Corrupted Votes and Rule Compliance*

Arno Apffelstaedt† and Jana Freundt‡

Abstract

We study experimentally how people's willingness to comply with elected rules is affected by voter manipulation and disenfranchisement. Groups of 100 subjects vote on a "code of conduct" regarding behavior in a dictator game. Introducing a voting fee, offering subjects money to change their votes, or excluding the votes of low-income subjects leads to a strong decline in voluntary compliance with elected rules that ask subjects to give. Rules that ask subjects to not give see no decline. Heterogeneity in behavioral reactions suggests that treatment effects are driven by preferences for democratic participation and by preferences for unbiased election procedures.

Keywords: endogenous institutions, democracy, corruption, rule compliance, procedural fairness, legitimacy

JEL Codes: D02, D72, D91, C92

*We thank Ned Augenblick, Alexander W. Cappelen, Shachar Kariv, Andreas Lange, Katherine L. Milkman, Francesco Nava, Louis Putterman, Alex Rees-Jones, Karl Schlag, Claudia Schwirplies, Dmitry Taubinsky, Bertil Tungodden as well as audiences at the ESA European Meetings 2016 and 2017, IMEBESS 2017, Behavioral Models of Politics Conference 2017, Econometric Society Winter Meeting 2017, SAEe 2017, NoBeC 2018, and at seminars at the Universities of Cologne, Hamburg, Kiel, and Vienna, the Wharton School of the University of Pennsylvania, Stanford University, Vienna University for Economics and Business, and the WZB Berlin Social Science Center for insightful comments and discussions. We gratefully acknowledge financial support from the University of Hamburg.

†University of Cologne, Faculty of Management, Economics and Social Sciences and Center for Social and Economic Behavior (C-SEB), Albertus-Magnus-Platz, 50931 Cologne, Germany. Email: apffelstaedt@wiso.uni-koeln.de.

‡University of Pennsylvania, School of Arts and Sciences, 311 Claudia Cohen Hall, 249 S 36th St, Philadelphia, PA 19104. Email: jfreundt@sas.upenn.edu.

1 Introduction

This paper studies the corruptive effects of electoral malpractice on rule compliance. In many instances around the world, promises of "free and fair" elections are undermined by practices ranging from systematic vote buying to arguably unintentional disenfranchisement of poor voters or racial minorities. In the 2008 Nicaraguan municipal election, for example, 24% of registered voters reported to have been offered some sort of gift for their vote (Gonzalez-Ocantos et al., 2012).[1] In the Russian parliamentary elections of 2011, electoral fraud is estimated to have increased the share of votes for the incumbent party by at least 11 percentage points (Enikolopov et al., 2013). In other countries, such as the UK and the US, allegations of voter fraud have recently been extensively discussed in the popular press (see, e.g., Cottrell, Herron and Westwood, 2018; UK Electoral Commision, 2018). Both, actual instances of electoral malpractice as well as allegations thereof—even if entirely unfounded—can lead voters to question the integrity of elections (Norris, 2014).[2]

While voters who are dissatisfied with how an election was conducted may challenge elections in court or participate in protests, the ramifications of electoral malpractice are likely to go beyond what is immediately observable as a direct expression of dissatisfaction.[3] Similar to how the introduction of democratic elections can increase the legitimacy of institutions and lead to higher compliance with, for example, tax laws (see, e.g., Frey, 1997; Dal Bó, Foster

[1]Similar numbers are reported for other countries. In a survey study in Argentina from 2002, for example, 35% of respondents reported to have observed the distribution of gifts by political parties in their neighborhoods during election campaigns and 12% of low-income respondents reported to have received something from a political party or candidate (Brusco, Nazareno and Stokes, 2004, pp. 69-70).

[2]Sometimes, allegations concern practices that may or may not be defined as electoral malpractice. In the US, for example, voter registration based on restrictive ID laws and the practice of disenfranchising prison inmates (both of which disproportionately affect racial minorities and the poor) are regarded by some to be a form of systematic malpractice which has the power to bias elections (De Alth, 2009; Manza and Uggen, 2008).

[3]Note, also, that in many countries where electoral fraud is rife, direct expressions of dissatisfaction are often not possible because demonstrations are oppressed by the state (or by party militias) and courts seldom accept cases relating to voter fraud (see, e.g., Enikolopov et al., 2013).

and Putterman, 2010), electoral malpractice may decrease the willingness of voters to comply with the same rules and laws. Suggestive evidence that this is the case can be found, for instance, in the 6th wave of the World Values Survey (WVS, 2014): In countries where people perceive a high level of malpractice in elections, they are also more likely to say that it is justifiable to break rules, ranging from wrongfully claiming government benefits to cheating on taxes, see Figure 1.

Correlations such as those presented in Figure 1 raise two important research questions: (1) Does there indeed exist a *causal* effect of electoral malpractice on actual compliance behavior ? (2) What are the psychological mechanisms that lead people to show negative reactions to malpractice? Can we say more, for instance, about how beliefs about the average rule compliance of others or the specific experience of individuals during elections affect reactions? To shed light on these questions, we have run a novel online experiment, the results of which we present in this paper. We consider an experiment to be particularly valuable not only because it can identify causal effects, but also because it allows us to construct a well-defined *behavioral* measure of voluntary rule compliance. In surveys, researchers mostly rely on self-reported attitudes (see, e.g. Berman et al., 2014), while in field studies, deterrence usually interferes with intrinsic motives to obey rules, for example when observing tax compliance as in Fjeldstad et al. (2018).

We explore the hypothesis that the exclusion or manipulation of voters may negatively affect rule compliance in the context of a referendum in which 100 subjects vote on a "code of conduct" regarding behavior in a subsequent dictator game. In the dictator game, subjects decide between giving and not giving 30 percent of their experimental income to another subject without income. In the referendum, subjects can vote either for a code that asks everyone to give (*Rule:Give*) or a code that asks everyone to not give (*Rule:Don't*). Before the referendum, we observe each subject's behavior in a dictator game without a code. After the referendum, we observe each subject's voluntary compliance with the elected rule.[4]

[4]Because we collect information on the giving behavior in the absence of a rule as well

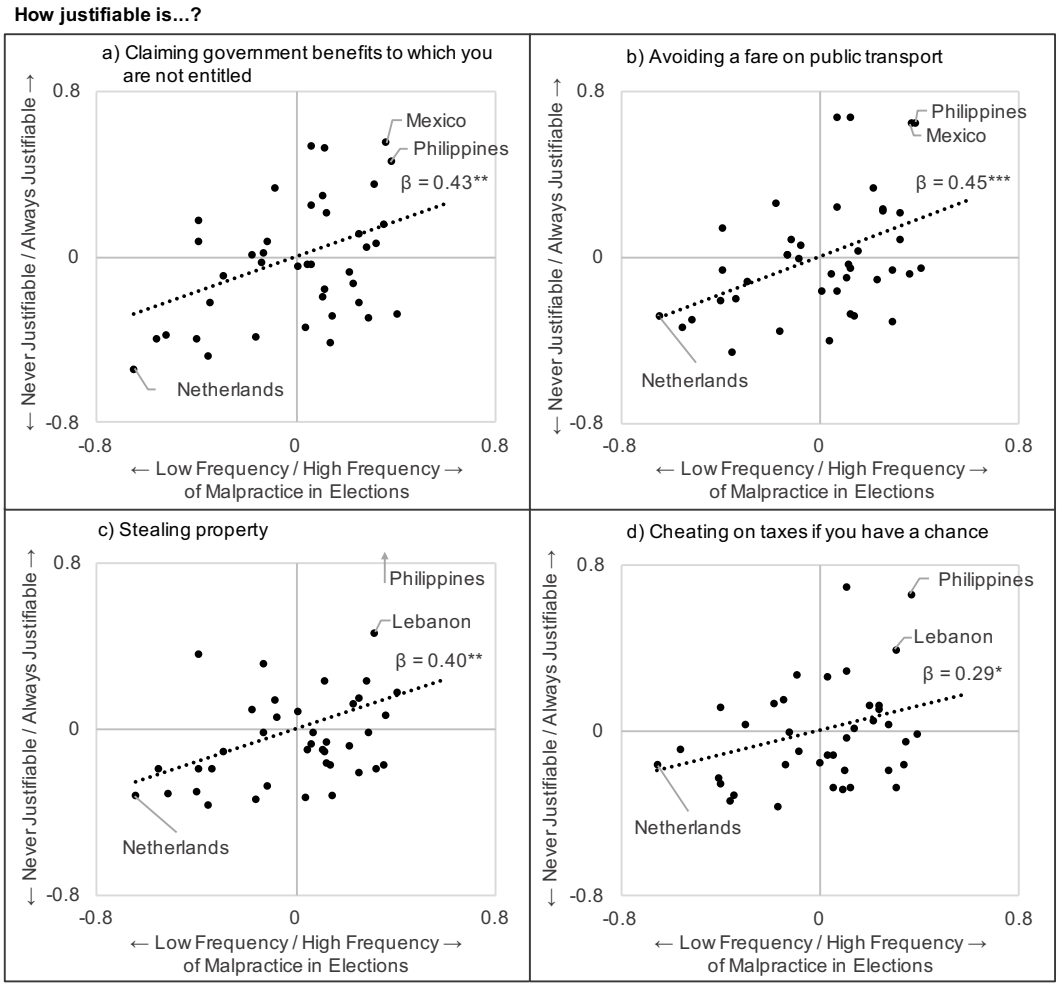


Figure 1: Country-level correlations between citizens' perceived frequency of malpractice in elections and their statements about the justifiability of violating rules and laws (country averages calculated from the WVS (2014)). Y-axis: Average answers in a country to questions V198-V201 ("*How justifiable is...?*"). X-axis: Index of perceived malpractice in elections, calculated from average answers in a country to questions V228 B,C,D,G, and H ("*How often do the following things occur in your country? B: Opposition candidates are prevented from running, C: TV news favor the governing party, D: Voters are bribed, G: Rich people buy elections, H: Voters are threatened with violence at the polls.*"). Data is normalized to show relative deviations from the average across all countries. Univariate OLS regressions without intercept: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

as on the votes of all subjects, our design allows us to control for selection effects that

To set a benchmark, we begin with a baseline treatment (*T_Baseline*), in which the code of conduct is selected by a simple majority vote among the 100 subjects. This treatment demonstrates that soft (i.e. non-binding) rules—when democratically elected—have considerable power to shift behavior: 85 percent (70 percent) of subjects voluntarily follow *Rule:Give* (*Rule:Don't*), compared to just 61 percent (39 percent) choosing to give (not give) in the absence of a code.

Using three additional treatments we then explore how compliance reacts to voter manipulation or disenfranchisement. In the real world, there exists many different forms of interventions to a referendum that can lead some voters to not cast their (true) vote. Examples range from relatively less intrusive practices like the introduction of voting cost to clearly manipulative practices such as vote buying or the explicit exclusion of parts of the electorate. We use the following between-subjects treatments: In treatment *T_Pay4Vote*, subjects are asked to pay a small fee to make their vote count (the votes of subjects who do not pay are not counted towards the referendum). In treatment *T_MoneyOffer*, subjects are offered a small bonus payment if they are willing to change their initial vote to the opposite rule. In treatment *T_ExcludePoor*, the votes of subjects with a household income weakly below GBP 40,000 are not counted in the ballot. As a result of these interventions a considerable share of voters in each treatment (between 35% and 50%) either cast a manipulated vote or no vote.

Our main result is as follows. All three interventions (*Pay4Vote*, *MoneyOffer*, *ExcludePoor*) lead to a significant and sizable loss in the share of subjects complying with *Rule:Give* compared to the baseline. On average, *Rule:Give* loses roughly 50 percent of its power to positively affect giving rates. In stark comparison, we find no systematic effects of our interventions on the power of *Rule:Don't*. We interpret this finding as evidence that the experience of malpractice during an election can indeed have strong, negative and causal effects

may result from different types being present in the different treatments. This allows us to establish clean, causal treatment effects on the population level (à la Dal Bó, Foster and Putterman, 2010).

on rule compliance, *in particular* when elected rules ask subjects to behave pro-socially.[5]

The experimental design allows us to gain insight into who is driving the average effect and into the psychological mechanisms underlying the observed behavioral reaction. Our identification strategy consists of two building blocks: First, in order to understand to what extent treatment differences are driven by strategic concerns, we first study how beliefs about the rule compliance of *other* subjects affects own compliance behavior. While there are no classical coordination incentives in the one-shot dictator game of our experiment, subjects may care about aligning their behavior with what others do or value (Bernheim, 1994; Bénabou and Tirole, 2012; Krupka and Weber, 2013). Note that elections that include voter manipulation or disenfranchisement lead to a noisier signal of the modal policy preferences in the population than a standard majority vote that aggregates *all* subjects' preferences over the two rules. Such a more noisy signal may lead to increased uncertainty if by following the rule one is following the behavior of the majority. Preferences to align one's behavior with others may then drive observed differences between treatments. Because any correlation between beliefs and behavior is likely to be endogenous, we exogenously manipulate beliefs by providing subjects with randomly varying information about the giving behavior in our pilot. Using this as an instrument for beliefs allows us to say whether strategic concerns causally affect behavior.

Second, we then exploit the heterogeneity among our subjects regarding whether they were *personally* affected by our interventions (that is, whether they were personally disenfranchised or their own vote manipulated) as well as regarding their beliefs about the effects of the intervention on the voting outcome. Research in psychology and behavioral economics suggests that procedural aspects of decision making can directly affect behavior. In particular, people seem to care about the "fairness" of decision-making processes (see, e.g., Tyler, 1990; Frey, Benz and Stutzer, 2004; Cappelen et al., 2013) as well as

[5]In fact, most rules and laws that one finds in the real world have a pro-social character, including those on which we display data in Figure 1.

about personally partaking in them (see, e.g., Bonin, Jones and Putterman, 1993; Bardhan, 2000; Bartling, Fehr and Herz, 2014).[6] In our experiment, we empirically identify participants who may have such procedural concerns about disenfranchisement and a lack of representativeness by two variables: We measure (1) whether a subjects' original vote was not counted because she did not pay the fee, accepted the bribe or had a low household income and (2) incentivized beliefs about the share of votes for *Rule:Give* before and after the intervention.

We find that while compliance with *Rule:Don't* is largely driven by strategic concerns, there is no evidence that this is the case for *Rule:Give*. This suggests that the negative effect of malpractice on compliance with pro-social rules may be attributed to immediate reactions to procedural differences rather than to peer effects. Indeed, we find that the heterogeneity in behavioral reactions to malpractice we find in our experiment reflects such immediate reactions that we suggest are linked to procedural concerns: A substantial part of the treatment effects on compliance with *Rule:Give* can be attributed to subjects who lost their vote due to the intervention and to subjects who believe that malpractice biased the final voting outcome.

While our controlled online experiment has the crucial advantage that it can identify causal effects and relate them to distinct psychological motives, its admittedly abstract environment and neutral language may raise concerns about the relation of our findings to real-world behavior. To mitigate these concerns, we complement our experiment with an extensive questionnaire on subjects' standpoints regarding various political issues such as redistribution, corruption, democratic values, and personal trust in institutions. The questionnaire is sent to subjects about two weeks after they participated in the experiment and is presented to them as an unrelated survey (using a different

[6]The idea that procedural concerns may lower the normative appeal of elected rules and thus directly affect the willingness of people to comply is also related to theories of "legitimate authority" (Weber, 1978; Tyler, 2006; Akerlof, 2017). Supporting this view, Besley, Jensen and Persson (2015) find that a change in property taxes in the UK—which was perceived as highly unfair by the public—led to an increase in tax evasion by about 10 percentage points. The authors suggest to attribute this increase to a shock in intrinsic motivation; they cannot, however, pin down the exact motives.

visual design and researcher profile) in order to prevent the risk of spillovers from exposure to different treatments to questionnaire answers. We find that treatment effects are more significant and of larger magnitude among subjects who live in (relatively) democratic countries, who self-report to have a high valuation for democratic institutions, and who perceive instances of bribing and lobbying in the real-world negatively. This finding indicates that our treatment effects relate to psychological domains that are also relevant in corresponding real-world decision making. Moreover, it corroborates our analysis of mechanisms in showing that it is indeed people with a preference for democratic elections who show negative reactions to our interventions.

Our paper is novel in several respects. Existing research in public and political economics has concentrated on comparing democratically elected with exogenously imposed institutions. This literature has established a robust positive effect of democracy. Early correlational evidence, for example by Pommerehne and Weck-Hannemann (1996) and Frey (1997)—who find that tax compliance is higher in Swiss cantons with more democratic participation—has been complemented by extensive experimental evidence. In the field, Bardhan (2000) finds that South Indian farmers are more likely to follow irrigation rules if they partake in crafting them. Grossman and Baldassarri (2012) show that Ugandan farmers contribute more to public goods in a lab-in-the-field experiment if the centralized sanctioning authority has been elected rather than exogenously assigned. In experimental settings with US American and European university students, Tyran and Feld (2006), Ertan, Page and Putterman (2009) and Sutter, Haigner and Kocher (2010) demonstrate that punishments and rewards can have a greater impact on contributions to public goods when they are implemented by majority vote rather than exogenously by a computer. Dal Bó, Foster and Putterman (2010) provide experimental evidence of a similar "democracy effect" in coordination games.[7]

[7]This list of studies is not meant to be exhaustive. See, e.g., Dal Bó (2014) for further studies. There is a related literature in organizational economics that studies the value of "democratic" decision-making mechanism within firms. Bonin, Jones and Putterman (1993), Black and Lynch (2001) and Zwick (2004), for example, provide empirical support that employee participation is associated with increased worker productivity.

A conclusion that can be drawn from the existing literature is that giving citizens decision rights through democratic elections can bring important efficiency gains to societies. We show that for such efficiency gains to materialize it matters *how* democratic institutions are introduced. In particular, we show that the positive dividend of democracy may be sensitive to the existence of attempts to manipulate or disenfranchise voters. Moreover, our design allows for a richer analysis of potential mechanisms that drive the observed behavioral reactions and can therefore shed new light on the psychological sources of "democracy effects".

