ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Apffelstaedt, Arno; Freundt, Jan

Working Paper

Corrupted votes and rule compliance

ECONtribute Discussion Paper, No. 137

Provided in Cooperation with:
Reinhard Selten Institute (RSI), University of Bonn and University of Cologne

Suggested Citation: Apffelstaedt, Arno; Freundt, Jan (2022) : Corrupted votes and rule compliance, ECONtribute Discussion Paper, No. 137, University of Bonn and University of Cologne, Reinhard Selten Institute (RSI), Bonn and Cologne

This Version is available at:
https://hdl.handle.net/10419/249875

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 137

Corrupted Votes and Rule Compliance

Arno Apffelstaedt Jana Freundt

January 2022 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged.

Corrupted Votes and Rule Compliance*

Arno Apffelstaedt† ⓡ Jana Freundt‡

This Version: 4th February 2020

Abstract

Allegations of voter fraud accompany many real-world elections. How does electoral malpractice affect the acceptance of elected institutions? Using an online experiment in which people distribute income according to majority-elected rules, we show that those who experience vote buying or voter disenfranchisement during the election are subsequently less willing to comply with the rule. On average, the detrimental impact of electoral malpractice on compliance is of the same magnitude as removing the election altogether and imposing a rule exogenously. Our experiment shows how corrupting democratic processes can impact economic behavior and sheds light on the behavioral mechanisms underlying "rule legitimacy".

Keywords: rule compliance, endogenous institutions, corruption, procedural fairness, legitimacy
JEL Codes: D02, D72, D91, C92

*Earlier versions of the paper have circulated under the title "Manipulated Votes and Rule Compliance". We thank Ned Augenblick, Berno Buechel, Alexander W. Cappelen, Holger Herz, 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 funding by the Center for a Sustainable University (KNU) and the Graduate School of Economics and Social Sciences of the University of Hamburg. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC2126/1 – 390838866.

†University of Cologne and ECONtribute. Address: Faculty of Management, Economics, and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50931 Cologne, Germany. Email: apffelstaedt@wiso.uni-koeln.de.

‡University of Fribourg, Department of Economics and University of Pennsylvania, School of Arts and Sciences. Address: University of Fribourg, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland. Email: jana.freundt@unifr.ch.

1 Introduction

People follow rules for different reasons. One reason is the existence of incentive and deterrence mechanisms such as implicit or explicit rewards for compliance, or punishment for non-compliance. Another reason, stressed by legal scholars and political scientists, comes from people accepting the procedure by which the rule came into force as *legitimate*: When the rule setting procedure is seen as being fair, people may change their behavior and follow the rule "*voluntarily out of obligation rather than out of the fear of punishment or anticipation of reward*" (Tyler, 2006, p.375). This paper is about such latter type of rule compliance.

An important source of rule legitimacy are thought to be democratic voting procedures. Consider, for instance, the introduction of a CO_2 tax or a policy that changes the rules of organ donation from an 'opt-in' to an 'opt-out' organ donation system. Intuition suggests that such policies will see higher acceptance and will be voluntarily complied with to a larger extent if people perceive the rule setting mechanism to be participatory and inclusive. Indeed, all but a handful of countries in today's world hold elections or referenda of some kind—often in an attempt to confer legitimacy on a policy addressing some critical political issue on which the electorate is divided (LeDuc, Niemi and Norris, 2014).[1] The extent to how well democratic procedures are implemented, however, varies widely. In many countries, promises of a "free and fair" vote are openly undermined by practices ranging from systematic vote buying to the outright exclusion of social groups, often minorities or poor voters. In other instances, unintentional disenfranchisement or alleged manipulation of parts of the electorate leads people to question the integrity of elections and referenda.[2]

When many people perceive a voting procedure to be "corrupt" or "flawed", legitimacy of the elected outcome may suffer, possibly leading citizens to show lower compliance with elected rules and policies. Suggestive evidence for this claim can be found in survey data, see Figure 1: In countries with higher perceived levels of electoral malpractice (X-axis), the average citizen is significantly more likely to say that it is justifiable to break social rules (Y-axis), ranging from wrongfully claiming government benefits to not paying the fare on public transport.

Causal evidence on whether voting procedures directly affect behavior can be gathered through experiments. However, no causal evidence exists on whether the power of democracy to change behavior is sensitive to electoral malpractice such as vote buying and voter disenfranchisement. In this paper, we investigate this question using a novel experiment. Modelling a typical referendum situation in which the electorate is split between two competing policies, our experiment allows us to systematically investigate whether and due to which psychological mechanisms the willingness

[1] Recent examples include the 2016 Brexit referendum, the 2017 constitutional referendum in Turkey, and the 2019 referendum in Romania about whether to prohibit amnesties and pardons for corruption offenses.

