

Rules, Rule-Following, and Cooperation *

Erik O. Kimbrough¹ and Alexander Vostroknutov²

^{1,2} Department of Economics (AE1), School of Business and Economics, Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands

August 9, 2011

Abstract

We demonstrate experimentally that individual willingness to follow costly rules predicts cooperation in social dilemmas. Subjects participate in a rule-following task in which they may incur costs to follow an arbitrary written rule in an individual choice setting. Without their knowledge, we sort them into groups according to their willingness to follow the rule. These groups then play repeated public goods or trust games. Rule-following groups sustain high public goods contributions over time, but in rule-breaking groups cooperation decays. Rule-followers also reciprocate more in trust games. Arbitrary rules may persist because of their value as screening mechanisms for identifying cooperators.

JEL Classifications: C9, D7, D03

Keywords: experimental economics, rules, social dilemmas, cooperation

Corresponding author: Erik O. Kimbrough, ekimbrough@gmail.com.

* The authors thank Dan Houser, Arno Riedl, Vernon Smith, and Bart Wilson for helpful comments and gratefully acknowledge funding from Maastricht University's METEOR research school and the European Union Marie Curie FP7 grant program. Some figures and data analysis produced using R: A Language and Environment for Statistical Computing. The data are available from the authors upon request. Any remaining errors are our own.

1. Introduction

Without this sacred regard to general rules, there is no man whose conduct can be much depended upon. It is this which constitutes the most essential difference between a man of principle and honor and a worthless fellow. (Adam Smith, 1759. *The Theory of Moral Sentiments*, §3.5.2)

There exists a crucial problem inherent to rule-governed behavior: namely, by following rules in all circumstances for which they are prescribed, individuals and societies will often incur avoidable costs. For example, proper application of legal procedure may occasionally lead to criminals being released even when their guilt is not in doubt. Nevertheless, the existence of a general system of rules is integral to the functioning of the social order because rules and institutions provide consistency and reduce transactions costs (Sowell 1980; Hayek 1988). In general rules develop and persist to the extent that the certainty and consistency they provide more than offsets the costs of creating them, including the costs of occasional misapplications (Demsetz 1967). However, this argument need not imply that effective rules are always the product of deliberate design. Instead, extant rules may have emerged historically as a product of circumstance and persist because they provide advantages to those groups that employ them (Gintis et al. 2001; Henrich et al. 2010).¹ In the case of cross-cultural prohibitions on theft and murder, the value of the rules is obvious, but research across the social sciences has also demonstrated an underlying utilitarian logic to many less obviously beneficial rules and institutions.²

We provide evidence that rules serve a second, complementary purpose beyond the practical wisdom they embody for the solution of social problems: The decision to follow a costly rule reveals information about an individual's type. Specifically, people who follow rules, when doing so is costly, reveal their propensity to (conditionally) cooperate. We design a two-stage laboratory experiment in which we first observe subjects' private willingness to follow an arbitrary and costly rule. Then, unbeknownst to the subjects, we sort them into groups based on the extent of their adherence to the rule. Individuals then

¹ As Hayek (1988) argues, "if we stopped doing everything for which we do not know the reason, or for which we cannot provide a justification (...) we would probably very soon be dead."

² See e.g. Cheung (1968) on sharecropping, Kaplan and Hill (1985) and Gurven (2004) on food sharing norms in hunter-gatherer groups, Ellickson (1989) on 19th century whaling norms, Leeson (2010) on medieval dispute resolution mechanisms, and Iannacone, Haight, and Rubin (2011) on the Oracle of Delphi and divination.

play a repeated social dilemma game with others who followed the rule to a similar degree. We find that groups that adhere to the rule in the first stage sustain cooperation in the second stage.

In one treatment, where the second stage consists of 10 periods of a four-person public goods game with voluntary contributions (Isaac and Walker 1988), individuals in rule-following groups begin by contributing an average of 57% of their endowment in period 1, and by period 10, contributions slightly increase to an average of 64%. On the other hand, in rule-breaking groups, 1st period contributions are nearly identical to the rule-followers at 58%, but by period 10, average contributions decline to 29%. Similarly, in a treatment where the second stage consists of a repeated trust game (Berg et al. 1995), we find that rule-following groups provide 20% greater returns on trust than rule-breakers. Thus, we argue that rule following is the mark of a cooperative individual.

This idea has precedent in the literature on the economics of religion. For example, religious strictures regarding the choice of food items and articles of clothing may act as screening mechanisms that allow members of religious groups to distinguish sincere prospective members from free riders (Iannacone 1992). By imposing a cost on entrants, these groups are able to maintain a high level of public (or club) good provision for their current members. To test this hypothesis, Aimone et al. (2010) design a public goods experiment with endogenous group formation in which the cost of joining various groups differs, and they find that individuals who join groups with higher entry costs also contribute more to the public good. However, our experiment differs from theirs in that our subjects do not choose their own groups, and neither do they know that they are being sorted into groups based on their willingness to endure a cost. Hence, we eliminate the possibility that free riders undertake the cost strategically.

Recent research has also demonstrated that experimental decisions can be used to identify behavioral types (Burnham et al. 2000; McCabe et al. 2001; Houser et al. 2004; Kurzban and Houser 2005; Wilson et al. 2010), and this information can be used to sustain cooperation among a subset of experimental subjects. For example, Gunnthorsdottir et al. (2007) regroup subjects in public goods games according to their initial contributions and find that assortative matching supports cooperation over time. Similarly, Rigdon et al. (2007) show that endogenous sorting of cooperative types in a repeated trust game

sustains cooperation among the positively sorted. In general, behavioral typing from experimental data relies on early-period decisions in the relevant experiment to classify types, which may confound interpretation of the results.³ We also sort our subjects by type without their knowledge, but instead of identifying types based on early decisions in the repeated game, we use apparently unrelated behavior to develop our classification. Subjects decide to what extent they will follow the rule in private, without knowledge of the behavior of others and without knowledge of the second stage of the experiment. We find that willingness to endure a cost in a completely unrelated task nevertheless predicts cooperation in both public goods and trust games. Rule following identifies cooperative individuals.

2. Experimental Design

The experiment consists of two decision-making stages and a questionnaire. In stage 1, which we call the Rule Following stage (RF), subjects control a stick figure walking across the computer screen. Each subject makes 5 decisions concerning the amount of time they wait at a sequence of red traffic lights, each of which will turn green 5 seconds after their arrival. Figure 1 shows the screen that the subjects see.

At the beginning of the RF stage, the stick figure is standing at the left border of the screen, and all traffic lights are red.⁴ Subjects initiate the RF stage by pressing the START button. At this moment, the stick figure starts walking towards the first traffic light. Upon reaching the first red light, the stick figure automatically stops. The light turns green 5 seconds after the stick figure stops; however, subjects are free to press a button labeled 'WALK' any time after the stick figure stops. When a subject presses 'WALK', the stick figure continues walking to the next red light before stopping again, and subjects must once again press 'WALK' to continue to the next light. Throughout the RF stage, the WALK button is shown in the middle of the screen. Subjects can press the WALK button at any time during

³ One interesting exception is Rietz et al. (2011) who implement a surprise restart of the experiment after a one-shot game and use behavior in the first game to type subjects in a repeated version of the same game.

⁴ Before subjects start the task, they see a short cartoon in which the traffic lights blink from red to green. This ensures that subjects understand that the lights can turn green.

the RF stage. However, it becomes functional only when the stick figure stops at a traffic light.

Subjects receive an endowment of 8 Euro, and they are told that for each second they spend in the RF stage they will lose 0.08 Euro. It takes 4 seconds to walk between each traffic light, and 4 seconds from the final light to the finish. Therefore, all subjects lose around 2 Euro walking, and if a subject waits for green at all 5 traffic lights, she will lose an additional 2 Euro waiting. Thus the most a subject can earn in the RF stage is 6 Euro (if she spends no time waiting at traffic lights), and the most she can earn if she waits is 4 Euro (if she waits exactly 5 seconds at each light).⁵ In the instructions for the RF stage (see Appendix A) subjects are told: “*The rule is to wait at each stop light until it turns green*”. No other information, apart from the payment scheme and a general description of the walking procedure, is provided in the instructions.⁶

The rule following task creates a situation, familiar to most subjects, in which they are asked to follow an arbitrary rule at some cost to themselves. Waiting at a stoplight when there are no other vehicles or individuals in sight is an example of seemingly ‘irrational’ obedience, in the sense that (barring the presence of traffic cameras) there is no cost to breaking the rule. In such circumstances, the usual justification for obeying traffic law – ensuring the safety of drivers and pedestrians – has no bite because there are no other drivers or pedestrians to protect or be protected from. Yet in our experience, it is quite common for people to stop and wait impatiently at traffic lights, even in the middle of the night. Why individuals are willing to incur these costs in service of a rule is an open question. One plausible interpretation is that rule following minimizes cognitive exertion; by following the rules of the road, people avoid investing effort in defining their own rules. However, we propose that the decision to follow costly rules is more psychologically and behaviorally revealing: those who incur costs in order to follow rules implicitly identify themselves as cooperators or ‘team players’.

