contributions in early periods. This demonstrates that rules, as artificial as they are in our experiment, can be a force that is capable to sustain redistribution.

<sup>1</sup>The numbers 40% and 80% were chosen since these are the thresholds below which subjects can be excluded in groups 50 and 100 in the Exclusion treatment. So, choosing contributing above 40% in group 50 and above 80% in group 100 has no consequences.

<sup>2</sup>In the Exclusion treatment the only cost is the potential lost future gains from free riding after exclusion. However, given that contributions decrease with time, the expected gains of this sort are not very high.

## C Analysis of the Rule-Following Propensities

Figure 8 shows the histograms of the rule-following choices as measured by the number of balls that subjects put into the blue bucket in the rule-following task. The pattern is very similar to that observed in the Netherlands, US, Canada, and Italy (Kimbrough and Vostroknutov, 2018). Namely, there is a considerable number of subjects who put less than 5 balls into the blue bucket (rule-breakers), a considerable number who put more than 95 balls into the blue bucket (rule-followers), and also many subjects in between. In our data, there are 15% of both extreme rule-followers and extreme rule-breakers. Rank-sum tests between treatments show no significant difference in distributions ( $p > 0.77$ ). This suggests that the subject pool was not contaminated by early sessions.



**Figure 8:** Histograms of rule-following in three treatments.

Table 3 in this Appendix shows the OLS and logit regressions of the number of balls that subjects put in the blue bucket as dependent on various demographic variables and preference estimates. Three variables have a significant effect on the number of balls in the blue bucket: 1) the number of economics courses taken decreases the rule-following propensity by the maximum of around 20%; 2) membership in organization increases the rule-following propensity by 8%; 3) risk preferences, measured by a Likert scale question *How willing are you to take risks in general?*, change the rule-following propensity by up to 20%: risk averse subjects seem to be more rule-following, though the significance level is at 10%.<sup>3</sup> It should be mentioned as well that none of the questions related to social preferences (see Appendix K) show any significant effect on rule-following.

<sup>3</sup>Dohmen et al. (2011) used the same question to elicit risk preferences and showed that it is a reliable measurement.



## Rule-Following Choice

VARIABLES	(1)	(2)	(3)	(4)
	Rule Following	2-Limit Tobit	Pr(Blue Balls=0)	Pr(Blue Balls=100)
Age	0.007 (0.011)	0.005 (0.013)	-0.001 (0.006)	-0.014 (0.009)
Gender	-0.040 (0.040)	-0.061 (0.045)	0.044** (0.019)	-0.057* (0.031)
Number of Younger Siblings	-0.011 (0.023)	-0.013 (0.025)	-0.000 (0.010)	-0.009 (0.016)
Number of Older Siblings	-0.024 (0.023)	-0.028 (0.028)	0.003 (0.007)	-0.018 (0.033)
Economics or Business Major	0.040 (0.050)	0.039 (0.058)	-0.020 (0.021)	-0.048 (0.039)
Number of Economics Courses	-0.043** (0.018)	-0.044** (0.021)	0.010 (0.006)	0.017 (0.016)
Organization Membership	0.080** (0.037)	0.090** (0.043)	-0.006 (0.018)	0.029 (0.032)
Risk Appetite	-0.191* (0.102)	-0.222* (0.117)	0.028 (0.046)	-0.076 (0.079)
Justification Tendency	0.028 (0.093)	0.022 (0.109)	-0.049 (0.053)	-0.087 (0.092)
Belief that the Poor are in need due to Unfair Society	-0.035 (0.076)	-0.029 (0.087)	-0.003 (0.030)	0.037 (0.059)
Belief that Hard Work Pays Off	-0.007 (0.072)	-0.016 (0.084)	-0.003 (0.035)	-0.044 (0.057)
Belief that Incomes Should be More Equal	-0.054 (0.077)	-0.081 (0.092)	-0.015 (0.027)	-0.136** (0.058)
Fairness Perception of the Income Distribution in the Country	-0.135 (0.157)	-0.144 (0.190)	0.052 (0.055)	0.085 (0.107)
Fairness Perception of Income Distribution by the Experiment	0.008 (0.071)	0.009 (0.085)	0.035 (0.031)	0.058 (0.065)
Constant	0.612** (0.248)	0.705** (0.279)		
Observations	336	336	336	336
$R^2$	0.066			
Tobit $\sigma$		0.372*** (0.015)		

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 3:** OLS, 2-limit tobit and logit regressions of the propensity to follow rules on various demographic variables. Variables are described in Appendix A.

