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

Mehmet Y. Gürdal† *Boğaziçi University*
Orhan Torul‡ *Boğaziçi University*
Alexander Vostroknutov§ *Maastricht University*

January, 2020

Abstract

We study the intrinsic 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 norms 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 institutions, which are defined by explicitly stated rules that require them to put all, half, or any amount of income in a tax pool for redistribution. The treatments differ in the level of enforcement of these rules. We find that contributions are sustained only when free riding is not possible. However, subjects with a generally strong propensity to follow norms persist in following the rules of redistribution even after experiencing many periods of losses due to free riding. We find that these subjects also perceive the same level of income inequality as fairer, when it was achieved without breaking the rules, and favor redistributive mechanisms with more stringent rules. This suggests that well-defined redistributive rules 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, H26, H41.*
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 our own.

†Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey.
Email: mehmet.gurdal@boun.edu.tr

‡Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey.
Email: orhan.torul@boun.edu.tr

§Address: Department of Economics (MPE) Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands
Email: 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 is suggested by standard economic theory. However, there are many examples of institutions that are threatened by free riding but which, 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 resource usage and with no cases of serious infringements while Alm and Torgler (2011) conclude from their review of empirical studies on tax compliance 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. That is, the enduring success of redistribution emanates from an *intrinsic desire* to adhere to social norms that guide ommunity-conscious behavior and, from a preference for the future stability of the institution over immediate gains from free riding (Proto et al., 2018). Studies in experimental economics support this view. Kimbrough and Vostroknutov (2016) demonstrate that some individuals, named rule-followers, are willing to abide by arbitrary rules at a cost to themselves. These same individuals are also more cooperative in various social dilemmas than others, called rule-breakers, who are not willing to follow costly rules. For example, groups of rule-followers can sustain contributions in a repeated Public Goods game, whereas, 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 ready to forgo any personal gain in order to stick to rules or norms: rather, when put in an institution with rule-breakers they act as conditional cooperators and stop contributing in response to others' free riding. Moreover, Kimbrough and Vostroknutov (2015) show that in a dynamic CPR game, the propensity to follow rules is a necessary, but not at all sufficient condition for the sustainability of the common resource because the low growth rate of CPR prevents groups of rule-followers from successfully managing the resource. Overall, experimental research on social norms suggests that heterogeneity in social behavior within a given context is explained by individual differences in the propensity to follow norms (Kimbrough and Vostroknutov, 2018), and, thus, by the trade-offs between monetary gains and following norms. In contrast, the difference in behavior between contexts results from variations in social norms (Krupka and Weber, 2013).

The picture presented above provides 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 the greater good and the abundance of examples of norm-following at a personal cost (Bicchieri, 2005) suggest that enforcement, at least under

some circumstances, is not necessary to sustain redistribution. It therefore remains unclear how all these ingredients—propensity to follow social norms, incentives to free ride, and degree of enforcement—interact and how this in turn determines success or failure of a redistributive institution.

The goal of this paper is to shed light on incentives that drive behavior in various redistributive mechanisms; determine how these incentives shape the outcome of redistribution for the group and group members' satisfaction with this outcome; and explain the role of enforcement in it. We are interested in the interaction of two opposing *intrinsic motivations*: the opportunistic incentive to use a redistributive institution for personal gain versus 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 a redistributive institution can be successful under the threat of free riding. That is, we wish to determine 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 there is no direct cost of violating them. Thus, we are interested in whether the desire to follow the institution's rules is overcome by free riding. Next, we compare the fairness ratings to understand how satisfaction with overall income inequality depends on redistribution rules and experience of opportunistic behavior. Finally, we investigate how varying enforcement levels affects the sustainability of the redistribution and, 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 (RF) task (Kimbrough and Vostroknutov, 2018) and the Institution Choice and Redistribution (ICR) task. In the RF task, subjects choose how much to follow an arbitrary costly rule set by the experimenter. Performance in this task allows us to quantify the individual degree of "norm compliance" or the "rule-following propensity." In the ICR task, prior to learning their randomly-generated income, subjects choose among three different redistributive institutions. These differ in terms of the announced tax rate, which can be 100%, 50%, or 0%. After choosing an institution and observing their income, subjects choose how much to contribute to the tax pool. After all contributions within an institution are collected, the tax pool is equally divided among all subjects who decided to join that institution. This task is repeated 20 times, which allows us to observe the dynamics of their institution choices and respective contributions.

The experiment has three treatments (between-subjects) that differ in the degree of enforcement of the rules of the institutions. In the Enforcement treatment, subjects have to abide by the rule of the institution that they choose to join and should contribute no less than the amount stated by the rule of the institution. Specifically, if the institution requires 100% contributions, then subjects can only contribute everything they have. In the other two institutions, subjects can contribute no less than 50% or 0% (but they are free to contribute more). This treatment is designed to test our idea that rule-following individuals prefer institutions with more redistribution, or "stricter" rules. This is not an ex ante obvious connection since a strong propensity to follow rules does not directly imply that stricter rules are preferred to weaker ones. In the No Enforcement treatment subjects are free to choose any contribution between zero and their entire initial income in all three institutions without any consequences. This is a "baseline" treatment to investigate the interaction of the two forces mentioned above: the opportunistic desire to free ride and

the motive to follow the rules of the institution. Finally, the Exclusion treatment is identical to the No Enforcement treatment except that there are random checks of the subjects' contributions. Those who do not follow the prescribed rule and get "caught" are excluded from future participation in that institution. This treatment is designed to check the effect of enforcement of rules on the balance of opportunistic and normative incentives. We chose exclusion, instead of fining subjects for violating the rules, to keep the behavior as close to the No Enforcement treatment as possible. This is because subjects may perceive fines as simply decreasing the expected profit from free riding, while not creating the feeling that rules are enforced.

We find, unsurprisingly, that in the No Enforcement and Exclusion treatments, where subjects can contribute any amount of their income, 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 despite the personal cost due to free riding. Moreover, we find that subjects classified as "rule-followers" by the Rule-Following task, contribute significantly more than "rule-breakers", even in the later periods where average contributions fall close to zero. Remarkably, rule-followers contribute more than rule-breakers even in the institution that does not require any contribution. This demonstrates that institutional rules are perceived by rule-followers, not as some arbitrary instructions that must be obeyed, but rather as guidelines on the appropriate contributions that the prevailing social norm prescribes.

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 the differential fairness judgements 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, the perceived fairness of an income distribution depends on whether others follow the rules than 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 they should contribute all their income to the tax pool whereas, rule-breakers prefer the "anarchic" institution that allows any contribution. In the No Enforcement treatment, the preference of rule-followers and rule-breakers reverses 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 rule-followers not only persist in following the rules in the egalitarian institution (or to a degree in case of free riders), but also 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 model proposed by Kessler and Leider (2012), which only accounts for the disutility of deviation from the norm.

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 experience 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 relationship between norm compliance and pro-social behavior in various settings. Here, studies that involve economic experiments commonly find that the 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 various versions of the Dictator Game. 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 institutions 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 principles correspond to the conditions in our experiment (see below), whereas third is a modified form of egalitarianism where subjects are held accountable for the choices they make, while excluding luck.[1] This is a principle constructed along the same lines as the accountability principle of Konow (2000). In a recent study, Starmans et al. (2017) show that people are mainly disturbed by the unfair processes that generate unequal incomes rather than the final inequality itself. Klor and Shayo (2010a) experimentally show that the knowledge about the mean income of a group influences each subject's preferences for 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 to make a novel connection between fairness perceptions and income inequality.

In some experimental studies, subjects are allowed to choose between different institutions during the experiment, as in our current paper. In Gürerk et al. (2006), subjects choose to play a public goods game in

[1] There are several papers in the literature that study the role of luck vs. effort/performance in shaping redistributive preferences (Cappelen et al., 2010; Krawczyk, 2010; Lefgren et al., 2016; Rey-Biel et al., 2018). For the sake of clarity and parsimony, in our experimental design income is determined purely by luck. While we acknowledge that relating income to effort or performance could yield novel results, we leave this interesting exercise for future research.

an environment where they can sanction other group members or in an environment where this option is not available. After a learning period, subjects overwhelmingly choose a sanctioning institution. Sutter et al. (2010) study a similar environment where subjects unanimously decide to introduce punishments, rewards, or neither. They report higher contributions when the institutions are chosen endogenously rather than 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. Although subjects earn more in the Harmony Game, a substantial fraction votes for the 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 organization's members are expected to contribute all their endowment and impose punishment on those who do not follow the rule. Subjects commonly choose to form these institutions and they induce higher contributions. Putterman et al. (2011), study a public goods game where subjects choose the 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 models of tax compliance that use an augmented utility function 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 well-defined rules of conduct.

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 I for instructions).[2] These two main tasks were followed by a questionnaire with items related to demographics and attitudes towards the fairness of hypothetical income distributions (see Appendix J). There are three treatments (between-subjects) that differ in the degree of enforcement in the Institution Choice and Redistribution task.