[2] Brusco, Nazareno and Stokes (2004) and Gonzalez-Ocantos et al. (2012) document vote buying schemes in Argentina (2002) and Nicaragua (2008), respectively. Enikolopov et al. (2013) presents data on the extent of electoral fraud during the Russian parliamentary elections of 2011. In the UK and the US, allegations of voter fraud have recently been extensively discussed in the popular press (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).

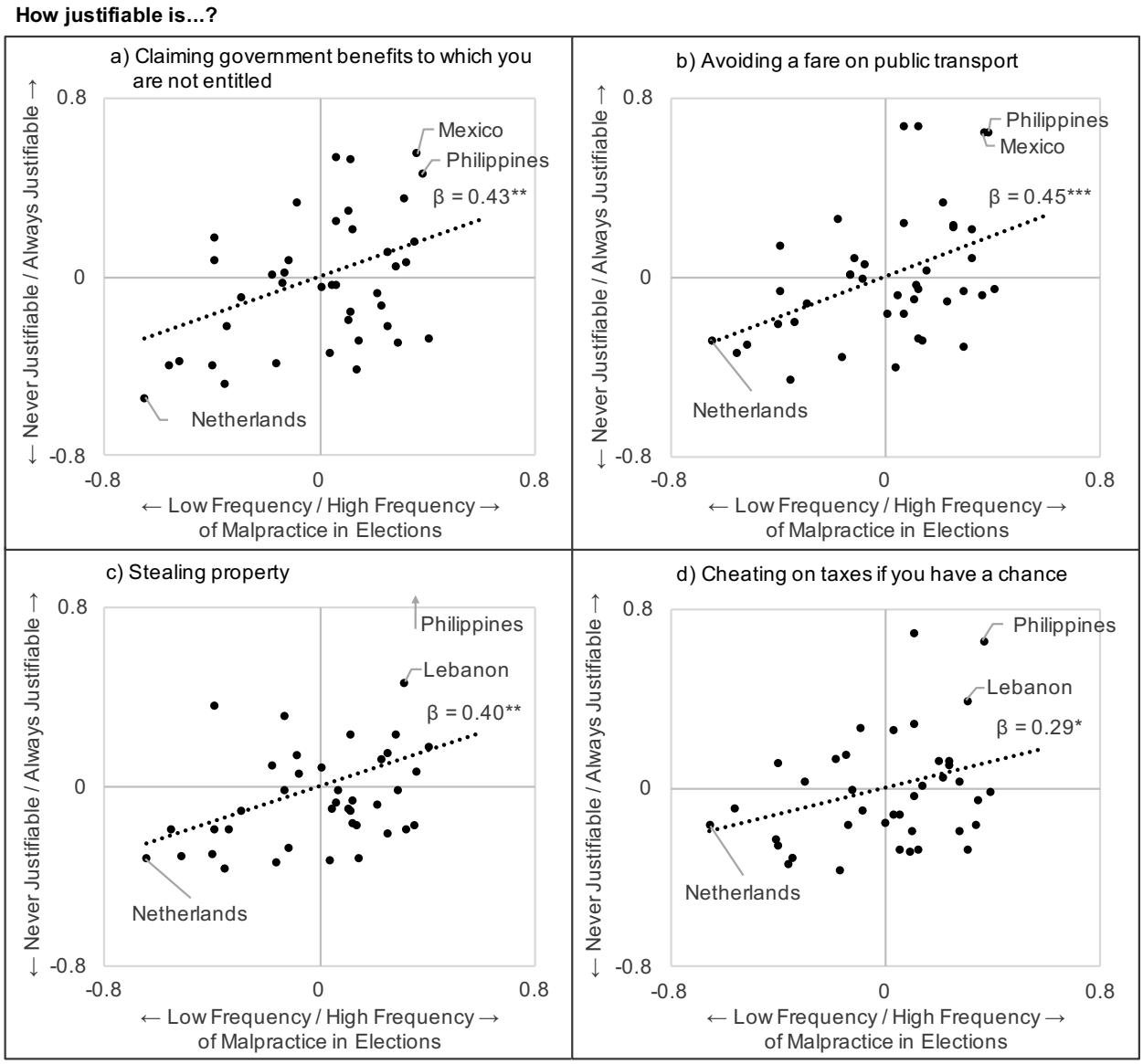


Figure 1: Country-level correlations between citizens' perceived frequency of malpractice in elections and their statements about the justifiability of violating rules and laws (Country averages calculated from the WVS (2014)). Y-axis: Average answers in a country to questions V198-V201 ("*How justifiable 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: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to accept and comply with democratically elected rules may suffer when people perceive the voting procedure to be "corrupt".