## D Additional Analysis of Rule Following and Contributions

Periods:	No Enforcement			Exclusion		
	Early	Late	All	Early	Late	All
constant	0.081*** (0.019)	0.050*** (0.018)	0.066*** (0.016)	0.188*** (0.030)	0.040*** (0.010)	0.115*** (0.018)
RF 2nd Quartile	0.072** (0.032)	0.008 (0.022)	0.041* (0.023)	0.066 (0.043)	0.025 (0.021)	0.045 (0.028)
RF 3rd Quartile	0.115*** (0.027)	0.017 (0.022)	0.066*** (0.021)	0.017 (0.045)	0.053* (0.029)	0.035 (0.033)
RF 4th Quartile	0.109*** (0.036)	0.050* (0.030)	0.081*** (0.029)	0.144*** (0.048)	0.101*** (0.032)	0.122*** (0.035)
<i>N</i>	1,564	1,562	3,126	713	709	1,422
<i>N</i> subjects	158	158	158	72	72	72

**Table 4:** Random effects regressions of contributions. Errors are clustered by subjects and robust. Variables are described in Appendix A. \* –  $p < 0.1$ ; \*\* –  $p < 0.05$ ; \*\*\* –  $p < 0.01$ .

	Contribution Ratio Choice	
	(1) Contribution Ratio	(2) Contribution Ratio
No Enforcement Treatment	-0.379*** (0.032)	-0.379*** (0.031)
Exclusion Treatment	-0.344*** (0.024)	-0.339*** (0.024)
Rule-Following	0.132*** (0.040)	0.121*** (0.039)
Rule-Following x No Enforcement	-0.075 (0.048)	-0.068 (0.047)
Rule-Following x No Enforcement x Group 100	0.116*** (0.028)	0.115*** (0.028)
Income	-0.002*** (0.000)	-0.002*** (0.000)
Period	-0.008*** (0.001)	-0.008*** (0.001)
Constant	0.586*** (0.031)	0.368*** (0.133)
Observations	6,656	6,656
Number of subjects	336	336
Controls	No	Yes

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 5:** Random effects GLS regressions of contribution choices. Errors are clustered by subject. Variables are described in Appendix A.

Periods:	Early	Late	All
constant	0.082*** (0.019)	0.050*** (0.018)	0.066*** (0.016)
Exclusion Treatment	0.106*** (0.036)	-0.010 (0.020)	0.049** (0.024)
RF 2nd Quartile	0.072** (0.032)	0.008 (0.022)	0.041* (0.023)
RF 3rd Quartile	0.115*** (0.027)	0.017 (0.022)	0.066*** (0.021)
RF 4th Quartile	0.109*** (0.036)	0.050* (0.030)	0.080*** (0.029)
Exclusion Treatment × RF 2nd Quartile	-0.006 (0.054)	0.017 (0.030)	0.004 (0.036)
Exclusion Treatment × RF 3rd Quartile	-0.097* (0.053)	0.036 (0.036)	-0.031 (0.039)
Exclusion Treatment × RF 4th Quartile	0.035 (0.059)	0.051 (0.044)	0.041 (0.046)
$N$	2,277	2,271	4,548
$N$ subjects	230	230	230

**Table 6:** Random effects regressions of contributions. Errors are clustered by subjects and robust. Variables are described in Appendix A. \* -  $p < 0.1$ ; \*\* -  $p < 0.05$ ; \*\*\* -  $p < 0.01$ .

## E Additional Analysis of Fairness Perceptions

To further support the findings reported in Figure 4 we run an OLS regression of fairness perception on average income and the frequency of being in group 100. The results are presented in Table 7. The regression of the Enforcement treatment data shows no significant effect of either variable on fairness perception. This is not surprising since in this treatment subjects join any group that they like by their own volition, thus, committing themselves to the rules of that group. We see a significant influence of income on fairness perception only among rule-followers in the No Enforcement treatment: higher income makes them feel that the distribution is fairer, whereas frequency of being in group 100 decreases the perceived fairness of income distribution. This happens because rule-followers, who try to contribute high amounts according to the rules of group 100, think that free riders are breaking the rule. This creates the feeling of an unfair distribution of income.

