Rules, Rule-Following, and Cooperation*

Erik O. Kimbrough\textsuperscript{a} and Alexander Vostroknutov\textsuperscript{b}

March 12, 2012

Abstract

Rules are thought to persist to the extent that the direct benefits of having them (e.g. reduced transactions costs) exceed the costs of enforcement and of occasional misapplications. We argue that a second crucial role of rules is as screening mechanisms for identifying cooperative types. Thus some apparently costly rules may persist because they allow third parties to screen out defectors. We demonstrate experimentally that costly rule-following can be used to screen for conditional cooperators. 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. However, when individuals are not sorted by type, we observe no differences in the behavior of rule-followers and rule-breakers.

\textit{JEL Classifications:} C91, C92, D70, D03

\textit{Keywords:} experimental economics, rules, social dilemmas, cooperation

\textsuperscript{*} The authors thank Dan Houser, Arno Riedl, Vernon Smith, Bart Wilson, participants in a seminar at the University of British Columbia, and conference participants at the 2011 North American meetings of the Economic Science Association and at the 2012 meetings of the Association for the Study for Religion, Economics, and Culture 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.

\textsuperscript{a} Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

\textsuperscript{b} Department of Economics (AE1), School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
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

---

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.”

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 play a repeated social dilemma game with others who followed the rule to a similar degree. We find that groups composed of individuals who 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 some arbitrary, costly rules may exist because they allow others to screen for cooperators.

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 (Iannaconne 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 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 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 while following a completely unrelated rule nevertheless predicts cooperation when groups are sorted by type; however, when individuals are randomly allocated to groups, rule-following has no predictive power.

\(^3\) 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.
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. 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 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

---

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.
at each light). 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.

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 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 as a first

---

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.
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

---

\(^7\) 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.
amount placed in the private account plus the individual return from the group account, which is 
\((0.5 \times \text{sum of all contributions})\).\(^9\) 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 \((80 \text{ tokens} – \text{tokens sent to the second mover} + \text{tokens sent back} 
\text{from the second mover})\). The earnings of the second mover are \((80 \text{ tokens} + \text{tokens received} 
\text{from the first mover} – \text{tokens sent back to the first mover})\). After each period subjects observe the

\(^9\) 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.
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 and February 2012. Overall 72 subjects participated in the PG 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 PG treatment in which subjects first performed the rule-following task and then played the Public Goods game with 3 randomly chosen individuals (64 subjects, 16 groups of 4); 2) 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); 3) 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); 4) a no-rule reverse-PG treatment combining (1) and (2) (24 subjects, 6 groups of 4); and 5) 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 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 \((W_{9,9} = 61, p\text{-value} = 0.039, \text{ one-sided test})\) and late \((W_{9,9} = 65, p\text{-value} = 0.017, \text{ one-sided test})\).\(^{11}\)

### 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 return on trust (defined as 100*(amount sent/amount returned) – 100), and mean waiting time, by group type. While the return on trust is negative for subjects in

---

\(^{11}\) 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 \((\beta_{\text{rule-breaker*period}} = -1.85, \text{ p-value} < 0.01)\) 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.
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).

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).

### 3.3 Robustness: The No Sorting Treatments

One possible explanation for the observed difference between types is that identified rule-followers are generally more likely to make decisions with reference to some rule or norm. In other words, people who follow an explicit rule in one situation are likely to follow implicit rules in others. The “role-rule” model of human social behavior argues that many decisions can be

---

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.
explained by assuming that people are trying to “play by the rules”, where the rules are
determined by the individual’s perceived role in the interaction (Harré and Secord 1972). Thus, a
rule-follower might make high contributions because she believes it is the appropriate behavior
in a game with potential gains from cooperation.