To test this hypothesis, there are two treatments in our experiment: the Public Goods treatment (PG) and the Trust Game treatment (TG). Stage 1 of both treatments is the

⁵ We substituted earnings from this task for a formal show-up payment.

⁶ If subjects asked what would happen if they pass through the red light, one of the experimenters explained that all information relevant to the experiment is given in the instructions.

Rule Following task as described above. In stage 2 of the PG treatment subjects play 10 periods of a repeated Public Goods game with a voluntary contributions mechanism in fixed groups of 4. In stage 2 of the TG treatment subjects play a repeated Trust game 6 times in fixed groups of 4. In particular, each subject plays the game twice with each other subject in the group, once a first mover and once as a second mover. The order is randomized, and subjects receive no identifying information about their partner.

Before making decisions in the RF stage, subjects only receive instructions for that stage. In particular, they are aware that the experiment will consist of several stages, but they know neither what they will do in the next stage(s) nor the connection between the RF stage and consecutive stages.⁷ Unknown to the subjects, their decisions in the RF stage determine their group membership in the PG and TG stages.

We employ an identical matching procedure in both treatments. First, we randomly divide subjects into groups of 8. Second, within each group of 8, we rank subjects according to the total time they spent waiting at traffic lights – at least 25 seconds for those subjects who waited for green at all traffic lights and close to 0 seconds for those who did not wait at any traffic light. Then, in each group of 8, we separate the top 4 subjects (Rule-Followers) and the bottom 4 subjects (Rule-Breakers) into two groups for stage 2. After we match subjects, there is no interaction between any groups of 4. Subjects are not informed about the matching procedure, and they are told only that they will now interact with a fixed group of three other participants (see Appendices B and C).⁸

In the PG treatment each subject receives an endowment of 50 tokens at the beginning of each of the 10 periods (1 token = 1 Euro cent), and she must choose how to divide her tokens between a *group account* and a *private account*. In each period, each subject earns the sum of the amount placed in the private account plus the individual

⁷ In particular, subjects see a label that reads “Part 1” at the top of the rule following instructions (see Appendix A). In dictator game experiments, knowledge of the existence of an unspecified second-stage has been shown to alter subjects’ behavior by making them more cooperative in expectation that their first-stage behavior may influence their second-stage reputation (Smith 2008). If subjects are concerned for their reputation and thus wait longer than they might in a treatment without an implicit ‘shadow of the future’ (or, similarly, with a double-blind protocol), this would dilute the information content of the rule-following task, thereby strengthening our results.

⁸ Note that we did not deceive our subjects. None of the statements in the instructions are false or misleading. As we discuss in the conclusion, it is a separate, and also potentially interesting, question whether subjects’ behavior would change if they had knowledge of the sorting procedure, but our purpose was to discover whether isolated rule-following behavior was sufficient to identify subjects as cooperators.

return from the group account, which is $(0.5 * (\text{sum of all contributions}))$.⁹ Thus, it is individually optimal to contribute nothing to the group account and Pareto optimal for all subjects to contribute their entire endowments. After each period, subjects learn their earnings in that period, the sum of group account contributions from all members of their group, and their total earnings through that period. To avoid end-game effects, subjects are informed only that they will participate in *several* periods of decision-making.

In each period of the TG treatment, we divide each group of 4 into pairs. During the 6 periods, pairs are re-matched so that no pair ever interacts in two consecutive periods. Each subject participates 3 times in the role of first mover (blue person) and 3 times as a second mover (red person, see Appendix C). As in the public goods game, subjects are informed only that they will make *several* decisions, but they are aware that they will participate in both roles.¹⁰

Each subject receives an endowment of 80 tokens in each period (1 token = 1 Euro cent). The first mover chooses to send any amount between 0 and 80 tokens, knowing that the amount sent will be multiplied by 3 and given to the second mover. The second mover then chooses to send back to the first mover any amount between 0 and the amount received. In each period the earnings of the first mover are (80 tokens – tokens sent to the second mover + tokens sent back from the second mover). The earnings of the second mover are (80 tokens + tokens received from the first mover – tokens sent back to the first mover). After each period subjects observe the amounts sent, received and returned as well as their own total earnings up to and including that period.

After stages 1 and 2, subjects answered the Moral Foundations Questionnaire, which was designed to measure the extent of subjects' concern for certain fundamental moral issues (Graham et al. 2008; see Appendix D). Then subjects received cash equal to the sum of money earned in stages 1 and 2. The experiments were conducted at Maastricht University's BEELab in May – June 2011. Overall 72 subjects participated in the PG

⁹ In the instructions subjects are told that all tokens contributed to group account are doubled and then equally divided among the 4 members of their group. It is well known that contributions are increasing in the marginal per capita return (MPCR). We chose an MPCR of 0.5 because it is easy to explain and because it occupies a middle ground between the MPCRs of 0.3 and 0.75 reported in Isaac and Walker (1988).

¹⁰ Burks et al. (2003) find that telling subjects that they will be playing both roles reduces both trust and reciprocity relative to a treatment in which they are unaware.

treatment (18 groups of 4) and 96 subjects participated in the TG treatment (24 groups of 4). As robustness checks which we will discuss in sections 3.3 and 4, we ran the following additional treatments: 1) a *reverse*-PG treatment in which the Public Goods game was played first with random matching into groups of 4, followed by the Rule Following task and the questionnaire (48 subjects, 12 groups of 4); 2) a *no-rule*-PG treatment in which the phrase “*The rule is to wait at each stop light until it turns green*” in the instructions for the RF stage was replaced by “*5 seconds after the stick figure reaches a stop light, it will turn from red to green*” (24 subjects, 6 groups of 4); 3) a *no-rule reverse*-PG treatment combining (1) and (2) (24 subjects, 6 groups of 4); and 4) our first Trust Game session, which fell prey to a software error and was dropped. No other data were collected for this experiment either in the form of pilots or other sessions/treatments. All experiments were programmed in z-Tree (Fischbacher 2007).

3. Results

In this section we analyze the Public Goods and Trust Game treatments in sequence, and after describing the data and summarizing our results from each treatment separately, we discuss the findings from both second-stage treatments together, along with the data from the rule-following task. In particular, we explore the relation between our data and previous findings from the experimental literature on cooperation, reciprocity, and behavioral typing in social dilemmas. We also perform two robustness checks on the rule-following task to confirm its predictive power and to identify the extent to which rule-following results from our explicit statement of the rule. We find no significant differences between experimental sessions, so we pool the data for analysis.

3.1 Public Goods Treatment

Table 1 displays average public goods contributions and red-light waiting times for individuals in rule-following and rule-breaking groups. On average, rule-followers wait 7 seconds longer at the red lights and contribute 17% more of their endowment to the public good than rule-breakers. Figure 2 displays time series of mean total contributions and associated standard errors in rule-following and rule-breaking groups. From the figure, it is

clear that contributions decline over time only among rule-breakers, and we find statistical support in Table 2 which reports Wilcoxon rank-sum tests of the hypothesis of equality of mean group-wise contributions by group type for each period. In 7 out of 10 periods, we reject the null hypothesis of equal mean contributions in favor of the alternative hypothesis that rule-followers contribute more to the public good. Furthermore, comparing average group contributions over the first 5 periods and last 5 periods, additional Wilcoxon tests indicate that mean group contribution is significantly higher in rule-following groups than in rule-breaking groups in both early periods ($W_{9,9} = 61$, p-value = 0.039, one-sided test) and late ($W_{9,9} = 65$, p-value = 0.017, one-sided test).¹¹

3.2 Trust Game Treatment

Figure 3 displays histograms of the amount sent by first movers of each type, and Figure 4 plots the average amount returned by second movers to first movers as a percent of the amount sent, for both group types in 3 bins.¹² Note that there is little difference in the amount of trust between rule-followers and rule-breakers. However, the percent returned is higher in the rule-following groups than in the rule-breaking groups. When rule followers receive a high number of tokens (between 61 and 80), they return an average of 102% of the amount sent, so that first movers suffer no loss due to trust. Rule breakers receiving between 61 and 80 tokens, on the other hand, return only 77% percent of the amount sent. Furthermore, the return on trust among rule followers increases noticeably with the amount sent, but among rule breakers, the average percent returned plateaus at 77% when the amount sent exceeds 20 tokens. To support these observations, Table 3 lists the mean percent of subjects' initial endowments sent to the responder, the mean percent

¹¹ As a robustness check, we estimate a panel regression where the dependent variable is mean group contribution to the public good, and the independent variables are a period trend, a dummy variable that takes a value of 1 if the subject was in a rule-breaking group and a value of 0 otherwise, an interaction between rule-breaking and the period trend, and a constant term. We include random effects for each group to control for repeated measures, and we estimate heteroskedasticity-robust standard errors. A table of estimation results is available in Appendix E, Table E1, column (1). As expected, a negative and significant coefficient on the interaction term between period and rule-breaking ($\beta_{rule-breaker*period} = -1.85$, p-value < 0.01) supports the evidence in Figure 2 that contributions decline over time among rule-breakers, and an insignificant effect of period indicates that contributions do not decline among rule-followers.