Regarding the specific design of our study, we deviate from the existing experimental literature on endogenous institutions by studying behavior in the non-strategic environment of a dictator game. Here, non-binding rules should mainly work by their normative appeal.[8] This allows us to create a situation in which people disagree about the "right" code of conduct and hence, potentially, prefer different rules. Note, importantly, that there is no efficiency-dominant rule; *Rule:Give* and *Rule:Don't* differ only in their distributive nature. While numerous studies show that people differ in their judgments regarding whether income received through luck should be redistributed (see, e.g., Cappelen et al., 2007; Almås, Cappelen and Tungodden, 2017), we are not aware of a study that investigates the potential of democratic institutions to unite people behind an "egalitarian" rule (*Rule:Give*) or a "libertarian" one (*Rule:Don't*).[9]

Finally, we want to point out that in comparison to earlier studies on compliance, we study an environment in which compliance with rules is entirely voluntary. There is no (formal or informal) punishment involved with deviation. We consider it of great importance to study compliance in the absence of

[8]Dictator games have been chosen in earlier studies for similar reasons, see, for example, Krupka and Weber (2013), albeit not to our knowledge in studies on the effects of democracy on behavior. Note also that, in comparison to other games—for example, games used to study cheating or lying behavior (Fischbacher and Föllmi-Heusi, 2013; Gächter and Schulz, 2016)—, dictator games do not entail the possibility that a subject can punish the *experimenter* for a procedure she perceives as unfair by not complying with the rules.

[9]Where, following a definition suggested by Almås, Cappelen and Tungodden (2017), "egalitarian" ("libertarian") would mean rejecting (accepting) income differences that are due to luck.

deterrence. First, as discussed by Feld and Tyran (2002, p.88), even in cases where compliance is legally enforced (e.g., in the case of tax rules), expected punishments are usually not high enough to explain the high levels of compliance observed. In other instances, enforcement might simply be too costly or practically unfeasible, as in the case of littering, for example. Second, in the presence of high expected punishments, intrinsic and extrinsic motives might interact (potentially leading to a crowding in or out). Without having established a reliable measure of the intrinsic component of rule compliance it will be difficult to correctly assess the impact of punishment. We thus see our study as also making a methodological contribution in this area.

The remainder of the paper is structured as follows. Section 2 explains the experimental design in detail. Sections 3 presents our results: We first estimate the effects of malpractice on rule compliance (section 3.1) and then study the behavioral mechanism that drive these effects (section 3.2). Our findings are discussed in section 4, before we conclude in section 5. Screenshots of the experimental instructions and the questionnaire can be found in the appendix.

2 Experimental Design

The main prediction guiding the design of our experiment and our analysis is as follows:

Prediction 1 (Effect of electoral malpractice on rule compliance)**.** *The exclusion or manipulation of voters lowers voluntary compliance with elected rules:*

$$E(\mathit{Compliance} \,|\, \mathit{Malpractice} = 1) < E(\mathit{Compliance} \,|\, \mathit{Malpractice} = 0).$$

For each treatment, 100 individual subjects are recruited on the online platform Prolific.ac with a small, fixed base payment and the prospect of a lottery that has one of them wining GBP 100.[10] The lottery is used to

[10]For details on recruitment see paragraph Implementation below. For demographics of the Prolific.ac subject pool, see `https://www.prolific.ac/demographics` (accessed December 3rd, 2018).

naturally form voting groups and to construct a binary dictator game with role uncertainty.

Dictator game. At the beginning of the experiment, subjects are informed that 500 lottery tickets are distributed among the 100 participants of which one is the winning ticket worth GBP 100. The winning ticket is only revealed after the experiment. They are also informed that lottery tickets will be distributed in the following way: 50 randomly chosen participants (called "*receivers*") get 10 lottery tickets each, while the remaining 50 participants (called "*nonreceivers*") get no tickets. Before learning whether she is a receiver of tickets, each subject is asked to decide whether—in case of being a receiver—she wants to $Give_i \in \{0, 1\}$ three out of her ten lottery tickets to a randomly selected non-receiver.[11] In other words, each subject decides whether she wants to redistribute chances to win to another participant who was unlucky and, thus, has zero chances to win the prize. Sharing lottery tickets may reflect social preferences of individual i such as inequality aversion or "warm glow" utility.[12]

Each session implements two rounds of the dictator game. Participants are informed that there will be two rounds but learn about the details of round 2 only after having completed round 1. One round is randomly drawn to determine the distribution of lottery tickets relevant for payment. All decisions are taken anonymously and in private. The timeline of a session is summarized in Figure 2.

Round 1: Giving behavior without a rule. In round 1, each subject decides to $(Give_i|NoRule) \in \{0, 1\}$ in the absence of a rule. Subjects do not receive feedback about the giving decisions of other participants. To ease

[11]Subjects are informed that in the case of being a receiver (50% probability), their decision is automatically implemented and determines the number of lottery tickets for them and for one random other. They are also informed that in the case of being a nonreceiver (50% probability), their decision does not play a role for the distribution of lottery tickets.

[12]Typical references for standard settings are Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Andreoni (1989, 1990). Inequity aversion over chances to win a prize has been modeled by, for example, Saito (2013). Experimental evidence showing how prosocial behavior extends to choices over risky payoffs can be found in Brock, Lange and Ozbay (2013) and Freundt and Lange (2017), among others.

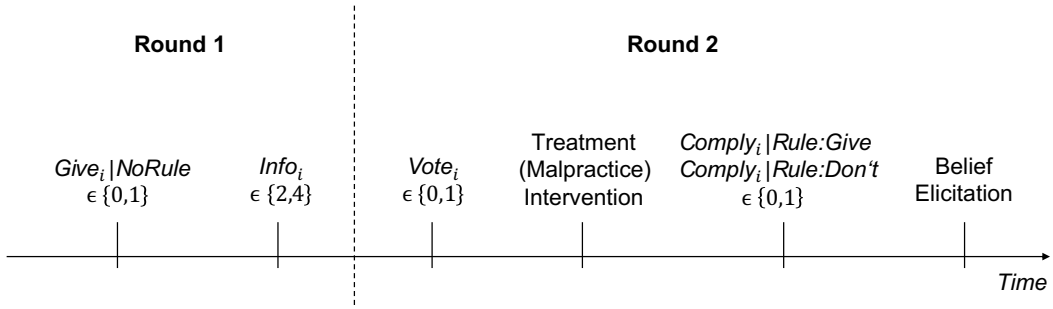


Figure 2: Timeline of experimental session

notation, we henceforth call subjects who choose to give in round 1 *Givers* and those who choose to not give *Non-Givers*.

Definition 1 (Giver and Non-Giver)**.** *If* $(Give_i|NoRule) = 1$*, we call individual* i *a* Giver*. If* $(Give_i|NoRule) = 0$*, we call individual* i *a* Non-Giver*.*

Round 2: Referendum. In round 2, before subjects play the the dictator game again, they are informed that in this round, a "*code of conduct*" will be implemented for all participants. Each subject is asked to vote for the code that she "*prefers to have implemented as the code of conduct for all participants.*" The subject can cast her vote either for *Rule:Give* ("*everybody should choose Give*") or for *Rule:Don't* ("*everybody should choose Don't Give*"). The decision of the subject is coded $Vote_i \in \{Rule{:}Give, Rule{:}Don't\}$. Subjects are not informed about how other participants voted until after the experiment.

Baseline. In the baseline treatment (*T_Baseline*), the rule is selected by simple majority vote among all 100 participants. After a subject has submitted her vote, she is informed that "*the rule that receives more votes in total will be implemented as the code of conduct.*" Each individual then decides whether she wants to $(Give_i|Rule{:}Give) \in \{0, 1\}$ conditional on *Rule:Give* being elected and whether she wants to $(Give_i|Rule{:}Don't) \in \{0, 1\}$ conditional on *Rule:Don't* being elected. Eliciting such state-dependent compliance choices has two major advantages: there is no selection into *Rule:Give* or *Rule:Don't* and the decision whether to give under each rule is made without yet know-

ing the voting outcome. There is no (monetary) punishment involved in not following the elected rule. These choices form our measure of rule compliance:

Definition 2 (Rule Compliance)**.** *We say that a subject complies with Rule:Give,* $(Comply_i|Rule{:}Give) = 1$*, if and only if* $(Give_i|Rule{:}Give) = 1$*. We say that a subject complies with Rule:Don't,* $(Comply_i|Rule{:}Don't) = 1$*, if and only if* $(Give_i|Rule{:}Don't) = 0$*.*

Malpractice. Using three additional (between-subjects) treatments, we explore how compliance reacts to voters being manipulated or disenfranchised. Each treatment introduces a simple intervention to the referendum of the baseline setup, see Table 1. The intervention takes place after subjects have made their initial voting decision, $Vote_i \in \{Rule{:}Give, Rule{:}Don't\}$, but before they decide whether to comply or not. In *T_Pay4Vote*, subjects learn that "*only the votes of participants who pay GBP 0.20 will be counted.*" If a subject decides to pay, her vote is counted toward the majority vote; otherwise, her vote is not counted. In *T_MoneyOffer*, subjects learn that "*all participants are offered an extra payment of GBP 0.20 to vote for the rule that is opposite to what they originally wanted to vote for.*" If a subject decides to accept, her vote is reversed and counts for the opposite rule. Otherwise, her original vote is counted. In *T_ExcludePoor*, subjects are informed that "*only the votes of participants with a household income above GBP 40,000 are counted.*" Each subject also learns whether this implies that her individual vote has not been counted toward the majority vote.[13]

In all treatments, participants know that everyone in their session is subject to the same voting mechanism. They are not informed, however, about the number of participants who decide to pay the fee in *T_Pay4Vote*, about the number of participants who accept the bonus payment in *T_MoneyOffer*, or about the number of participants whose votes are excluded due to their household income in *T_ExcludePoor*. After the intervention, subjects proceed with playing the dictator game in the same way as in *T_Baseline*: Each sub-

[13]To identify a subject as having a household income above or (weakly) below GBP 40,000, we use self-declared information provided by Prolific.ac.

ject decides whether she wants to $(Give_i|Rule{:}Give) \in \{0,1\}$ conditional on *Rule:Give* being elected and whether she wants to $(Give_i|Rule{:}Don't) \in \{0,1\}$ conditional on *Rule:Don't* being elected.

Treatment	Malpractice?	Description	n
T_Baseline	No	Standard majority vote	100
T_Pay4Vote	Yes	Subjects have to pay GBP 0.20 to make vote count	100
T_MoneyOffer	Yes	Subjects are offered GBP 0.20 to reverse their vote	100
T_ExcludePoor	Yes	Only the votes of subjects with annual household income > GBP 40K are counted in the referendum	100

Table 1: Overview of Treatments

Belief Elicitation and Manipulation. At the end of all treatments, we ask participants to state their beliefs about how many of the other 99 participants in their treatment (a) voted for *Rule:Give*, (b) decided to comply with *Rule:Give*, and (c) decided to comply with *Rule:Don't*. Subjects give their answer by indicating a bracket in the set [(0-9), (10-19),..., (90-99)], following Schlag and Tremewan (2016). In order to incentivize agents to state their true empirical expectations, a GBP 0.50 bonus payment is awarded for each correct answer.[14] In *T_Pay4Vote*, *T_MoneyOffer* and *T_ExcludePoor*, we additionally elicit beliefs about the impact of the intervention on final voting outcomes. In *T_Pay4Vote* we ask participants to guess (d) what share of *Rule:Give*-voters in their session were willing to pay for their vote, and (e) what share of *Rule:Don't*-voters in their session were willing to pay. We do the same regarding the share of *Rule:Give*-voters (*Rule:Don't*-voters) who accept the monetary offer in In *T_MoneyOffer*. Finally, in *T_ExcludePoor*, we ask subjects to guess the share of votes for *Rule:Give* separately among high income

[14]Simply put, the subject is asked to guess (up to a certain precision) an empirical frequency that is observed by the experimenter. A prize is then awarded if and only if her guess coincides with the realized frequency. Schlag and Tremewan (2016) show that this method is not only easy to implement, but also particularly robust: Inference does not require postulating any assumptions on the utility function beyond assuming that the subject strictly prefers the prize.

(income > GBP 40,000) and low income participants (income ≤ GBP 40,000) in their session.

To learn how beliefs about the rule compliance of others *causally* affect the decision of a subject to comply, we randomly shock beliefs using exogenous information about others' dictator game giving. At the end of round 1, subjects are presented with a screen that shows them how five other people in "*an earlier study*" (= participants in the pilot) decided in the exact same situation. A random draw (with equal probability) determines whether a subject sees a sample where two out of five participants chose to give ($info_i = 2$) or one where four out of five participants chose to give ($info_i = 4$).

After all participants have finished the experiment, random draws are executed, subjects are matched into pairs and decisions are implemented. Subjects receive all payments and an e-mail with a summary of the outcomes through the online survey platform Prolific.ac within two days after the experiment.

Post-Experimental Questionnaire. In a post-experimental questionnaire, we ask participants about their experience with and attitudes toward, e.g., redistribution, corruption and democratic institutions. Most of the questions in this part are either directly taken or adapted from questions featuring in the 6th wave of the World Value Survey (WVS, 2014). We also collect data on personality characteristics such as risk preferences (self-reported and hypothetical lottery choice), trust, and the Big Five personality traits (using the question format in Gosling, Rentfrow and Swann (2003)). The questionnaire was posted on Prolific.ac as an unrelated survey using a different visual design and researcher profile no earlier than two weeks after a subject had participated in the experiment. These measures are meant to minimize the risk of spillovers from decisions in the experiment and especially from exposure to the different treatments to questionnaire answers. Only subjects who participated in our experiment were able to enter the survey. The full list of questions can be found in Appendix A.3.

Implementation. The experiment was implemented in February and March 2017 online using a subject pool of international participants on the platform Prolific.ac based in Oxford, UK. We programmed the experiment using the software *LimeSurvey* (Schmitz et al., 2012). Detailed instructions and screenshots can be found in Appendix A.4. To ensure understanding and common knowledge thereof, control questions at the end of each screen had to be answered correctly in order to proceed with the experiment. Registered participants on Prolific.ac have a unique ID that is used to identify subjects, to prevent repeated participation and to process payments. When selecting into the experiment, *all* subjects see that they will take part in a lottery that pays GBP 100 to one out of 100 participants and that they will receive a fixed base payment of GBP 1.30 for completing the study which takes roughly 15 minutes to complete.[15] Additional payments are announced during the course of the experiment.

For completing the 10 minute post-experimental questionnaire, subjects receive a compensation of GBP 1. The follow-up-rate is close to 100 percent.[16] In addition, subjects' unique Prolific-ID allows us to access an extensive set of self-reported socio-demographic data, including gender, nationality and income. Our population sample differs in several respects from the typical subject pool at Western university labs: The mean age is 31, almost two thirds of the participants are not students (64%), and about one third have a non-Western nationality (32%).

3 Experimental Results

To set the stage for the analysis of how rule compliance varies with treatments, we begin by providing summary statistics of how subjects behave in round 1, how they vote in round 2, and how the interventions *Pay4Vote*, *MoneyOffer*,

[15] In the case of *T_Pay4Vote*, we increase the base payment by GBP 0.20 to counter adverse wealth effects when subjects pay to make their vote count. This is only announced after they selected into the study; the base payment announced on the prolific website is the same across all treatments.

[16] Of 400 subjects, 387 filled out the questionnaire, i.e. 96.75 percent.

and *ExludePoor* affect the voting process. This information is summarized in Table 2.

	Base-line	*Pay 4 Vote*	*Money Offer*	*Exclude Poor*
Round 1				
Share of subjects choosing...				
$(Give_i \| NoRule) = 1$	.57	.57	.71	.60
Round 2				
Share of...				
initial votes cast = *Rule:Give*	.64	.75	.81	.71
if $(Give_i \| NoRule) = 0$	.35	.47	.45	.38
if $(Give_i \| NoRule) = 1$	.86	.97	.96	.93
subjects paying for vote		.65		
if $Vote_i$ = *Rule:Give*		.69		
if $Vote_i$ = *Rule:Don't*		.52		
subjects accepting money offer			.39	
if $Vote_i$ = *Rule:Give*			.31	
if $Vote_i$ = *Rule:Don't*			.74	
subjects excluded by income $\leq 40K$				.50
if $Vote_i$ = *Rule:Give*				.52
if $Vote_i$ = *Rule:Don't*				.45
final votes counted = *Rule:Give*	.64	.80	.70	.68
Measures of Election Bias				
Outcome_Bias[a]	0	.05	.11	.03
Lost_Votes[b]	0	.35	.39	.50
Observations	100	100	100	100

[a] |(Share of initial votes cast = *Rule:Give*) − (Share of final votes counted = *Rule:Give*)|
[b] $Lost\_Vote_i = 1$ if i does not pay for vote, accepts money offer, or has income $\leq 40K$

Table 2: Summary Statistics. Giving in round 1, voting behavior, and measures of election bias by treatment.

In the absence of a rule, subjects are roughly split between giving and non-giving: On average, 61% of subjects (245/400) choose to give in round 1 (row 1 of Table 2).[17] Voting behavior in round 2 (summarized in the second

[17] While the specific set-up of our dictator game is atypical (role uncertainty, binary decisions, risky prospects with a small probability to win a high price, online participant pool), observed behavior in round 1 of our experiment does not deviate much from typical

to fourth rows) strongly correlates with giving behavior in round 1: Among *Givers* ($(Give_i|NoRule) = 1$), an overwhelming majority (93% on average) vote for *Rule:Give*. Among *Non-Givers* ($(Give_i|NoRule) = 0$), *Rule:Don't* always receives more than half of the votes (59% on average). Overall, between 64% and 81% of the 100 subjects in a treatment group cast their vote for *Rule:Give*.

As a result of the treatment interventions, a considerable share of votes are either not counted or reversed: 35% of participants in *T_Pay4Vote* refuse to pay a fee to make their vote count, 39% of participants in *T_MoneyOffer* are willing to reverse their vote in exchange for the small bonus payment, and, by design, 50% of voters are excluded due to a low household income in *T_ExcludePoor*, see the second to last row of Table 2. We introduce the variable $Lost\_Vote_i \in \{0, 1\}$ to identify a subject whose vote is either uncounted (*T_Pay4Vote* and *T_ExcludePoor*) or reversed (*T_MoneyOffer*) due to the intervention as one of our measures of election bias. Intuitively, excluding a substantial fraction of voters can affect the voting outcome. We measure *Outcome_Bias* as the (absolute) difference between the share of votes for *Rule:Give* before and after the intervention. While a large share of participants lose their vote, the effects on voting outcomes are relatively minor: *Outcome_Bias* ranges between three and eleven percentage points, see the third to last row of Table 2.