One implication of simple rule-governed choice is that observed choices should be less
variable over time. This is implied, e.g., by the definition of a social norm (read: informal rule) in
Elster (1989) as “either unconditional or, if conditional, not future-oriented.” In the extreme case,
rule-governed contributions to a public good may be completely insensitive to the behavior of
others. In the case of our sorting public goods treatment, the variability of rule-following group
contributions is obviously less than that of rule-breaking groups since contributions do not
decline over time. Figure 5, showing time series of individual contributions to the public good in
each group of four in our sorting treatment, makes this point clearly. Comparing the bottom and
top halves of the figure, one can observe that rule-following groups are much more likely to
contain individuals who never alter their contributions.

While we cannot directly observe whether a subject chose in accordance with a perceived
contribution rule (we did not collect data on subjects’ stated reasons for their contribution
decisions), we are able to exploit our no-sorting treatments to explore the hypothesis of rule-
governed contributions indirectly. If our rule-following task captures individuals whose behavior
adheres to a rule without regard to the future, then the contributions of rule-breakers should
systematically differ over time from those of rule-followers even in the absence of sorting. Thus,
we also collected data on 16 groups of 4 in a Public Goods treatment without sorting and 18
groups of 4 subjects each in a 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 6 shows mean group contribution by period for the No Sorting- and Reverse-PG treatments as well as for Rule-Followers and Rule-Breakers. When subjects are matched randomly into groups, the well-known pattern of cooperative decay reappears. In period 1 of the reverse-PG and no sorting- treatments, the mean contribution is 60% of 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 and no sorting-PG contributions fall to 44%, 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 the pooled no sorting-PG and reverse-PG groups in favor of the alternative that contributions are higher among rule-followers (W_{9,34} = 262, p-value = 0.028, one-sided test), but we cannot reject the null hypothesis of equal mean contributions between rule-breakers and no sorting-PG and reverse-PG groups (W_{9,34} = 159, p-value = 0.244, two-sided test).\textsuperscript{15}

Moreover, we find no significant relationship between individual rule-following and contributions to the public good. For the pooled no sorting and reverse-PG treatments, we estimate a panel regression of individual contributions on a dummy variable that indicates whether the subject broke the rule, a period trend and an interaction term; we include random effects for each individual to control for repeated measures, and we cluster standard errors at the group level. The regression, reported in Table 4 reveals no significant relationship between individual rule-following and contributions, though the period trend is negative and significant. Therefore, we conclude that the sorting procedure in the PG treatment eliminates cooperative decay in Rule-Following groups. This suggests that norm driven behavior, to the extent that we

\textsuperscript{15} These results are unchanged if we perform separate tests for the No Sorting and Reverse-PG treatments.
observe it, is group-contingent. Rule-following individuals, when matched with other individuals who follow rules, are able to endogenously generate and conform to norms of high contribution, but in the absence of other rule-followers, such norms do not emerge.

### 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 Gachter 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. However, when individuals are not sorted according to their rule-following behavior, individual waiting time is uncorrelated with contributions, which suggests that we have not simply identified those individuals who are willing to incur costs to cooperate.
Notably, excluding the No Rule treatments, 61% of subjects spend at least 25 seconds waiting (5 seconds per light) indicating that they obey the rule without exception, though it costs them at least €2. Furthermore, average waiting time is 22.2 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.
3.5 Individual Differences in Rule-Following

A final feature of our design allows us to explore the determinants of rule-following at an individual level. 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 Likert 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.

Specifically, we report logistic regression analysis explaining the decision to break the rule in terms of subjects’ moral foundation scores with controls for 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.16

Table 5 reports the estimation results, which are unchanged if we exclude the No-Rule treatment. 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.17 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. Surprisingly, none of the moral values scores are correlated with rule-following.18

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

---

16 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.

17 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.

18 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 E4 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 public goods treatments, fairness positively and significantly impacts contributions. In the PG treatment (Table E1, col. 2), authority also increases contributions and purity has a negative and significant effect, but these effects are not observed in the no sorting-PG treatments (Table E3, col. 3). In the no sorting treatments, we also observe a positive and significant impact of in-group; when we also include the data from the no-rule reverse-PG treatment (Table E3, col. 4), only purity has a negative and marginally significant effect on contributions. In the trust games, we observe no significant effect of moral values on returns to trust (Table E2, col. 3).
composed of rule-followers are far more cooperative than those containing rule-breakers. Follow-up treatments in which subjects are randomly assigned to groups indicate that cooperation by rule-followers is contingent on interacting with others of the same type.