The key take away of our study is that the power of an election to increase the acceptance of

a rule can be significantly reduced when democratic voting procedures are tempered with. In fact, we find that electoral malpractice can wipe out the entire democracy premium, meaning that rule compliance after a "corrupt" election is equivalent to imposing the rule exogenously. Voluntary compliance decreases especially among people who are personally excluded from the ballot as well as among people who believe the voting outcome to be biased. This does not mean, however, that malpractice always reduces rule compliance: In our experiment, if people follow a rule for reasons other than its perceived legitimacy (i.e., if there does not exist a democracy premium in the first place), electoral malpractice leaves behavior unaffected.

The setting of our experiment is as follows. In each session, 100 subjects have to decide, each individually, whether to share one's experimental income with another subject who is less well of. Before subjects make that decision, they vote on whether to introduce a policy that asks everyone in the session to voluntarily share (*Rule:Give*) or to introduce a policy that asks everyone to not share (*Rule:Don't*). We measure the strength of the elected rule by its power to convince people to change their behavior relative to a setting without a rule. In the baseline treatment, a majority vote among all 100 subjects selects the rule. With three further treatments, we measure the causal effect of electoral malpractice on rule compliance: In one treatment, we demand that subjects pay for their vote, excluding everyone from the ballot who does not pay. In another, we manipulate votes by paying subjects for reversing their initial vote. In a third, we exclude subjects with a low household income from the ballot.

Our main result is that electoral malpractice drastically reduces the power of democracy to convince people to follow *Rule:Give*, but does not affect the power of *Rule:Don't*. In the baseline treatment, the election of *Rule:Give* has the power to decrease non-giving rates by more than 60%. Malpractice reduces this power by half ($p < .01$). With the help of an additional (fifth) treatment, we show that this "malpractice effect" on compliance with *Rule:Give* is equivalent to the effect of removing the election altogether and imposing the rule exogenously. In other words, electoral malpractice wipes out the democracy premium entirely. For *Rule:Don't*, on the other hand, we find that there is neither a malpractice effect nor a democracy premium: Across all treatments, the power of *Rule:Don't* is strong yet constant (its implementation decreases giving rates by roughly 50%).

To shed light on the psychological reasons why individual behavior may respond to how a rule has been selected, we study two possible mechanisms. We first explore the role of beliefs about how other subjects behave under the same rule. The rationale is as follows. Notice that an election in which parts of the electorate have been excluded (or when their votes have been manipulated) leads to a noisier signal of the modal policy preferences in the population than an unbiased majority vote. Hence, if people care to align their behavior with what others do or value, then a "corrupt" voting procedure may lead to a weaker response to the election result, and thus, to lower individual rule compliance.[3] We explore this idea in our experimental framework by analyzing elicited beliefs

[3]While there are no monetary coordination incentives in our experiment, it is reasonable to assume that some subjects may nonetheless care about aligning their behavior with what others do or value (see, for instance, Bernheim, 1994; Bénabou and Tirole, 2012; Krupka and Weber, 2013).

about the behavior of others across treatments. Using an exogenous shock to these beliefs, we measure their causal effect on the willingness to comply. While we do not find evidence for this idea when studying behavior under *Rule:Give*, compliance with *Rule:Don't*—the rule for which we do *not* find an effect of malpractice–is to a large extent driven by preferences for following the behavior of others.

The second mechanism we investigate are "intrinsic" concerns about the fairness of the voting procedure. In particular, we study whether the effect of electoral malpractice on compliance is associated with (1) subjects who have been personally excluded from taking part in the election and (2) subjects who believe that the voting outcome is biased.[4] Indeed, we find that roughly 80% of the treatment variance under *Rule:Give* is captured with these two variables, suggesting that people intrinsically care about personal participation as well as about the overall unbiasedness of the procedure.

Our experiment is conducted online with subjects from different countries and demographic backgrounds. Using a post-experimental questionnaire, this variance allows us to investigate how treatment effects relate to standpoints on various political issues such as redistribution, corruption, democratic values, and trust in institutions. We find that treatment effects are more significant and of larger magnitude among subjects who live in (relatively) democratic countries and among those who self-report to have stronger concerns for democratic values. This finding indicates that the effect of malpractice we identify in our experimental game relates 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 electoral manipulation.

Related Literature. To our knowledge, this paper is the first to provide causal evidence for the negative effects of electoral malpractice on the acceptance of democratic institutions. We complement earlier research in public and political economics that has provided evidence for the positive effects of democratic compared to exogenously imposed institutions. For instance, Pommerehne and Weck-Hannemann (1996) and Frey (1997) show that tax compliance is higher in Swiss cantons that see more democratic participation. Subsequent experiments, for example by Tyran and Feld (2006), Ertan, Page and Putterman (2009), Sutter, Haigner and Kocher (2010), Grossman and Baldassarri (2012), and Dal Bó, Foster and Putterman (2010), have shown that in social dilemma situations, punishments and rewards work better when endogenously elected rather than exogenously imposed.[5] Note that these experiments compare cooperation rates under an endogenously

[4]Measured as the belief about the difference between the share of votes for a given rule with and without malpractice.