3.1 Rule Compliance

Because compliance with either rule likely depends on whether the individual is a *Giver* ($(Give_i|NoRule) = 1$) or *Non-Giver* ($(Give_i|NoRule) = 0$), as well as on whether the individual voted for *Rule:Give* or *Rule:Don't* we take a type-weighted approach to studying rule compliance.[18] We first assess, for each $Type_i = (Give_i|NoRule) \times Vote_i$, the level of rule compliance in the baseline treatment and the effect of interventions *Pay4Vote*, *MoneyOffer* and *Exclude-*

findings on dictator game behavior in the literature. For instance, in a meta-study of 129 dictator game studies covering 41,433 observations, Engel (2011, p.6) finds a share of 63.89% of subjects giving non-zero amounts.

[18]In Appendix A.1 we provide a theoretical framework supporting the claim that rule compliance and voting behavior likely depends on the intrinsic giving preferences (i.e., $(Give_i|NoRule)$) of the individual.

Poor against this benchmark. We then weight types according to the relative frequency with which they appear in our sample. This approach, which closely follows Dal Bó, Foster and Putterman (2010), prevents a misestimation of compliance that can result from an unbalanced distribution of types across our four treatments and can hide or exaggerate actual changes in behavior.

Below, we present estimates for rule compliance on the population level as well as for subgroups defined by giving behavior in round 1 ($Give_i|NoRule$) and voting behavior in the referendum ($Vote_i$). Type-level estimates can be found in Appendix A.2.

3.1.1 Baseline Rule Compliance

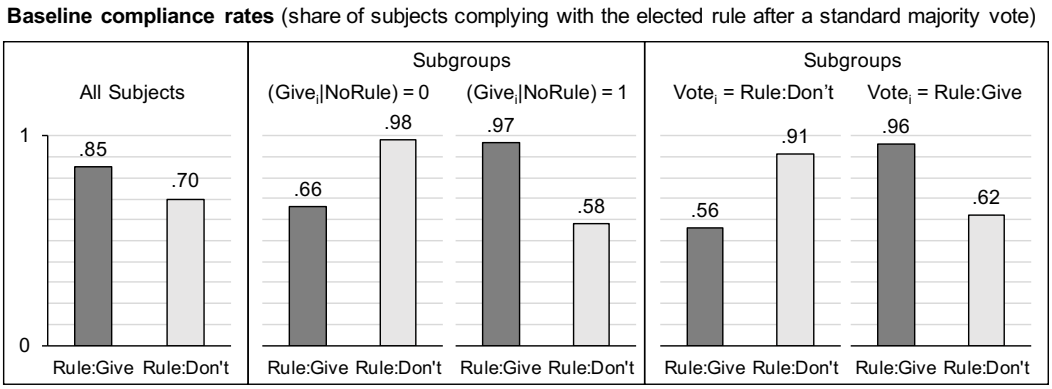


Figure 3: Share of subjects complying with majority-elected rules. Graphs show type-weighted averages. For details see Table A.1 in Appendix A.2.

We observe high compliance with both *Rule:Give* and *Rule:Don't* when rules are selected by a standard majority vote, see Figure 3. As expected, a subject is more likely to follow *Rule:Give* if she is a *Giver* rather than a *Non-Giver* and if she voted for *Rule:Give* rather than for *Rule:Don't*. A symmetric observation holds for compliance with *Rule:Don't*. These differences notwithstanding, the probability with which subjects comply with rules that are opposite to their innate behavior and/or stated preference is striking: 66% of *Non-Givers* (56% of *Rule:Don't*-voters) voluntarily follow *Rule:Give* when it is elected by the majority of participants. Similarly, 58% of *Givers* (62% of *Rule:Give*-voters) comply with *Rule:Don't*. Taking the weighted average

across all types, we find that the unconditional probability of compliance is .85 for *Rule:Give* and .70 for *Rule:Don't*. This compares to a probability of giving (non-giving) in the absence of a rule of only .61 (.39).

In fact, the average difference between an individual's choice in round 2 ($Give_i|Rule{:}Give$ and $Give_i|Rule{:}Don't$, respectively) and the same individual's choice in round 1 ($Give_i|NoRule$) can be used as an estimator of the power of the majority-elected rule to change individual behavior. Analyzing $\Delta Give_i|Rule := (Give_i|Rule) - (Give_i|NoRule)$ in *T_Baseline* we find:

Result 1 (Rules selected by majority vote shift behavior)**.** *When selected by standard majority vote, the share of subjects complying with Rule:Give (Rule:Don't) is substantially larger than the share of subjects choosing to give (to not give) in the absence of a rule.*

Support. Within *T_Baseline*, the average of $\Delta Give_i|Rule{:}Give := (Give_i|Rule{:}Give) - (Give_i|NoRule)$ is $+.24$, which implies a large (24 percentage points) and highly significant ($p < 0.001$, one-sample t-test, two-tailed) increase in giving rates under *Rule:Give*. Similarly, the average of $\Delta Give_i|Rule{:}Don't := (Give_i|Rule{:}Don't) - (Give_i|NoRule)$ is $-.29$, which implies a large (29 percentage points) and highly significant ($p < 0.001$, one-sample t-test, two-tailed) decrease in giving rates under *Rule:Don't*.[19] Confirming these results, non-parametric McNemar tests of the null hypotheses that subjects are equally likely to choose to give in round 1 and round 2 are rejected for both rules ($p < 0.001$). On average, majority-elected non-binding rules thus shift aggregate behavior in the dictator game towards the action promoted by the rule.

3.1.2 Treatment Effects

How does malpractice affect compliance with elected rules? Table 3 reports the estimated difference between the share of subjects complying with *Rule:Give*

[19] These estimates control for correlation in error terms that are due to unobserved individual fixed effects when comparing the behavior of the same group of individuals in round 1 and round 2.

(*Rule:Don't*) after intervention *Pay4Vote*/*MoneyOffer*/*ExcludePoor* and the share of subjects complying with *Rule:Give* (*Rule:Don't*) in the baseline. The results are visualized in Figure 4.

Share of n complying with...	All Subjects	Subgroups $Give_i \mid NoRule = 0$	$Give_i \mid NoRule = 1$	$Vote_i$ = *Rule: Don't*	$Vote_i$ = *Rule: Give*
...**Rule:Give**					
Pay4Vote	-.10	-.25	.00	-.07	-.10
	(.053)	(.021)	(.982)	(.587)	(.020)
MoneyOffer	-.12	-.24	-.05	-.03	-.16
	(.013)	(.049)	(.146)	(.817)	(.000)
ExcludePoor	-.09	-.21	-.02	-.11	-.09
	(.059)	(.052)	(.656)	(.387)	(.056)
Malpractice	-.11	-.23	-.03	-.09	-.11
(Pooled)	(.008)	(.009)	(.384)	(.383)	(.002)
Constant	.85	.66	.97	.56	.96
(*T_Baseline*)	(.000)	(.000)	(.000)	(.000)	(.000)
...**Rule:Don't**					
Pay4Vote	-.10	-.07	-.11	-.14	-.08
	(.111)	(.188)	(.237)	(.054)	(.312)
MoneyOffer	-.02	-.12	.04	-.02	-.02
	(.768)	(.055)	(.624)	(.827)	(.813)
ExcludePoor	.06	-.00	.10	.05	.06
	(.316)	(.963)	(.283)	(.421)	(.435)
Malpractice	-.01	-.06	-.02	-.01	-.01
(Pooled)	(.823)	(.188)	(.802)	(.843)	(.869)
Constant	.70	.98	.58	.91	.62
(*T_Baseline*)	(.000)	(.000)	(.000)	(.000)	(.000)
Observations	400	155	245	109	291

p-values (two-tailed t-test) in parentheses.

Table 3: Effect of interventions *Pay4Vote*, *MoneyOffer*, and *ExcludePoor* on compliance rates. Average treatment effects (ATE) calculated as the weighted average of treatment effects by $Type_i = (Give_i|NoRule) \times Vote_i$ assuming normally distributed standard errors. See Table A.1 in Appendix A.2 for treatment effects on type-level.

We see strong, systematic, and statistically significant effects on compliance with *Rule:Give*. When subjects are asked to pay for their vote (*T_Pay4Vote*),

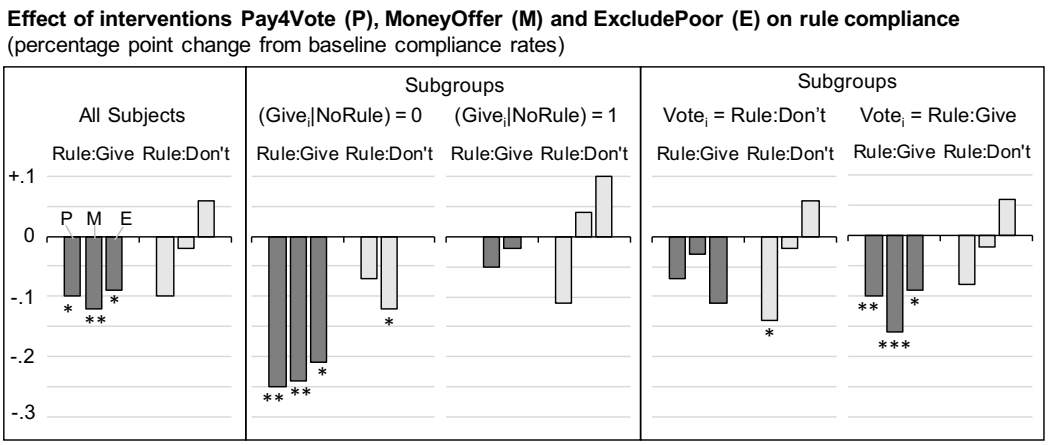


Figure 4: Effect of interventions *Pay4Vote* (P), *MoneyOffer* (M) and *ExcludePoor* (E) on rule compliance. Graphs show type-weighted averages, see Table 3. Stars denote statistically significant differences to the baseline compliance rate: * $p < .1$, ** $p < .05$, *** $p < .01$.

when they are offered money to reverse their vote (*T_MoneyOffer*), or when a large share of them is excluded from the ballot due to household income (*T_ExcludePoor*), compliance with the prosocial rule decreases between 9 and 12 percentage points in the overall population (see column 1 in Table 3 as well as the first panel of Figure 4). The second column in Table 3 (the second panel in Figure 4, respectively) shows that this effect is largely driven by *Non-Givers*: Only roughly 40% of *Non-Givers* follow *Rule:Give* after an election that saw one of the three interventions, compared to roughly 65% in the baseline. This is intuitive: First and foremost, malpractice should be affecting those subjects who need to be convinced to follow the behavior promoted by the rule.[20] In the case of *Rule:Give*, these are the *Non-Givers*. The strongest effect is found for *Non-Givers* who voted for *Rule:Give*. It is their response which drives the significant decrease among *Rule:Give*-voters (displayed in the third panel of Figure 4).[21] While other types show smaller effects, the negative impact on compliance with *Rule:Give* is systematic across the entire population. Although the introduced interventions are quite different, their effect on

[20]See Appendix A.1 for a theoretical framework which formalizes this claim.
[21]For treatment effects on type-level see Table A.1 in Appendix A.2.

compliance with *Rule:Give* is strikingly similar.

Regarding subjects' compliance with *Rule:Don't*, Table 3 and Figure 4 show smaller, inconsistent, and mostly insignificant effects. Given the systematic changes we observe for the opposite rule, this might be surprising. Intuition suggest that *Givers* in particular should be affected. However, no systematic or significant treatment effects can be observed in the respective subgroup $((Give_i|NoRule) = 1)$. We conclude:

Result 2 (**Main Result**) (Electoral malpractice decreases compliance with *Rule:Give* but not with *Rule:Don't*)**.** *Subjects display strong, systematic, and statistically significant reductions in compliance with Rule:Give when the rule is elected in the presence of interventions Pay4Vote, MoneyOffer, and Exlude-Poor. We observe smaller, inconsistent, and mostly insignificant effects of the same interventions on compliance with Rule:Don't.*

Support. Using a type-weighted approach, we find that the population average treatment effect (ATE) of interventions *Pay4Vote*, *MoneyOffer*, and *ExludePoor* on compliance with *Rule:Give* is $-.10$ ($p = 0.053$), $-.12$ ($p = 0.013$), and $-.09$ ($p = 0.059$), respectively (see Table 3, column 1). When pooling interventions, the ATE on compliance with *Rule:Give* is $-.11$ ($p = 0.008$). While *Non-Givers* show the strongest decline, a weakly negative effect is found for all subgroups (see columns 2 to 5). Treatment effects on *Rule:Don't*, on the other hand, are sometimes positive and sometimes negative, mostly insignificant and generally smaller. On average, the interventions are estimated to have little to no effect on compliance with *Rule:Don't*: The pooled ATE is $-.01$ ($p = 0.823$). Identical effects as those reported in Table 3 (usually with higher levels of significance) are found with other methods that account for type-dependent treatment effects, for example, inverse probability weighting or regression adjustment.[22]

Figure 5 provides an insightful illustration of our main result: The figure

[22]Note that the type-weighted approach we follow is identical to a matching estimator with exact matching on (discrete) type covariates.

explicitly depicts the power of elected rules to change behavior: A standard majority vote for *Rule:Give* has the power to decrease non-giving rates by more than 60% relative to the share of subjects choosing not to give in the absence of a rule. This power to reduce selfish behavior is reduced by nearly half (to roughly 30%) by interventions *Pay4Vote*, *MoneyOffer*, or *ExludePoor* (average effect when pooling treatments). At the same time, the power of *Rule:Don't* to decrease giving rates stays roughly constant at about 50%.

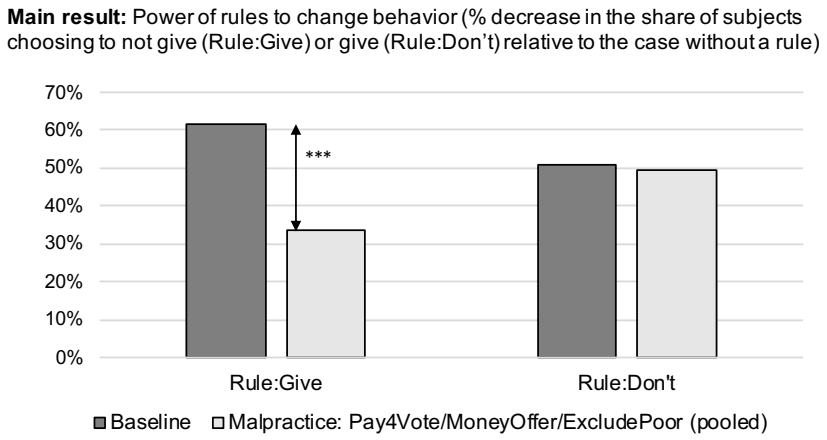


Figure 5: Effect of interventions *Pay4Vote*, *MoneyOffer* and *ExludePoor* (pooled) on the power of elected rules to change behavior. Bars show decrease (in %) in the share of subjects choosing to not give (give) after the election of *Rule:Give* (*Rule:Don't*) relative to the share of subjects choosing to not give (give) in the absence of a rule. The graph is based on type-weighted averages, stars denote significance of population average treatment effect of *Malpractice* (Pooled) on compliance with *Rule:Give*, $p < .01$, see Table 3.

3.2 Understanding Rule Compliance and Treatment Effects

We exploit several features of our design to shed light on potential psychological mechanisms underlying the treatment effects we find. We first analyze elicited beliefs about the rule compliance of other participants to be able to say more about the potential role of "peer effects" in compliance decisions. In particular, it might be that subjects change their behavior as a reaction to

our interventions because the intervention changed their beliefs about what others will do. We also explore two explanations that are directly related to procedural preferences subjects may have about rule setting mechanisms: Are people less willing to comply with rules if they did not *personally* participate in selecting them? Does compliance vary with beliefs about a potential election bias?

3.2.1 Beliefs About the Rule Compliance of Other Subjects

Do subjects follow rules because they want to follow others? While the dictator game does not provide strategic incentives in a strictly economic sense, subjects may still be inclined to condition their compliance choices on the expected behavior of the 99 other participants in their group. Following this conjecture, we study to what extent beliefs about the compliance behavior of other subjects can explain rule compliance in general and treatment differences in particular. Recall that during the experiment subjects are not informed about what other participants choose. At the end of each session, we elicit beliefs about (a) how many of the other 99 subjects vote for *Rule:Give*, (b) how many of the other 99 subjects comply with *Rule:Give*, and (c) how many of the other 99 subjects comply with *Rule:Don't*. Figure 6 displays the frequency of these beliefs (pooled across all treatments) by answer bracket.

Comparing the distributions of individual beliefs about the behavior of other participants in treatment *T_Baseline* with *T_Pay4Vote*, *T_MoneyOffer* and *T_ExcludePoor*, we do not observe systematic differences.[23] This makes beliefs about others an unlikely candidate to explain the treatment differences we find. Nonetheless, they may be an important determinant of rule compliance in general: Understanding the causal effect of beliefs about others on the decision to comply with *Rule:Give* and *Rule:Don't*, respectively, may help us to better understand the overall pattern of choices observed in the experiment.

In a regression of beliefs on behavior, beliefs are very likely to be endogenous, i.e., correlated with the error term. In the case of rule compliance, for

[23]Figure 6 plots the distribution of these beliefs when pooling all four treatments. Beliefs in each individual treatment follow very much the same distribution.