Job applications, military training, religious ordination, and gang initiation, to name just a few examples, all require that potential group members display a willingness to follow arbitrary rules and procedures. Our research indicates that one important function of such rules is to screen out non-cooperative individuals, and this fact suggests an important reason why rules, norms, and conventions may tend to outlive their other, more apparent, uses. If rules sometimes appear silly, outdated, or inefficient, their continued existence may be explained by their value as screening mechanisms.

An important policy implication of our results is concerned with the “broken windows” hypothesis mentioned above (Keizer et al. 2008). Since rule-following individuals are more prone to cooperate, but only when they interact with other cooperators, 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 direction for future research will be to test 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 their decisions will impact their access to one or more groups. 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:


Tables

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

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

Standard errors in parentheses.

Table 2. Wilcoxon Tests of Mean Group Contribution, μ, by Period

<table>
<thead>
<tr>
<th>Period</th>
<th>Test Statistic ($W_{0.9}$)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>0.53</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>61.5</td>
<td>0.035</td>
</tr>
<tr>
<td>4</td>
<td>55.5</td>
<td>0.100</td>
</tr>
<tr>
<td>5</td>
<td>63</td>
<td>0.026</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
<td>0.021</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>0.047</td>
</tr>
<tr>
<td>8</td>
<td>60.5</td>
<td>0.042</td>
</tr>
<tr>
<td>9</td>
<td>61</td>
<td>0.039</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Bolded entries statistically significant with p-value < 0.05. One-sided tests.

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

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

Standard errors in parentheses.
Table 4: Regression Explaining PG Contributions, No-Sorting and Reverse Treatments

<table>
<thead>
<tr>
<th>Individual Contribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period</strong></td>
<td>-1.333***</td>
</tr>
<tr>
<td></td>
<td>(0.327)</td>
</tr>
<tr>
<td><strong>Rule-Breaker</strong></td>
<td>2.912</td>
</tr>
<tr>
<td></td>
<td>(3.179)</td>
</tr>
<tr>
<td><strong>Period*Rule-Breaker</strong></td>
<td>-0.129</td>
</tr>
<tr>
<td></td>
<td>(0.419)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>32.56***</td>
</tr>
<tr>
<td></td>
<td>(2.599)</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R²</strong></td>
<td>0.049</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1120</td>
</tr>
</tbody>
</table>

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

Robust standard errors in parentheses.
Table 5. Determinants of Rule-Breaking, Logistic Regression

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Rule-Breaker = {0,1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reverse</td>
<td>-0.545</td>
</tr>
<tr>
<td></td>
<td>(1.163)</td>
</tr>
<tr>
<td>NoRule</td>
<td>2.074***</td>
</tr>
<tr>
<td></td>
<td>(0.536)</td>
</tr>
<tr>
<td>Reverse*NoRule</td>
<td>1.092</td>
</tr>
<tr>
<td></td>
<td>(1.230)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.615*</td>
</tr>
<tr>
<td></td>
<td>(0.283)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0294</td>
</tr>
<tr>
<td></td>
<td>(0.0633)</td>
</tr>
<tr>
<td>Harm</td>
<td>-0.00638</td>
</tr>
<tr>
<td></td>
<td>(0.0373)</td>
</tr>
<tr>
<td>Fairness</td>
<td>0.00427</td>
</tr>
<tr>
<td></td>
<td>(0.0405)</td>
</tr>
<tr>
<td>Ingroup</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>(0.0378)</td>
</tr>
<tr>
<td>Purity</td>
<td>0.00627</td>
</tr>
<tr>
<td></td>
<td>(0.0324)</td>
</tr>
<tr>
<td>Authority</td>
<td>-0.0573</td>
</tr>
<tr>
<td></td>
<td>(0.0359)</td>
</tr>
<tr>
<td>Economics</td>
<td>0.418</td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
</tr>
<tr>
<td>Law</td>
<td>1.026+</td>
</tr>
<tr>
<td></td>
<td>(0.580)</td>
</tr>
<tr>
<td>Psych</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>(0.513)</td>
</tr>
<tr>
<td>Other</td>
<td>-0.151</td>
</tr>
<tr>
<td></td>
<td>(0.402)</td>
</tr>
<tr>
<td>Non-European</td>
<td>0.0912</td>
</tr>
<tr>
<td></td>
<td>(0.397)</td>
</tr>
<tr>
<td>Mean Contribution*Reverse</td>
<td>0.0401</td>
</tr>
<tr>
<td></td>
<td>(0.0277)</td>
</tr>
<tr>
<td>Mean Others Contribution*Reverse</td>
<td>0.00660</td>
</tr>
<tr>
<td></td>
<td>(0.0132)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.684</td>
</tr>
<tr>
<td></td>
<td>(1.664)</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-189.159</td>
</tr>
<tr>
<td>N</td>
<td>328</td>
</tr>
</tbody>
</table>