[5]This list of studies is not meant to be exhaustive. See, e.g., Dal Bó (2014) for further studies. A related literature in organizational economics studies the value of "democratic" compared to "autocratic" decision-making mechanisms within firms and organizations. For example, Bonin, Jones and Putterman (1993), Black and Lynch (2001) and Zwick (2004) provide empirical support that higher levels of employee participation are associated with increased worker productivity, leading to potentially large efficiency gains. Similarly, Fehr, Herz and Wilkening (2013) show that giving away decision-rights leads to an under-provision of working effort.

elected versus an exogenously selected institution instead of directly measuring individual rule compliance as we do. Overall, the existing literature suggests that giving citizens decision rights through majority votes can bring important efficiency gains to societies. We show that for such efficiency gains to materialize it matters *how* these institutions are introduced. More specifically, we provide evidence that the positive dividend of democracy is sensitive to interventions in the voting procedure that disenfranchise or manipulate voters. Because our design allows us to isolate and study the effect of endogenous institutions on the *intrinsic* component of preferences better than earlier studies, we also generate new insights into the psychological mechanisms driving democracy effects.

Probably closest to the aim of our study, Dickson, Gordon and Huber (2015) experimentally show that people are more willing to actively help (and less willing to actively hinder) the punishment authority in a public good game if this authority has been elected by a majority vote rather than exogenously imposed. They interpret their finding as showing differences in the perceived legitimacy of the authority.[6] We study the (indirect) behavioral consequences of legitimate procedures that can affect the efficiency of the working of institutions rather than direct expressions of support for an authority.

With this, we add to a different stream of research in psychology and behavioral economics suggesting that procedural aspects of decision making can affect behavior. In particular, studies have shown that people seem to care about the "fairness" of decision-making processes in a more general sense (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). 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. The authors suggest to attribute this increase to a shock in intrinsic motivation; they cannot, however, pin down the exact motives.

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 3.2) and then study the behavioral mechanisms that drive these effects (section 3.3). 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.

[6] A similar approach is followed by Berman et al. (2014). Here, the authors measure the effect of an election fraud intervention in the field on multiple survey measures of attitudes toward government, including the willingness to report insurgent behaviors to security forces.

2 Experimental Design

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

Prediction (Malpractice Effect)**.** *Electoral malpractice lowers voluntary compliance with the elected policy:*

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

Our goal is to design an experiment which can (1) identify a causal effect of malpractice on compliance and that (2) can shed light on the psychological mechanisms driving this effect. Satisfying this goal comes with different requirements for our design.

First, we want to make sure that the effect we measure is a general malpractice effect and not a feature of a specific malpractice intervention. Our experiment for that reason implements three malpractice treatments in order to mimic the variation in corruptive practices in the real world. With this, we can robustly test the hypothesis that—independent of the particular practice—compliance with elected policies will decrease if democratic principles are violated.

Second, in order to identify a causal effect of malpractice, we need to control for possibly unbalanced treatment groups as different people may have different inclinations *ex-ante* to prefer and therefore follow a given policy (see Dal Bó, Foster and Putterman, 2010; Dal Bó, Foster and Kamei, 2019). By eliciting individual giving choices and votes *before* the introduction of each treatment, we are able to control for different distributions of types across treatment groups.

Third, we aim to set up an environment in which people disagree about what is the "right" thing to do and therefore vote for different policies. We achieve this by letting subjects vote on policies in the environment of a (binary) dictator game. Numerous studies show that people differ in their judgements regarding whether income received through luck should be redistributed or not (see, e.g., Cappelen et al., 2007; Almås, Cappelen and Tungodden, 2017). Our design allows us to measure the power of elected rules to change behavior away from what people ex-ante preferred as an action or as a policy.

Finally, there are a few design elements that we require in order to study the psychological mechanism driving behavior. To make sure that we measure *voluntary* compliance with elected policies, we do not implement any form of punishment or reward for certain behavior. That is, subjects are free to choose to follow (or not follow) the elected rule without having to fear any monetary consequences. Because there are no classical coordination incentives in a one-shot dictator game (and no reputation effects),[7] voting mechanisms can then only work by their normative appeal.[8] As outlined in the introduction, the normative appeal of a democratic election may be

[7]From the perspective of standard game theory, treatment effects cannot be driven by people adjusting their behavior to "equilibrium effects" (as, for example, in Dal Bó, Dal Bó and Eyster, 2017).

[8]The possibility to construct a well-defined *behavioral* measure of voluntary compliance is a major advantage of using an experiment. With surveys, researchers have to rely on the self-reported willingness to comply (see, e.g.,

due to (a) people having intrinsic preferences for a fair and unbiased election procedure or due to (b) the election producing a good signal about what other people do and value, making it easier to "do what others do". To be able to shed light on these two mechanisms, we elicit subjects' beliefs about what other subjects do and introduce an exogenous shock to these beliefs in order to estimate a causal effect of these beliefs on behavior.