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) while 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.

[2]The Rule-Following task was always preceding the Institution Choice and Redistribution task. This order was chosen both because the measure of rule-following propensity is central to our experiment and to ensure the consistency with the main treatments of Kimbrough and Vostroknutov (2016), who also tested order effects of the Rule-Following task and several other tasks and found none. Some studies use the same Rule-Following task *after* the main task (e.g., Panizza et al., 2019; Thomsson and Vostroknutov, 2017) and obtain results consistent with the general hypothesis that prosocial behavior is explained by rule-following propensity, which shows that the opposite order does not change the general nature of the findings.

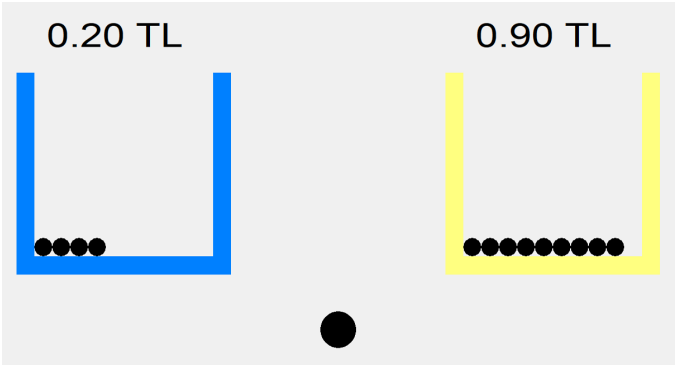

Figure 1: The Rule-Following task. The blue bucket is on the left and the yellow bucket is on the right.

The instructions explicitly state that *"the rule is to put the balls into the blue bucket"* (see Appendix I). Subjects have 100 balls to allocate, so 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.[3]

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. There are three institutions to choose from with different redistribution rules. The rules are described by the following normative statements:

- Institution 100: We should all our income (100%) to the tax pool;
- Institution 50: We should all put half of our income (50%) to the tax pool;
- Institution 0: We can contribute any amount to the tax pool at our own discretion.[4]

After making their 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 tax pool. For each institution, the contributions to the tax pool are summed and divided equally among all the subjects who have chosen to join that institution.[5] We decided to give subjects a random income instead of a fixed known one to make sure that the choice of institution is not influenced by their wealth. Another reason was to create a perception among subjects that Institutions 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 institution. In particular, they observed their initial income and the average initial income of other institution members; their contribution to the tax pool and the average contribution of other institution members; the amount

[3] 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.

[4] In the experiment, the three institutions were called A, B, and C.

[5] In case only one subject has chosen an institution, the contribution choice was not provided.

they and each other member of the institution received from the tax pool (everyone gets the same amount); as well as their final income and the average final income of other institution members.

As outlined earlier, the experiment has three treatments that 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. That is, subjects who chose Institution 100 must contribute all their initial income to the tax pool; subjects in Institution 50 must contribute at least 50% of their initial income (any amount at or above 50%); and subjects in Institution 0 can choose any amount between zero and their initial income as their contribution to the tax pool. In the No Enforcement treatment subjects are free to choose any contribution between zero and their entire initial income in all three institutions. 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 external checks, with a probability of 20%. In the case of Institutions 100 and 50, a subject is not allowed to join either of these institutions ever again if the external check reveals that the contribution was lower than 80% of the respective norm (100% or 50% requirement). By introducing this cost to free riding, we intended to test if compliance with the rules of the institutions can be sustained.[6]

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 tax-pool redistribution (for each institution in each period, the sum of all contributions in the institution's tax pool divided by the number of subjects in the institution). 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. The experiment was programmed in z-Tree (Fischbacher, 2007).

4 Results

4.1 Contribution Dynamics

We first present 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 income is random in each period in our setup. Thus, looking at absolute contributions can be misleading.

Figure 2 shows the dynamics of the contribution ratios (Contribution) in the No Enforcement, Exclusion, and Enforcement treatments. In the Enforcement treatment, the picture is clear: subjects cannot contribute

[6] 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 for an external check was set at 20% so that it is small enough to not disincentivize all cheating, while large enough so that in 20 periods all cheaters are excluded: For example, with 20% check probability, the probability of undetected cheating for 20 periods is 1%, whereas it is 12% (too high) with 10% check probability. Finally, we set the threshold of cheating detection at 80% of the required tax rate to distinguish between subjects who genuinely follow the norm from those who want to comply just because of the threat of punishment.

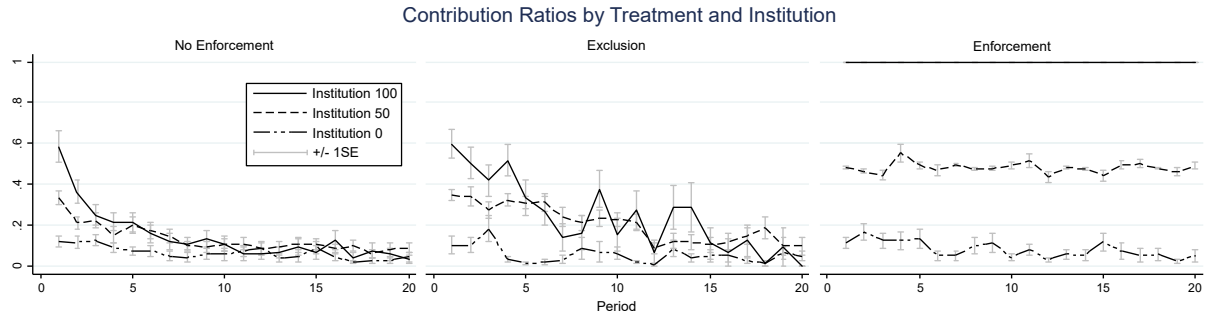

Figure 2: The dynamics of contribution ratios in the three treatments divided by institution. The spikes are ±1SE.

less than a fixed percentage of their income.[7] In the No Enforcement treatment the contributions decay fast in all three institutions, 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 note that the average contributions in the first four periods are ranked according to the percentage of income that the institution's rule requires subjects to share. This demonstrates that some subjects initially do try to follow the rule. However, they eventually decrease their contributions as a conditional response because the institution becomes overwhelmed by free riders. In Appendix A we support this conclusion with an additional analysis, which shows that around half of subjects start with abiding by the announced rules at the beginning of the task.

In the Exclusion treatment, average contributions also fall to zero, as in the No Enforcement treatment. However, this happens much more slowly: while in the No Enforcement treatment the average contributions fall below 20% after period 5, in the Exclusion treatment, the averages remain above 20% for Institutions 100 and 50 until around period 12. This demonstrates that exclusion does affect average contributions.

Result 1. *In the first 4 periods of the No Enforcement treatment and the first 12 periods of the Exclusion treatment contributions are ranked according to the announced tax rate of the institutions: around half of subjects start with following the rule 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 investigate whether subjects' contributions depend on their rule-following propensity.[8] Table 1 shows the random-effects regressions of Contribution on Rule-Following, defined as the number of balls in the blue bucket normalized to $[0, 1]$, dummies for Institution 50 and Institution 0, randomly determined Pre-Tax Income, and Period (detailed descriptions all variables used in the regressions can be found in Appendix C).

For both the No Enforcement and Exclusion treatments we run separate regressions for the first 10 periods (early periods) and the last 10 periods (late periods), as well as all periods. The coefficient on Rule-Following is significant for all-periods regressions in both treatments. This means that in Institution 100

[7] The averages sometimes fall below the required contribution of 50% in Institution 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%).

[8] The descriptive statistics for the behavior in the rule-following task are provided in Appendix B.

Treatment:	No Enforcement			Exclusion		
Periods:	Early	Late	All	Early	Late	All
Rule-Following	0.195***	0.095*	0.151***	0.173*	0.279***	0.179**
	(0.069)	(0.051)	(0.055)	(0.090)	(0.082)	(0.077)
Institution 50	–0.022	0.023	0.003	–0.085	0.054	–0.066*
	(0.034)	(0.021)	(0.023)	(0.053)	(0.034)	(0.037)
Institution 0	–0.112***	–0.018	–0.064**	–0.206***	0.006	–0.127***
	(0.043)	(0.033)	(0.032)	(0.058)	(0.037)	(0.036)
Rule-Following × Institution 50	–0.056	–0.032	–0.054	–0.025	–0.149**	–0.032
	(0.061)	(0.043)	(0.046)	(0.082)	(0.070)	(0.064)
Rule-Following × Institution 0	–0.090	–0.066	–0.081	–0.151	–0.231**	–0.135*
	(0.072)	(0.057)	(0.057)	(0.093)	(0.091)	(0.080)
Pre-Tax Income	–0.169***	–0.129***	–0.146***	–0.250***	–0.179***	–0.216***
	(0.025)	(0.024)	(0.019)	(0.037)	(0.031)	(0.025)
Period	–0.024***	–0.003**	–0.009***	–0.020***	–0.003	–0.012***
	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)
Constant	0.342***	0.155***	0.256***	0.518***	0.153**	0.419***
	(0.047)	(0.035)	(0.033)	(0.059)	(0.060)	(0.048)
N observations	1,566	1,563	3,129	713	713	1,426
N subjects	158	158	158	72	72	72

Table 1: Random-effects regressions of Contribution. Standard errors in parentheses. Errors are clustered by subject and robust. * – $p < 0.1$; ** – $p < 0.05$; *** – $p < 0.01$.