Standard errors in parentheses, + p<0.10, * p<0.05, ** p<0.01, *** p<0.001
Figures

Figure 1. Screenshot of the Rule-Following Stage.

Figure 2. Time Series of Mean Group Public Good Contributions by 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.
Figure 5: Time Series of Individual Contributions to the Public Good, Rule-Breakers and Rule-Followers
Figure 6. Time Series of Mean Group Public Good Contributions by Treatment
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

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 \((2 \times 150) / 4 = 75\) 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 \((2 \times 60) / 4 = 30\) 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 \((2 \times 120) / 4 = 60\) 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:

\[
\text{Earnings} = \text{Starting tokens} - \text{Amount Sent to Red} + \text{Amount Received from Red}
\]

A Red person’s earnings for a period are:
Earnings = 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:

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Strongly Disagree</td>
<td>1 Moderately Disagree</td>
</tr>
</tbody>
</table>

_____ 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.

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 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

<table>
<thead>
<tr>
<th></th>
<th>Contribution (1)</th>
<th>Contribution (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>32.27***</td>
<td>-87.47**</td>
</tr>
<tr>
<td></td>
<td>(2.680)</td>
<td>(31.30)</td>
</tr>
<tr>
<td><strong>Period</strong></td>
<td>-0.064</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>(0.473)</td>
<td>(0.480)</td>
</tr>
<tr>
<td><strong>Rule-Breaker</strong></td>
<td>1.381</td>
<td>1.420</td>
</tr>
<tr>
<td></td>
<td>(3.004)</td>
<td>(3.592)</td>
</tr>
<tr>
<td><strong>Period*Rule-Breaker</strong></td>
<td>-1.851**</td>
<td>-1.851**</td>
</tr>
<tr>
<td></td>
<td>(0.619)</td>
<td>(0.627)</td>
</tr>
<tr>
<td><strong>Authority</strong></td>
<td></td>
<td>4.991**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.563)</td>
</tr>
<tr>
<td><strong>Fairness</strong></td>
<td></td>
<td>4.206**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.493)</td>
</tr>
<tr>
<td><strong>Ingroup</strong></td>
<td></td>
<td>-0.797</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.884)</td>
</tr>
<tr>
<td><strong>Purity</strong></td>
<td></td>
<td>-2.834**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.060)</td>
</tr>
<tr>
<td><strong>Harm</strong></td>
<td></td>
<td>-0.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.607)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Return on Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-55.05***</td>
</tr>
<tr>
<td></td>
<td>(13.35)</td>
</tr>
<tr>
<td><strong>Rule-Breaker</strong></td>
<td>-21.495^</td>
</tr>
<tr>
<td></td>
<td>(12.43)</td>
</tr>
<tr>
<td><strong>Amount Sent</strong></td>
<td>0.976***</td>
</tr>
<tr>
<td></td>
<td>(0.206)</td>
</tr>
<tr>
<td><strong>Period</strong></td>
<td>-16.53**</td>
</tr>
<tr>
<td></td>
<td>(5.89)</td>
</tr>
<tr>
<td><strong>Period*Rule-Breaker</strong></td>
<td>7.45</td>
</tr>
<tr>
<td></td>
<td>(6.89)</td>
</tr>
<tr>
<td><strong>Authority</strong></td>
<td>2.595</td>
</tr>
<tr>
<td></td>
<td>(4.770)</td>
</tr>
<tr>
<td><strong>Fairness</strong></td>
<td>1.216</td>
</tr>
<tr>
<td></td>
<td>(3.015)</td>
</tr>
<tr>
<td><strong>Ingroup</strong></td>
<td>-1.221</td>
</tr>
<tr>
<td></td>
<td>(3.170)</td>
</tr>
<tr>
<td><strong>Purity</strong></td>
<td>-3.948</td>
</tr>
<tr>
<td></td>
<td>(5.715)</td>
</tr>
<tr>
<td><strong>Harm</strong></td>
<td>4.844</td>
</tr>
<tr>
<td></td>
<td>(3.214)</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.394</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>72</td>
</tr>
</tbody>
</table>