¹² Appendix F contains an additional figure showing, for each observation, the amount received by second-movers and the corresponding amounts returned and kept by group type.

return on trust (defined as $100 * (\text{amount sent} / \text{amount returned}) - 100$), and mean waiting time, by group type. While the return on trust is negative for subjects in both rule-breaking (-10.4%) and rule-following groups (-29.1%), it is substantially higher among rule-followers.

This finding is supported by Wilcoxon rank-sum tests of the null hypothesis of equal mean return on trust in rule-following and rule-breaking groups for each period (1-3), where the first period is defined as the first time a subject was in the role of first-mover, and observations are excluded where the first mover sent 0. In the first two periods, we reject the null hypothesis in favor of the alternative hypothesis that mean return on trust is higher in rule-following groups ($W_{43,42} = 752.5$, p-value = 0.089 and $W_{42,37} = 633.5$, p-value = 0.071, one-sided tests); however we cannot reject the null hypothesis in the third period ($W_{39,38} = 698.5$, p-value = 0.331, one-sided test).¹³ Pooling the data and taking the mean return on trust for each subject over all three periods, another Wilcoxon test rejects the null hypothesis of equal mean returns ($W_{48,47} = 950.5$, p-value = 0.092, one-sided test).¹⁴

3.3 Reverse-PG Treatment

To ensure that the sorting mechanism is robust, we also ran 18 groups of 4 subjects each in the Reverse Public Goods treatment in which the order of the two stages was reversed; that is, subjects first participated in a repeated public goods game in randomly assigned groups and *then* participated in the Rule Following task. Figure 2 also shows mean group contribution by period and associated standard errors for the Reverse-PG treatment. When subjects are matched randomly into groups, the well-known pattern of cooperative decay reappears. In period 1 of the reverse-PG treatment, the mean contribution is 60% of

¹³ The number of observations changes because we only consider cases where first movers sent a positive amount.

¹⁴ This finding is also supported by linear panel regressions in which the dependent variable is the mean percent return on trust in group k in period t , and the independent variables include a dummy variable that takes a value of 1 if the group was composed of rule-breakers and 0 otherwise, a control for the amount sent, and a constant term. We include a random effects error structure by group to control for repeated measures and estimate heteroskedasticity-robust standard errors. The results are reported in Appendix E, Table E2, columns 1-2. A negative and weakly significant coefficient on rule-breaking supports our claim that rule-breakers reciprocate less than rule followers (p-value = 0.084). A second regression including a period trend and an interaction between the period trend and the rule-breaking dummy yields a stronger negative coefficient on rule-breaking (p-value = 0.048); however, we also find a significant negative coefficient on the period trend indicating that returns decline over time among rule-followers. This is consistent with Bohnet and Huck (2004) who find that in repeated trust games, with both stranger and fixed matching, returns on trust tend to decline over time.

the endowment whereas in the PG treatment both rule-followers and rule-breakers average 58%. However, by period 10, reverse-PG mean contributions decline to 41% of the endowment, while rule-followers contribute 64% and rule-breakers contribute 29%. A Wilcoxon test rejects the null hypothesis of equal mean contributions in period 10 between rule-followers and reverse-PG groups in favor of the alternative that contributions are higher among rule-followers ($W_{9,18} = 118.5$, $p\text{-value} = 0.028$, one-sided test), but we cannot reject the null hypothesis of equal mean contributions between rule-breakers and reverse-PG groups ($W_{9,18} = 105.5$, $p\text{-value} = 0.216$, two-sided test). Therefore, we conclude that *the sorting procedure* in the PG treatment eliminates cooperative decay in Rule Following groups.

3.4 Discussion

While cooperative deviations from Nash equilibrium play are well-documented in the literature on social dilemmas (See e.g. Andreoni 1995; Henrich et al. 2001; Houser and Kurzban 2002), and the cognitive mechanisms underlying cooperation and reciprocity are being slowly revealed by neuroeconomics (McCabe et al. 2001; Kosfeld et al. 2005; Knoch et al. 2006), the absence of cooperative decay in public goods experiments has generally been observed in outlying cases or with the introduction of communication and/or punishment (Isaac and Walker 1988; Fehr and Gächter 2000; Bochet et al. 2006; Kosfeld et al. 2009; Xiao and Houser 2011). Exceptions to this rule exist; for example, subjects can achieve sustained cooperation when they are sorted according to their contributions (Gunthorsdottir et al. 2007) or when they make binding, incremental, publicly observable contributions in real-time (Kurzban et al. 2001). Similarly, allowing individuals to form their own groups increases average contributions, but there is still a tendency for contributions to decline over time (Page et al. 2005). Yet by simply screening our subjects according to how much cost they will incur to follow an arbitrary rule, we are able, in our first treatment, to identify cooperative types whose contributions to the public good never decline, and in our second treatment, to identify reciprocal types in a trust game.

Notably, 66% of subjects in the PG and TG treatments spend at least 25 seconds waiting (5 seconds per light) indicating that they obey the rule without exception, though it costs them €2. Furthermore, average waiting time is 23.2 seconds, and many subjects who

break the rule while waiting at one or two of the five stoplights nevertheless follow the rule in general. Obedience to arbitrary rules in experimental environments is well-known in social psychology, even when following a rule consists of administering “painful” punishment to others, as in the famous Milgram experiment and numerous replications (Milgram 1963; Zimbardo 2007). The fact that many individuals in our “rule-breaking” groups were not themselves gross violators of the rule suggests that the “broken windows” effect (in which individuals who observe violations of a rule or social norm are more likely to violate the same norm) may be operating in our environment (Wilson and Kelling 1982; Keizer et al. 2008).

Our results in the PG treatment are also consistent with evidence that individuals tend to conform to the (implicit or explicit) norms established by those whose actions they observe (Frey and Meier 2004; Bardsley and Sausgruber 2005; Alpizar et al. 2008; Bicchieri and Xiao 2009; Korth and Reiss 2011). High levels of contribution to the public good, which we initially observe in both the rule-following and rule-breaking groups, are gradually crowded out, but only among rule-breakers. This is also consistent with evidence that the presence of one or more free riders in a population largely composed of conditional and unconditional cooperators is sufficient to induce cooperative decay in a voluntary contributions public goods game (Fischbacher et al. 2001; Kurzban and Houser 2005; Gunnthorsdottir et al. 2007). However, our mechanism allows us to identify these types prior to observing their play in the public goods game, and it also predicts reciprocal play in trust games.

Our results emphasize the importance of understanding what distinguishes rule followers from rule breakers. To identify the impact of individual differences on rule following and cooperation, at the end of each session all subjects answered the Moral Foundations Questionnaire designed to measure the strength of their respect for various moral values (Haidt and Joseph 2004; Graham et al. 2008). While the list is not necessarily exhaustive, the purpose is to measure moral intuitions about the following five values: 1) aversion to doing *harm*; 2) concerns for justice or *fairness*; 3) love of country, family, and the *ingroup*; 4) respect for *authority*; and 5) the desire for cleanliness and *purity*. Subjects answer 6 questions about each of these values using a Lichert scale. We construct a score between 0 and 30 that represents the strength of their respect for each value. Table E4 in

Appendix E summarizes the distribution of individual moral foundation scores pooled across treatments. As an additional control we also ran one No-Rule public goods session with 24 subjects and a No-Rule *reverse*-PG session with 24 subjects in which subjects in the first stage were *not* told “the rule is to wait...” Appendix F contains a figure displaying histograms of waiting time by Rule/No-Rule treatment. We pool the data from all experimental sessions and explore whether any of the moral foundations predict the extent to which individuals follow the rule.