(baseline) subjects with a high propensity to follow rules contribute more than those with a low propensity although free riding is ubiquitous. What is even more remarkable is that even in the late periods of the No Enforcement treatment, the coefficient for Rule-Following is still significant (albeit at 10% level), which means that subjects a with high propensity to follow rules try to contribute more in Institution 100 even when average contributions are essentially zero. The same is true for Institution 50 in the late periods of the No Enforcement treatment: the sum of coefficients on Rule-Following and Rule-Following×Institution 50 is .064*. Importantly, the coefficients for rule-following in Institution 50 are significant at $p < 0.005$ in all other regressions even though the interactions Rule-Following×Institution 50 are all negative. This suggests that rule-followers contribute less in Institution 50 than in Institution 100, but still more than rule-breakers. Rule-followers contribute more than rule-breakers even in Institution 0 of the No Enforcement treatment. The sum of coefficients on Rule-Following and Rule-Following×Institution 0 is equal to 0.07*** for all periods (0.10*** in the early periods and 0.03 in the late periods). This demonstrates that rule-followers contribute more than rule-breakers in the No Enforcement treatment even when they join Institution 0 where the rule explicitly says that any level of contribution is acceptable.[9] This implies that rule-followers are intrinsically motivated to be prosocial even when they do not have to be, which is in line with the results of Kimbrough and Vostroknutov (2016) who connect rule-following to cooperation in various social dilemmas where no explicit rules are stated. Taken together, these results provide strong evidence that rule-followers consider explicitly stated rules of Institutions 50 and 100 not simply as mechanical instructions that they must

[9]In the Exclusion, treatment the sum of coefficients Rule-Following and Rule-Following×Institution 0 is not significant in all three regressions. This is most likely due to the fact that in this treatment there is a higher proportion of rule-breakers in Institution 0 than in the same institution of the No Enforcement treatment: rule-breakers are gradually getting excluded from the other two institutions in the Exclusion treatment.

follow for no specific reason, but as *clarifications of the extent* to which they should abide by the social norm that they perceive is already in place in the redistribution task.

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. In the No Enforcement treatment, rue-followers contribute more even in Institution 0, where the rule states that any amount can be contributed. This shows that the announced rules are perceived as norms.*

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 they are being constantly exploited by free riders, it is interesting to investigate their perceptions of the fairness of redistribution. We now 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 8 in Appendix D shows the distributions of answers in the three treatments). Rank-sum tests show a significant difference in the distributions between the No Enforcement and Enforcement treatments ($p < 0.0001$), and in the Exclusion and Enforcement treatments ($p = 0.0010$).

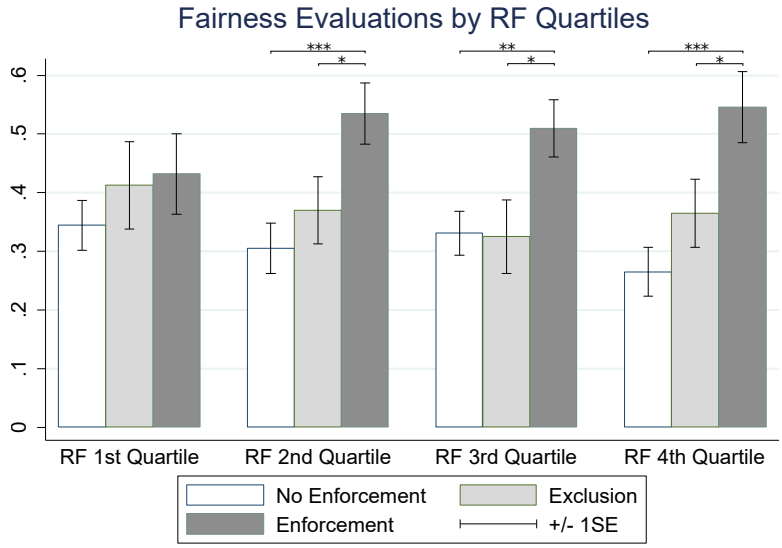

Figure 3: Average fairness estimates for the four quartiles of rule-following propensity by treatment. The significance levels reported correspond to the rank-sum tests, Benjamini-Hochberg corrected for 12 comparisons: * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

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 3 shows the average fairness ratings for the four quartiles of rule-following propensity. For the rule-breakers in the first 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 between treatments.[10] That is, rule-followers, but not rule-breakers, find the income distribution in the No Enforcement and Exclusion treatments less fair than in the Enforcement treatment. This is because, 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, whereas they do not in the Enforcement treatment. We provide a more detailed regression analysis that supports this conclusion in Appendix E.

These findings are related to a recent discussion on fairness perception of income inequality. Deaton (2017) claims that inequality is mapped onto unfairness by its origin rather than its level: that is , only when inequality stems from deeply and obviously unjust foundations, is it regarded as 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.[11] In Table 2 we illustrate the distributional properties of pre-tax and post-tax incomes. We find that in the No Enforcement treatment, post-tax income inequality surpasses pre-tax inequality. In addition, the supposedly egalitarian Institution 100 generates the most unequal post-tax income distribution of all three institutions. In other words, *de facto* redistribution across institutions not only blurs but also clashes with the announced rules, particularly in the case of Institution 100.[12] Rule-followers in Institution 100, who try hard to play by the rules and conform with the rules by contributing sizeably, realize that they are the constant losers due to the free riding of those who contravene the announced rule.[13] As a result of their legitimate resentment, they evaluate as unfair the unequal distribution resulting from others bending the announced norm.[14]

In order to unveil subjects' mapping of economic inequality onto fairness, Table 2 reports the fairness scores of subjects who spent more than half of the experiment in one of the three institutions. Our results are mostly in accordance with claims by Deaton (2017) and Starmans et al. (2017): comparing between-institution variations by treatment, we find 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, the enforcement rule is decisive beyond inequality: while Institution 0 in the Enforcement treatment generates moderately more 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.

[10] 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.

[11] See also Rustichini and Vostroknutov (2014) for an experimental investigation.

[12] A careful reader could notice that the post-tax income Gini coefficient for Institution 100 in the Enforcement treatment in Table 2 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 using the post-tax incomes of subjects in Institution 100 over *all* periods. As such, depending on the period-specific income draws of subjects in Institution 100, the average amount collected in its tax pool varies over periods, especially during periods when the number of Institution 100 subjects is limited. As a result, post-tax incomes of Institution 100 subjects in the Enforcement treatment are not fully equalized, and the resulting Gini coefficient is small but different from zero.

[13] We discuss this in more detail in Section 4.4. Figure 5 provides support for our point: it shows gains and losses due to contribution and redistribution by rule-following propensity.

[14] For the literature investigating individual characteristics affecting fairness evaluations and redistributive preferences see Esarey et al. (2012), Agranov and Palfrey (2015), Klor and Shayo (2010b), Ku and Salmon (2013), Großer and Reuben (2013), and Höchtl et al. (2012).

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

Table 2: Inequality and fairness perception. Fairness perception of the institutions refers to the average fairness perception scores of subjects who spent more than half of 20 periods in that institution.

To summarize, our results on the perceived fairness of redistribution demonstrate the importance of norms for perceptions of inequality and attitudes towards the redistribution mechanism. In the 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 the opposite result to what might have been intended. This is a tangible consequence of people's the intrinsic rule-following propensity. Our experiment thus shows that established norms of redistribution can strongly influence actual and perceive inequality 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 norm violations.*

4.4 Choice of an Institution

Our results regarding the perceived fairness of redistribution suggest that rule-following subjects may gravitate in their choice of the institution towards the one that they perceive as fairest under the given enforcement rule. To test this idea, we analyze how subjects choose an institution (or redistribution rules), depending on their characteristics and treatment. As we hypothesized in the introduction, we expect to find a connection between rule-following propensity and preference regarding redistribution rules. Specifically, in our study, we expect rule-followers to prefer a more "structured" institution, represented by the institution where subjects should contribute 100% of their income (Institution 100). Conversely, rule-breakers should prefer the least-structured institution with a 0% contribution requirement (Institution 0). However, if subjects also care about the fairness of redistribution – that is, how often the redistribution rules are broken – then their choice may be influenced by the level of enforcement.