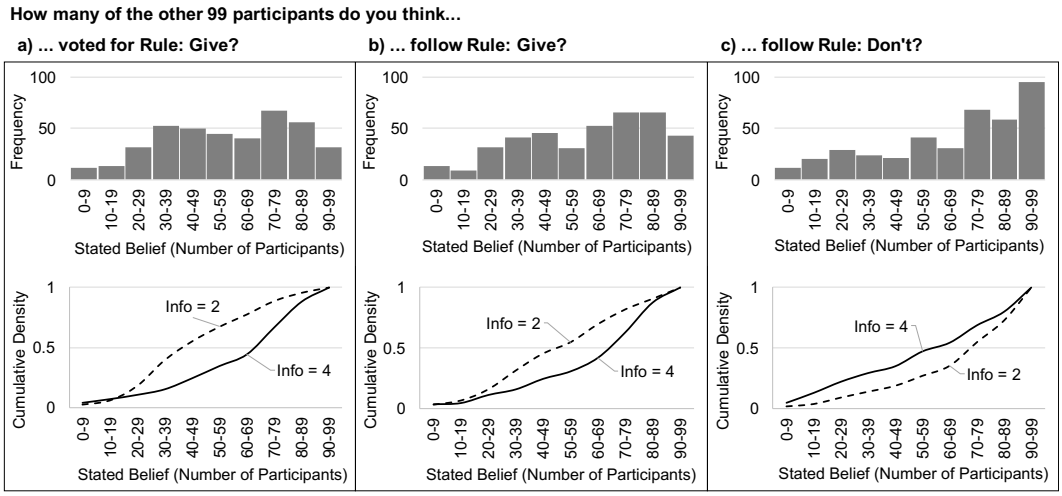


Figure 6: Beliefs about the choices of other participants (data from all treatments pooled, N=400). Top: Frequency of beliefs by answer bracket. Bottom: Cumulative density of answers among subjects having received *info*= 2 and *info*= 4, respectively.

example, attitudes about how one "ought" to behave (injunctive social norms) will most likely affect both how an individual behaves herself and what the individual believes about how others will behave (see also the discussion in Costa-Gomes, Huck and Weizsäcker, 2014). Likewise, other unobserved individual characteristics can lead to an omitted variable bias. To overcome the endogeneity issue and to estimate a causal effect of beliefs on behavior, we use variable $info_i \in \{2, 4\}$ as an instrument for beliefs. Variable $info_i$ records whether, at the end of round 1, individual i was i.i.d. randomly shown a sample in which four out of five subjects chose to give in the dictator game ($info_i = 4$) or, alternatively, a sample in which two out of five subjects chose to give ($info_i = 2$). As this is the only information that participants receive about the behavior of others throughout the entire experiment, $info_i$ is very likely to have a strong effect on subjects' beliefs about the distribution of pro-social types in the population. Figure 6 (bottom panel) confirms this intuition: Subjects who randomly received $info_i = 4$ have consistently higher beliefs about the number of other subjects (a) voting for or (b) complying with *Rule:Give*, as well as consistently lower beliefs about (c) the number of other subjects complying with *Rule:Don't*.

	(a) ***Rule:Give***					(b) ***Rule:Don't***				
	$E_i(Comply_{-i})$	$Comply_i = 1$				$E_i(Comply_{-i})$	$Comply_i = 1$			
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	2SLS	OLS	OLS	OLS	OLS	2SLS	OLS	OLS
$info_i$	.13			-.04	-.04	-.13			-.11	-.09
	(.000)			(.256)	(.316)	(.000)			(.009)	(.044)
$E_i(Comply_{-i})$		.46	-.32				.51	.87		
		(.000)	(.299)				(.000)	(.008)		
Malpractice	-.02	-.10	-.11	-.11	-.10	.01	-.02	-.02	-.01	-.05
	(.735)	(.009)	(.012)	(.009)	(.017)	(.631)	(.741)	(.694)	(.901)	(.367)
Constant	.50	.31	.74	.58	.30	.79	.57	.30	.99	.59
	(.000)	(.000)	(.000)	(.000)	(.070)	(.000)	(.000)	(.230)	(.000)	(.003)
Control for $Type_i$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Add. Controls					Yes					Yes
Observations	400	400	400	400	375	400	400	400	400	375

p-values in parentheses.

Table 4: The role of others in guiding behavior. $E_i(Comply_{-i})$ is individual i's belief about the share of other participants complying with the rule. *Malpractice* = 1 if individual i is in treatment *T_Pay4Vote*, *T_MoneyOffer* or *T_ExcludePoor*. IV regressions are 2SLS with $E_i(Comply_{-i})$ being instrumented by $1.[info_i = 4]$. Control for $Type_i$ includes $Give_i|NoRule$, $Vote_i$, and $(Give_i|NoRule) \times Vote_i$. Additional controls in (5) are: $Female_i$, $Risk\_Seeking_i$, $Betrayal\_Aversion_i$, $Western_i$, $Student_i$, $UGrad_i$, number of mistakes in control questions, factor variables measuring political and social values in questionnaire, as well as *Big Five* personality test measures. $Female_i$ and $Risk\_Seeking_i$ (answer on 11-point Likert-scale to "Are you a person who is generally willing to take risks (10) or do you try to avoid taking risks (0)?") are weakly significant for compliance with *Rule:Give* (.08 and .02, respectively, $p < .10$). $Betrayal\_Aversion_i$ (answer on 11-point Likert-scale to "Do you think that most people would try to take advantage of you if they got the chance (10), or would they try to be fair (0)?") is highly significant for compliance with *Rule:Don't* (.04, $p < .01$). All other demographic and questionnaire controls are insignificant.

Table 4 presents the results of an instrumental variable approach to estimating the role of beliefs about others' behavior in guiding a subject's own choices under *Rule:Give* (panel a) and *Rule:Don't* (panel b). The main covariate of interest in this analysis is $E_i(Comply_{-i})$, which is the share of the 99 other participants whom individual i believes to comply with *Rule:Give* or *Rule:Don't*, respectively.[24] Columns (1) in Table 4 present the results of OLS regressions on $E_i(Comply_{-i})$, using $info_i$, a binary variable *Malpractice* (equal to one if individual i is a subject in treatment *T_Pay4Vote*, *T_MoneyOffer* or *T_ExcludePoor*, zero otherwise), and type controls $(Give_i|NoRule) \times Vote_i$ as covariates. The large and highly significant coefficients on $info_i$ confirm the observation from Figure 6 that variable $info_i$ is a powerful instrument to assess the causal effect of beliefs on behavior under both rules.

Columns (2) report results of OLS regressions of $E_i(Comply_{-i})$ on compliance with *Rule:Give* (panel a) and with *Rule:Don't* (panel b), respectively. The strong and highly significant coefficients on $E_i(Comply_{-i})$ show that beliefs about the behavior of others and individual compliance decisions are highly correlated. To identify the causal effect of beliefs on behavior, we use an IV (2SLS) estimator with $info_i$ instrumenting for $E_i(Comply_{-i})$ in columns (3). Columns (4) and (5) present variations on the same scheme: Columns (4) show the result of an OLS regression using $info_i$ directly as an explanatory variable instead of using it as an instrument for $E_i(Comply_{-i})$. This way, we control for *any* systematic dependency between individual behavior and beliefs about the share of pro-social agents in the population that are shifted by $info_i$. Columns (5) include individual characteristics and questionnaire answers as controls. The following result summarizes our findings:

Result 3 (Beliefs about others only affect compliance with *Rule:Don't*)**.** *Variance in subjects' beliefs about the rule compliance of others cannot explain the negative effect of interventions Pay4Vote, MoneyOffer, and ExcludePoor on*

[24] We ask subjects to state their belief about the *number* of compliant others in their treatment. The response of individual i identifies a bracket, $E_i(\#Compliers_{-i}) \in \{0\text{-}9, 10\text{-}19, ..., 90\text{-}99\}$. $E_i(Comply_{-i})$ is the median of this bracket divided by 99. For example, if $E_i(\#Compliers_{-i}) = 40\text{-}49$, then the median is 44.5 and $E_i(Comply_{-i}) = 44.5/99 \approx 0.45$.

compliance with Rule:Give. Moreover, there is no evidence that beliefs about others' compliance causally affect compliance with Rule:Give. A subject's compliance with Rule:Don't, on the other hand, is strongly and positively affected by higher beliefs about the rule following of others.

Support. Two-sample Kolmogorov-Smirnov tests cannot reject the equality of belief distributions across treatments regarding the number of other subjects who vote for *Rule:Give* (smallest p-value is $p = .468$), comply with *Rule:Give* (smallest p-value is $p = .813$), or comply with *Rule:Don't* (smallest p-value is $p = .699$). In line with these results, variable *Malpractice* is insignificant in an OLS regression on $E_i(Comply_{-i})$, both for *Rule:Give* and for *Rule:Don't*, see Table 4, columns (1). Also, variance in $E_i(Comply_{-i})$ cannot explain the negative effect of interventions *Pay4Vote*, *MoneyOffer*, and *ExcludePoor* on compliance with *Rule:Give*: Irrespective of whether one includes beliefs directly as a control (Table 4, column (2)) or via instrument $info_i$ (column (3)), *Malpractice* is identified to have virtually the same average treatment effect (ATE) on rule compliance as in Table 3. That is, it reduces compliance with *Rule:Give* by approximately 10 percentage points.

Regarding rule compliance in general, Table 4 column (3) shows that beliefs about the rule compliance of others causally impact compliance with *Rule:Don't* but do not affect compliance with *Rule:Give*. Specifically, using $info_i$ as an instrument for $E_i(Comply_{-i})$, a 1 percentage point increase in $E_i(Comply_{-i})$ is estimated to increase the probability of individual i to comply with *Rule:Don't* by 0.87 percentage points ($p < 0.01$). Accounting for this effect, no other explanatory variable is significant at the 5 percent level. For compliance with *Rule:Give*, on the other hand, the effect of $E_i(Comply_{-i})$ (when instrumented with $info_i$) is insignificant. Our results are robust to using $info_i$ directly as an explanatory variable (columns 4 of Table 4) and to including a battery of individual characteristics and questionnaire answers as controls (columns 5).

3.2.2 Lost Votes and Beliefs about Outcome Bias

While treatments *T_Pay4Vote*, *T_MoneyOffer*, and *T_ExcludePoor* differ in the particular form of electoral malpractice, they have in common that due to the intervention many votes are not counted or not counted for the rule the individual originally preferred. In the beginning of this section, we observed that a substantial fraction of participants are excluded from having their vote count due to the intervention in each treatment (35%, 39% and 50%, see binary variable $Lost\_Vote_i$ in Table 2 and Figure 7 panel (a)). If between-treatment differences in rule compliance vary with $Lost\_Vote_i$, this can be an indication that part of the malpractice effect we see can be explained by subjects disregarding rules that were elected without their *personal* vote being accounted for.

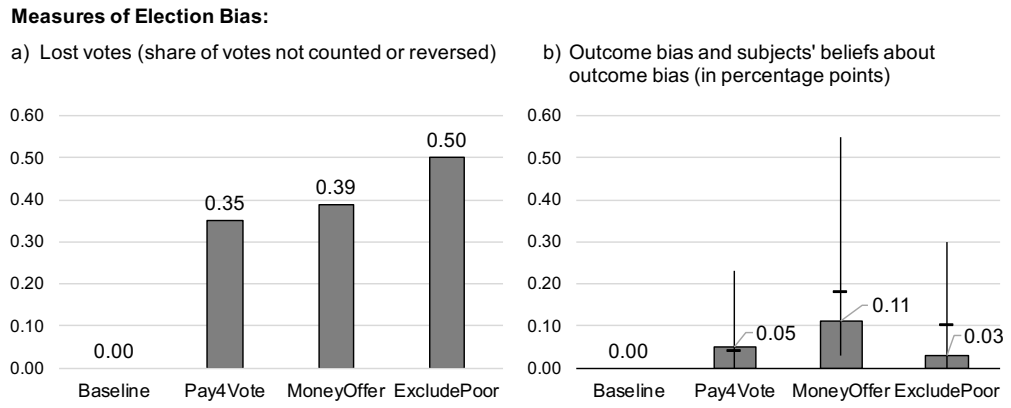


Figure 7: Measures of Election Bias. Panel a): Share of votes not counted or reversed due to interventions *Pay4Vote*, *MoneyOffer*, *ExcludePoor*. Panel b): Outcome bias (percentage point change in vote shares due to intervention, bar plot) and subjects' beliefs about outcome bias (10th to 90th percentile with median, whisker plot).

Intuitively, the exclusion or manipulation of votes can lead to vote shares being shifted relative to a standard majority vote without interventions. The absolute shift in vote shares in our treatments, which we call *Outcome_Bias*, is minor (5 (*T_Pay4Vote*), 11 (*T_Bribe*) and 3 percentage points (*T_ExcludePoor*), respectively, see figure 7 panel (b)) and is never critical in

shifting the voting outcome to the other rule. Because subjects are not informed about how other participants behave during the experiment or how many votes were lost due to the intervention, however, individuals' *beliefs* about the outcome bias may vary. Figure 7 panel (b) plots the median and the 10th to 90th percentile of beliefs about this bias for each of our treatments.[25] A relatively large proportion of subjects expresses beliefs implying that they expect vote shares to shift by more than 10 percentage points (26%, 70%, and 53%, respectively). We can exploit the variance in $E_i[Outcome\_Bias]$ to explore in how far beliefs about the referendum's overall representativeness may explain the shift in rule compliance observed across our treatments.

$Lost\_Vote_i$ and $E_i[Outcome\_Bias]$ thus form our two measures of (perceived) election bias. Can these two measures explain the differences in compliance with *Rule:Give* between treatments? To explore this idea, we now move to analyzing whether the variance in $Lost\_Vote_i$ and $E_i[Outcome\_Bias]$ can account for the variance in rule compliance with *Rule:Give*. Table 5 presents results from OLS regressions of binary treatment variables and controls on $Comply_i|Rule{:}Give$, to which we successively add $Lost\_Vote_i$ (column (2)) and $E_i[Outcome\_Bias]$ (column (3)) as additional explanatory variables; column (4) includes both. We also run analyses of variance (ANOVA) to learn more about the share of variance in treatment effects that is captured by variance in $Lost\_Vote_i$ and $E_i[Outcome\_Bias]$.

[25]Note that to avoid responses that are influenced by social desirability, we do not ask subjects to directly report their beliefs about a potential outcome bias. Instead, we compute $E_i[Outcome\_Bias]$ from elicited beliefs regarding the share of subjects accepting to pay for their vote ($T\_Pay4Vote$), the share of subjects accepting the monetary offer (T_Bribe), or the voting behavior among "poor" and "rich" subjects ($T\_ExcludePoor$). In particular, we calculate individual i's belief about the outcome bias as $E_i[Outcome\_Bias]$

$$:= \begin{cases} 0 & \text{if } i \text{ is in } T\_Baseline, \\ \left| \dfrac{E_i[Accept\_Pay_j | Vote_j = 1] E_i[Vote_j]}{E_i[Accept\_Pay_j]} \right| & \text{if } i \text{ is in } T\_Pay4Vote, \\ \big| E_i[Accept\_MoneyOffer_j | Vote_j = 1] E_i[Vote_j] & \\ \quad + E_i[Accept\_MoneyOffer_j | Vote_j = 0](1 - E_i[Vote_j]) \big| & \text{if } i \text{ is in } T\_MoneyOffer, \\ |E_i[Vote_j | Income_j > 40\text{K}] - E_i[Vote_j]| & \text{if } i \text{ is in } T\_ExcludePoor. \end{cases}$$

	$Comply_i$|**Rule:Give**			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
$Lost\_Vote_i$		-.11		-.10
		(0.011)		(.014)
$E_i[Outcome\_Bias]$			-.34	-.33
			(.005)	(.007)
Pay4Vote	-.11	-.07	-.08	-.05
	(.031)	(.175)	(.099)	(.357)
MoneyOffer	-.12	-.08	-.04	.00
	(.016)	(.150)	(.530)	(.955)
ExcludePoor	-.09	-.04	-.06	-.01
	(.064)	(.487)	(.270)	(.923)
Constant	.58	.59	.58	.59
	(.000)	(.000)	(.000)	(.000)
Control for $Type_i$, $info_i$	Yes	Yes	Yes	Yes
Observations	400	400	400	400

p-values in parentheses.

Table 5: Explaining treatment variance in compliance with *Rule:Give* with variance in $Lost\_Vote_i \in \{0,1\}$ and with variance in subjects' beliefs about outcome bias $E_i[Outcome\_Bias] \in [0,1]$. Controls are: $Give_i|NoRule$, $Vote_i$, $(Give_i|NoRule) \times Vote_i$, and $info_i$.

Analyzing Table 5 and ANOVA results, we find:

Result 4 (Explanatory power of lost votes and beliefs about outcome bias.)**.** *Subjects whose (original) vote is not counted and subjects who hold the belief that the referendum is not representative drive the decline in compliance with Rule:Give in treatments T_Pay4Vote, T_MoneyOffer, and T_ExcludePoor.*

Support. Table 5 shows that the addition of $Lost\_Vote_i$ (column (2)), $E_i[Outcome\_Bias]$ (column (3)), or both (column (4)) as explanatory variables for compliance with *Rule:Give* considerably lowers the explanatory power of binary treatment variables for treatments *T_Pay4Vote*, *T_MoneyOffer*, and *T_ExcludePoor*. Column (1) reproduces our main finding that all three forms of malpractice (*T_Pay4Vote*, *T_MoneyOffer*, and *T_ExcludePoor*) significantly

reduce compliance with *Rule:Give* by roughly 10 percentage points. Including just one of the two variables in the regression (columns (2) and (3)) lowers the estimated coefficients on treatment variables to roughly one third to two thirds of their original effect. Including both variables simultaneously (column (4)) leads to the average residual effects of the treatment variables being further reduced to an estimated residual effect of -.05 ($p = 0.36$) for *T_Pay4Vote* and effects close to zero for the other two treatments. When running the same regression with the pooled treatment indicator *Malpractice* instead of including each treatment separately, the average residual effect amounts to -.03 ($p = .57$). Analysis-of-variance (ANOVA) models suggest that including $Lost\_Vote_i$ and $E_i[Outcome\_Bias]$ as explanatory variables for rule compliance decreases the variance in behavior explained by binary treatment variables by roughly 80%.

Table 5 confirms our expectation that both $Lost\_Vote_i$ and $E_i[Outcome\_Bias]$ are associated with significantly lower rates of rule compliance.[26] Interestingly though, our analysis shows that it is *not* only the subjects losing their vote who show negative responses to interventions *Pay4Vote*, *MoneyOffer*, and *ExcludePoor*: In column (2), residual treatment effects are smaller but remain consistently negative. This suggests that the experience of malpractice alone—even without one's personal vote being directly affected—can negatively affect compliance rates. One important channel through which this experience might lower compliance is the generation of beliefs that the election is no longer representative of voting preferences in the underlying population (see columns (3) and (4)).