To understand the determinants of rule following, we report logistic regression analysis explaining the decision to break the rule in terms of subjects’ moral foundation scores with controls for subjects’ demographic characteristics and our various treatments. The dependent variable is a binomial variable called “Rule-Breaker” that takes a value of 1 if the subject waited less than 25 seconds and 0 otherwise, and the independent variables are subjects’ scores for each of the five moral values, age, gender, dummy variables for the reverse-PG and No-Rule treatments, an interaction dummy between No-Rule and reverse-PG, field of study dummies, a dummy for non-European subjects, and a constant term. In the reverse-PG treatment, we also control for subjects’ own mean contribution to the public good as well as the mean contribution of others in their group.¹⁵

Table 4 reports the estimation results. First, we note that subjects are substantially more likely to break the rule in the No-Rule treatment than in the other treatments, which indicates that an explicitly stated verbal rule, with no strings attached, is sufficient to induce rule following. Second, we find that female subjects are less likely to break the rule than their male counterparts, and that age has no noticeable effect on rule breaking.¹⁶ Only law students show a significant increase in the likelihood of breaking the rule. Other field of study dummies and the non-European dummy are insignificant. More important for our purposes are the effects of the moral values on rule breaking. We observe only one

¹⁵ Most of our subjects are business majors, so the field of study dummies indicate differences from the average business major. Note that we do not include a dummy for the Trust Game treatment since all Public Goods and Trust Game subjects were unaware of the second stage when making their rule-following decisions. Our results are unchanged when we include a random effects error structure by session.

¹⁶ Women are also less likely to cross at red lights in observational studies of pedestrian behavior in Amman, Jordan and Tel-Aviv, Israel. (Hamed 2000; Rosenbloom 2009) In Amman, age is also negatively correlated with crossing, but because our sample consists of university students, our data may lack the variability necessary to identify an effect. On the other hand, age may matter less in a simulated environment because age no longer correlates with the ability to quickly cross the road.

significant effect. Perhaps unsurprisingly, respect for *authority* is positively and significantly correlated with waiting time.¹⁷

4. Conclusions

We design an experiment that highlights the value of rules as screening mechanisms for identifying cooperative and reciprocal types. Subjects who follow costly rules are sorted, without their knowledge, into groups that participate in repeated social dilemma games. The groups composed of rule-followers are far more cooperative than those containing rule-breakers. We argue that an unremarked value of rules, beyond their application to the solution of various social problems, is that the people who follow them are revealed to be cooperative. This suggests an important reason why rules, norms, and conventions may tend to outlive their other, more apparent, uses. If rules sometimes appear silly or outdated, their continued existence may be explained by the simple fact that, by following them, individuals reveal their willingness to cooperate in a variety of other situations. As Adam Smith put it,

“That reverence for the rule which past experience has impressed upon him, checks the impetuosity of his passion, and helps him to correct the too partial views which self-love might otherwise suggest, of what was proper to be

¹⁷ For the curious reader, Appendix E also reports re-estimations of the regressions in sections 3.1 – 3.2 including average moral value scores within each group as additional independent variables, reported in tables E1 and E2. We also report a third regression table E3 identifying the impact of the period trend and moral values on contributions in the reverse-PG treatment. As before we include random effects for each group and estimate heteroskedasticity-robust standard errors. In both public goods treatments, respect for *authority* positively and significantly impacts contributions and *purity* has a negative and significant effect. In the PG treatment (Table E1, col. 2), *fairness* also increases with contributions, but this effect is not observed in the reverse-PG treatment (Table E3, col. 3); when we also include the data from the no-rule reverse-PG treatment (Table E3, col. 4), the effect of average fairness score on contributions is actually *negative* and marginally significant. In the trust games, we observe no significant effect of moral values on returns to trust (Table E2, col. 3). Also, we note the negative correlation of mean waiting time with mean group contributions in the reverse-PG treatment (Table E3, cols. 1-3). One plausible explanation for this effect is that subjects who restrain their impulse to free ride in the PG game in stage 1 are less able to restrain their impulse to cheat when later asked to follow the rule. See e.g. Mauraven et al. (1998) for a discussion of self-control as a limited resource, and see also Thaler and Shefrin (1981) and Fudenberg and Levine (2006) for attempts to model self-control. A second explanation is that subjects break the rule as a form of (misdirected) retaliation or desire for economic retribution. In the reverse-PG treatment, even high contributors are likely to have been victims of free riding over time. Thus, they may desire to earn back some of their losses by breaking the rule, even though in normal circumstances they otherwise would not. A similar effect has been observed in which those who are treated, by their own judgments, ‘unfairly’, in a dictator game are more likely to cheat when faced with a follow-up task where their performance is self-reported (Houser et al. 2011).

done in his situation.” ~ Adam Smith, 1759. *The Theory of Moral Sentiments*, §3.4.12

An important policy implication of our results is concerned with the “broken windows” hypothesis mentioned above (Keizer et al. 2008). If it is true that rule-following individuals are more prone to cooperate, then encouraging rule following in one social domain might improve prospects for cooperation in other domains. The opposite is also true: rule breaking in one domain might degrade respect for rules and prospects for cooperation in general.

One important direction for future research will be to explore whether costly rules can effectively screen for cooperators when individuals also observe the rule-following decisions of others and/or when they are aware that they will be sorted into groups based on their choices. For example, when forming groups for the provision of public goods, if individuals believe that groups of rule-followers will be more cooperative, then some individuals may strategically follow rules to gain access to those groups. Their free-riding behavior would then likely reduce the benefits of assortative matching. Similarly, it is important to determine whether rule following predicts cooperation when there is no cost to following the rule.

Recent research also demonstrates that greater exposure to impersonal exchange (markets) and to large-scale institutions such as organized religion are both correlated with experimental measures of other-regarding and cooperative behavior (Henrich et al. 2010). Although our subject pool contains individuals from a large number of nations, the preponderance of subjects hail from European nations and were raised according to the rules and norms common to European culture(s). For this reason it will also be important to explore the applicability of our results to a subject pool from a broader range of cultural backgrounds.

References:

- Aimone, Jason A., Laurence R. Iannaccone, Michael D. Makowsky, and Jared Rubin, 2010. "Endogenous Group Formation via Unproductive Costs," ICES Working Paper, George Mason University, Arlington.
- Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. "Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica," *Journal of Public Economics*, 92(5-6), 1047-1060.
- Andreoni, James. 1995. "Cooperation in Public-Goods Experiments: Kindness or Confusion?," *American Economic Review*, 85(4), 891-904.
- Bardsley, Nicholas, and Rupert Sausgruber. 2005. "Conformity and Reciprocity in Public Good Provision," *Journal of Economic Psychology*, 26(5), 664-681.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122-142.
- Bicchieri, Christina, and Erte Xiao. 2009. "Do the Right Thing: But Only if Others Do So," *Journal of Behavioral Decision Making*, 22, 191-208.
- Bochet, Olivier, Talbot Page, and Louis Putterman. 2006. "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior & Organization*, 60(1), 11-26.
- Bohnet, Iris, and Steffen Huck. 2004. "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change," *American Economic Review*, 94(2), 362-366.
- Burks, Stephen V., Jeffrey P. Carpenter, and Eric Verhoogen. 2003. "Playing Both Roles in the Trust Game," *Journal of Economic Behavior & Organization*, 51, 195-216.
- Burnham, Terence, Kevin A. McCabe and Vernon L. Smith. 2000. "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game," *Journal of Economic Behavior and Organization* 43, 57-73.
- Cheung, Steven N. S. 1968. "Private Property Rights and Sharecropping," *Journal of Political Economy*, 76(6), 1107-1122.
- Demsetz, Harold. 1967. "Towards a Theory of Property Rights," *American Economic Review*, 59(2), 347-359.

- Ellickson, Robert C. 1989. "A Hypothesis of Wealth-Maximizing Norms: Evidence from the Whaling Industry," *Journal of Law, Economics & Organization*, 5(1), 83-97.
- Fehr, Ernst and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), 980-994.
- Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10, 171-178.
- Fischbacher, Urs, Simon Gächter and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from Public Goods Experiment," *Economics Letters*, 71, 397-404.
- Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment," *American Economic Review*, 94(5), 1717-1722.
- Fudenberg, Drew, and David K. Levine. 2006. "A Dual-Self Model of Impulse Control," *American Economic Review*, 96(5), 1449-1476.
- Gintis, Herbert, Eric A. Smith and Samuel Bowles. 2001. "Costly Signalling and Cooperation," *Journal of Theoretical Biology*, 213, 103-119.
- Graham, Jesse, Jonathan Haidt, and Brian Nosek. 2008. The Moral Foundations Quiz, www.yourmorals.org
- Gunthorsdottir, Anna, Daniel Houser, and Kevin McCabe. 2007. "Disposition, History and Contributions in Public Goods Experiments," *Journal of Economic Behavior & Organization*, 62(2), 304-315.
- Gurven, Michael. 2004. "To Give and to Give Not: The Behavioral Ecology of Human Food Transfers," *Behavioral and Brain Sciences*, 27, 543-583.
- Haidt, Jonathan, and Craig Joseph. 2004. "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," *Daedalus*, 133(4), 55-66.
- Hamed, Mohammed M. 2001. "Analysis of Pedestrians' Behavior at Pedestrian Crossings," *Safety Science*, 38, 63-82.
- Hayek, Friedrich. 1991. *The Fatal Conceit: The Errors of Socialism*, London, Routledge.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, 91(2), 73-78.

- Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolynatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesogorol, Frank Marlow, David Tracer, and John Ziker. 2010. "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment," *Science*, 327(5972), 1480-1484.
- Houser, Daniel, Michael Keane, and Kevin McCabe. 2004. "Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm," *Econometrica*, 72(3), 781-822.
- Houser, Daniel and Robert Kurzban. 2002. "Revisiting Kindness and Confusion in Public Goods Experiments," *American Economic Review*, 92(4), 1062-1069.
- Houser, Daniel, Stefan Vetter, and Joachim Winter. 2011. "Fairness and Cheating," Working Paper, George Mason University.
- Iannaccone, Laurence R. 1992. "Sacrifice and Stigma: Reducing Free-riding in Cults, Communes, and Other Collectives," *Journal of Political Economy*, 100(2), 271-291.
- Iannaccone, Laurence R., Colleen E. Haight, Jared Rubin, 2011. "Lessons from Delphi: Religious Markets and Spiritual Capitals", *Journal of Economic Behavior & Organization*, 77(3), 326-338.
- Isaac, R. Mark, and James M. Walker. 1988. "Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism," *Economic Inquiry*, 26(4), 585-608.
- Kaplan, Hillard and Kim Hill. 1985. "Food Sharing among Ache Foragers: Tests of Explanatory Hypotheses," *Current Anthropology*, 26(2), 223-246.
- Keiser, Kees, Siegwart Lindenbergh, and Linda Steg. 2008. "The Spreading of Disorder," *Science*, 322, 1681-1685.
- Knoch, Daria, Alvaro Pascual-Leone, Kaspar Meyer, Valerie Treyer, and Ernst Fehr. 2006. "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex," *Science*, 314, 829-832.
- Korth, Christian and J. Philipp Reiss. 2011. "Vacuous Information Affects Bargaining," Working Paper, Maastricht University.
- Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans," *Nature*, 435, 673-676.
- Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. "Institution Formation in Public Goods Games," *American Economic Review*, 99(4), 1335-1355.

- Kurzban, Robert, and Daniel Houser. 2005. "Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations," *Proceedings of the National Academy of Sciences*, 102(5), 1803-1807.
- Kurzban, Robert, Kevin McCabe, Vernon L. Smith, and Bart J. Wilson. 2001. "Incremental Commitment and Reciprocity in a Real-Time Public Goods Game," *Personality and Social Psychology Bulletin*, 27(2), 1662-1673.
- Leeson, Peter. Forthcoming. "Trial by Battle," *Journal of Legal Analysis*.
- Muraven, Mark, Dianne M. Tice, and Roy F. Baumeister. 1998. "Self-Control as a Limited Resource: Regulatory Depletion Patterns," *Journal of Personality and Social Psychology*, 74(3), 774-789.
- McCabe, Kevin A., Daniel Houser, Lee Ryan, Vernon Smith, and Theodore Trouard. 2001. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange," *Proceedings of the National Academy of Sciences*, 98(20): 11832-11835.
- Milgram, Stanley. 1963. "Behavioral Study of Obedience," *Journal of Abnormal & Social Psychology*, 67(4), 371-378.
- Page, Talbot, Louis Putterman and Bulent Unel. 2005. "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Economic Journal*, 115, 1032-1053.
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Reitz, Thomas A., Roman M. Sheremeta, Timothy W. Shields and Vernon L. Smith. 2011. "Transparency, Efficiency and the Distribution of Economic Welfare in Pass-Through Investment Trust Games," Economic Science Institute Working Paper, Chapman University, Orange, CA.
- Rigdon, Mary L., Kevin A. McCabe, and Vernon L. Smith. 2007. "Sustaining Cooperation in Trust Games," *Economic Journal*, 117, 991-1007.
- Rosenbloom, Tova. 2009. "Crossing at a Red Light: Behaviour of Individuals and Groups," *Transportation Research Part F*, 12, 389-394.
- Smith, Adam. 1759. *The Theory of Moral Sentiments*, Indianapolis, Liberty Fund. (1982)
- Smith, Vernon L. 2008. *Rationality in Economics: Constructivist and Ecological Forms*, New York, Cambridge University Press.

Sowell, Thomas. 1980. *Knowledge and Decisions*, New York, Basic Books.

Thaler, Richard H., and H. M. Shefrin. 1981. "An Economic Theory of Self-Control," *Journal of Political Economy*, 89(2), 392-406.

Wilson, James Q. and George Kelling. 1982. "Broken Windows," *The Atlantic Monthly*. URL: <http://www.theatlantic.com/doc/198203/broken-windows>

Wilson, Bart J., Taylor A. Jaworski, Karl Schurter, and Andrew Smyth. 2010. "The Ecological and Civil Mainsprings of Property: An Experimental Economic History of Whalers' Rules of Capture," Working Paper, Chapman University, Orange, CA.

Xiao, Erte and Daniel Houser. 2011. "Punish in Public," *Journal of Public Economics*, 95, 1006-1017.

Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil*, New York, Random House.



Figure 1. Screenshot of the Rule-Following Stage.

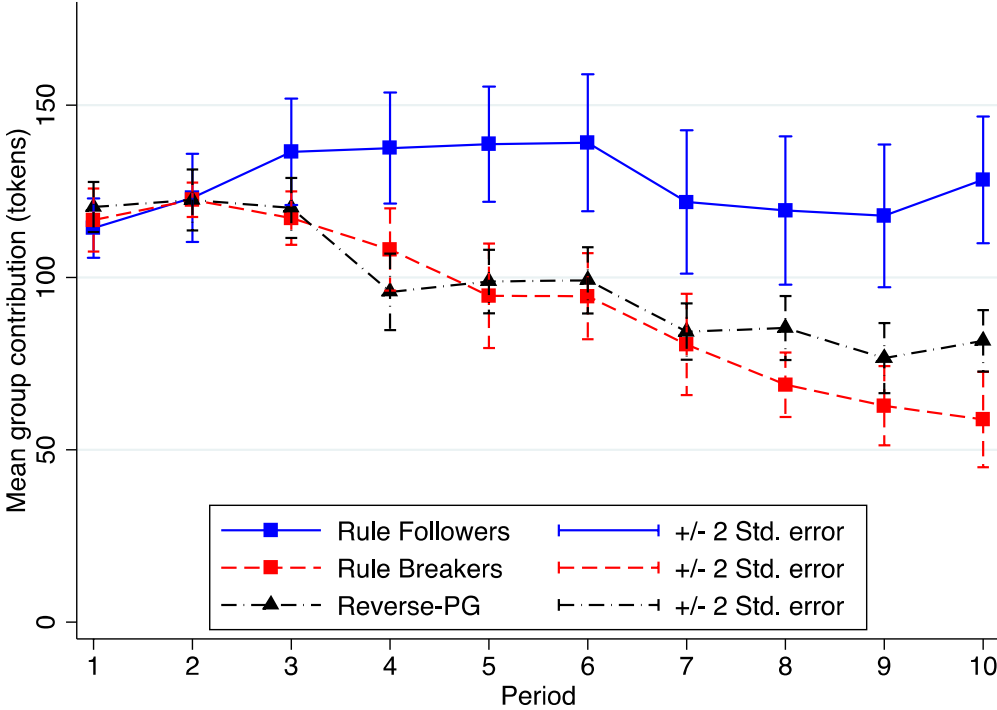


Figure 2. Time Series of Mean Public Good Contributions by Group and Type

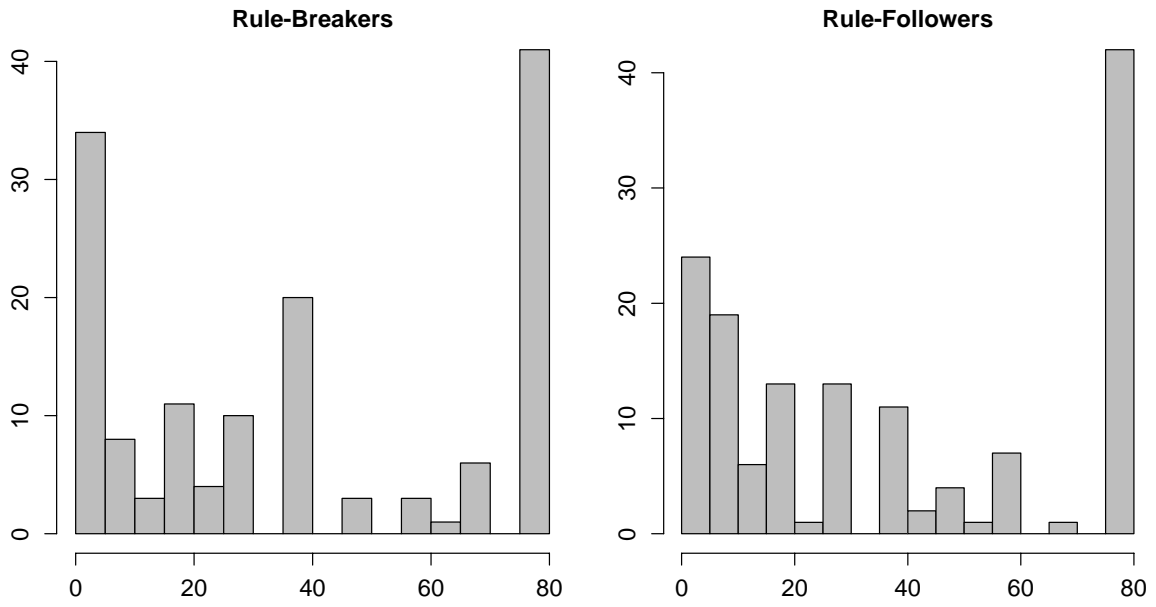


Figure 3. Histograms of Amount Sent in the TG treatment.