Figure 4 shows average rule-following propensities by treatment and institution. In the No Enforcement treatment rule-followers are congregated in Institution 0, where there are no rules of redistribution, whereas

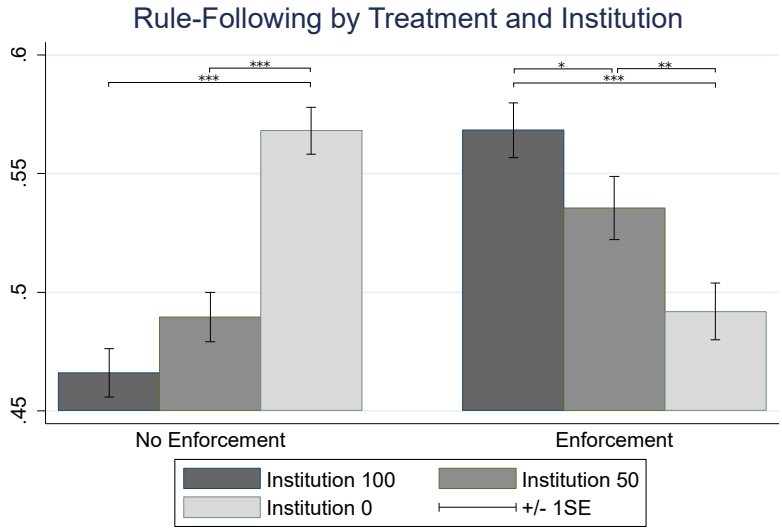

Figure 4: Average rule-following propensity by treatment and institution.

rule-breakers go to Institution 100, where the rules are to contribute 100% of income. In the Enforcement treatment, the picture is the opposite: rule-followers prefer Institution 100 and rule-breakers Institution 0. Permutation tests on differences of means show significant differences between Institution 0 and institutions 100 and 50 in both treatments.[15] The same observation can be made if we look at the average rule-following propensities in Institutions 100 and 0 period by period. Figure 9 in Appendix D shows that the average rule-following propensity in the No Enforcement treatment is always higher in Institution 0 whereas it is always higher in Institution 100 in the Enforcement treatment . In Appendix F we discuss several regression specifications that support these findings.

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 4. Specifically, we show that the risk and social preferences elicited in the questionnaire do not correlate with the choice of institution in the Enforcement treatment.[16] Thus, we propose an alternative explanation. It is easier to start with the Enforcement treatment, which gives subjects only a pure choice between institutions uncontaminated by their contribution choice. We conjecture that the observed pattern of institution choices shown in the right-hand panel of Figure 4 can only be explained by assuming a separate *preference over redistribution rules*. In particular, rule-followers prefer institutions that have demanding 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 Institution 100 than Institution 0 in the Enforcement treatment.

[15]For the No Enforcement treatment: $p = 0.0003$ for comparison between institutions 0 and 100; $p = 0.0005$ between institutions 0 and 50. In the Enforcement treatment: $p = 0.0002$ for comparison between institutions 0 and 100; $p = 0.0258$ between institutions 0 and 50, and $p = 0.0754$ between institutions 100 and 50. All p-values are Benjamini-Hochberg corrected for 6 comparisons. We chose to use permutation tests since most subjects appear in each institution multiple times, so observations are not independent. 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 invalid due to the non-independence of observations.

[16]The evidence on social preferences that we present in Appendix G is based on a question in the post-experimental questionnaire. Given the experiment's time limitations, we could not afford to elicit more individual characteristics related to social preferences. Therefore, it cannot be ruled out that some other forms of redistributive concerns, which we did not think about, play a role in the choice of institution.

In the No Enforcement treatment subjects choose both their institution and contribution amount. We can see that the pattern of rule-following by institution reverses. In light the postulated preference over rules, this means that rule-followers, as much as they like being in Institution 100, nevertheless choose to switch to Institution 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 Institution 0 to Institution 100. Another is fairness considerations: everything else being equal rule-followers prefer the institution where fewer people break the norm, whichever norm that is. The contributions are roughly the same and close to zero in all three institutions in the No Enforcement treatment. Thus, rule-followers prefer Institution 0, which is "fairer" than others in the sense that at least no one is breaking the norm by contributing very little.

To see that rule-followers have a tangible incentive to avoid Institution 100 due to free riding we analyze the differences between the pre-tax income, the amount that subjects receive at the beginning of each period, and the post-tax income, equal to the pre-tax income minus tax contribution plus the return from the tax pool redistribution. Figure 5 shows the difference between post-tax and pre-tax income for the four quartiles of rule-following propensity in Institution 100 of the No Enforcement treatment. In the early periods (from 1 to 10), rule-breakers in the 1st quartile earn more income than they receive at the beginning of the period, while rule-followers in the 4th quartile lose money, which may force them to leave Institution 100 and join Institution 0.[17] In the late periods (from 11 to 20), when subjects decrease their contributions (see Figure 2), the difference expectedly becomes zero for rule-breakers and slightly negative for the subjects in the fourth quartile of rule-following.

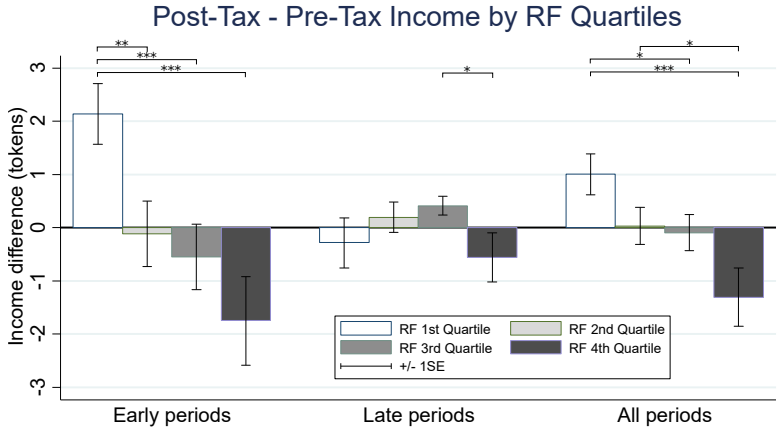

Figure 5: Differences between post-tax and pre-tax incomes in Institution 100 of the No Enforcement treatment in early and late periods. Permutation tests: significance levels Benjamini-Hochberg corrected for 18 comparisons: * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

To support our argument that rule-followers leave Institution 100 in the No Enforcement treatment due to perceived unfairness, we notice that Institutions 100 and 50 are close in terms of average rule-following propensity but different from Institution 0 (Figure 4). Since contributions are very low in all institutions in late periods, this observation is consistent with the idea that rule-followers choose Institution 0 because no one breaks its norms, which makes it different from Institutions 100 and 50.

[17] We use permutation tests to compare the means of income differences on Figure 5. 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.

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 4. Specifically, we show that the self-selection of rule-followers and rule-breakers into different institutions in the No Enforcement and Enforcement treatments can only come about if rule-followers receive utility from being in the institution where others abide by the norm (or disutility from being in the institution 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 a 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 being 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 direct evidence to support this hypothesis. We find that rule-followers, defined as people who follow an arbitrarily imposed rule at personal cost, contribute more income to the tax pool than rule-breakers, defined as people who are less inclined to trade-off money for following an arbitrary rule. This is true even after rule-followers have experienced free riding by rule-breakers (see Table 1). Thus, we confirm that pro-sociality is indeed learned by docile individuals (in our terminology rule-followers) as a 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). However, one of the goals of our design was to go further by exploring 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 that is governed by strict rules to the one with no rules at all. This idea is in line with Richerson et al. (2016) who hypothesize 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 4). This finding may have important implications for 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 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, the erosion is slower in the latter case). This happens regardless of the specified redistribution "norm." Thus, given the choice to join one of the institutions—the egalitarian one with 100% contributions, the semi-egalitarian with 50%, or the libertarian one with 0%—many subjects, who would otherwise stay in the libertarian institution, 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 institution. This leads to an overall decrease of the tax contributions to almost zero in all institutions reaching an equilibrium in which rule-breakers end up in the institution with 100% contributions and rule-followers in the institution with zero required contributions. Conversely, when the rules of the institutions are enforced, the preferences for the redistribution rules discussed above are expressed: rule-followers join the most-regulated institution whereas 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 their conditional response to free riding, which leads to a decline in contributions and a failure of the redistribution mechanism. Nevertheless, our results suggest that full enforcement, which in our design prevents free riding, may not be necessary as long as the community consists of norm-abiding individuals who are aware that other members are alike. Under these conditions, the preference for a regulated society can solve the social dilemma and create an environment where people pay taxes voluntarily.

6 Conclusion

We conduct an experiment 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 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.

Agranov, M. and Palfrey, T. R. (2015). Equilibrium tax rates and income redistribution: A laboratory study. *Journal of Public Economics*, 130:45–58.

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.

Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Responsibility for What? Fairness and Individual Responsibility. *European Economic Review*, 54(3):429–441.