VARIABLES	Fairness Perception				
	(1) All	(2) Enforcement	(3) No Enforcement	(4) No Enf. Rule-Breakers	(5) No Enf. Rule-Followers
Average Post-Tax/Pre-Tax Income	0.153 (0.206)	-0.262 (0.270)	0.406* (0.215)	0.061 (0.478)	0.612*** (0.232)
Frequency in Group 100	-0.032 (0.086)	0.144 (0.118)	-0.230** (0.095)	-0.150 (0.147)	-0.287** (0.125)
Constant	0.247 (0.199)	0.724*** (0.264)	-0.021 (0.206)	0.308 (0.472)	-0.214 (0.214)
Observations	264	106	158	71	87
R-squared	0.003	0.019	0.045	0.015	0.088

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

**Table 7:** OLS regressions of fairness perception as measured by the answer to the question “*How fair do you find the income distribution that resulted in the experiment?*” on a 0 to 10 Likert-scale. Variables are described in Appendix A.

Finally, it should be emphasized that we find no effect of any variables, except gender on the perception of fairness of rule-breakers (Table 8 below). This suggests an intimate connection between fairness judgments and rule-following propensity. In particular, it seems that rule-followers do associate rule violations with unfair distribution of income whereas rule-breakers do not. It remains an open question whether the lack or inconsistency of moral judgment makes rule-breakers not responsive to rules or vice versa. We leave this question for the future investigation.

VARIABLES	Fairness Perception				
	(1) All	(2) Enforcement	(3) No Enforcement	(4) No Enf. Rule-Breakers	(5) No Enf. Rule-Followers
Average Post-Tax/Pre-Tax Income	0.201 (0.209)	-0.161 (0.323)	0.376 (0.228)	0.083 (0.502)	0.560*** (0.210)
Frequency in Group 1	-0.059 (0.085)	0.164 (0.131)	-0.230** (0.096)	-0.153 (0.142)	-0.204 (0.133)
Age	0.026** (0.012)	0.018 (0.020)	0.023 (0.014)	0.035 (0.025)	0.019 (0.021)
Gender	0.080** (0.037)	0.042 (0.064)	0.039 (0.046)	0.153** (0.074)	-0.047 (0.057)
Number of Younger Siblings	-0.014 (0.022)	-0.041 (0.032)	-0.012 (0.026)	-0.031 (0.032)	0.025 (0.039)
Number of Older Siblings	-0.021 (0.020)	-0.045 (0.048)	-0.000 (0.025)	-0.008 (0.042)	0.003 (0.033)
Economics or Business Major	0.067 (0.051)	0.097 (0.077)	0.023 (0.063)	0.139* (0.079)	-0.103 (0.095)
Number of Economics Courses	-0.020 (0.018)	-0.022 (0.030)	0.005 (0.022)	-0.051 (0.032)	0.052 (0.034)
Organization Membership	0.022 (0.035)	0.055 (0.059)	-0.021 (0.040)	-0.001 (0.065)	-0.047 (0.055)
Constant	-0.376 (0.355)	0.225 (0.628)	-0.494 (0.394)	-0.482 (0.892)	-0.614 (0.471)
Observations	264	106	158	71	87
R-squared	0.050	0.056	0.078	0.154	0.141

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

**Table 8:** OLS regressions of fairness perception as measured by the answer to the question “*How fair do you find the income distribution that resulted in the experiment?*” on a 0 to 10 Lickert-scale. Variables are described in Appendix A.

## F Additional Analysis of a Choice of an Institution

In this section we report several regression specifications that support the findings shown in Figure 5.

Frequency of times in	(1) Group 100	(2) Group 100	(3) Group 50	(4) Group 50	(5) Group 0	(6) Group 0
No Enforcement Treatment	0.112** (0.052)	0.119** (0.054)	0.051 (0.059)	0.051 (0.059)	-0.163** (0.068)	-0.170** (0.070)
Rule-Following	0.120* (0.070)	0.109 (0.070)	0.013 (0.076)	0.014 (0.076)	-0.133 (0.085)	-0.123 (0.089)
Rule-Following x No Enforcement	-0.252*** (0.086)	-0.238*** (0.088)	-0.073 (0.092)	-0.082 (0.092)	0.324*** (0.110)	0.321*** (0.115)
Constant	0.278*** (0.041)	0.276 (0.216)	0.288*** (0.050)	0.348* (0.181)	0.434*** (0.054)	0.376* (0.222)
Observations	264	264	264	264	264	264
Controls	No	Yes	No	Yes	No	Yes

**Table 9:** OLS regressions of the frequency of group choices. Robust standard errors in parentheses. Significance levels: \* -  $p < 0.1$ ; \*\* -  $p < 0.05$ ; \*\*\* -  $p < 0.01$ .