[26]Note that the exact coefficients on $Lost\_Vote_i$ should be interpreted with caution: While the decrease in treatment effect size implies that part of the effect *must* be causal (because treatment exposure is random on the individual), the variable is very likely to also capture selection effects in treatments *T_Pay4Vote* and *T_MoneyOffer*. In these two treatments, whether a subject's vote is counted in the ballot is endogenous to her decision of whether to pay the fee or to accept the bribe, respectively. We included *T_ExcludePoor* in our experiment in order to have one treatment with an exogenous exclusion criterion where subjects do not select into "being treated".

4 Discussion

Relation of malpractice and democracy effects. Our paper shows that experimentally induced "malpractice" during the election of a rule governing voluntary social behavior can lead to lower compliance with the elected rule. One way to interpret the result is that malpractice erodes the positive "democracy effect" that earlier studies have found in experimental games in which subjects can vote for similar institutions. Dal Bó, Foster and Putterman (2010), for example, study the effect on cooperation when subjects endogenously—i.e., through voting—choose to convert a prisoners' dilemma game into a coordination game compared to the effect of changing the game exogenously (by random choice of the computer). They find an endogeneity premium in cooperation of roughly 14 percentage points.

How does the "malpractice effect" we find compare to a potential "democracy premium" in the same game? To answer this question, we discuss the results of an additional treatment, *T_Exo*.[27] In this treatment, everything is equal to our baseline treatment except that the rule (*Rule:Give* or *Rule:Don't*) is now exogenously implemented. Before playing the second round of the dictator game, participants are informed that "*(t)he code of conduct will be randomly selected by the computer*" using a "*coin flip*" with equal probabilities. We find that in *T_Exo*, 75% of subjects comply with *Rule:Give* and 70% with *Rule:Don't*. Compared to our baseline treatment, this amounts to a decline in compliance of $-.10$ ($p = .037$) and ± 0.00 (i.e., no significant reduction, $p = .96$), respectively. In other words, measured against the implementation of an exogenous rule, we find a democracy premium of $+10$ percentage points for *Rule:Give* when the rule is selected by standard majority vote, but no such premium for *Rule:Don't*.

Strikingly, the positive democracy premium for *Rule:Give* that we establish against *T_Exo* is virtually identical to the negative malpractice effect we find in treatments *T_Pay4Vote*, *T_MoneyOffer*, and *T_ExcludePoor* (-10, -12 and

[27] The treatment was run with 100 new participants in summer 2018 on *Prolific.ac*. The mean age of participants is 29 years, 53% are female, and 37% are students.

−9 percentage points). At the same time, for *Rule:Don't*, where malpractice on average does not affect compliance rates, *T_Exo* can also not establish a democracy effect. This finding suggests that, indeed, the mechanism by which malpractice erodes compliance is by undermining the democracy premium on domains in which such a premium exists.

Do treatment effects relate to how people perceive violations of democratic principles in the real world? Our experiment establishes how personal disenfranchisement and voters' beliefs about biases in the voting outcome affect subsequent compliance with elected rules of behavior in a neutrally framed experimental setup. With this, we aim to establish a finding that relates to the behavioral consequences of electoral malpractice in real world elections. One way to find suggestive evidence for this relation to behavior in real world institutions is to study whether treatment effects are more likely to be found among participants who place a high value on democratic institutions and who are sensitive to mechanisms (such as bribing and lobbying) that may corrupt these institutions. If this is the case, then the reactions of these participants to instances of real world malpractice can be thought to be governed by similar concerns as their reactions in our experiment. In Table 6, we perform this exercise by exploiting the variation in demographic characteristics in our online subject pool as well as in participants' answers in the post-experimental questionnaire to empirically identify types with a relatively lower or higher value for—or expectation of—democratic procedures.[28]

Table 6 demonstrates that interventions *Pay4Vote*, *MoneyOffer*, and *ExcludePoor* tend to produce treatment effects of larger magnitude and higher statistical significance among participants who have more experience with democratic institutions (1,2), among participants who self-identify as placing

[28]Recall that the questionnaire is sent to subjects using a different researcher profile and visual design more than two weeks after they have taken part in the experiment, making spillovers from our treatments to the questionnaire answers highly unlikely. Indeed, we find that the probability for a subject to be identified as "High" or "Low" in Table 6 does not significantly depend on the treatment to which the subject was assigned. There is only one exception: In column (3), a subject is more likely to be identified as "High Dem_Importance=1" if she participated in treatment *T_Pay4Vote*.

	$Comply_i$|**Rule:Give**													
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Western		High Dem_Country		High Dem_Importance		High Dem_Voice		Low Just_Bribes		Low Just_Lobby		Low Just_Money	
	1	0	1	0	1	0	1	0	1	0	1	0	1	0
Pay4Vote	-.14	-.08	-.17	-.09	-.13	-.06	-.12	-.15	-.16	-.06	-.17	-.09	-.14	-.03
	(.022)	(.373)	(.057)	(.293)	(.086)	(.479)	(.142)	(.066)	(.026)	(.425)	(.040)	(.222)	(.069)	(.702)
MoneyOffer	-.12	-.12	-.22	-.10	-.18	-.00	-.19	-.11	-.22	-.02	-.18	-.07	-.19	-.07
	(.054)	(.155)	(.016)	(.208)	(.018)	(.994)	(.018)	(.160)	(.001)	(.754)	(.025)	(.317)	(.008)	(.405)
ExcludePoor	-.11	-.01	-.19	-.11	-.08	-.09	-.19	.-04	-.13	-.05	-.12	-.06	-.11	-.06
	(.051)	(.921)	(.034)	(.172)	(.275)	(.307)	(.023)	(.581)	(.065)	(.464)	(.132)	(.375)	(.117)	(.457)
Constant	.57	.63	.63	.52	.54	.56	.51	.58	.59	.54	.63	.50	.61	.59
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	272	128	139	156	183	129	140	184	214	170	168	179	189	142

p-values in parentheses

Table 6: Treatment effects on compliance with *Rule:Give* (OLS estimates) by nationality and by questionnaire responses to the following questions: (2): "*How democratic do you think your country is overall?*" (median = 7/10), (3): "*How important is it for you to live in a country that is governed democratically*?" (median = 9/10); (4): "*How important is it for you to personally express your voice when it comes to political decision making?*" (median = 8/10); (5), (6), and (7): "*Please indicate to what extent you think the following actions can be justified:*" (5) "*Accepting a bribe in the course of one's duties.*" (median = 0/10), (6) "*Lobbying politicians to influence legislation.*" (median = 3/10), (7) "*Influencing the actions of people by giving them money.*" (median = 2/10). Except for column (5), "High" and "Low" identifies subjects with answers strictly above or strictly below the median (subjects with answers on the median are not included in regressions). In column (5), "Low" identifies subjects with answer = 0 (median) and "High" subjects with answers > 0. Controls are: $Give_i|NoRule$, $Vote_i$, $(Give_i|NoRule) \times Vote_i$ and $info_i$.

high value on democratic decision-making processes (3,4), and, finally, among subjects who believe that it is never justifiable to offer or take a bribe, or to lobby politicians (5,6,7). Column (5) provides maybe the strongest support for our claim: Those who indicate a very high sensitivity to bribery in the real world also react very sensitively to electoral malpractice in our experiment, the strongest negative effect being found in treatment *T_MoneyOffer*. Overall, the observations in Table 6 suggest that, indeed, our findings in the (context-free) online experiment relate to psychological domains that are also relevant in corresponding real-world decision making.

Discussion of behavioral mechanisms. Our results in section 3.2 suggest that procedural concerns about the inclusiveness and unbiasedness of the election procedure might be underlying the decline in compliance observed for *Rule:Give*. This finding resonates with theories of "legitimate authority" (e.g., Weber, 1978; Tyler, 2006; Akerlof, 2017), which suggest that if a rule has come into force by a fair procedure, "*people feel that they ought to defer [its] decisions and rules, following them voluntarily out of obligation rather than out of fear of punishment or anticipation of reward.*" (Tyler, 2006, p.375). More generally, it connects to earlier literature in psychology and behavioral economics which suggests that procedural aspects of decision making affect preferences directly (Tyler, 1990; Frey, Benz and Stutzer, 2004; Cappelen et al., 2013; Bartling, Fehr and Herz, 2014, among others). In particular, we confirm the idea that people seem to care about the "fairness" of decision making processes (see, e.g., Tyler, 1990; Frey, Benz and Stutzer, 2004; Cappelen et al., 2013) as well as about personally partaking in them (see, e.g., Bonin, Jones and Putterman, 1993; Bardhan, 2000; Bartling, Fehr and Herz, 2014) in our particular setup. In line with our findings, the previously established "democracy effect" in Dal Bó, Foster and Putterman (2010) (see, in particular, p.2222f) also does not seem to work via differences in informational content (of the election) and strategic motives, but rather by the appeal of the endogenous institution itself.

Interestingly, we find that interventions that undermine democratic election procedures do not affect rule compliance *per-se*: While compliance with

the "egalitarian" rule promoting pro-social behavior (*Rule:Give*) deteriorates strongly, average compliance with *Rule:Don't* is largely unaffected by electoral malpractice. We also find an important asymmetry in the mechanisms that drive behavior under each of the two rules. In the case of *Rule:Don't*, compliance is strongly driven by peer-effects; in the case of *Rule:Give*, beliefs about the rule compliance of others do not play a significant role. We thus speculate that democracy effects as well as malpractice effects might not be effective in domains where peer effects drive compliance choices but rather occur primarily in situations where intrinsic motivation to follow rules and procedural concerns play a role. Whether this speculation holds true in a more general sense and outside of our experimental setup will need to be uncovered by future research.

Relation to wider literature. The relationship of corruption and "rule compliance" in the wider sense is also the main theme of Fisman and Miguel (2007), DeBacker, Heim and Tran (2015), and Gächter and Schulz (2016). These papers study how the level of corruption in the home country of an individual determines voluntary compliance with *exogenously* given rules. Fisman and Miguel (2007) show that UN diplomats from high-corruption countries are less likely to voluntarily follow New York City parking rules than diplomats from low-corruption countries. DeBacker, Heim and Tran (2015) document a similar effect on the tax compliance among foreign-owned corporations in the U.S. Finally, Gächter and Schulz (2016) find that even in entirely neutrally framed lab experiments, corruption-indices on country level can predict the level of honesty observed among participants. Note that in all of these cases—in contrast to the choice situation in our experiment—rules are exogenous in the sense that the decision makers whose behavior is studied did not take part (and never meant to take part) in selecting the rules they are supposed to follow. It is exciting, then, to ask why it is that "corruption corrupts". Extrapolating from the findings of our study, two possible mechanisms come to mind. The first is that corruption might have behavioral spillover effects: If an individual is exposed to corrupt authorities for a long time—for example, if she grows up in a country where corruption is deeply rooted in the political

system—norms of non-compliance may be internalized and drive behavior also in unrelated situations. The second is that—maybe due to a different value system being in place in high-corruption countries compared to low-corruption countries (see, e.g., Enke, 2018; Schulz et al., 2018)—the exogenous rule-giver (say, U.S. authorities or the experimenter) is regarded a non-legitimate authority and therefore, his rules are simply not "ought" to be followed. Further exploring and disentangling these two possibilities is a fascinating task for future research.

Giving in dictator games and donation decisions have been shown to be sensitive to nudge interventions such as defaults (for donations see, e.g., Altmann et al. forthcoming; Damgaard and Gravert 2018; for dictator games, e.g., Krupka and Weber 2009). The non-binding rules in our experiment do not affect the available choice options or payoffs and in this respect, can be regarded a particular case of nudges. We show that a democratic vote can increase the efficiency of such types of nudges as more subjects are willing to follow the action they promote. Likewise, recent experimental studies (see, e.g., Krupka and Weber, 2013) demonstrate the importance of social norms for explaining variation in giving across dictator games. While inconclusive in this respect, our analysis suggests that *Rule:Give* may evoke an injunctive pro-social norm and that the normative appeal of this rule is undermined by the 'corrupt' interventions during the majority vote. Further testing this idea might be an exciting avenue for future research. We speculate that such an effect can be of great importance for the ability of laws and social rules to manifest pro-social norms in society. Laws where compliance might be difficult or impossible to enforce as well as nudge policies with no material consequences both (partly) work via their normative appeal.[29] Our analysis shows that this power might be weakened if the procedure that implements the respective policy is perceived as manipulated or "unfair".

[29]This also relates to the discussion around the expressive function of the law; for example (Cooter, 1998, 586) discusses that laws "can change the individual values of rational people".

5 Conclusion

In this paper, we demonstrated how introducing a voting fee, offering subjects money to reverse their vote, or excluding low-income voters from the ballot during a referendum causally impact subsequent compliance with elected rules of behavior in a (modified) dictator game. We find strong and significant reductions in voluntary compliance with elected rules that ask subjects to behave pro-socially compared to a democratic majority vote. In contrast, compliance with rules promoting selfish behavior is generally high but not sensitive to our interventions. There is no (causal) evidence for the treatment effects being driven by subjects' beliefs about other participants' compliance behavior. Rather, the behavior of subjects who are excluded from the ballot and of those who believe the voting outcome to not be representative due to the respective intervention can explain differences in average compliance rates with *Rule:Give* across treatments.

Overall, the experimental results presented in this paper imply that the positive behavioral effects of democratic procedures that earlier studies have established (for example, Frey, 1997; Tyran and Feld, 2006; Ertan, Page and Putterman, 2009; Sutter, Haigner and Kocher, 2010; Dal Bó, Foster and Putterman, 2010) may be sensitive to the manipulation of votes. We see this study as a first step towards understanding the effects of electoral malpractice on behavior for democratically elected institutions; more research is needed to draw general conclusions. We chose to study rule compliance in the domain of redistribution for its important economic and social role. Extending the analysis to other domains such as cheating and tax evasion, as well as to other forms of centralized and de-centralized manipulation (such as ballot box stuffing and subject-to-subject bribes), will allow us to establish results about compliance with social rules in general.

References

Akerlof, Robert. 2017. "The Importance of Legitimacy." *The World Bank*

Economic Review, 30: 157–165.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. 2017. "Cutthroat Capitalism Versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" HCEO Working Paper 2017-003.

Altmann, Steffen, Armin Falk, Paul Heidhues, Rajshri Jayaraman, and Marrit Teirlinck. forthcoming. "Defaults and Donations: Evidence from a Field Experiment." *Review of Economics and Statistics*.

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447–1458.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal*, 100(401): 464–477.

Bardhan, Pranab. 2000. "Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communitiesin South India." *Economic Development and Cultural Change*, 48(4): 847–865.

Bartling, Björn, Ernst Fehr, and Holger Herz. 2014. "The Intrinsic Value of Decision Rights." *Econometrica*, 82(6): 2005–2039.

Bénabou, Roland, and Jean Tirole. 2012. "Laws and Norms." IZA Discussion Paper No. 6290.

Berman, Eli, Michael Callen, Clark Gibson, and James D. Long. 2014. "Election Fairness and Government Legitimacy in Afghanistan." NBER Working Paper No. 19949.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 8.

Besley, Timothy, Anders Jensen, and Torsten Persson. 2015. "Norms, Enforcement, and Tax Evasion." CEPR Working paper No. DP10372.

Black, Sandra E., and Lisa M. Lynch. 2001. "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity." *Review of Economics and Statistics*, 83(3): 434–445.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–193.

Bonin, John P., Derek C. Jones, and Louis Putterman. 1993. "Theoretical and Empirical Studies of Producer Cooperatives: Will Ever the Twain Meet?" *Journal of Economic Literature*, 31(3): 1290–1320.

Brock, J.Michelle, Andreas Lange, and Erkut Y. Ozbay. 2013. "Dictating the Risk -Experimental Evidence on Giving in Risky Environments." *American Economic Review*, 103(1): 415–437.

Brusco, Valeria, Marcelo Nazareno, and Susan Carol Stokes. 2004. "Vote Buying in Argentina." *Latin American Research Review*, 39(2): 66–88.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review*, 97(3): 818–827.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden. 2013. "Just Luck: An Experimental Study of Risk-Taking and Fairness." *American Economic Review*, 103(4): 1398–1413.

Cooter, Robert. 1998. "Expressive law and economics." *The Journal of Legal Studies*, 27(S2): 585–607.

Costa-Gomes, Miguel A, Steffen Huck, and Georg Weizsäcker. 2014. "Beliefs and Actions in the Trust game: Creating Instrumental Variables to Estimate the Causal Effect." *Games and Economic Behavior*, 88: 298–309.

Cottrell, David, Michael C. Herron, and Sean J. Westwood. 2018. "An Exploration of Donald Trump's Allegations of Massive Voter Fraud in the 2016 General Election." *Electoral Studies*, 51: 123–142.

Dal Bó, Pedro. 2014. "Experimental Evidence on the Workings of Democratic Institutions." In *Institutions, Property Rights, and Economic Growth: The Legacy of Douglass North.* , ed. Sebastian Gallani and Itai Sened. New York:Cambridge University Press.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100: 2205–2229.

Damgaard, Mette Trier, and Christina Gravert. 2018. "The hidden costs of nudging: Experimental evidence from reminders in fundraising." *Journal of Public Economics*, 157: 15–26.

De Alth, Shelley. 2009. "ID at the Polls: Assessing the Impact of Recent State Voter ID Laws on Voter Turnout." *Harvard Law and Policy Review*, 3(1): 185–202.

DeBacker, Jason, Bradley T Heim, and Anh Tran. 2015. "Importing corruption culture from overseas: Evidence from corporate tax evasion in the United States." *Journal of Financial Economics*, 117(1): 122–138.

Engel, Christoph. 2011. "Dictator Games: A Meta Study." *Experimental Economics*, 14(4): 583–610.

Enikolopov, Ruben, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov. 2013. "Field experiment estimate of electoral fraud in Russian parliamentary elections." *Proceedings of the National Academy of Sciences*, 110(2): 448–452.

Enke, Benjamin. 2018. "Kinship, Cooperation, and the Evolution of Moral Systems." Working Paper (This Version: October 5, 2018).

Ertan, Arhan, Talbot Page, and Louis Putterman. 2009. "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem." *European Economic Review*, 53(5): 495–511.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 817–868.

Feld, Lars P, and Jean-Robert Tyran. 2002. "Tax evasion and voting: An experimental analysis." *Kyklos*, 55(2): 197–221.

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2013. "Lies in Disguise—An Experimental Study on Cheating." *Journal of the European Economic Association*, 11: 525–547.