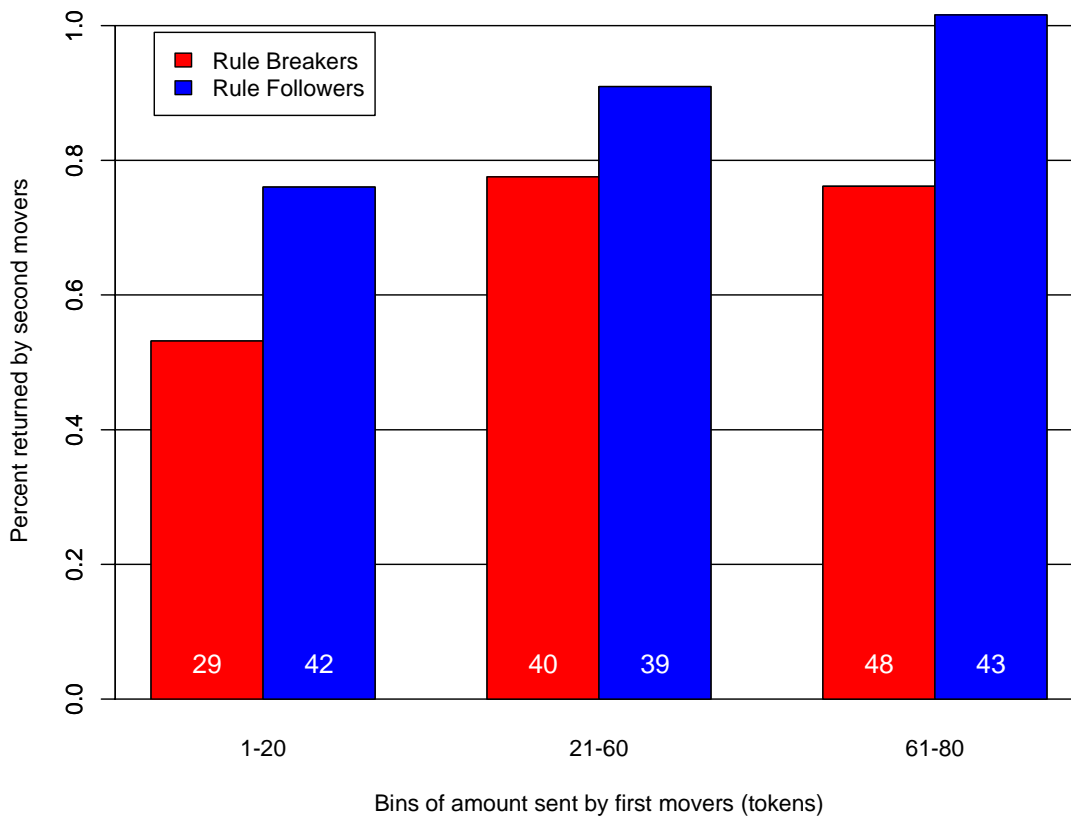


Figure 4. Barplots of Percent Returned by Second Mover in the TG treatment, by Group Type. # of observations listed within each bar.

Table 1. Mean Public Goods Contributions and Waiting Time by Type

Variables	Public Goods	
	Rule-Followers	Rule-Breakers
<i>Percent Contributed (All Periods)</i>	63.84 (2.020)	46.24 (1.890)
<i>Percent Contributed (Periods 1-5)</i>	65.01 (2.742)	55.92 (2.569)
<i>Percent Contributed (Periods 6-10)</i>	62.67 (2.972)	36.56 (2.585)
<i>Waiting Time (Seconds)</i>	27.19 (0.090)	20.39 (0.438)

Standard errors in parentheses.

Table 2. Wilcoxon Tests of Mean Group Contribution, μ , by Period ($H_a: \mu_{Followers} > \mu_{Breakers}$)

<i>Period</i>	1	2	3	4	5	6	7	8	9	10
<i>Test Statistic ($W_{9,9}$)</i>	40	47	61.5	55.5	63	64	60	60.5	61	67
<i>p-value</i>	0.53	0.30	0.035	0.100	0.026	0.021	0.047	0.042	0.039	0.011

Bolded entries statistically significant with p-value < 0.05.

Table 3. Percent of Endowment Sent, Return on Trust, and Waiting Time by Type

Variables	Trust Game	
	Rule-Followers	Rule-Breakers
<i>Percent of Endowment Sent</i>	48.66 (4.868)	49.07 (4.970)
<i>Percent Return on Trust</i>	-10.402 (6.396)	-29.073 (6.017)
<i>Waiting Time (seconds)</i>	27.50 (0.598)	17.94 (1.341)

Standard errors in parentheses.

Table 4. Determinants of Rule-Breaking, Logistic Regression

Independent Variable	Rule-Breaker = {0,1}
<i>Reverse</i>	-0.386 (1.226)
<i>NoRule</i>	2.485*** (0.595)
<i>Reverse*NoRule</i>	0.750 (1.271)
<i>Female</i>	-0.862** (0.324)
<i>Age</i>	-0.000236 (0.000186)
<i>Harm</i>	0.0334 (0.0430)
<i>Fairness</i>	-0.0555 (0.0461)
<i>Ingroup</i>	0.0589 (0.0431)
<i>Purity</i>	0.0302 (0.0395)
<i>Authority</i>	-0.103* (0.0427)
<i>Economics</i>	0.491 (0.353)
<i>Law</i>	1.413* (0.631)
<i>Psych</i>	-0.352 (0.622)
<i>Other</i>	0.0239 (0.490)
<i>Non-European</i>	0.102 (0.442)
<i>Mean Contribution*Reverse</i>	0.0430 (0.0285)
<i>Mean Others Contribution*Reverse</i>	0.00884 (0.0136)
<i>Constant</i>	0.432 (0.910)
Log Likelihood	-143.192
N	264

Clustered standard errors in parentheses, + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Appendix A: Instructions for the Rule Following Task

General information

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

This set of instructions is for your private use only. **During the experiment you are not allowed to communicate with anybody.** In case of questions, please raise your hand. Then we will come to your seat and answer your questions. Any violation of this rule excludes you immediately from the experiment and all payments. The research organization METEOR has provided funds for conducting this experiment.

Part I

In Part I of this experiment, you control a stick figure that will walk across the screen.

Once the experiment begins, you can start walking by clicking the “**Start**” button on the left of the screen.

Your stick figure will approach a series of stop lights and will stop to wait at each light. To make your stick figure walk again, click the “**Walk**” button in the middle of the screen.

The rule is to wait at each stop light until it turns green.

Your earnings in Part I are determined by the amount of time it takes your stick figure to walk across the screen. Specifically, **you begin with an initial endowment of 8 Euro.** Each second, this endowment will decrease by **0.08 Euro.**

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

Appendix B: Instructions for the Public Goods Game

Part II

This part of the experiment will consist of several decision making periods. In each period, you are given an endowment of **50 tokens**. Your task is to decide how to divide these tokens into either or both of two accounts: a **private** account and a **group** account.

Each period you receive the sum of your earnings from your private account plus your earnings from the group account.

There are **4** people, including yourself, participating in your group. You will be matched with the same people for all of Part II.

Each token you place in the **private** account generates a cash return to you (and to you alone) of **one cent (0.01 Euro)**.

Tokens placed in the **group** account yield a different return.

Every member of the group receives the same return for each token you place in the **group** account. Similarly, you receive a return for every token that the other members of the group place in the **group** account.

Thus, your earnings in each decision period are the number of tokens you place in your **private** account, plus the return from all tokens you and the other members of the group place in the **group** account.