Consider the OLS regressions presented in Table 9. For each subject the dependent variable is the frequency of times that she chose group 100, 50, and 0. Independent variables are: dummy for the No Enforcement treatment, rule-following propensity (number of balls in the blue bucket normalized to  $[0, 1]$ ), and their interaction. The regressions show that more subjects overall choose group 100 and less subjects group 0 in the No Enforcement treatment (significant coefficient on the dummy, columns 2 and 6). Conversely, the significance of the interaction term suggests that in the No Enforcement treatment subjects with high rule-following propensity leave group 100 and join group 0: exactly the effect that the averages above show.

The same effects are found in other regression specifications. Ordered logit regressions of the probabilities of choosing groups 100, 50, and 0 show the same effect of the No Enforcement treatment dummy and its interaction with rule-following (Table 10 below). The same is true for the logit regressions of probabilities of choosing a specific group (Table 11 below). Finally, the logit regressions in Table 12 show that rule-followers in the No Enforcement treatment, after joining group 0, tend to switch to other groups with a lower probability than rule-breakers (negative coefficient on the interaction in column 6). These auxiliary regressions provide support for our interpretation of the results in Section 4.4.

Probability of Group Choice		
	(1)	(2)
	Prob(Group)	Prob(Group)
No Enforcement Treatment	-0.781** (0.339)	-0.814** (0.350)
Rule-Following	-0.807* (0.423)	-0.725 (0.442)
Rule-Following x No Enforcement	1.748*** (0.553)	1.686*** (0.577)
Period	0.007 (0.007)	0.007 (0.007)
Constant Cut 1	-1.220*** (0.276)	-0.891 (1.246)
Constant Cut 2	0.361 (0.278)	0.690 (1.249)
$\sigma_u^2$	1.592 (0.249)	1.548 (0.244)
Observations	5,280	5,280
Number of subjects	264	264
Controls	No	Yes

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

**Table 10:** Random effects ordered logit regressions of the probability of group choice in the No Enforcement and Enforcement treatments. Groups are ordered as follows: group 100, group 50, group 0. So, positive coefficients mean the increase in probability of joining group 0. Errors are clustered by subject. Variables are described in Appendix A.

	Group Choice Probability					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(Group 100)	Pr(Group 100)	Pr(Group 50)	Pr(Group 50)	Pr(Group 0)	Pr(Group 0)
No Enforcement Treatment	0.121** (0.055)	0.127** (0.057)	0.058 (0.060)	0.059 (0.060)	-0.167** (0.071)	-0.177** (0.073)
Rule-Following	0.125* (0.072)	0.112 (0.074)	0.024 (0.078)	0.025 (0.078)	-0.139 (0.087)	-0.136 (0.091)
Rule-Following x No Enforcement	-0.256*** (0.089)	-0.242*** (0.092)	-0.082 (0.094)	-0.093 (0.094)	0.339*** (0.114)	0.344*** (0.119)
Period	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Observations	5,280	5,280	5,280	5,280	5,280	5,280
Controls	No	Yes	No	Yes	No	Yes

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

**Table 11:** Random effects logit regressions of the probabilities of choosing a group in the No Enforcement and Enforcement treatments. Errors are clustered by subject. Variables are described in Appendix A.

VARIABLES	Group Switch Choice Conditional on Last Period Groups					
	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(Switch   100)	Pr(Switch   100)	Pr(Switch   50)	Pr(Switch   50)	Pr(Switch   0)	Pr(Switch   0)
No Enforcement Treatment	-0.033 (0.063)	-0.043 (0.063)	0.016 (0.084)	0.001 (0.084)	0.176* (0.090)	0.181* (0.095)
Rule-Following	-0.051 (0.083)	-0.065 (0.083)	0.038 (0.110)	0.027 (0.104)	0.146 (0.104)	0.124 (0.113)
Rule-Following x No Enforcement	0.183* (0.108)	0.188* (0.107)	0.017 (0.131)	0.040 (0.129)	-0.372*** (0.143)	-0.361** (0.152)
Period	0.002 (0.002)	0.002 (0.002)	-0.012*** (0.002)	-0.013*** (0.002)	-0.005* (0.002)	-0.005* (0.002)
Observations	1,666	1,666	1,507	1,507	1,843	1,843
Controls	No	Yes	No	Yes	No	Yes

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

**Table 12:** Random effects logit regressions of the probability of leaving a group. Errors are clustered by subject. Variables are described in Appendix A.