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets." *Journal of Political Economy*, 115(6): 1020–1048.

Fjeldstad, Odd-Helge, Cecilia Kagoma, Ephraim Mdee, and Ingrid Hoem Sjursen. 2018. "The Customer is King: Evidence on VAT Compliance in Tanzania."

Freundt, Jana, and Andreas Lange. 2017. "On the Determinants of Giving under Risk." *Journal of Economic Behavior & Organization*, 142: 24–31.

Frey, Bruno S. 1997. "A Constitution For Knaves Crowds Out Civic Virtues." *The Economic Journal*, 107(443): 1043–1053.

Frey, Bruno S., Matthias Benz, and Alois Stutzer. 2004. "Introducing Procedural Utility: Not Only What, But Also How Matters." *Journal of Institutional and Theoretical Economics*, 160: 377–401.

Gächter, Simon, and Jonathan F. Schulz. 2016. "Intrinsic Honesty and the Prevalence of Rule Violations Across Societies." *Nature*, 531: 496–499.

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet De Jonge, Carlos Meléndez, Javier Osorio, and David W Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science*, 56(1): 202–217.

Gosling, Samuel D, Peter J Rentfrow, and William B Swann. 2003. "A Very Brief Measure of the Big-Five Personality Domains." *Journal of Research in Personality*, 37(6): 504–528.

Grossman, Guy, and Delia Baldassarri. 2012. "The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda." *American Journal of Political Science*, 56(4): 964–985.

Krupka, Erin, and Roberto A. Weber. 2009. "The focusing and informational effects of norms on pro–social behavior." *Journal of Economic Psychology*, 30: 307–320.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying Social Norms Using Simple Coordination games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Manza, Jeff, and Christopher Uggen. 2008. *Locked Out: Felon Disenfranchisement and American Democracy.* Oxford University Press.

Norris, Pippa. 2014. *Why Electoral Integrity Matters.* Cambridge University Press.

Pommerehne, Werner W, and Hannelore Weck-Hannemann. 1996. "Tax rates, tax administration and income tax evasion in Switzerland." *Public Choice*, 88(1-2): 161–170.

Saito, Kota. 2013. "Social Preferences under Risk: Equality of Opportunity versus Equality of Outcome." *American Economic Review*, 103(7): 3084–3101.

Schlag, Karl H., and James Tremewan. 2016. "Simple Belief Elicitation." mimeo.

Schmitz, Carsten, et al. 2012. "LimeSurvey: An Open Source Survey Tool." LimeSurvey Project Hamburg, Germany. URL http://www. limesurvey. org.

Schulz, Jonathan, Duman Bahrami-Rad, Jonathan Beauchamp, and Joseph Henrich. 2018. "The Origins of WEIRD Psychology." Working Paper (This Version: June 22, 2018).

Sutter, Matthias, Stefan Haigner, and Martin G Kocher. 2010. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *The Review of Economic Studies*, 77(4): 1540–1566.

Tyler, Tom. 1990. *Why People Obey Rules.* Yale University Press.

Tyler, Tom R. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology*, 57: 375–400.

Tyran, Jean-Robert, and Lars P. Feld. 2006. "Achieving Compliance when Legal Sanctions are Non-Deterrent." *The Scandinavian Journal of Economics*, 108(1): 135–156.

UK Electoral Commision. 2018. "Analysis of Cases of Alleged Electoral Fraud in the UK in 2017." Technical report (Summary of data recorded by police forces).

Weber, Max. 1978. *Economy and Society.* Berkeley: University of California Press.

WVS. 2014. "World Values Survey Wave 6 2010-2014 Official Aggregate v.20150418." World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: Asep/JDS, Madrid SPAIN.

Zwick, Thomas. 2004. "Employee Participation and Productivity." *Labour Economics*, 11(6): 715–740.

Appendix

A.1 Theoretical Framework

We provide a simple theoretical framework to guide the analysis of giving behavior and compliance rates across treatments. Consider first the decision to give in the absence of a code of conduct. Let $u_i(Give_i)$, $Give_i \in \{0, 1\}$ denote individual i's utility when deciding to give or not give, respectively. Define $\Delta u_i = u_i(Give_i = 1) - u_i(Give_i = 0)$. It follows that

$$(Give_i | NoRule) = 1 \Leftrightarrow \Delta u_i \geq 0.$$

A positive Δu_i may reflect social preferences of individual i such as inequality aversion or "warm glow" utility.[30] Let Δu_i be distributed in the population with cumulative density function $F[\cdot]$. The share of *Givers* in the population is then given by $1 - F[0]$ as illustrated in Figure A.1, panel a), below.

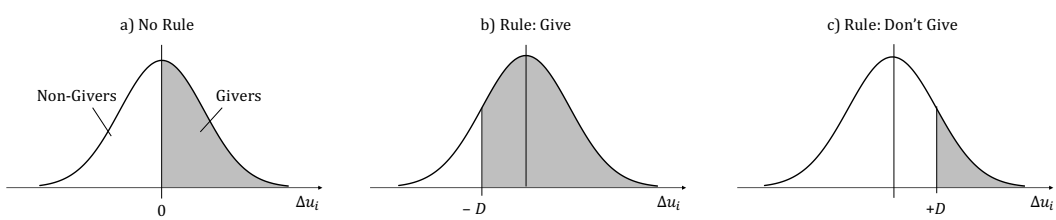


Figure A.1: Theory: Illustration of population shares choosing to give ($Give_i = 1$) and not to give ($Give_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with a standard majority vote (panels b and c).

Consider next the situation with a code of conduct, either *Rule:Give* or *Rule:Don't*. If the code has come into force with a standard majority vote (*T_Baseline*) we assume that it adds fixed utility $D \geq 0$ to the action that

[30] Typical examples in standard settings are Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Andreoni (1989, 1990). Inequity aversion over chances to win a prize has been modeled by, for example, Saito (2013). Experimental evidence showing how prosocial behavior extends to choices over risky payoffs can be found in Brock, Lange and Ozbay (2013) and Freundt and Lange (2017), among others.

is prescribed by the code. This constant can be interpreted as an emotional utility some people derive from following a rule elected by the majority. It follows that

$$\begin{aligned} \text{If } \textit{Malpractice} = 0, \quad & (\textit{Comply}_i | \textit{Rule:Give}) = 1 \Leftrightarrow \Delta u_i \geq -D, \\ \text{and} \quad & (\textit{Comply}_i | \textit{Rule:Don't}) = 1 \Leftrightarrow \Delta u_i < +D. \end{aligned}$$

Compared to the case without a code, the share of subjects choosing to give increases or decreases, see Figure A.1, panel (b) and (c), respectively. Note, importantly, that rules only affect the behavior of those individuals who in the absence of a code would have chosen the opposite action. While *Rule:Give* may convince a *Non-Giver* to give, it will leave the behavior of a *Giver* ($\Delta u_i \geq 0$) unaffected. Similarly, *Rule:Don't* may induce some *Givers* to stop giving, but will not affect the choice of *Non-Givers* ($\Delta u_i < 0$). We assume that electoral malpractice (in our experiment, *Pay4Vote*, *MoneyOffer*, *ExcludePoor*) alters the value some people derive from obeying the elected code. Instead of generating utility D, rule compliance is now associated with a lower utility $D - M$. Constant $M \geq 0$ measures the loss in utility induced by malpractice. As a result, individual i's propensity to comply with the elected rule is reduced. In particular,

$$\begin{aligned} \text{If } \textit{Malpractice} = 1, \quad & (\textit{Comply}_i | \textit{Rule:Give}) = 1 \Leftrightarrow \Delta u_i \geq -(D - M), \\ \text{and} \quad & (\textit{Comply}_i | \textit{Rule:Don't}) = 1 \Leftrightarrow \Delta u_i < +(D - M). \end{aligned}$$

First and foremost, we thus expect that malpractice leads people to revert back to their individually preferred behavior: As M increases, a lower share of *Non-Givers* will follow *Rule:Give*, see Figure A.2, panel b). Similarly, a lower share of *Givers* will be willing to follow *Rule:Don't* (Figure A.2, panel c)). As M becomes sufficiently large such that $D - M$ turns negative, people may even turn against rules that match their individual giving preferences. For example, it is theoretically possible that giving under *Rule:Give* will deteriorate below rates observed in the absence of a code, although such a strong reaction might

be unlikely to be observed in the experiment.

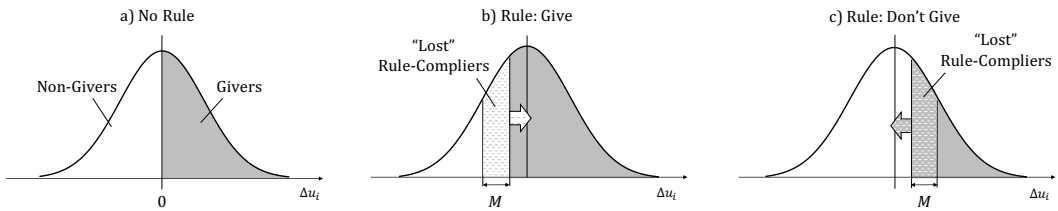


Figure A.2: Theory: Illustration of population shares choosing to give ($Give_i = 1$) and not to give ($Give_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with malpractice during the election (panels b and c).

Voting Behavior. We can extend above theory to yield predictions about voting behavior. Note that in all treatments, subjects vote before interventions take place that may undermine the democratic election. Voting decisions are therefore unbiased by the exposure to a particular treatment. We assume that each subject votes *sincerely* in the sense that she chooses to vote for the outcome that yields her a higher expected utility. Let $U_i[Rule]$ denote i's expected utility given $Rule \in \{Rule{:}Give, Rule{:}Don't\}$. When voting, individual i takes into account how her own giving behavior will be affected by the rule as well as how the behavior of *other* subjects will be affected. Conditional on i not receiving tickets from the computer (which happens with probability 0.5), let $\Delta u(Receive) > 0$ denote the difference in utility between receiving three tickets from another subject and not receiving any tickets. Because the average subject in the population is more likely to give under *Rule:Give* than under *Rule:Don't*, the conditional probability that i will receive three tickets from another subject increases by

$$\Delta F[D] = F[+D] - F[-D]$$

when going from *Rule:Don't* to *Rule:Give*. In our setup, voting behavior depends on the individual's giving preferences $\Delta u_i(Give)$ as follows:

1. *Unconditional Givers:* If $\Delta u_i \geq +D$, individual i will choose $Give_i = 1$

irrespective of the rule. Individual i will then always vote for *Rule:Give*:

$$U_i[\textit{Rule:Give} \,|(\textit{Give}_i|\textit{Rule}) = 1)] \geq U_i[\textit{Rule:Don't} \,|(\textit{Give}_i|\textit{Rule}) = 1]$$

$$0.5 \cdot [u_i(\textit{Give}_i = 1) + D] + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(\textit{Receive}) \geq 0.5 \cdot u_i(\textit{Give}_i = 1)$$

$$\Leftrightarrow \underbrace{\Delta F(D)}_{>0} \geq \underbrace{-\frac{D}{\Delta u(\textit{Receive})}}_{<0}.$$

2. *Unconditional Non-Givers:* If $\Delta u_i < -D$, individual i will choose $\textit{Give}_i = 0$ irrespective of the rule. Individual i will then vote for *Rule:Give* if

$$U_i[\textit{Rule:Give} \,|(\textit{Give}_i|\textit{Rule}) = 0] \geq U_i[\textit{Rule:Don't} \,|(\textit{Give}_i|\textit{Rule}) = 0]$$

$$0.5 \cdot u_i(\textit{Give}_i = 0) + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(\textit{Receive}) \geq 0.5 \cdot [u_i(\textit{Give}_i = 0) + D]$$

$$\Leftrightarrow \; -D \geq -\Delta F(D) \cdot \Delta u(\textit{Receive})$$

$$\Leftrightarrow \Delta F(D) \geq \frac{D}{\Delta u(\textit{Receive})}$$

and otherwise will vote for *Rule:Don't*.

3. *Rule-Followers:* If $-D \leq \Delta u_i < +D$, individual i will choose $\textit{Give}_i = 1$ under *Rule:Give* and $\textit{Give}_i = 0$ under *Rule:Don't*. Individual i will then vote for *Rule:Give* if

$$U_i[\textit{Rule:Give} \,|(\textit{Give}_i|\textit{Rule}) = 1] \geq U_i[\textit{Rule:Don't} \,|(\textit{Give}_i|\textit{Rule}) = 0]$$

$$0.5 \cdot [u_i(\textit{Give}_i = 1) + D] + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(\textit{Receive}) \geq 0.5 \cdot [u_i(\textit{Give}_i = 0) + D]$$

$$\Leftrightarrow \; \Delta u_i \geq -\Delta F(D) \cdot \Delta u(\textit{Receive})$$

$$\Leftrightarrow \; \Delta F(D) \geq -\frac{\Delta u_i}{\Delta u(\textit{Receive})},$$

and otherwise will vote for *Rule:Don't*. Note that this implies that *Givers* ($\Delta u_i \geq 0$) always vote for *Rule:Give*, while Non-Givers ($\Delta u_i < 0$) do the

same if and only if $\Delta F(D)$ is sufficiently large.

We can see that there is a monotonic relation between $\Delta u_i(Give)$ and the tendency to vote for *Rule:Give*. *Givers* always vote for *Rule:Give*. This is true for both, unconditional givers and rule-followers. *Non-Givers*, on the other hand, only vote for *Rule:Give* if they expect that rules have sufficiently large effect on the giving behavior of others. Otherwise, they vote for *Rule:Don't*. If $\Delta F[D]$ is close to zero, all *Non-Givers* vote for *Rule:Don't*. This case is illustrated in Figure A.3, panel a). Increasing $\Delta F[D]$ shifts voting preferences of non-givers in favor of *Rule:Give*. This first affects rule-following *Non-Givers* who indeed would choose to give under the pro-social rule, i.e., those individuals who satisfy $-D \leq \Delta u_i(Give) < 0$, see Figure A.3, panel (b). Only once $\Delta F(D) \geq \frac{D}{\Delta u(Receive)}$, also unconditional non-givers (and thus, all individuals) vote for *Rule:Give*, see Figure A.3, panel c).

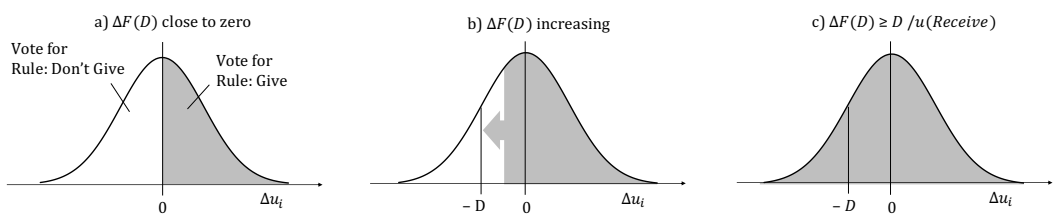


Figure A.3: Theory: Share of Population voting for Rule: Give

A.2 Type-level analysis

(a) All treatments: n by $Type_i$

	$Give_i \mid NoRule$		
$Vote_i$	0	1	Σ
Rule:Don't	92	17	109
Rule:Give	63	228	291
Σ	155	245	**400**

(b) *T_Baseline*: Share of n complying with...

	...**Rule:Give**			...**Rule:Don't**		
	$Give_i \mid NoRule$			$Give_i \mid NoRule$		
$Vote_i$	0	1	w.avg.	0	1	w.avg.
Rule:Don't	.57	.50	.56	.96	.63	.91
Rule:Give	.80	1	.96	1	.51	.62
w.avg.	.66	.97	**.85**	.98	.58	**.70**

(c) Treatment Effects (vs. *T_Baseline*):

		...**Rule:Give**			...**Rule:Don't**		
		$Give_i \mid NoRule$			$Give_i \mid NoRule$		
	$Vote_i$	0	1	w.avg.	0	1	w.avg.
T_Pay4Vote	*Rule:Don't*	-.18	.50	-.07	-.05	-.63	-.14
		(.14)	(.42)	(.14)	(.07)	(.30)	(.08)
	Rule:Give	-.35	-.04	-.10	-.10	-.07	-.08
		(.16)	(.03)	(.04)	(.08)	(.10)	(.08)
	w.avg.	-.25	.00	**-.10**	-.07	-.11	**-.10**
		(.11)	(.04)	(.05)	(.05)	(.09)	(.06)
T_MoneyOffer	*Rule:Don't*	-.01	-.17	.-03	-.09	.38	-.02
		(.16)	(.36)	(.14)	(.08)	(.26)	(.08)
	Rule:Give	-.57	-.04	-.16	-.15	.02	-.02
		(.18)	(.03)	(.05)	(.09)	(.09)	(.08)
	w.avg.	-.24	-.05	**-.12**	-.12	.04	**-.02**
		(.12)	(.04)	(.05)	(.06)	(.09)	(.06)
T_ExcludePoor	*Rule:Don't*	-.13	.00	-.11	.00	.38	.06
		(.14)	(.33)	(.13)	(.07)	(.23)	(.07)
	Rule:Give	-.33	-.02	-.09	.00	.08	.06
		(.17)	(.03)	(.04)	(.09)	(.10)	(.08)
	w.avg.	-.21	-.02	**-.09**	.00	.10	**.06**
		(.11)	(.04)	(.05)	(.05)	(.09)	(.06)
Pooled	*Rule:Don't*	-.12	.06	-.09	-.04	.15	-.01
		(.11)	(.26)	(.10)	(.06)	(.23)	(.06)
	Rule:Give	-.40	-.03	-.11	-.08	.01	-.01
		(.14)	(.03)	(.04)	(.07)	(.08)	(.07)
	w.avg.	-.23	-.03	**-.11**	-.06	.02	**-.01**
		(.09)	(.03)	(.04)	(.04)	(.08)	(.05)

Standard errors in parentheses.

Table A.1: Number of subjects (a), baseline compliance rates (b) and treatment effects by $Type_i = (Give_i|NoRule) \times Vote_i$. Gray cells in (b) and (c) show weighted averages. Weights follow the type-distribution in panel (a). Weighted standard errors calculated assuming normally distributed standard errors (Delta method).

A.3 Questionnaire

Questionnaire: Politics

Overall, there are 15 questions. The first 10 questions relate to your views on politics.