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.

Esarey, J., Salmon, T. C., and Barrilleaux, C. (2012). What motivates political preferences? self-interest, ideology, and fairness in a laboratory democracy. *Economic Inquiry*, 50(3):604–624.

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.

Großer, J. and Reuben, E. (2013). Redistribution and market efficiency: An experimental study. *Journal of Public Economics*, 101:39–52.

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.

Höchtl, W., Sausgruber, R., and Tyran, J.-R. (2012). Inequality aversion and voting on redistribution. *European economic review*, 56(7):1406–1421.

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. (2010a). Social identity and preferences over redistribution. *Journal of Public Economics*, 94(3-4):269–278.

Klor, E. F. and Shayo, M. (2010b). 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.

Krawczyk, M. (2010). A Glimpse Through the Veil of Ignorance: Equality of Opportunity and Support for Redistribution. *Journal of Public Economics*, 94(1-2):131–141.

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.

Ku, H. and Salmon, T. C. (2013). Procedural fairness and the tolerance for income inequality. *European Economic Review*, 64:111–128.

Lefgren, L. J., Sims, D. P., and Stoddard, O. B. (2016). Effort, Luck, and Voting for Redistribution. *Journal of Public Economics*, 143(C):89–97.

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.

Panizza, F., Vostroknutov, A., and Coricelli, G. (2019). Meta-context and choice-set effects in mini-Dictator games. mimeo, University of Trento and University of Southern California.

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.

Rey-Biel, P., Sheremeta, R., and Uler, N. (2018). When income depends on performance and luck: The effects of culture and information on giving. In *Experimental Economics and Culture*, pages 167–203. Emerald Publishing Limited.

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.

Thomsson, K. M. and Vostroknutov, A. (2017). Small-world conservatives and rigid liberals: Attitudes towards sharing in self-proclaimed left and right. *Journal of Economic Behavior & Organization*, 135:181–192.

Appendix

A 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 institution, which subjects join voluntarily. In particular, in Institution 100, subjects know that in this institution *"We should all put our whole income (100%) into the common pool"* while in Institution 50 *"We should all put half of our income (50%) into the common pool."* Thus, the norm of each institution 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 institution, in standard experiments subjects might be unsure about what the appropriate level of contributions is and "discover" it as they choose repeatedly.

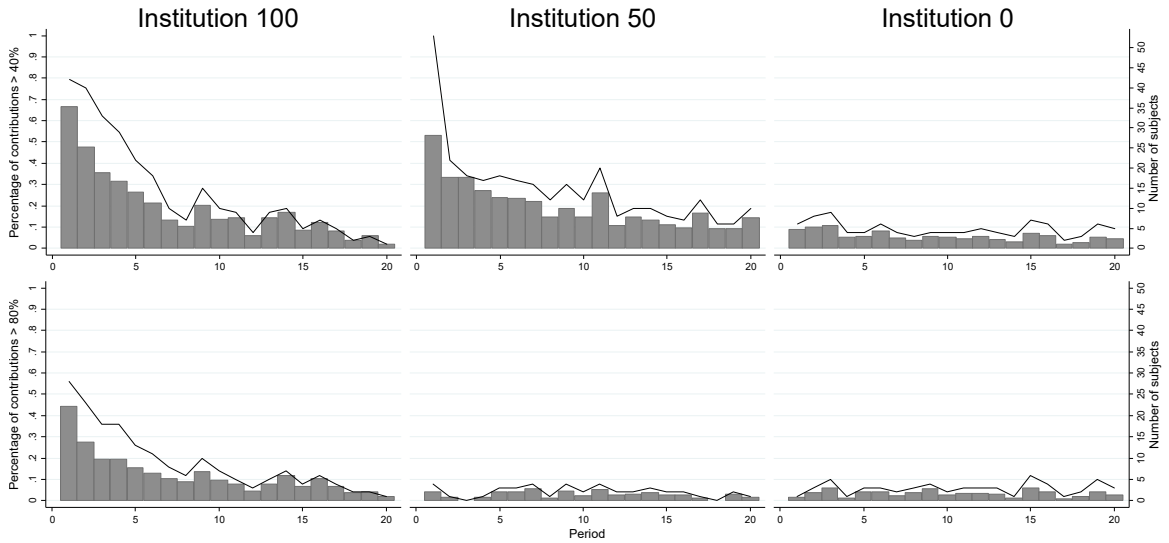

Figure 6: The percentages of contributions above 40% and 80% of income by institution in the No Enforcement and Exclusion treatments (gray bars, left y-axis). Number of subjects who contributed above 40% and 80% of income (black lines, right y-axis).

We introduced institution rules to study how opportunism and the desire to follow rules interact and influence redistribution choices. While free riding and conditional cooperation eventually decrease contributions in the No Enforcement and Exclusion treatments, the pre-specified rules still have a sizeable effect on contribution choices. Figure 6 shows the percentages and numbers of subjects in institutions 100, 50, and 0 who contributed above 40% or 80% of their income (in the No Enforcement and Exclusion treatments together).[1] The differences are very noticeable: in Institution 0, virtually no one makes contributions above 40%; in Institution 50, many subjects make contributions above 40%, but almost no one above 80%; and in Institution 100, a significant number of subjects contribute above 80%.

The rules of the three institutions are simple announcements that are non-binding and bear no violation costs.[2] Nevertheless, they have a strong effect on contributions in the early periods. This demonstrates that rules, as artificial as they are in our experiment, are capable of sustaining redistribution.

[1] The numbers 40% and 80% were chosen since these are the thresholds below which subjects can be excluded in institutions 50 and 100 in the Exclusion treatment. So, choosing contributing above 40% in Institution 50 and above 80% in Institution 100 has no consequences.

[2] 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.

B Analysis of Rule-Following Propensities

Figure 7 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 many subjects in between. In our data, 15% of subjects were 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.

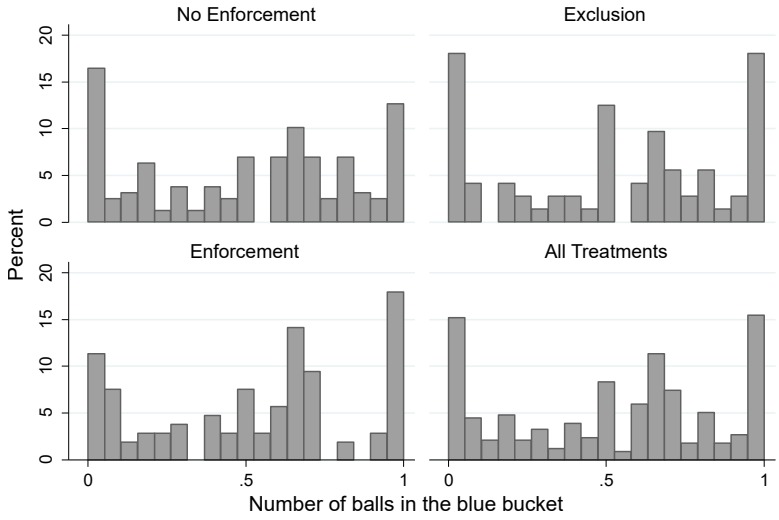

Figure 7: Histograms of rule-following in three treatments.

Table 3 below 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 up to 20%; 2) membership of an organization increases the rule-following propensity by 8%; 3) risk preferences, measured by the Likert-type scale question, *How willing are you to take risks in general?*, change the rule-following propensity by up to 20%. That is risk averse subjects seem to be more rule-following, though the significance level is only at 10%.[3] It should also be noted that none of the questions related to social preferences (see Appendix J) have any significant effect on rule-following.

[3] Dohmen et al. (2011), who used the same question to elicit risk preferences, showed that it is a reliable measurement.