## G Alternative Hypotheses about the Choice of Institution

In this section we falsify several alternative hypotheses that can partially account for the patterns observed in Figure 5. One possible explanation is risk preferences. In the Enforcement treatment subjects cannot contribute less than 100% of income in group 100 and can contribute any amount in group 0, which leads to almost zero contributions (see Figure 2). This entails different degrees of variance of payoffs in the groups. Variance is the highest in group 0 and the lowest in group 100. Before we reported a connection between risk preferences and rule-following (see first regression in Table 3). Namely, rule-followers seem to be more risk-averse than rule-breakers. Thus, the pattern on the right panel of Figure 5 can, in principle, be explained by risk preferences. However, the comparisons of distributions of elicited risk preferences between groups in the Enforcement treatment reveal no significant differences (rank-sum tests, all  $p > 0.18$ ). In the No Enforcement treatment, the variance of payoffs in all three groups is approximately the same since all subjects contribute the amounts close to zero after period 10. Thus, according to the risk preferences hypothesis, we should not observe any difference in rule-following propensity among groups. This is not the case. Therefore, we can conclude that risk preferences are not responsible for the observed effect.

Another possible explanation is social preferences. Indeed, in group 100 of the Enforcement treatment the payoffs are the same for all subjects who joined it, and in group 0 payoffs are random and unequal. Thus, inequality averse subjects might be willing to join group 100 and selfish subjects group 0. In the post-experimental questionnaire we asked subjects to agree or disagree with the statement “*Incomes should be made more equal*” on a 10 options Likert-scale. We find no significant difference in distributions of answers between three groups (rank-sum tests, all  $p > 0.72$ ). Thus, the differences between groups in the Enforcement treatment cannot be explained by self-selection due to distributional concerns. The only explanation that remains is, therefore, the one involving rule-following propensity.<sup>4</sup>

---

<sup>4</sup>It should be noted that the choices to put balls in the blue bucket in the rule-following task cannot be attributed to any social preferences. It is an individual task with no one else but the subject herself being influenced by the choices. Subjects without norm preferences, but with social preferences, should never put any balls in the blue bucket.



# H Model

Individuals have a norm-dependent utility where a “norm” is prescribed by the rule of the group that they choose to join (contribute 100% of income, 50%, or any amount). Let  $r_i \in [0, 1]$  denote the rule-following propensity of individual  $i$ , and  $T_s \in [0, 1]$  be the exogenously assigned tax rate in society  $s$ . For simplicity, we assume there are only two types of individuals,  $H$  and  $L$  (i.e., **H**igh and **L**ow rule-followers) with  $r_H > r_L$ .

There are two societies which differ by the redistribution norms: **E**galitarian and **F**ree, with  $T_E = 1$  and  $T_F = 0$ , i.e., the norm in the egalitarian society is to contribute all income to the public tax pool and receive the same rebate as everyone in the group. The free society presumes no one has to pay taxes, hence, the norm is not to contribute to the pool at all, so that taxation and redistribution become redundant.

We define the utility of individual  $i$  from choosing to belong to society  $s$  to be

$$V_i(s) = \underbrace{-r_i(t_i - T_s)^2}_{\text{Rule-following}} \underbrace{-(1 - r_i)t_i^2}_{\text{Self interest}} \underbrace{+\delta(t_{-i}^s)}_{\text{Public good interest}}, \quad (1)$$

where  $t_i$  is the amount of tax chosen by type  $i$  and  $t_{-i}^s = \frac{\sum_{k \neq i}^{N_s} t_k}{N_s - 1}$  is the average tax rate chosen by other members of the society  $s$  ( $N_s$  is the number of members of  $s$ ).

The first term in (1) is the disutility from deviating from the prescribed norm  $T_s$  with rule-followers being more sensitive to it than rule-breakers. The second term in (1) is the consumption utility, or disutility from paying taxes. Notice that the consumption utility decreases with the rule-following propensity  $r_i$ . This is a modelling choice that we will discuss at the end of this section. The third term is a positive externality from other members of the group paying taxes. For the sake of generality, we assume that  $\delta(\cdot)$  is independent of the type and increases with the average amount of taxes paid by other members of the group.

We assume the following:

- (i)  $0.5 > r_H > r_L > 0$
- (ii)  $\delta(0) = 0$
- (iii)  $r_H(1 - r_H) > \delta(r_L) > r_L(1 - r_L)$
- (iv)  $1 - r_L > \delta(1) > 1 - r_H$

First assumption simply states that rule-followers care more about following the rule of the society than rule-breakers. The upper and lower bounds are for the sake of technical simplicity. Second assumption is merely a matter of normalization: if other members of the group do not pay any taxes, i.e.,  $t_{-i}^s = 0$ , then individual  $i$  does not have any utility from externality. Third assumption provides a technical lower bound on the benefit from externality for the rule-breakers. Note that if the externality is simply a numeraire,  $\delta(x) = x$ , then  $\delta(r_L) > r_L(1 - r_L)$  holds automatically given assumption (i). Furthermore, while  $r_H(1 - r_H) > r_L(1 - r_L)$  directly follows from (i),  $r_H(1 - r_H) > \delta(r_L)$  imposes an upper bound on the externality when other members contribute  $r_L$  to the tax pool. Fourth assumption imposes a limit on  $\delta(1)$  when all other members of the society contribute all their income.