1. In political matters, people talk of "the left" and "the right". On a scale from 0 to 10, where would you place your views, generally speaking?
(Scale: 0 = Left, 10 = Right)

2. On a scale from 0 to 10, how important is it for you to live in a country that is governed democratically?
(Scale: 0 = not at all important, 10 = extremely important)

3. How democratic do you think your country is overall?
(Scale: 0 = not at all democratic, 10 = completely democratic)

4. How important is it for you to personally express your voice when it comes to political decision making?
(Scale: 0 = not at all important, 10 = extremely important)

5. It is important that you pay attention to this study. Please tick number 7 to show that you pay attention. The scale below does not play a role.
(Scale: 0 = not at all important, 10 = very important)

6. On a scale from 0 to 10, where 0 means "no trust at all" and 10 means "very much trust", how much do you personally trust...
...politicians?
...large corporations?
...the results of elections?

7. Please indicate for each of the following actions to what extent you think that action can be justified:
(Scale: 0= can never be justified, 10= can always be justified)

- Violating the instructions of one's superiors (for example at work or school).
- Accepting a bribe in the course of one's duties.
- Cheating on taxes if one has the chance.
- Influencing the actions of people by giving them money.
- Lobbying politicians to influence legislation.

8. Below you find two opposing statements on redistribution. How would you place your personal standpoint between the two statements *(0 means that you agree completely with the statement on the left, 10 means that you agree completely with the statement on the right)*

0:	10:
"The rich have an obligation to subsidize the poor. If necessary, they have to be forced to do so."	"Everybody is responsible for himself. Forcefully taking from the rich to subsidize the poor is theft."

9. Below you find two opposing statements on inequality. How would you place your personal standpoint between the two statements *(0 means that you agree completely with the statement on the left, 10 means that you agree completely with the statement on the right)*

0:	10:
"For a society to be fair, the incomes of all people should be equal."	"There is nothing unfair in having more money than somebody else, no matter how large the difference."

10. When elections take place, do you vote always, usually, or never?

Never Rarely Usually Almost always Always

Questionnaire: General questions

These are the final 5 questions of our study. They concern your views in general and your personality.

1. How do you see yourself: Are you a person who is generally willing to take risks, or do you try to avoid taking risks?
(Scale: 0 = Completely unwilling to take risks, 10 = Very willing to take risks)

2. How much do you agree with the following statement: "Money brings out the worst in people."?
(Scale: 0 = Do not agree at all, 10 = Agree completely)

3. Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?
(Scale: 0 = All people would try to be fair, 10 = All people would try to take advantage of you)

4. Assume that you had the opportunity to take part in the following gamble: There are 100 balls in an urn. Of these balls, 99 are black and 1 is red. One ball is randomly drawn from the urn. If it is red you win 1000 GBP. If it is black you win 0 GBP. What would be the maximal amount of money you would be willing to pay in order to take part?
Would be willing to pay at most... (dropdown menu with answer choices from 0 GBP to 20 GBP in steps of 1)

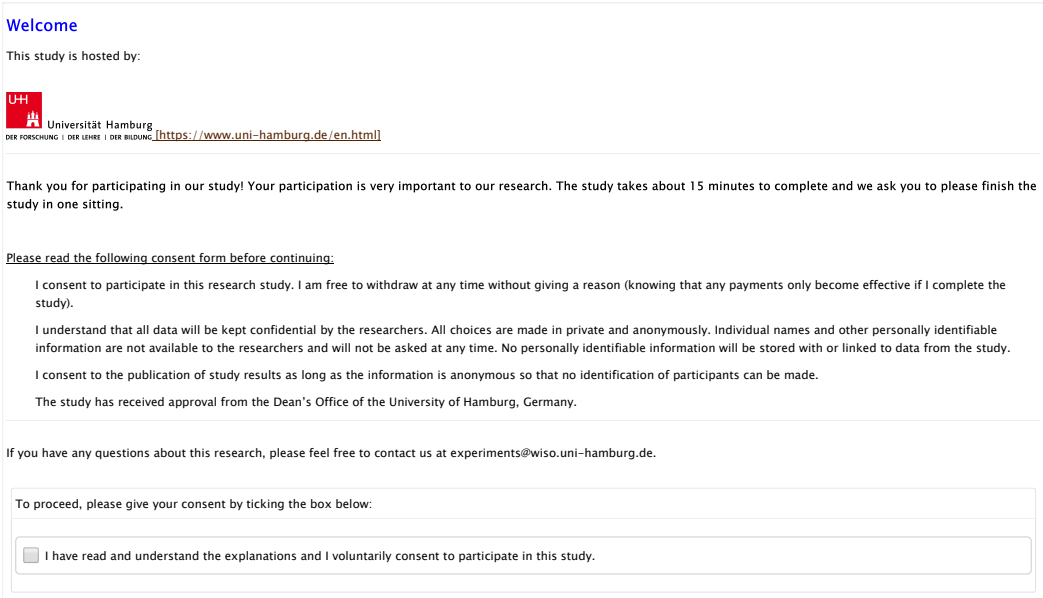
5. Here are a number of personality traits that may or may not apply to you. Please indicate to what extent <u>you agree or disagree</u> that these personality traits apply to you.

Note: You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.
I see myself as...

- Extraverted, enthusiastic (NOT reserved or shy)
- Agreeable, kind (NOT quarrelsome or critical)
- Dependable, self-disciplined (NOT careless or disorganized)
- Emotionally stable, calm (NOT anxious or easily upset/stressed)
- Open to new experiences, creative (NOT conventional)

(Scale: 1 = Disagree strongly, 2 = Disagree moderately, 3 = Disagree a little, 4 = Neither agree nor disagree, 5 = agree a little, 6 = agree moderately, 7 = agree strongly)

A.4 Instructions and Screenshots



Welcome

This study is hosted by:

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG. [https://www.uni-hamburg.de/en.html]

Thank you for participating in our study! Your participation is very important to our research. The study takes about 15 minutes to complete and we ask you to please finish the study in one sitting.

Please read the following consent form before continuing:

I consent to participate in this research study. I am free to withdraw at any time without giving a reason (knowing that any payments only become effective if I complete the study).

I understand that all data will be kept confidential by the researchers. All choices are made in private and anonymously. Individual names and other personally identifiable information are not available to the researchers and will not be asked at any time. No personally identifiable information will be stored with or linked to data from the study.

I consent to the publication of study results as long as the information is anonymous so that no identification of participants can be made.

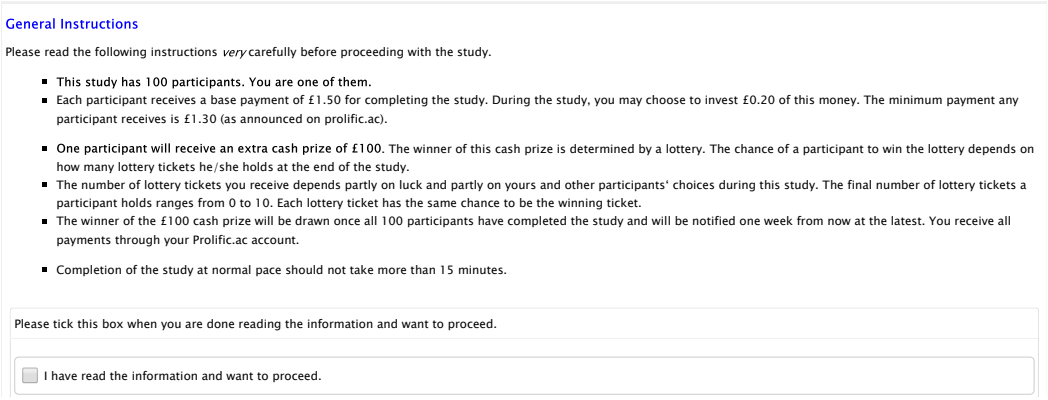
The study has received approval from the Dean's Office of the University of Hamburg, Germany.

If you have any questions about this research, please feel free to contact us at experiments@wiso.uni-hamburg.de.

To proceed, please give your consent by ticking the box below:

☐ I have read and understand the explanations and I voluntarily consent to participate in this study.

Figure A.4: Screenshot: Welcome and Consent Form



General Instructions

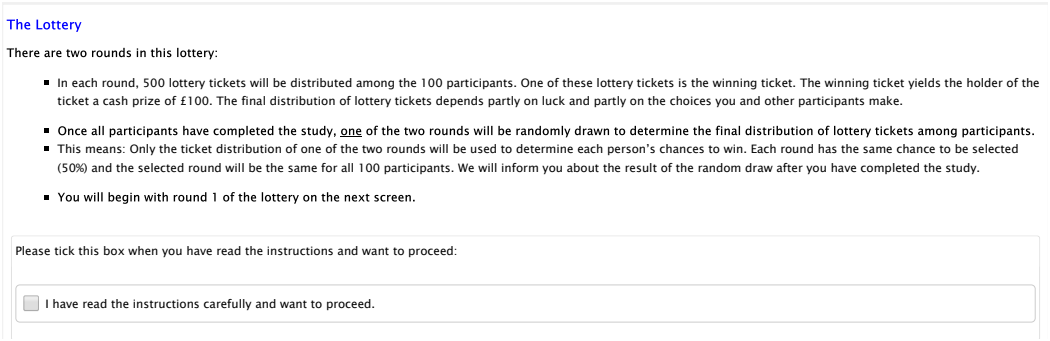
Please read the following instructions *very* carefully before proceeding with the study.

- This study has 100 participants. You are one of them.
- Each participant receives a base payment of £1.50 for completing the study. During the study, you may choose to invest £0.20 of this money. The minimum payment any participant receives is £1.30 (as announced on prolific.ac).
- One participant will receive an extra cash prize of £100. The winner of this cash prize is determined by a lottery. The chance of a participant to win the lottery depends on how many lottery tickets he/she holds at the end of the study.
- The number of lottery tickets you receive depends partly on luck and partly on yours and other participants' choices during this study. The final number of lottery tickets a participant holds ranges from 0 to 10. Each lottery ticket has the same chance to be the winning ticket.
- The winner of the £100 cash prize will be drawn once all 100 participants have completed the study and will be notified one week from now at the latest. You receive all payments through your Prolific.ac account.
- Completion of the study at normal pace should not take more than 15 minutes.

Please tick this box when you are done reading the information and want to proceed.

☐ I have read the information and want to proceed.

Figure A.5: Screenshot: General Instructions ($T\_Pay4Vote$)



The Lottery

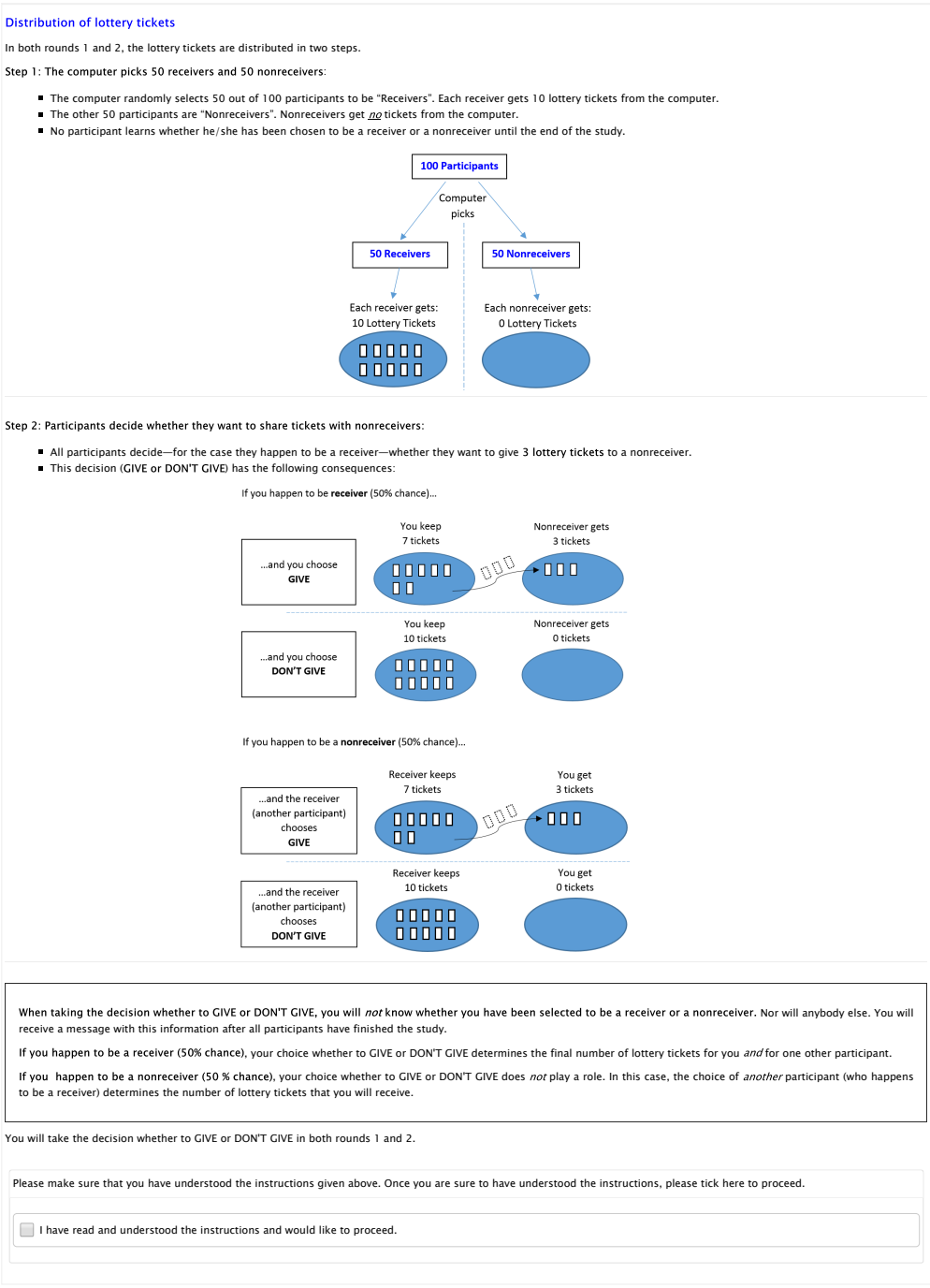
There are two rounds in this lottery:

- In each round, 500 lottery tickets will be distributed among the 100 participants. One of these lottery tickets is the winning ticket. The winning ticket yields the holder of the ticket a cash prize of £100. The final distribution of lottery tickets depends partly on luck and partly on the choices you and other participants make.
- Once all participants have completed the study, one of the two rounds will be randomly drawn to determine the final distribution of lottery tickets among participants.
- This means: Only the ticket distribution of one of the two rounds will be used to determine each person's chances to win. Each round has the same chance to be selected (50%) and the selected round will be the same for all 100 participants. We will inform you about the result of the random draw after you have completed the study.
- You will begin with round 1 of the lottery on the next screen.

Please tick this box when you have read the instructions and want to proceed:

☐ I have read the instructions carefully and want to proceed.

Figure A.6: Screenshot: Instructions about the Lottery



Distribution of lottery tickets

In both rounds 1 and 2, the lottery tickets are distributed in two steps.

Step 1: The computer picks 50 receivers and 50 nonreceivers:

- The computer randomly selects 50 out of 100 participants to be "Receivers". Each receiver gets 10 lottery tickets from the computer.
- The other 50 participants are "Nonreceivers". Nonreceivers get *no* tickets from the computer.
- No participant learns whether he/she has been chosen to be a receiver or a nonreceiver until the end of the study.

Step 2: Participants decide whether they want to share tickets with nonreceivers:

- All participants decide—for the case they happen to be a receiver—whether they want to give 3 lottery tickets to a nonreceiver.
- This decision (GIVE or DON'T GIVE) has the following consequences:

When taking the decision whether to GIVE or DON'T GIVE, you will *not* know whether you have been selected to be a receiver or a nonreceiver. Nor will anybody else. You will receive a message with this information after all participants have finished the study.

If you happen to be a receiver (50% chance), your choice whether to GIVE or DON'T GIVE determines the final number of lottery tickets for you *and* for one other participant.

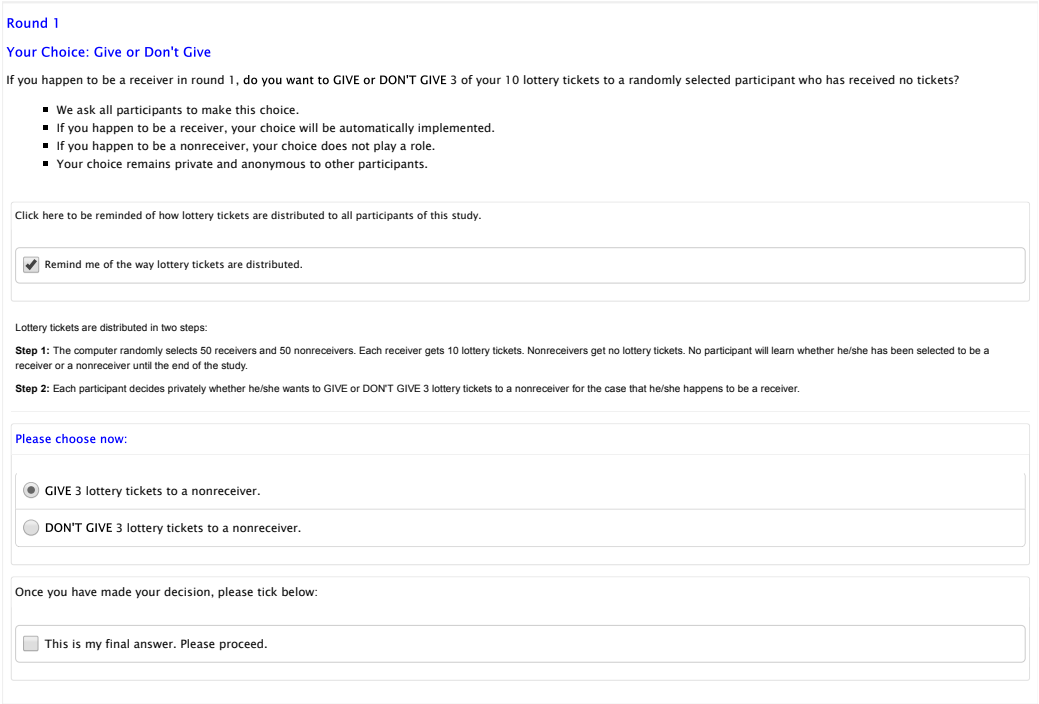
If you happen to be a nonreceiver (50 % chance), your choice whether to GIVE or DON'T GIVE does *not* play a role. In this case, the choice of *another* participant (who happens to be a receiver) determines the number of lottery tickets that you will receive.