2.1 The Experiment in Detail

For each session, 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.[9] The lottery is used to 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.[10] 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.[11]

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.

Berman et al., 2014), while in field studies, deterrence mechanisms often interfere with clean measures of intrinsic motives (see, e.g., Fjeldstad et al., 2018). Being able to measure voluntary compliance is not merely a technically desirable feature. In many cases where deterrence mechanisms are in place in the real world, expected punishments are usually not high enough to explain the high levels of compliance observed (e.g., in the case of tax compliance, see Feld and Frey, 2002, p.88). In some instances deterrence might not be feasible to implement, as for the case of littering. In situations with deterrence mechanisms, it is important to understand the intrinsic component of rule compliance in order to properly isolate and understand the effect of punishments on compliance (see, e.g., Dwenger et al., 2016).

[9] 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).

[10] 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.

[11] 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.

All decisions are taken anonymously and in private. The timeline of a session is summarized in Figure 2.

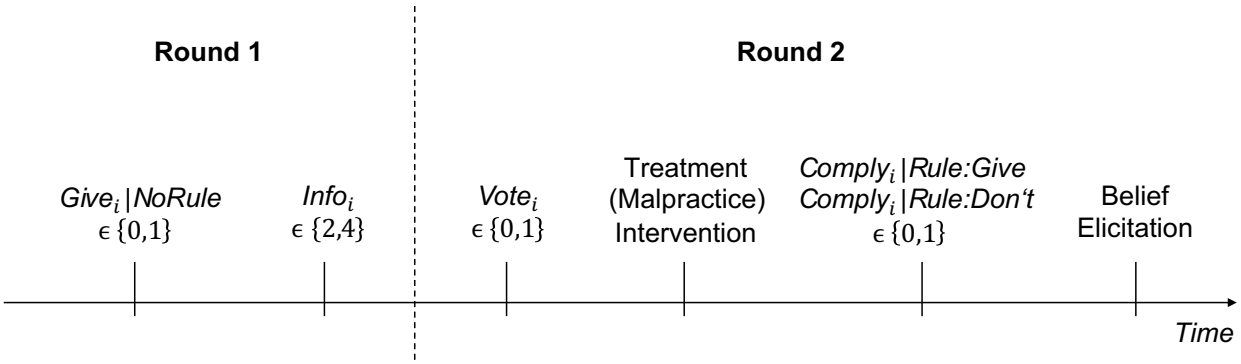


Figure 2: Timeline of experimental session

Round 1. Round 1 consists of two stages: A choice stage and and information stage.

Choice stage. Round 1 implements the dictator game without a rule for behavior being in place. Each subject decides individually whether to give, $(Give_i|NoRule) = 1$, or not give, $(Give_i|NoRule) = 0$. To ease notation, we introduce the following definition:

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

Information stage. After a subject has made her choice in round 1, she is presented with a screen that shows her information on how five other people in "*an earlier study*" (participants in our pilot sessions) decided in the exact same situation. With probability one half we show the subject a pre-selected sample that features two Givers and three Non-Givers ($info_i = 2$), and with probability one half a sample that features four Givers and one Non-Giver ($info_i = 4$). We introduce $info_i \in \{2, 4\}$ in order to generate exogenous variance to the beliefs of a subject about how other subjects will behave in round 2 of the dictator game.

Round 2. Round 2 consists of four stages: A voting stage, a treatment stage, a rule compliance stage, and a belief elicitation stage.

Voting Stage. At the beginning of round 2, subjects are informed that they will shortly play the dictator game again. They are also informed that in this round, a "*code of conduct*" will be implemented for all participants. Each subject is then 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*"). All participants of a lottery decide in one large voting group of 100 subjects on the rule they prefer to have implemented for everyone. With this, subjects are very unlikely to cast a pivotal vote (which would potentially lead to strategic voting considerations) and our

results are thus scalable to larger societies. The decision of the subject in the voting stage is coded $Vote_i \in \{Rule{:}Give, Rule{:}Don't\}$. Subjects are not informed about how other participants voted until after the experiment.

Treatment Stage. Treatments are introduced after the voting stage. We employ a between-subject design making it difficult for subjects to infer our research question. There are four treatments, see Table 3.

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

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.*" In treatment *T_Pay4Vote*, subjects learn that "*only the votes of participants who pay GBP 0.20 will be counted.*" Each subject can decide whether or not to pay. 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.*" Each subject can decide whether or not to accept the offer. 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 learns whether her individual vote has not been counted toward the majority vote.[12] 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*.

Rule Compliance Stage. After the treatment stage, subjects play the dictator game a second time. Each subject 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. Thus, all subjects make the decision whether or not to follow *each* rule conditional on it being elected.[13] These two choices form our measure of rule compliance:

[12] 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.