It is straightforward to see that regardless of the degree of enforcement, the first-order condition of individual  $i$  is  $t_i = r_i T_s$ .

Next, we analyze the model and check if there are equilibria that qualitatively resemble our experimental observations. We study two cases of enforcement:

### Case 1: Enforcement

In this scenario, individuals have to set  $t_i = 1$  in society  $E$  and  $t_i = 0$  in society  $F$  due to enforcement. This implies that the utility of rule-followers who choose the egalitarian society is

$$V_H(E) = -(1 - r_H) + \delta(1) \quad (2)$$

and the utility of rule-breakers who choose the egalitarian society is

$$V_L(E) = -(1 - r_L) + \delta(1). \quad (3)$$

Assumption (iv) implies that the right-hand-side of (2) is positive, and the right-hand-side of (3) is negative. Since  $V_i(F) = -r_i(0 - 0)^2 - (1 - r_i) \cdot 0 + \delta(0) = 0$ , i.e., both types receive utility level of 0 from choosing the free society, only rule-breakers, i.e., individuals with low rule-following propensity  $r_L$ , choose the free society, as 0 is higher than the utility they would have gotten from the egalitarian society. Accordingly, in equilibrium, only rule-followers choose the egalitarian society and only rule-breakers choose the free society, as seen in the data.

The separation of types in equilibrium happens because, in the interpretation of this model, the two types receive different disutility from paying taxes, which depends on the rule-following propensity. This might be not a particularly desirable property since consumption utility from taxes and from public good are weighed differently. Below we show how to reinterpret the model so that the disutility from paying taxes is the same for both types.

### Case 2: No enforcement

In this scenario, when type  $i$  chooses the egalitarian society, she will set  $t_i = r_i T_E = r_i$  in accordance with her first-order condition. In the same vein, she will choose  $t_i = 0$  in the free society. As no type is interested in contributing a positive amount in the free society,  $t_{-i}^F = \frac{\sum_{k \neq i}^{N_F} t_k}{N_F - 1} = \frac{\sum_{k \neq i}^{N_F} 0}{N_F - 1} = 0$ , the utility of both types choosing the free society is  $V_i(F) = -r_i(0 - 0)^2 - (1 - r_i) \cdot 0 + \delta(0) = 0$ .

We look for an equilibrium in which rule-breakers choose the society  $E$ , and rule-followers choose  $F$ . Note that in such an equilibrium each rule-breaker pays  $r_L$  in taxes in society  $E$  and rule-followers pay 0 in society  $F$ . Hence, rule-breakers receive a utility of  $V_L(E) = -r_L(1 - r_L) + \delta(r_L)$ , which is strictly positive given assumption (iii). Therefore, rule-breakers do not have any incentive to unilaterally deviate to society  $F$ . At the same time,  $V_H(E) = -r_H(1 - r_H) + \delta(r_L)$ . By assumption (iii) this is negative. Thus, rule-followers have no incentive to unilaterally deviate from the free society to the egalitarian one. Hence, this is indeed an equilibrium as no type has an incentive to deviate. Therefore, in No Enforcement

case, rule-breakers choose the egalitarian society and rule-followers choose the free society, as seen in the experimental data.