Rule-Following Choice

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

C Variables

Variable	Range	Definition
Contribution	$[0, 1]$	contribution ratio to the tax pool (1 if 100% contribution)
Rule-Following	$[0, 1]$	normalized number of balls in the blue bucket (1 if all balls are in the blue bucket)
Institution X	$0/1$	dummy for joining Institution $X \in \{0, 50, 100\}$
Pre-Tax Income	$[0, 50]$	income before contribution and redistribution
Post-Tax Income	$[0, 50]$	income after contribution and redistribution
Average Income	$[0, 50]$	equals (Pre-Tax Income + Post-Tax Income) / 2
Income Change	$[-42, 33]$	equals Post-Tax Income – Pre-Tax Income
Period	$[0, 20]$	period in which the decision was made
Frequency in Institution X	$[0, 1]$	normalized number of periods spent in Institution $X \in \{0, 50, 100\}$; equals 1 if Institution X was joined in all 20 periods
Switch	$0/1$	equals 1 if a subject switched institution in the current period
No Enforcement	$0/1$	dummy for the No Enforcement treatment
Enforcement	$0/1$	dummy for the Enforcement treatment
Exclusion	$0/1$	dummy for the Exclusion treatment
Demographics and Controls		
Age	$[18, 29]$	age
Gender	$0/1$	gender (1 if male)
Number of Siblings	$[0, 7]$	number of siblings
Number of Younger Siblings	$[0, 7]$	number of younger siblings
Number of Older Siblings	$[0, 8]$	number of older siblings
Economics or Business Major	$0/1$	university major (1 if economics or business)
Number of Economics Courses	$[0, 4]$	number of economics courses taken
Organization Membership	$0/1$	membership status of the subject (1 if subject is a member of some organization)
Risk Appetite	$[0, 1]$	normalized self-reported willingness to take risk (1 if risk-seeker)
Justification Tendency	$[0, 1]$	normalized belief that cheating is justifiable (1 if justifiable)
Belief that the Poor are in Need due to Unfair Society	$[0, 1]$	normalized belief about why the poor are in need (1 if due to unfair society)
Belief that Incomes Should be More Equal	$[0, 1]$	normalized preference for equality of incomes (1 if should be equal)
Belief that Hard Work Pays Off	$[0, 1]$	normalized belief on the role of luck versus effort (1 if hard work pays off)
Fairness Perception of the Income Distribution in the Country	$[0, 1]$	normalized fairness evaluation (1 if fair)
Fairness Perception of the Income Distribution in the Experiment	$[0, 1]$	normalized fairness evaluation (1 if fair)

Table 4: Variables used in the analyses and regressions.

D Additional Graphs

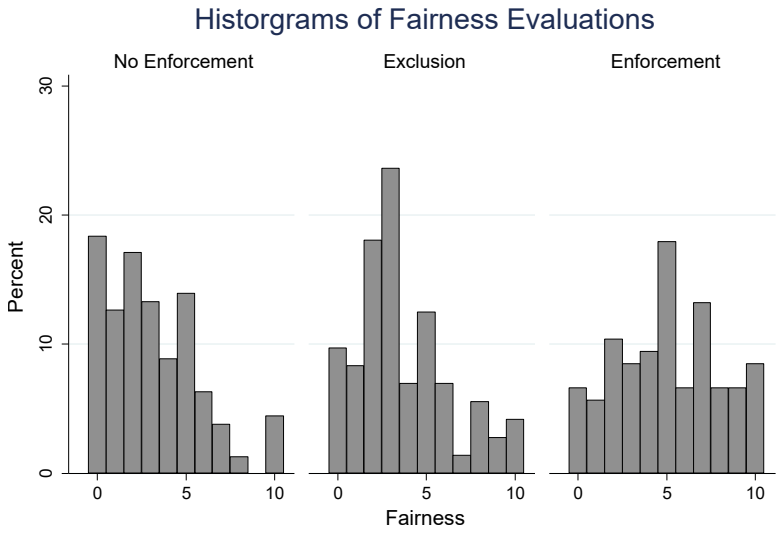

Figure 8: The distributions of the answers to the question *"How fair do you find the income distribution that resulted in the experiment?"* on a 0 to 10 Likert-scale.

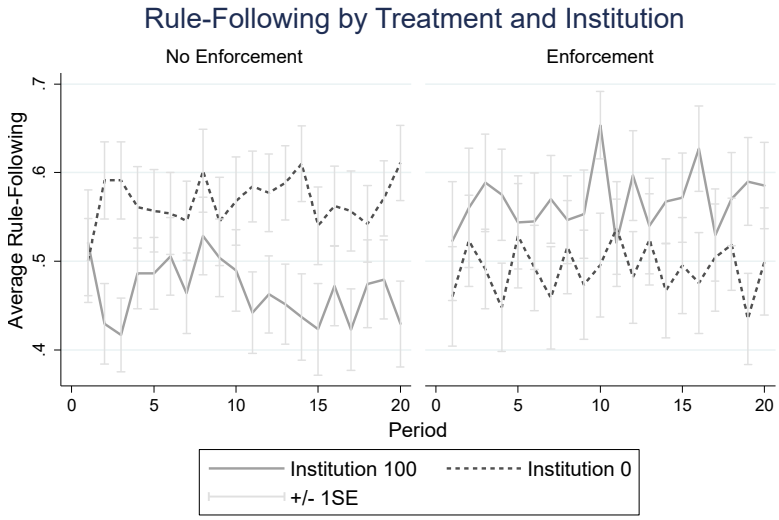

Figure 9: Rule-following measured as a percentage of balls in blue bucket by treatment and institution.

E Additional Analysis of Fairness Perceptions

To further support the findings reported in Figure 3, we report the results of an OLS regression of fairness perception on average income and the frequency of being in Institution 100, which is defined as the number of periods that a subject spent in Institution 100 normalized to $[0, 1]$. The results are presented in Table 5. The regression of the Enforcement treatment data shows no significant effect of either variable on fairness perception (column 2). This is not surprising since subjects in this treatment join freely any institution that they like, thereby committing themselves to the rules of that institution. Income significantly influences fairness perception only among rule-followers—subjects with an above-median rule-following propensity—in the No Enforcement treatment (column 5). That is, a higher income makes them feel that the distribution is fairer whereas the frequency of being in Institution 100 decreases the perceived fairness of income distribution. This happens because rule-followers, who try to contribute high amounts according to the rules of Institution 100, think that free riders are breaking the rule. This creates the feeling of an unfair distribution of income. The fairness evaluations of rule-breakers—subjects with a below median-rule-following propensity—do not follow this pattern (column 4). Overall, however, when we look at all subjects (column 3), the influence of Frequency in Institution 100 is significant.[4]

	Fairness Perception				
	(1)	(2)	(3)	(4)	(5)
	All	Enforcement	No Enforcement	No Enf. Rule-Breakers	No Enf. Rule-Followers
Average Income	0.153	-0.262	0.406*	0.061	0.612***
	(0.206)	(0.270)	(0.215)	(0.478)	(0.232)
Frequency in Institution 100	-0.032	0.144	-0.230**	-0.150	-0.287**
	(0.086)	(0.118)	(0.095)	(0.147)	(0.125)
Constant	0.247	0.724***	-0.021	0.308	-0.214
	(0.199)	(0.264)	(0.206)	(0.472)	(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 5: 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 C.

Finally, it should be emphasized that only gender affects the perception of fairness of rule-breakers (Table 6 below). This suggests an intimate connection between fairness judgments and rule-following propensity. That is, rule-followers do associate rule violations with unfair income distribution whereas rule-breakers do not. It remains an open question whether a lack or inconsistency of moral judgment makes rule-breakers unresponsive to rules or vice versa. We leave this question for future investigation.

[4] We do not analyze the Exclusion treatment here because the variable Frequency in Institution 100 is not well-defined for this treatment, given that subjects can be, and many are, excluded from participation in Institution 100. In other words, some subjects, especially rule-breakers, cannot choose freely to join Institution 100 once they have been excluded, which creates a bias in the relationship between this variable and rule-following propensity.

	Fairness Perception				
	(1)	(2)	(3)	(4)	(5)
	All	Enforcement	No Enforcement	No Enf. Rule-Breakers	No Enf. Rule-Followers
Average Income	0.201	-0.161	0.376	0.083	0.560***
	(0.209)	(0.323)	(0.228)	(0.502)	(0.210)
Frequency in Institution 100	-0.059	0.164	-0.230**	-0.153	-0.204
	(0.085)	(0.131)	(0.096)	(0.142)	(0.133)
Age	0.026**	0.018	0.023	0.035	0.019
	(0.012)	(0.020)	(0.014)	(0.025)	(0.021)
Gender	0.080**	0.042	0.039	0.153**	-0.047
	(0.037)	(0.064)	(0.046)	(0.074)	(0.057)
Number of Younger Siblings	-0.014	-0.041	-0.012	-0.031	0.025
	(0.022)	(0.032)	(0.026)	(0.032)	(0.039)
Number of Older Siblings	-0.021	-0.045	-0.000	-0.008	0.003
	(0.020)	(0.048)	(0.025)	(0.042)	(0.033)
Economics or Business Major	0.067	0.097	0.023	0.139*	-0.103
	(0.051)	(0.077)	(0.063)	(0.079)	(0.095)
Number of Economics Courses	-0.020	-0.022	0.005	-0.051	0.052
	(0.018)	(0.030)	(0.022)	(0.032)	(0.034)
Organization Membership	0.022	0.055	-0.021	-0.001	-0.047
	(0.035)	(0.059)	(0.040)	(0.065)	(0.055)
Constant	-0.376	0.225	-0.494	-0.482	-0.614
	(0.355)	(0.628)	(0.394)	(0.892)	(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 6: 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. The variables are described in Appendix C.

F Additional Analysis of Choice of Institution

In this section, we report several regression specifications that support the findings shown in Figure 4.

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

Table 7: OLS regressions of Frequency in Institution X. Robust standard errors in parentheses. Significance levels: * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.