[13] Eliciting such state-dependent compliance choices has major advantages for us: There is no selection into

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* $(Give_i|Rule{:}Don't) = 0$*.*

Belief Elicitation Stage. At the end of round 2, 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 (income > GBP 40,000) and low income participants (income ≤ GBP 40,000).

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 follow-up-rate is close to 100 percent.[15] The full list of questions can be found in Appendix A.3.

Rule:Give or *Rule:Don't* and the decision whether to give under each rule is made without yet knowing the voting outcome. The latter is important for eliciting beliefs at the end of the experiment. Importantly, having large voter groups of 100 subjects—which we prioritize for external validity and scalability of the results—makes a real-time matching of all votes a practical problem that we avoid with this design choice.

[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.

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

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. 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%). 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. In addition, subjects' unique Prolific-ID allows us to access an extensive set of self-reported socio-demographic data, including gender, nationality and income. Everyone who filled out information on at least gender, nationality and country of birth was eligible to participate.[16] 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.[17] Additional payments are announced during the course of the experiment. Subjects receive all payments and an e-mail with a summary of all outcomes through the online survey platform Prolific.ac within two days after the experiment. For completing the 10 minute post-experimental questionnaire, subjects receive a compensation of GBP 1. Subjects had to give informed consent before they were able to enter the experiment and questionnaire, respectively.

3 Experimental Results

Figure 3 nicely summarizes the key insight of our experiment. We find that the power of rules to change behavior can be strongly and significantly reduced by the presence of electoral malpractice. When implemented by a fair majority vote (*Baseline*), *Rule:Give* (left-hand side) has the power to decrease non-giving rates by more than 60% relative to the situation without a rule. Malpractice reduces this power by nearly half to roughly 30%. Malpractice has no effect on the power of *Rule:Don't* (right-hand side).

We thus confirm our prediction that malpractice in an election can substantially impact compliance decisions, however, only for one type of rule. Why do we observe such an asymmetry and who are the people whose compliance decisions are sensitive to the procedure that implements *Rule:Give*? In the remainder of this section we will provide the results that lead to the above general finding. First, sections 3.1 and 3.2.1 set the stage by summarizing giving and voting behavior as well as baseline rule compliance. In 3.2.2, we explain in detail how we compute treatment effects

[16] In treatment *T_ExludePoor* we additionally required that participants had filled out information on household income.

[17] 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.

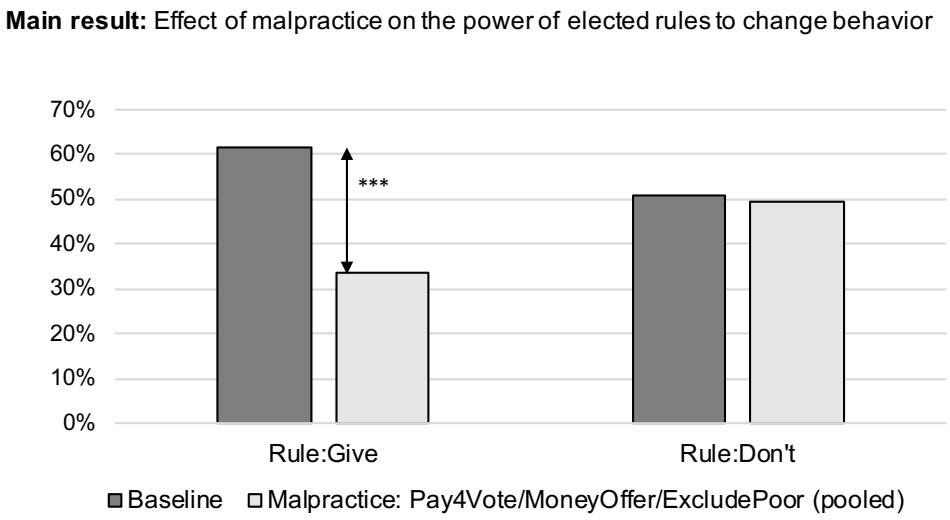


Figure 3: Effect of interventions *Pay4Vote*, *MoneyOffer* and *ExcludePoor* (pooled) on the power of elected rules to change behavior. Bars show decrease (in %) in the share of subjects choosing to not give (give) after the election of *Rule:Give* (*Rule:Don't*) relative to the share of subjects choosing 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.

to then continue analysing the behavioral mechanisms driving these effects in 3.3.

3.1 Setting the Stage

We begin by providing summary statistics of how subjects behave in round 1, how they vote in round 2, and how interventions *Pay4Vote*, *MoneyOffer*, and *ExludePoor* affect the voting process. This information is summarized in Table 2.

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).[18] Voting behavior in round 2 (summarized in the second 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

[18] 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 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.