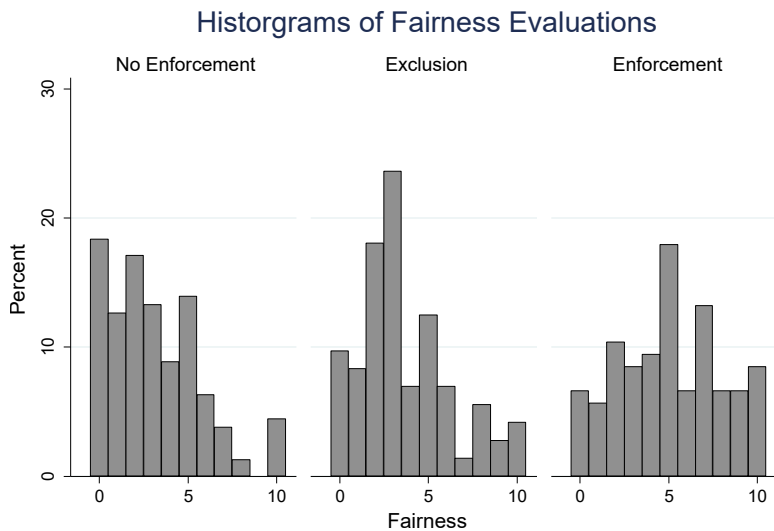
### Reinterpretation

The specification of the utility function with convex combination of norm and consumption terms was taken from [Acemoglu and Jackson \(2017\)](#). However, in our setting, there is an additional utility received from public good. The problem, as was mentioned above, is that in this case, the two parts of consumption utility end up being weighed differently: the tax part is weighed with  $1 - r_i$  whereas the public good part has weight 1. It is very simple to reinterpret the model without any changes to equilibrium analysis by dividing the utility by  $1 - r_i$ . The transformed utility becomes:

$$V_i(s) = -\phi_i(t_i - T_s)^2 - t_i^2 + (1 + \phi_i)\delta(t_{-i}^s) = -t^2 + \delta(t_{-i}^s) + \phi_i(\delta(t_{-i}^s) - (t_i - T_s)^2), \quad (4)$$

where  $\phi_i = r_i/(1 - r_i)$  is a monotonic increasing transformation of  $r_i$ . This way we have consumption utility  $-t^2 + \delta(t_{-i}^s)$  with two terms weighed equally and norm-dependent utility  $\phi_i(\delta(t_{-i}^s) - (t_i - T_s)^2)$ . The first term in the norm-dependent utility,  $\phi_i\delta(t_{-i}^s)$ , is a novel one: it represents the utility that rule-followers obtain from *others following the norm*. Notice that the higher the rule-following propensity the more utility is obtained from observing others abide by the rules. It is exactly this term that makes the equilibria described above work. Without this term, or, in other words, with the standard norm-dependent utility ([Kessler and Leider, 2012](#)), it would be impossible to construct equilibria that would explain the behavior we observe in the experiment.

# I Additional Graphs



**Figure 9:** The distributions of the answers to the question “How fair do you find the income distribution that resulted in the experiment?” on a 0 to 10 Likert-scale.



**Figure 10:** Rule-following measured as a percentage of balls in blue bucket by treatment and group.

# J Instructions

## J.1 Rule-Following Task Instructions

**General information** You are now participating in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in CASH at the end of the experiment

This set of instructions is for your private use only. During the experiment you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then we will come to you and answer your questions privately. Any violation of this rule excludes you immediately from the experiment.

**Part I** In Part I of this experiment, you will decide how to allocate 100 balls between two buckets.

Your task is to put each of the balls, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket. The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice. For each ball you put in the blue bucket, you will receive 5 cents, and for each ball you put in the yellow bucket, you will receive 10 cents.

The rule is to put the balls in the blue bucket.

Once the experiment begins, you will have 10 minutes to put the balls into the buckets. When you are finished, please wait quietly until the end of the 10-minute period.

Your payment from Part I will be based on your decisions: it is the sum of payments from the blue and yellow buckets.

This is the end of the instructions for Part I. If you have any questions, please raise your hand and an experimenter will answer them privately. Otherwise, please wait quietly for the experiment to begin.

## J.2 Institution Choice and Redistribution Task Instructions

### **General Information:**

You will get an income determined by chance in every round of the experiment. At the beginning of this experiment and every subsequent round you will be able to take one of three groups according to your preference before your income is determined. These groups will differ in terms of earnings sharing principles.

In the first group (A) the principle of division is that all your income is collected in a common tax pool and distributed equally among all participants. In the second group (B) the principle of division is that you transfer half of your income to the common pool, and then the amount in the common pool is evenly distributed among all participants. In the third group (C), the contribution you make to your pool from your earning is entirely on volunteerism.

As can be seen from the above explanations, in all three groups the final income of the participants will be determined by deducting the contribution they made to the common pool from their initial income and adding the equal share from the common pool.

Each round, you will have the opportunity to change your preferred group before your income is set, so the number of participants in each of the three groups will be determined by your preference in each round. During a tour, participants in one group will only be influenced by their decisions from the other participants in that group and will not be affected by the decisions of the participants in the other group.

**Roadmap:**

This part of the experiment consists of 20 rounds. Each round contains 2 stages. In the first stage, participants will choose their group. In the second, participants' initial incomes will be determined by chance, they will then choose their contributions to the pool and then their final incomes will be determined.

**Stage 1 - Group Selection:**

In stage 1, each participant decides on the group she wants to join. You can join one of three different groups (A, B, C).