Consider the OLS regressions presented in Table 7. For each subject, the dependent variable is Frequency in Institution X with $X \in \{100, 50, 0\}$, which is equal to the normalized number of periods that subjects spend in Institution X. Independent variables are the 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 Institution 100 and fewer subjects Institution 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 a with high rule-following propensity leave Institution 100 and join Institution 0: exactly the effect that the averages in Figure 4 show.

The same effects are found in other regression specifications. Ordered logit regressions of the probabilities of choosing Institutions 100, 50, and 0 show the same effect of the No Enforcement treatment dummy and its interaction with rule-following (Table 8 below). The same is true for the logit regressions of the probabilities of choosing a specific institution (Table 9 below). Finally, the logit regressions in Table 10 show that rule-followers in the No Enforcement treatment, after initially joining Institution 0, are less likely switch to other institutions 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 Institution Choice

	(1) Prob(Institution)	(2) Prob(Institution)
No Enforcement	-0.781**	-0.814**
	(0.339)	(0.350)
Rule-Following	-0.807*	-0.725
	(0.423)	(0.442)
Rule-Following × No Enforcement	1.748***	1.686***
	(0.553)	(0.577)
Period	0.007	0.007
	(0.007)	(0.007)
Constant Cut 1	-1.220***	-0.891
	(0.276)	(1.246)
Constant Cut 2	0.361	0.690
	(0.278)	(1.249)
σ_u^2	1.592	1.548
	(0.249)	(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 8: Random-effects ordered-logit regressions of the probability of institution choice in the No Enforcement and Enforcement treatments. Institutions are ordered as follows: Institution 100, Institution 50, Institution 0. So, positive coefficients mean the increase in probability of joining Institution 0. Errors are clustered by subject. The variables are described in Appendix C.

Institution Choice Probability

	(1) Pr(Inst 100)	(2) Pr(Inst 100)	(3) Pr(Inst 50)	(4) Pr(Inst 50)	(5) Pr(Inst 0)	(6) Pr(Inst 0)
No Enforcement	0.121**	0.127**	0.058	0.059	-0.167**	-0.177**
	(0.055)	(0.057)	(0.060)	(0.060)	(0.071)	(0.073)
Rule-Following	0.125*	0.112	0.024	0.025	-0.139	-0.136
	(0.072)	(0.074)	(0.078)	(0.078)	(0.087)	(0.091)
Rule-Following × No Enforcement	-0.256***	-0.242***	-0.082	-0.093	0.339***	0.344***
	(0.089)	(0.092)	(0.094)	(0.094)	(0.114)	(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 9: Random-effects logit regressions of the probabilities of choosing an institution in the No Enforcement and Enforcement treatments. Errors are clustered by subject. The variables are described in Appendix C.

Institution Switch Choice Conditional on Last Period Institution

	(1) Pr(Switch | 100)	(2) Pr(Switch | 100)	(3) Pr(Switch | 50)	(4) Pr(Switch | 50)	(5) Pr(Switch | 0)	(6) Pr(Switch | 0)
No Enforcement	-0.033	-0.043	0.016	0.001	0.176*	0.181*
	(0.063)	(0.063)	(0.084)	(0.084)	(0.090)	(0.095)
Rule-Following	-0.051	-0.065	0.038	0.027	0.146	0.124
	(0.083)	(0.083)	(0.110)	(0.104)	(0.104)	(0.113)
Rule-Following × No Enforcement	0.183*	0.188*	0.017	0.040	-0.372***	-0.361**
	(0.108)	(0.107)	(0.131)	(0.129)	(0.143)	(0.152)
Period	0.002	0.002	-0.012***	-0.013***	-0.005*	-0.005*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(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 10: Random-effects logit regressions of the probability of leaving an institution. Errors are clustered by subject. The variables are described in Appendix C.

G Alternative Hypotheses about Choice of Institution

In this section, we falsify several hypotheses that can potentially account for the patterns observed in Figure 4. One possible explanation is risk preferences. In the Enforcement treatment, subjects cannot contribute less than 100% of income in Institution 100 whereas they can contribute any amount in Institution 0, which leads to almost zero contributions (see Figure 2). This entails different degrees of variance of payoffs in the institutions. Variance is highest in Institution 0 and lowest in Institution 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-hand panel of Figure 4 can, in principle, be explained by risk preferences. However, the comparisons of the distributions of elicited risk preferences between institutions 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 institutions is approximately the same since all subjects contribute almost nothing after period 10. Thus, according to the risk preferences hypothesis, we should not observe any difference in rule-following propensity across institutions. However, this is not the case, so we can conclude that risk preferences are not responsible for the observed effect.

Another possible explanation is social preferences. Indeed, in Institution 100 of the Enforcement treatment, the payoffs are the same for all subjects who joined whereas payoffs are random and unequal in Institution 0. Thus, inequality averse subjects might prefer to join Institution 100 and selfish subjects Institution 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-point Likert-type scale. We found no significant difference in the distributions of answers between the three institutions (rank-sum tests, all $p > 0.72$). Thus, the differences in rule-following propensities between institutions in the Enforcement treatment cannot be explained by self-selection due to distributional concerns related to this question. It should also be noted that the choices to put balls in the blue bucket in the rule-following task cannot be attributed to any social preferences because it is an individual task in which only the subject is influenced by the choices. That is, without norm preferences but only social preferences, subjects should never put any balls in the blue bucket. Therefore, no model of social preferences can explain *both* the behavior in the Rule-Following task and the differences in the rule-following propensities between institutions in the Enforcement treatment. The only explanation that remains is thus the one involving rule-following propensity.

H Model

Suppose that individuals have a norm-dependent utility where the norm is equated with the rule of the institution that they choose to join (contribute 100% of their 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 that 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 institution. In the free society, no one has to pay taxes, hence, the norm is not to contribute to the pool at all, making taxation and redistribution become redundant.

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

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

where t_i is the amount of tax chosen by type i and $t^s_{-i} = \frac{\sum_{k \neq i}^{N_s} t_k}{N_s - 1}$ 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 modeling choice that we will discuss at the end of this section. The third term is a positive externality from other members of the institution 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 institution.

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) $1 - r_L > \delta(1) > 1 - r_H$

The first assumption simply states that rule-followers care more about following the rules of society than rule-breakers. The upper and lower bounds are chosen for the sake of technical simplicity. The second assumption is merely a matter of normalization: if other members of the institution do not pay any taxes, i.e., $t^s_{-i} = 0$, then individual i does not have any utility from the externality. The third assumption provides a technical lower bound on the benefit from the externality for 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. The fourth assumption imposes a limit on $\delta(1)$ when all other members of 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

$$V_H(E) = -(1 - r_H) + \delta(1) \tag{2}$$

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

$$V_L(E) = -(1 - r_L) + \delta(1). \tag{3}$$

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 because 0 is higher than the utility they would gain 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 experimental 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 their rule-following propensity. This might be not a particularly desirable property since consumption utility from taxes and from the 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, $t^F_{-i} = \frac{\sum_{k\neq i}^{N_F} t_k}{N_F - 1} = \frac{\sum_{k\neq i}^{N_F} 0}{N_F - 1} = 0$, 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 each rule-follower pays 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, $V_H(E) = -r_H(1-r_H) + \delta(r_L)$. According to assumption (iii), this is negative. Thus, rule-followers have no incentive to unilaterally switch from the free society to the egalitarian one. Hence, this is indeed an equilibrium as no type has an incentive to deviate. Therefore, in the No Enforcement case, rule-breakers choose the egalitarian society and rule-followers choose the free society, as seen in the experimental data.

Reinterpretation

Our specification of the utility function with a 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 the public good. The problem, as was mentioned above, is that, in this case, the two parts of the consumption utility end up being weighted differently: the tax part is weighted with the coefficient $1 - r_i$, whereas the public good part has a weight of 1. It is very simple to reinterpret the model without any changes to the equilibrium analysis by dividing the utility by $1 - r_i$. The transformed utility becomes:

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

where $\phi_i = r_i/(1-r_i)$ is a monotonic increasing transformation of r_i. In this way, we get the consumption utility $-t^2 + \delta(t^s_{-i})$ with two terms weighted equally and the norm-dependent utility $\phi_i(\delta(t^s_{-i}) - (t_i - T_s)^2)$. The first term in the norm-dependent utility, $\phi_i\delta(t^s_{-i})$, 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 can explain the behavior we observe in our experiment.

I Instructions

Below is the set of instructions we handed to participants. The original instructions were in Turkish as the native language of participants. Here, we present just the English translation for brevity.

As discussed, the experiment featured three treatments: i) No-Enforcement, ii) Exclusion, iii) Enforcement. For the sake of brevity, we do not repeat the instructions common across treatments. Instead, we highlight treatment-specific instructions explicitly.

I.1 Rule-Following Task Instructions

> The **Rule-Following Task Instructions** are identical across treatments.