*** 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, No Sorting and Reverse Treatments

<table>
<thead>
<tr>
<th>Mean Group Contributions</th>
<th>No Sorting and Reverse-PG</th>
<th>All Treatments, No Rule Data Included</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Constant</td>
<td>31.90**</td>
<td>35.59**</td>
</tr>
<tr>
<td></td>
<td>(12.06)</td>
<td>(12.62)</td>
</tr>
<tr>
<td>Period</td>
<td>-1.392***</td>
<td>-2.063*</td>
</tr>
<tr>
<td></td>
<td>(0.245)</td>
<td>(0.915)</td>
</tr>
<tr>
<td>Mean Waiting Time</td>
<td>0.125</td>
<td>-0.0501</td>
</tr>
<tr>
<td></td>
<td>(0.477)</td>
<td>(0.519)</td>
</tr>
<tr>
<td>Period*Mean Waiting Time</td>
<td>0.0318</td>
<td>0.0318</td>
</tr>
<tr>
<td></td>
<td>(0.0464)</td>
<td>(0.0468)</td>
</tr>
<tr>
<td>Reverse</td>
<td>-1.516</td>
<td>-1.516</td>
</tr>
<tr>
<td></td>
<td>(4.509)</td>
<td>(4.518)</td>
</tr>
<tr>
<td>Harm</td>
<td>-0.362</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>(1.000)</td>
<td>(1.033)</td>
</tr>
<tr>
<td>Purity</td>
<td>-0.793</td>
<td>-1.302^</td>
</tr>
<tr>
<td></td>
<td>(0.833)</td>
<td>(0.722)</td>
</tr>
<tr>
<td>Authority</td>
<td>-0.118</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td>(0.787)</td>
<td>(0.799)</td>
</tr>
<tr>
<td>Fairness</td>
<td>2.318*</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>(1.142)</td>
<td>(1.514)</td>
</tr>
<tr>
<td>Ingroup</td>
<td>1.780^</td>
<td>1.187</td>
</tr>
<tr>
<td></td>
<td>(0.995)</td>
<td>(1.004)</td>
</tr>
<tr>
<td>R²</td>
<td>0.116</td>
<td>0.118</td>
</tr>
<tr>
<td>N</td>
<td>280</td>
<td>280</td>
</tr>
</tbody>
</table>

*** 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)

<table>
<thead>
<tr>
<th>Moral Foundation</th>
<th>Authority</th>
<th>Fairness</th>
<th>Harm</th>
<th>Ingroup</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>16.08</td>
<td>21.16</td>
<td>20.68</td>
<td>17.06</td>
<td>13.82</td>
</tr>
<tr>
<td>(Std. Deviation)</td>
<td>(4.73)</td>
<td>(4.11)</td>
<td>(4.66)</td>
<td>(4.13)</td>
<td>(5.11)</td>
</tr>
</tbody>
</table>

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