	Base-line	*Pay 4 Vote*	*Money Offer*	*Exclude Poor*
Round 1				
Share of subjects choosing...				
$(Give_i \mid NoRule) = 1$	.57	.57	.71	.60
Round 2				
Share of...				
initial votes cast = *Rule:Give*	.64	.75	.81	.71
if $(Give_i \mid NoRule) = 0$	.35	.47	.45	.38
if $(Give_i \mid NoRule) = 1$	.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.

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.

3.2 Rule Compliance

Because compliance with either rule likely depends on whether the individual is a *Giver* ($(Give_i \mid NoRule) = 1$) or *Non-Giver* ($(Give_i \mid NoRule) = 0$), 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.[19]

[19] 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., $(Give_i \mid NoRule)$) of the individual.

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 *ExcludePoor* 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, which can hide or exaggerate actual changes in behavior.

In this section, 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 of all treatment effects can be found in table A.1 in Appendix A.2.

3.2.1 Baseline Rule Compliance

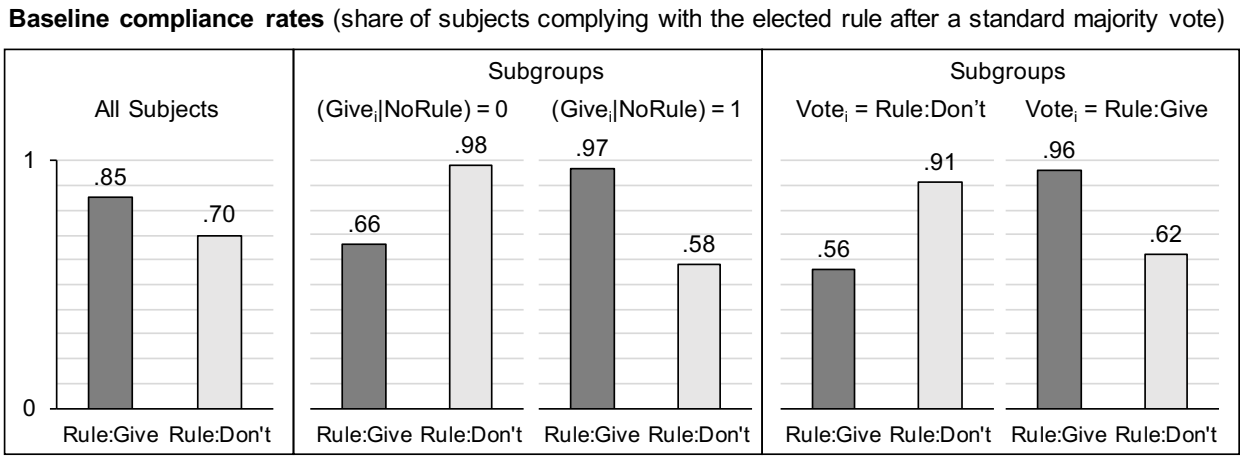


Figure 4: Share of subjects complying with majority-elected rules. Graphs show type-weighted averages. For details see Table A.1 in Appendix A.2.

We observe high compliance with both *Rule:Give* and *Rule:Don't* when rules are selected by a standard majority vote, see Figure 4. As expected, a subject is more likely to follow *Rule:Give* if she is a *Giver* and if she voted for *Rule:Give*. A symmetric observation holds for *Rule:Don't*. The probability with which subjects comply with rules that are opposite to their original choice 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).

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 ($Give_i|NoRule$) is 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 a 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*.[20] 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 ($p < 0.001$).

3.2.2 Treatment Effects

How does malpractice affect compliance with elected rules? Table 3 and Figure 5 report the estimated difference between the share of subjects complying with *Rule:Give* (*Rule:Don't*) after intervention *Pay4Vote*/*MoneyOffer*/*ExcludePoor* and the share of subjects complying with *Rule:Give* (*Rule:Don't*) in the baseline.