You will take the decision whether to GIVE or DON'T GIVE in both rounds 1 and 2.

Please make sure that you have understood the instructions given above. Once you are sure to have understood the instructions, please tick here to proceed.

I have read and understood the instructions and would like to proceed.

Figure A.7: Screenshot: Instructions about the Distribution of Lottery Tickets



Round 1

Your Choice: Give or Don't Give

If you happen to be a receiver in round 1, do you want to GIVE or DON'T GIVE 3 of your 10 lottery tickets to a randomly selected participant who has received no tickets?

- We ask all participants to make this choice.
- If you happen to be a receiver, your choice will be automatically implemented.
- If you happen to be a nonreceiver, your choice does not play a role.
- Your choice remains private and anonymous to other participants.

Click here to be reminded of how lottery tickets are distributed to all participants of this study.

☑ Remind me of the way lottery tickets are distributed.

Lottery tickets are distributed in two steps:

Step 1: The computer randomly selects 50 receivers and 50 nonreceivers. Each receiver gets 10 lottery tickets. Nonreceivers get no lottery tickets. No participant will learn whether he/she has been selected to be a receiver or a nonreceiver until the end of the study.

Step 2: Each participant decides privately whether he/she wants to GIVE or DON'T GIVE 3 lottery tickets to a nonreceiver for the case that he/she happens to be a receiver.

Please choose now:

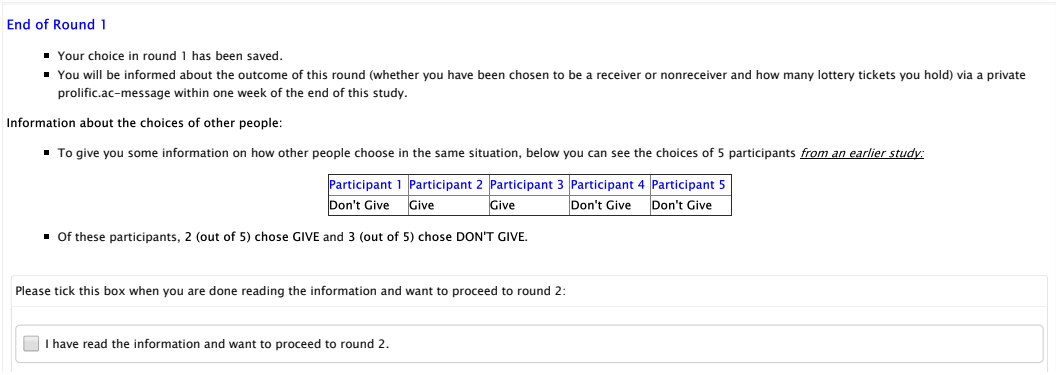
◉ GIVE 3 lottery tickets to a nonreceiver.

○ DON'T GIVE 3 lottery tickets to a nonreceiver.

Once you have made your decision, please tick below:

☐ This is my final answer. Please proceed.

Figure A.8: Screenshot: Choice $Give_i \in \{0, 1\}$ (Round 1)



End of Round 1

- Your choice in round 1 has been saved.
- You will be informed about the outcome of this round (whether you have been chosen to be a receiver or nonreceiver and how many lottery tickets you hold) via a private prolific.ac-message within one week of the end of this study.

Information about the choices of other people:

- To give you some information on how other people choose in the same situation, below you can see the choices of 5 participants *from an earlier study:*

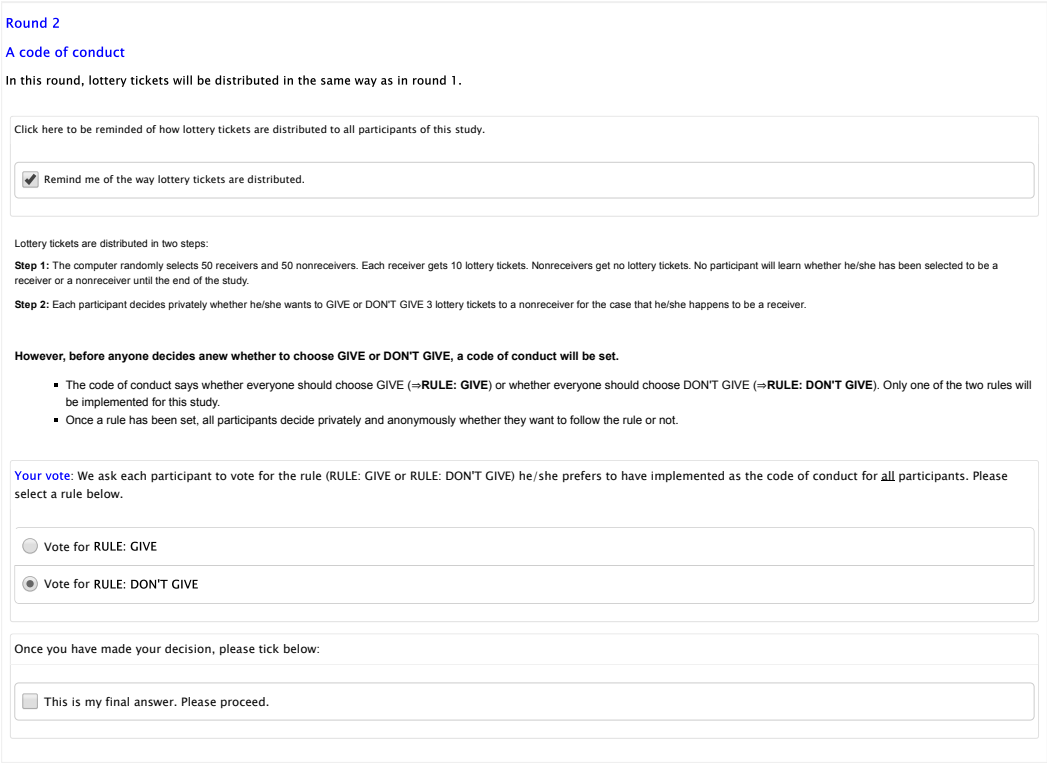
Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Don't Give	Give	Give	Don't Give	Don't Give

- Of these participants, 2 (out of 5) chose GIVE and 3 (out of 5) chose DON'T GIVE.

Please tick this box when you are done reading the information and want to proceed to round 2:

☐ I have read the information and want to proceed to round 2.

Figure A.9: Screenshot: Information $info_i \in \{2, 4\}$ (following Round 1)



Round 2

A code of conduct

In this round, lottery tickets will be distributed in the same way as in round 1.

Click here to be reminded of how lottery tickets are distributed to all participants of this study.

☑ Remind me of the way lottery tickets are distributed.

Lottery tickets are distributed in two steps:

Step 1: The computer randomly selects 50 receivers and 50 nonreceivers. Each receiver gets 10 lottery tickets. Nonreceivers get no lottery tickets. No participant will learn whether he/she has been selected to be a receiver or a nonreceiver until the end of the study.

Step 2: Each participant decides privately whether he/she wants to GIVE or DON'T GIVE 3 lottery tickets to a nonreceiver for the case that he/she happens to be a receiver.

However, before anyone decides anew whether to choose GIVE or DON'T GIVE, a code of conduct will be set.

- The code of conduct says whether everyone should choose GIVE (⇒**RULE: GIVE**) or whether everyone should choose DON'T GIVE (⇒**RULE: DON'T GIVE**). Only one of the two rules will be implemented for this study.
- Once a rule has been set, all participants decide privately and anonymously whether they want to follow the rule or not.

Your vote: We ask each participant to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) he/she prefers to have implemented as the code of conduct for all participants. Please select a rule below.

○ Vote for RULE: GIVE

◉ Vote for RULE: DON'T GIVE

Once you have made your decision, please tick below:

☐ This is my final answer. Please proceed.

Figure A.10: Screenshot: $Vote_i \in \{Rule{:}Give, Rule{:}Don't\}$ (Round 2)

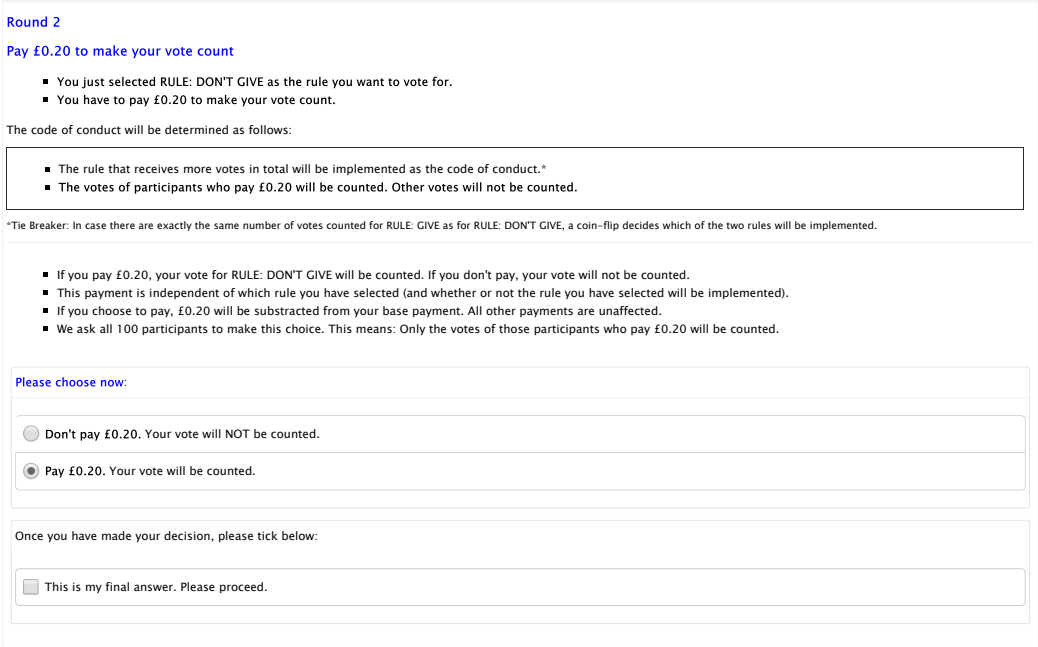


Figure A.11: Screenshot: $Accept\_Pay4Vote \in \{0, 1\}$ (Round 2, $T\_Pay4Vote$)

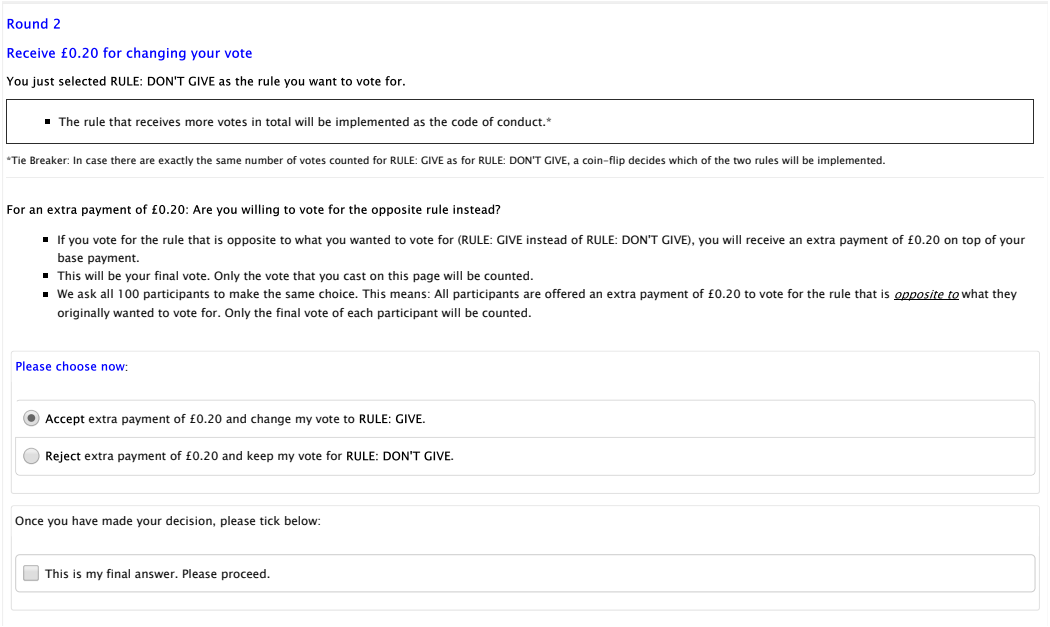


Figure A.12: Screenshot: $Accept\_MoneyOffer \in \{0, 1\}$ (Round 2, $T\_MoneyOffer$)

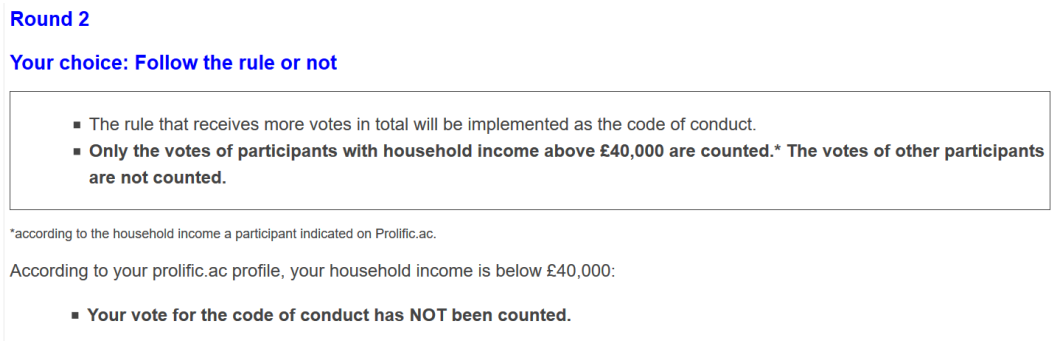


Figure A.13: Screenshot: Information about intervention *Exclude_Poor* (Round 2)

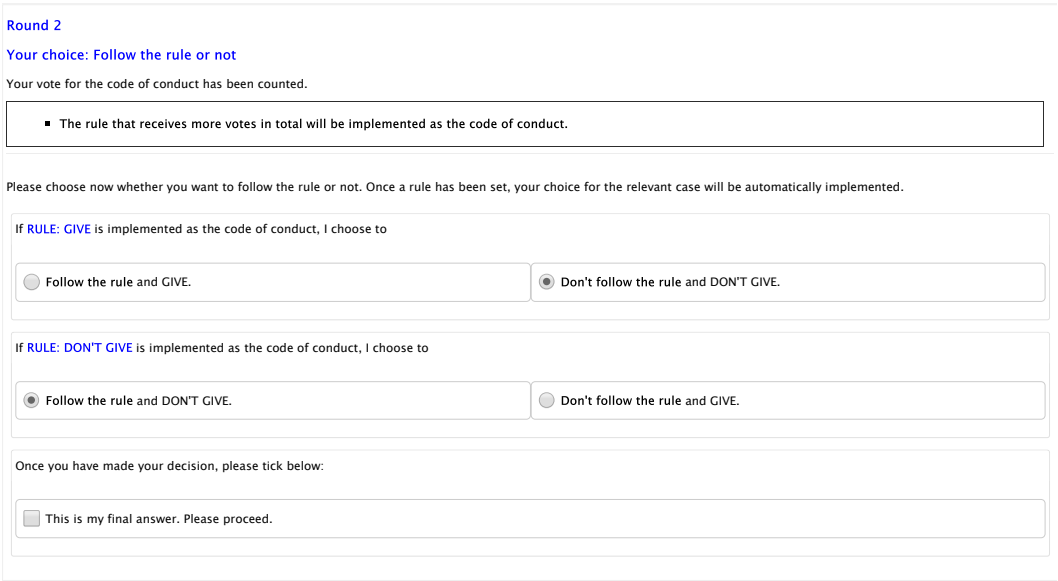


Figure A.14: Screenshot: $Give_i|Rule \in \{0, 1\}$ (Round 2, *T_Baseline*)

Round 2

Your belief about other participants

Your choice has been saved and will be implemented accordingly.

As a final step, we are interested in your belief about the behavior of *other* participants in this round:

- All other participants make the same choices as you just did.
- For each question where your belief about the behavior of other participants is correct, you will receive an extra payment of £0.50 on top of your base payment. In total, you can earn up to £1.50 in extra payment on this page.

Click here to be reminded of how lottery tickets are distributed or of how the code of conduct is determined.

☐ Remind me of how lottery tickets are distributed.

☑ Remind me of how the code of conduct is determined.

How is the code of conduct determined?

• The rule that receives more votes in total will be implemented as the code of conduct.

1. How many of the other participants follow the rule?

a) If RULE: GIVE is implemented as the code of conduct, how many of the other 99 participants do you think follow the rule and GIVE?

0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
○	○	●	○	○	○	○	○	○	○

b) If RULE: DON'T GIVE is implemented as the code of conduct, how many of the other 99 participants do you think follow the rule and DON'T GIVE?

0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
○	○	○	○	○	○	●	○	○	○

2. How do the other participants vote?

Of all other 99 participants, how many do you think have voted for RULE: GIVE to become the code of conduct?

0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
○	○	○	●	○	○	○	○	○	○

Once you have made your decisions, please tick below:

☑ These are my final answers. Please proceed.

Figure A.15: Screenshot: Beliefs about Others (Round 2, *T_Baseline*)

3. How many of the other participants pay £0.20 to make their vote count?

a) Of those participants who voted for **RULE: GIVE**, what share do you think **paid £0.20** to make their vote count?

	% 0-9	% 10-19	% 20-29	% 30-39	% 40-49	% 50-59	% 60-69	% 70-79	% 80-89	% 90-100
	○	○	○	○	○	○	○	○	○	○

b) Of those participants who voted for **RULE: DON'T GIVE**, what share do you think **paid £0.20** to make their vote count?

	% 0-9	% 10-19	% 20-29	% 30-39	% 40-49	% 50-59	% 60-69	% 70-79	% 80-89	% 90-100
	○	○	○	○	○	○	○	○	○	○

Once you have made your decisions, please tick below:

Figure A.16: Screenshot: Beliefs about Intervention (Round 2, *T_Pay4Vote*)