Specifically, the total amount of tokens in the group account, that is, your **group** account tokens and the tokens placed in the **group** account by other members of the group, is doubled and then equally divided among **4** members of the group.

Here are two examples to make this clear:

(1) Suppose you place **0** tokens in the **group** account and the other members of your group place a total of **150** tokens in the group account. Your earnings from the group account would be $(2 * 150) / 4 = 75$ cents. Other members of the group would also receive **75** cents from the group account.

(2) Suppose you place **45** tokens in the **group** account and the other members of your group place a total of **15** tokens in the group account. The total group contribution is **60**.

Your earnings from the group account would be $(2 * 60) / 4 = 30$ cents. Other members of the group would also receive **30** cents from the group account.

Each period proceeds as follows:

First, decide on the number of tokens to place in the private and in the group accounts by entering numbers into the boxes labeled private and group. Your entries must sum to your token endowment which is always **50**.

While you make your decision, the **3** other members in your group will also divide their token endowments between the private and group accounts.

Second, after everyone has made a decision, your earnings for that decision period are the sum of your earnings from the private and group accounts.

As an example, suppose the total contribution to the group account at the end of the period was **120**. Your contribution to the group account was **30**, which means your contribution to the private account was **20**. You would earn **80** cents this period, **20** from private account and $(2 * 120) / 4 = 60$ from the group account.

While you are deciding how to allocate your tokens, everyone else in your group will be doing so as well. When the period is over the computer will display your earnings for that period and your total earnings up to and including that period.

This is the end of the instructions. If you have any questions please raise your hand and an experimenter will come by to answer them.

Appendix C: Instructions for the Trust Game

Part II

This part of the experiment will consist of several periods.

In this part, there will be two types of people, **Red** and **Blue**. You will be both a **Red** person and a **Blue** person depending on the period.

Each period you will be randomly paired with a person of the other type.

In this experiment you will interact with **3** other people in the room.

Instructions for Blue People

Each **Blue** person begins each period with **80** tokens. A **Blue** person may choose to send some, all, or none of these tokens to a **Red** person he/she is paired with by typing the amount into a box in the center of the screen and then clicking "**OK**".

Any tokens that a **Blue** person sends to a **Red** person will be subtracted from the **Blue** person's account, multiplied by **3** and transferred to the **Red** person. Any tokens that a **Blue** person chooses *not* to send to the **Red** person remain the **Blue** person's earnings. (Only **Blue** people will be able to send tokens and have them multiplied.)

Instructions for Red People

Each **Red** person enters a period with **80** tokens.

After the **Blue** person makes a decision, the **Red** person will see how many tokens were sent from the **Blue** person.

The amount sent by the **Blue** person will be multiplied by **3** and added to the **Red** person's account. Then the **Red** person decides to send some, all or none of these tokens to the **Blue** person by typing the amount into a box in the center of the screen and then clicking "**OK**". (Only **Red** people will make this decision.)

In each period, each **Red** person is paired with one **Blue** person for the entire period. (One "period" consists of one **Blue** person deciding how many tokens to send to one **Red** person and that **Red** person deciding how many of the multiplied tokens to send to the paired **Blue** person.)

Summary

A **Blue** person's earnings for a period are:

$$\text{Earnings} = \text{Starting tokens} \\ \text{minus Amount Sent to Red}$$

plus Amount Received from Red

A **Red** person's earnings for a period are:

$$\text{Earnings} = \text{Starting tokens} \\ \text{plus Amount Received from Blue} \times 3 \\ \text{minus Amount Sent to Blue}$$

At the end of the experiment the sum of your tokens from all periods will be converted to Euros at a rate of **100 tokens = 1 Euro** and paid to you privately in cash, along with your earnings from Part 1 of the experiment.

This is the end of the instructions. If you have any questions please raise your hand and an experimenter will come by to answer them.

Appendix D: Moral Foundations Questionnaire

Part 1. When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking? Please rate each statement using this scale:

[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong)

[1] = not very relevant

[2] = slightly relevant

[3] = somewhat relevant

[4] = very relevant

[5] = extremely relevant (This is one of the most important factors when I judge right and wrong)

___ Whether or not someone suffered emotionally

___ Whether or not some people were treated differently than others

___ Whether or not someone's action showed love for his or her country

___ Whether or not someone showed a lack of respect for authority

___ Whether or not someone violated standards of purity and decency

___ Whether or not someone was good at math

___ Whether or not someone cared for someone weak or vulnerable

___ Whether or not someone acted unfairly

___ Whether or not someone did something to betray his or her group

___ Whether or not someone conformed to the traditions of society

___ Whether or not someone did something disgusting

___ Whether or not someone was cruel

___ Whether or not someone was denied his or her rights

___ Whether or not someone showed a lack of loyalty

___ Whether or not an action caused chaos or disorder

___ Whether or not someone acted in a way that God would approve of

Part 2. Please read the following sentences and indicate your agreement or disagreement:

[0]	[1]	[2]	[3]	[4]	[5]
Strongly disagree	Moderately disagree	Slightly disagree	Slightly agree	Moderately agree	Strongly agree

___ Compassion for those who are suffering is the most crucial virtue.

___ When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.

___ I am proud of my country's history.

___ Respect for authority is something all children need to learn.

___ People should not do things that are disgusting, even if no one is harmed.

___ It is better to do good than to do bad.

___ One of the worst things a person could do is hurt a defenseless animal.

___ Justice is the most important requirement for a society.

___ People should be loyal to their family members, even when they have done something wrong.

___ Men and women each have different roles to play in society.

___ I would call some acts wrong on the grounds that they are unnatural.

___ It can never be right to kill a human being.

___ I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.

___ It is more important to be a team player than to express oneself.

___ If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.

___ Chastity is an important and valuable virtue.

The Moral Foundations Questionnaire (full version, July 2008) by Jesse Graham, Jonathan Haidt, and Brian Nosek. For more information about Moral Foundations Theory and scoring this form, see: www.MoralFoundations.org

**Moral Foundations Questionnaire: 30-Item Full Version
Item Key, July 2008**

--Below are the items that compose the MFQ30. Variable names are IN CAPS
--Besides the 30 test items there are 2 “catch” items, MATH and GOOD
--For more information about the theory, or to print out a version of this scale formatted for participants, or to learn about scoring this scale, please see: www.moralfoundations.org

PART 1 ITEMS (responded to using the following response options: not at all relevant, not very relevant, slightly relevant, somewhat relevant, very relevant, extremely relevant)

MATH - Whether or not someone was good at math [This item is not scored; it is included both to force people to use the bottom end of the scale, and to catch and cut participants who respond with last 3 response options]

Harm:

EMOTIONALLY - Whether or not someone suffered emotionally
WEAK - Whether or not someone cared for someone weak or vulnerable
CRUEL - Whether or not someone was cruel

Fairness:

TREATED - Whether or not some people were treated differently than others
UNFAIRLY - Whether or not someone acted unfairly
RIGHTS - Whether or not someone was denied his or her rights

Ingroup:

LOVECOUNTRY - Whether or not someone’s action showed love for his or her country
BETRAY - Whether or not someone did something to betray his or her group
LOYALTY - Whether or not someone showed a lack of loyalty

Authority:

RESPECT - Whether or not someone showed a lack of respect for authority
TRADITIONS - Whether or not someone conformed to the traditions of society
CHAOS - Whether or not an action caused chaos or disorder

Purity:

DECENCY - Whether or not someone violated standards of purity and decency
DISGUSTING - Whether or not someone did something disgusting
GOD - Whether or not someone acted in a way that God would approve of

PART 2 ITEMS (responded to using the following response options: strongly disagree, moderately disagree, slightly disagree, slightly agree, moderately agree, strongly agree)

GOOD – It is better to do good than to do bad. [Not scored, included to force use of top of the scale, and to catch and cut people who respond with first 3 response options]

Harm:

COMPASSION - Compassion for those who are suffering is the most crucial virtue.

ANIMAL - One of the worst things a person could do is hurt a defenseless animal.

KILL - It can never be right to kill a human being.

Fairness:

FAIRLY - When the government makes laws, the number one principle should be ensuring that everyone is treated fairly.

JUSTICE – Justice is the most important requirement for a society.

RICH - I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.

Ingroup:

HISTORY - I am proud of my country's history.

FAMILY - People should be loyal to their family members, even when they have done something wrong.

TEAM - It is more important to be a team player than to express oneself.

Authority:

KIDRESPECT - Respect for authority is something all children need to learn.

SEXROLES - Men and women each have different roles to play in society.

SOLDIER - If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.

Purity:

HARMLESSDG - People should not do things that are disgusting, even if no one is harmed.

UNNATURAL - I would call some acts wrong on the grounds that they are unnatural.