- (A) We should all put our income in the common pool so that income is distributed equally among all the individuals
- (B) We should all put 50% of our income in the common pool so that income inequality is (remains/stays) at a reasonable level.
- (C) No one should have to contribute to the pool and must do so only at their own discretion and at any amount they choose.

In the first stage, the participants choose one of these three groups. Participants can not observe their income before the 1st stage group selection.

**Stage 2 - Determination of Income:**

After the group selection, the participants are assigned their initial income , which is a value between 0-50 TL determined by chance. Participants will see their first income on a screen and choose how much they will contribute to the common pool on this screen. This contribution must be at least as prescribed by the xxx principle. So if you choose group A, you contribute to the common pool all of your initial income, if you choose group B, you must contribute at least 50% of your initial income. If you choose group C, there is no lower limit for the contribution to the common pool. All members of the group make their contributions simultaneously and without observing each other's contribution decisions.

The final income in stage 2 is derived from contributions made to the first gained and common pool obtained.

Briefly, the final income of a participant is calculated as follows:

Participant's Initial Income
– Participant's contribution to the common pool
+ Amount received from the common pool

Amount received from the common pool is calculated as follows:

$$\frac{\text{Total Contribution to the Pool}}{\text{Number of people in the group}}$$

**Special Conditions:** If you're the only member of your group, your initial and final income will be the same and you will not have to make a contribution decision.

**Briefing at the end of the tour:** At the end of each round, you will see a detailed screen about the income distribution in your group. On the left side of the screen you will see your initial income, your contribution to the common pool, the amount you receive from the common pool and your final income. On the right side of the screen, you will see the respective averages of these for the group members other than you.

**Total earnings:** One of the 20 rounds you play with your total earnings at the end of the experiment consists of randomly selected. All tours have the same chance of being selected. Therefore, you should be able to pay maximum attention in every round.

Your earnings in this section will be based on a randomly selected one of 20 rounds. Your final income on the selected round will be your earning on this part of the experiment. All rounds have the same chance of being selected. Therefore, it is your best interest to decide every round as if that round will be the randomly selected one.

**Please Note:** During the experiment, you are not allowed to talk to each other, or to use mobile phones to distribute and attention. If you have a question, please raise your hand. We'll come to you and answer your question. All decisions will be made anonymously. No other participant will learn your individual decisions. The payment will also be made anonymously. Other participants can not learn about your earnings.

# K Questionnaire

## Demographic questions

*Age:* in years (integer valued).

*Sex:* 1=female, 0=male.

*Living:* living arrangement for the subject (0=student housing, 1=with family, 2= with friends, 3=alone).

*Siblings:* number of siblings of subject.

*Older siblings:* number of siblings who are older than the subject.

*Major:* 2=economics, 1=other business, 0=other.

*Econ:* number of economics classes (0, 1, 2, 3, 4+).

*Friends:* number of friends in the same session

*Member:* membership in a social group (student club, charity, political party etc.)

*Rely:* subjects' self assessment of the reliability of her data in the experiment. 0 ... 10

## Risk:

How willing are you to take risks in general? 0 ... 10

## Trust:

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people.

## Attitudinal questions

**Q:** How fair do you find the income distribution that resulted in the experiment?

Very unfair 0 ... 10 Very fair

**Q:** How fair do you find the income distribution in the country of your residence?

Very unfair 0 ... 10 Very fair

**Q:**How would you place your views on this scale? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between.

Incomes should be made more equal

0 ... 10

**Q:** Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between

Never justified 0 ... 10 Always justified

- Claiming government benefits which you are not entitled to
- Avoiding a fare on public transport
- Cheating on taxes if you have the chance



- Keeping money that you have found
- Failing to report damage you've done accidentally to a parked vehicle.

**Q:** Now I'd like you to tell me your views on various issues. How would you place your views on this scale? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between.

*"In the long run, hard work usually brings a better life" vs."*

*"Hard work doesn't generally bring success—it's more a matter of luck and connections"*

Hard work 0 . . . 10 Luck+Connections

**Q:** Why, in your opinion, are there people in this country who live in need? Here are two opinions: Which comes closest to your view?

Statement 1: *"They are poor because of laziness and lack of will power "*

Statement 2: *"They are poor because society treats them unfairly"*

Poor because of laziness and lack of will power 0 . . . 10 Poor because of unfair society

## References

- Acemoglu, D. and Jackson, M. O. (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association*, 15(2):245.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *The Journal of the European Economic Association*, 9(3):522–550.
- Kessler, J. B. and Leider, S. (2012). Norms and contracting. *Management Science*, 58(1):62–77.
- Kimbrough, E. and Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, forthcoming.