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 you have questions, please raise your hand. Then we will come and answer your questions privately. Any violation of this rule excludes you immediately from the experiment.

Part 1

In Part 1 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 kuruş (0.05 Turkish Lira), and for each ball you put in the yellow bucket, you will receive 10 kuruş (0.10 Turkish Lira).

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 1 will be based on your decisions: it is the sum of the payments from the blue and yellow buckets.

This is the end of the instructions for Part 1. 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.

I.2 Institution Choice and Redistribution Task Instructions

> Some items in the **Institution Choice and Redistribution Task Instructions** are different across treatments. We discuss the similarities and differences below.

Part 2 General Information:

> The **General Information** instructions are identical across treatments.

In each round of the experiment, you will earn an initial income of 0-50 Turkish Liras, which is determined by luck. After that, based on your decisions and the decisions of other participants your final income will be realized. At the

beginning of the experiment and each subsequent round, you can choose to join one of the three groups according to your preference. These groups will differ in terms of their principles of redistribution.

The redistribution principle in the first group (A) is that you transfer all of your initial income (100%) to the common pool. Then the amount collected in the common pool is equally distributed to the participants in the group. If everyone in the group follows this principle, then everyone's final income will be equal. The redistribution principle in the second group (B) is that you transfer half of your initial income (50%) to the common pool and then the amount collected in the common pool is equally distributed to the participants in the group. If everyone in the group follows this principle, then the final income of the participants will be the sum of half of their initial income and their equal share from the common pool, and the inequality among the final incomes of the participants will be more reasonable. The redistribution principle in the third group (C) is that your contribution to the common pool is entirely voluntary. It is envisaged that everyone in the group will only contribute to the common pool at their own discretion and at any amount they choose.

As it 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.

In each round, you will have the opportunity to change your group before your initial income is realized, so the number of participants in each of the three groups will be determined at your discretion in each round. During a given round, participants in each group will only be affected by the decisions of the other participants in that group and will not be affected by the decisions of the participants in the other groups.

Roadmap:

> The **Roadmap** instructions in the **Enforcement** and **No Enforcement** are identical as follows:

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

> The **Roadmap** instructions in the **Exclusion** treatment are as follows:

This part of the experiment consists of 20 rounds. Each round contains 3 stages. In the first stage, participants will choose only their group. In the second stage, participants' initial incomes will be determined by luck. After this, participants will choose their contributions to the common pool and then their final incomes will be determined. In the third stage, the contributions of some participants will be randomly checked. Depending on the amount of their contributions, some participants will be excluded from some groups.

Stage 1 - Group Selection:

> The initial part of the **Stage 1 - Group Selection** instructions is identical as follows:

In stage 1, each participant decides on the group he wants to enter. You can join one of three different groups (A, B, C), as we explained in detail above. The redistribution principles in these three groups are as follows:

A. We should all put our entire income (100%) into the common pool.
B. We should all put half of our income (50%) into the common pool.
C. We should only contribute to the common pool at our own discretion and at any amount we choose.

> The following statements in the **Stage 1 - Group Selection** instructions are common in the **Enforcement** and **No Enforcement** treatments:

In the first stage, participants choose one of these three groups. Participants cannot observe their income before the group selection in the first stage.

> These statements in the **Stage 1 - Group Selection** instructions of the **Exclusion** treatment are slightly different as follows:

In the first stage, participants choose one of these three groups. Note that, as the experiment unfolds, some of the groups might become unavailable to you. This depends on your previous choices (see explanations below). Participants cannot observe their income before the group selection in the first stage.

Stage 2 - Determination of Income:

> The first part of the **Stage 2 - Determination of Income** instructions differs across treatments. In the **No Enforcement** and **Exclusion** treatments, it is as follows:

After the group selection, participants receive an initial random income between 0-50 Turkish Liras, which is determined purely by luck. After observing their initial income on a computer screen, participants choose how much they will contribute to the common pool on this screen. This contribution can take a value between 0 and their initial income. All members of the group decide on their contribution at the same time without seeing each other's decisions. The determination of the final income depends on the initial income and the contributions to the common pool.

> The first part of the **Stage 2 - Determination of Income** instructions in the **Enforcement** treatment is slightly different, as follows:

After the group selection, participants receive an initial random income between 0-50 Turkish Liras, which is determined by purely by luck. After observing their initial income on a computer screen, participants choose how much they will contribute to the common pool on this screen. This contribution must be no less than the prescribed group principle. So, if you choose group A, your contribution to the common pool must be your entire initial income. If you choose group B, you must contribute at least 50% of your initial income to the common pool. If you choose group C, there is no lower limit for the contribution to the common pool. All members of the group decide on their contribution at the same time without seeing each other's decisions. The determination of the final income depends on the initial income and the contributions to the common pool.

> The rest of the **Stage 2 - Determination of Income** instructions is identical across treatments:

In short, 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

The amount received from the common pool is calculated as follows:

Total contribution to the common pool ÷ Number of people in the group

Special Conditions:

> The **Special Conditions** instructions are identical across treatments.

If you are the only member of your group, your initial and final income will be the same regardless of the group you choose, and you will not need to make a decision to contribute to the pool.

Briefing at the End of the Round:

> The **Briefing at the End of the Round** instructions are identical across treatments.

At the end of each round, you will see a detailed screen about the income distribution in your group. On the left-hand side of the screen, you will see your initial income, your contribution to the common pool, the amount you received from the common pool, and your final income. On the right-hand side of the screen, you will see the averages of these values for the group members other than yourself.

Stage 3 - Contribution Checking and Exclusion from Groups

> Stage 3 is only valid for the **Exclusion** treatment. Participants of the other two treatments do not receive the following message:

The choices of the participants will be subjected to a random check. This means that with a probability of 20% each choice of each participant after each round will be compared to the redistribution principle of the group in which this choice is made. In particular, in group A, where the redistribution principle is to contribute 100% of income to the common pool, if your (or any other participant's) contribution was selected for inspection, then it passes the check if your contribution is no less than 80% of the income. If the contribution is less than 80% of the income then you (or other participants) will be excluded from group A until the end of the experiment. This means that you (or others) will not be able to join group A in all remaining rounds. For group B, where the redistribution principle is to contribute 50% of income to the common pool, if your (or others') contribution is selected for inspection and the contribution rate is less than 40%, then you (or others) will be excluded from group B until the end of the experiment. No exclusion rules apply to group C, since it has no binding redistribution principle.

For example, suppose you are in group A and your income is 10. Then, if you choose to contribute 8, 9, or 10 points, you will be able to join group A in the future even if your contribution was subject to inspection. However, if you choose to contribute an amount less than 8 (say 7) and the computer randomly (with a probability of 20%) chooses to check your contribution, then you will be excluded from group A and you will not be able to join again in remaining rounds.

Similarly for group B, if your income is 10, then choosing any amount from 4 to 10 will not affect your ability to join group B in the future. However, if you contribute an amount less than 4, then, with a probability of 20% you will not be able to join group B again in all future rounds.

If your contribution gets selected for inspection and you get excluded from either group A or B, you will see a screen informing you about it. Otherwise, no information will be provided to you.

To make sure that you do not miscalculate the 80% and 40% part of your income, this calculation will be done for you and you will see the appropriate number on the computer screen together with your initial income.

Total Earnings:

> The **Total Earnings** instructions are identical across treatments.

Your earnings in the second part of the experiment will be based on one randomly selected round out of 20 rounds. Your final income on the selected round will be your earnings in this part of the experiment. All of the rounds have the same chance of being selected. So, every round you should decide as if that round is going to be actually selected for payment.

Please Note:

> The **Please Note:** instructions are identical across treatments.

Talking, distracting others, and using mobile phones are forbidden throughout the experiment. If you have a question, please raise your hand quietly. We will come and answer your question in private. All decisions will be made anonymously. In other words, no participant can identify you based on your decisions. The payment will also be made anonymously. Other participants can not learn your earnings.

J Questionnaire

> The **Questionnaire** is identical across treatments.

Demographic questions
Age: in years (as integers).
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 the 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 taken (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.) 1=member, 0=non-member.
Rely: subjects' self-assessment of the reliability of their data in the experiment. Not reliable 0 . . . 10 Very reliable.

Risk:
How willing are you to take risks in general?
Not willing at all 0 . . . 10 Very willing
Trust:
Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people.
People can be trusted 0 . . . 10 No harm in being too cautious

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? 0 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 Greater income differences are needed to give right incentives

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 that 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 have 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? 0 means you agree completely with the first statement on the left; 10 means you agree completely with the second statement on the right; and if your views fall somewhere in between, you can choose any number in between.

Statement 1: *"In the long run, hard work usually brings a better life."*
Statement 2: *"Hard work does not generally bring success; it is more a matter of luck and connections."*

Hard work 0 . . . 10 Luck and 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 an 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.