CHASTITY - Chastity is an important and valuable virtue.

Appendix E: Additional Regression Tables

Table E1. Mean Group Contributions to the Public Good

	Contribution	
	(1)	(2)
<i>Constant</i>	32.27*** (2.680)	-87.47** (31.30)
<i>Period</i>	-0.064 (0.473)	-0.064 (0.480)
<i>Rule-Breaker</i>	1.381 (3.004)	1.420 (3.592)
<i>Period*Rule-Breaker</i>	-1.851** (0.619)	-1.851** (0.627)
<i>Authority</i>		4.991** (1.563)
<i>Fairness</i>		4.206** (1.493)
<i>Ingroup</i>		-0.797 (0.884)
<i>Purity</i>		-2.834** (1.060)
<i>Harm</i>		-0.178 (1.607)
R²	0.218	0.495
N	180	180

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05
Robust standard errors in parentheses.

Table E2. Mean Percent Return on Trust by Group

	Return on Trust		
	(1)	(2)	(3)
<i>Constant</i>	-55.05*** (13.35)	-22.23^ (12.41)	-114.5* (56.93)
<i>Rule-Breaker</i>	-21.495^ (12.43)	-36.39* (18.37)	-29.08 (18.91)
<i>Amount Sent</i>	0.976*** (0.206)	0.982*** (0.199)	0.948*** (0.240)
<i>Period</i>		-16.53** (5.89)	-16.49** (6.170)
<i>Period*Rule-Breaker</i>		7.45 (6.89)	7.45 (7.158)
<i>Authority</i>			2.595 (4.770)
<i>Fairness</i>			1.216 (3.015)
<i>Ingroup</i>			-1.221 (3.170)
<i>Purity</i>			-3.948 (5.715)
<i>Harm</i>			4.844 (3.214)
R²	0.394	0.447	0.517
N	72	72	72

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05, ^ p-value < 0.1
Robust standard errors in parentheses.

Table E3. Mean Group Contributions to the Public Good, Reverse Treatment

	Mean Group Contribution			
	PG Treatment			Reverse-PG and No-Rule Data Included
	(1)	(2)	(3)	(4)
<i>Constant</i>	43.80*** (7.150)	51.88*** (9.834)	64.86 (40.81)	62.92^ (38.24)
<i>Period</i>	-1.524*** (0.303)	-2.994** (0.987)	-2.994** (1.009)	-1.332*** (0.266)
<i>Mean Waiting Time</i>	-0.567^ (0.337)	-1.008* (0.498)	-1.025* (0.398)	
<i>Period*Mean Waiting Time</i>		0.080 (0.057)	0.080 (0.058)	
<i>Mean Authority</i>			1.250** (0.464)	2.346** (0.865)
<i>Fairness</i>			-1.062 (1.733)	-2.977^ (1.560)
<i>Ingroup</i>			0.261 (1.538)	0.413 (1.524)
<i>Purity</i>			-1.663* (0.649)	-2.261*** (0.663)
<i>Harm</i>			0.394 (1.415)	1.432 (4.533)
R²	0.287	0.300	0.440	0.397
N	120	120	120	180

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05, ^ p-value < 0.1

Robust standard errors in parentheses.

Column 4 excludes the Waiting Time and interaction terms because the No-Rule treatment alters the interpretation of those variables.

Table E4. Average Moral Foundation Scores (out of 30)

	Moral Foundation				
	Authority	Fairness	Harm	Ingroup	Purity
<i>Mean</i>	16.13	21.22	20.64	17.11	13.71
<i>(Std. Deviation)</i>	(4.88)	(4.17)	(4.66)	(4.14)	(5.18)

Appendix F: Additional Figures

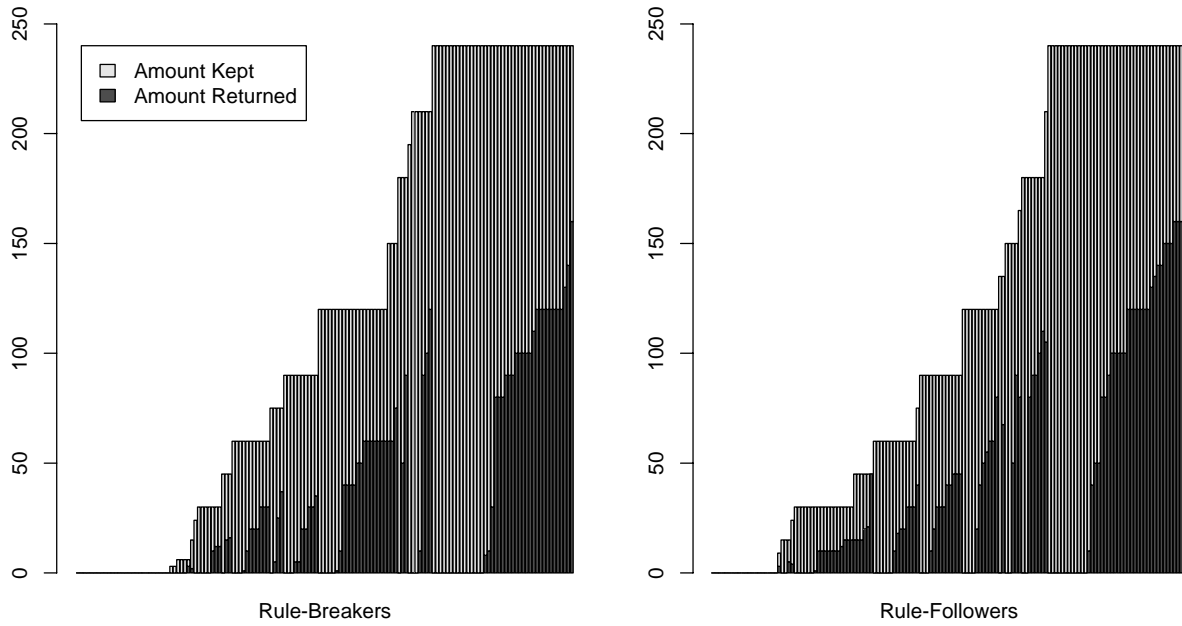


Figure F1. Amount Received, Kept and Returned in the TG Treatment, by Group Type

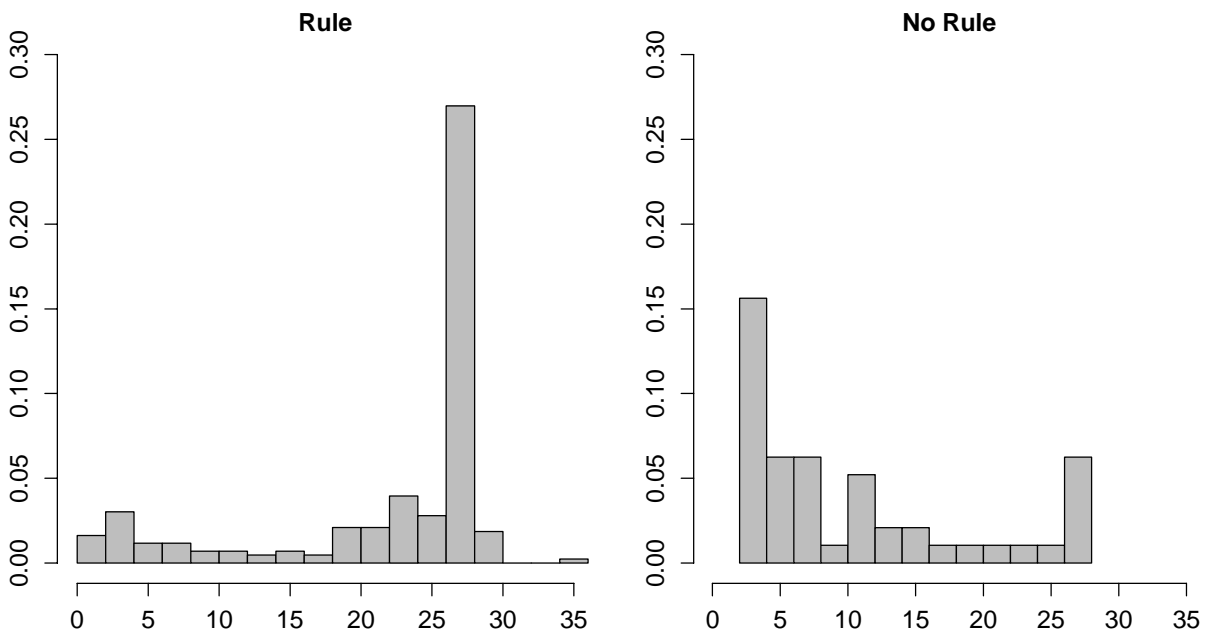


Figure F2. Histograms of Waiting Time in Seconds, Rule vs. No-Rule Treatments