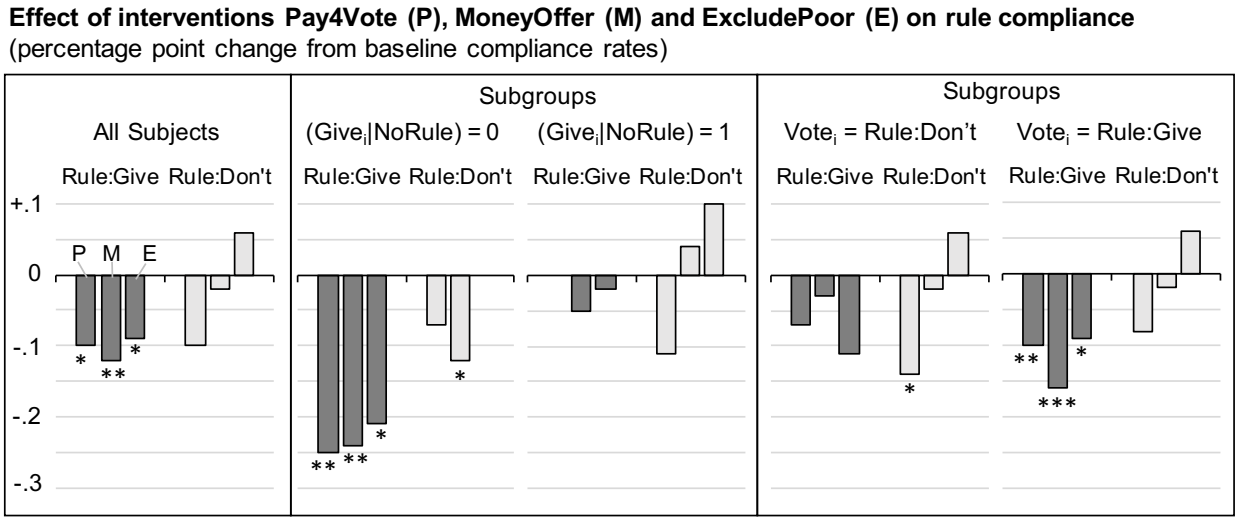


Figure 5: Effect of interventions *Pay4Vote* (P), *MoneyOffer* (M) and *ExcludePoor* (E) on rule compliance. Graphs show type-weighted averages, see Table 3. Stars denote statistically significant differences to the baseline compliance rate: * $p < .1$, ** $p < .05$, *** $p < .01$.

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

[20] 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.

Share of n complying with...	All Subjects	Subgroups $Give_i \mid NoRule = 0$	$= 1$	$Vote_i = Rule:$ *Don't*	*Give*
...**Rule:Give**					
Pay4Vote	-.10**	-.25**	.00	-.07	-.10**
	(.05)	(.11)	(.04)	(.14)	(.04)
MoneyOffer	-.12**	-.24**	-.05	-.03	-.16***
	(.05)	(.12)	(.04)	(.14)	(.05)
ExcludePoor	-.09*	-.21**	-.02	-.11	-.09*
	(.05)	(.11)	(.04)	(.13)	(.04)
Malpractice (Pooled)	-.11***	-.23***	-.03	-.09	-.11***
	(.04)	(.09)	(.03)	(.10)	(.04)
Constant (*T_Baseline*)	.85***	.66***	.97***	.56***	.96***
	(.03)	(.08)	(.03)	(.09)	(.03)
...**Rule:Don't**					
Pay4Vote	-.10	-.07	-.11	-.14*	-.08
	(.06)	(.05)	(.09)	(.08)	(.08)
MoneyOffer	-.02	-.12*	.04	-.02	-.02
	(.06)	(.06)	(.09)	(.08)	(.08)
ExcludePoor	.06	-.00	.10	.05	.06
	(.06)	(.05)	(.09)	(.07)	(.08)
Malpractice (Pooled)	-.01	-.06	.02	-.01	-.01
	(.05)	(.04)	(.08)	(.06)	(.07)
Constant (*T_Baseline*)	.70***	.98***	.58***	.91***	.62***
	(.04)	(.04)	(.07)	(.05)	(.06)
Observations	400	155	245	109	291

Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

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.

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 5). The second column in Table 3 (the second panel in Figure 5, 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.[21] The strongest effect is found for

[21] See Appendix A.1 for a theoretical framework which formalizes this claim.

Non-Givers who voted for *Rule:Give*.While other types show smaller effects, the negative impact on compliance with *Rule:Give* is systematic across the entire sample. Although the nature of the interventions is quite different, their effect on compliance with *Rule:Give* is strikingly similar.

Regarding subjects' compliance with *Rule:Don't*, Table 3 and Figure 5 show smaller, inconsistent, and mostly insignificant treatment effects. Given the systematic changes we observe for the opposite rule, this might be surprising. 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 ExludePoor. We observe smaller, inconsistent, and insignificant effects of the same interventions on compliance with Rule:Don't.*

Support. Using a type-weighting approach (see also Dal Bó, Foster and Kamei, 2019), 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).[22] 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 $-.01$ ($p = 0.823$).[23]

3.3 Understanding Rule Compliance and Treatment Effects

In order to shed light on potential psychological mechanisms underlying the treatment effects we find, we now analyze elicited beliefs about the rule compliance of other participants. With this, we can 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. In section 3.3.2, we 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? And, does compliance vary with beliefs about a potential bias in the voting outcome?

[22] For treatment effects on type-level see Table A.1 in Appendix A.2.

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

3.3.1 Beliefs About the Rule Compliance of Other Subjects

Do subjects follow rules because they want to follow others? While in the dictator game payoffs are not interdependent, subjects may still be inclined to condition their compliance choices on the expected behavior of the 99 other participants in their group, for example due to preferences for conditional cooperation or conformity. Following this conjecture, we study to what extent beliefs about the voting and compliance behavior of other subjects can explain rule compliance in general and treatment differences in particular. Figure 6 displays the frequencies of beliefs (pooled across all treatments) by answer bracket.

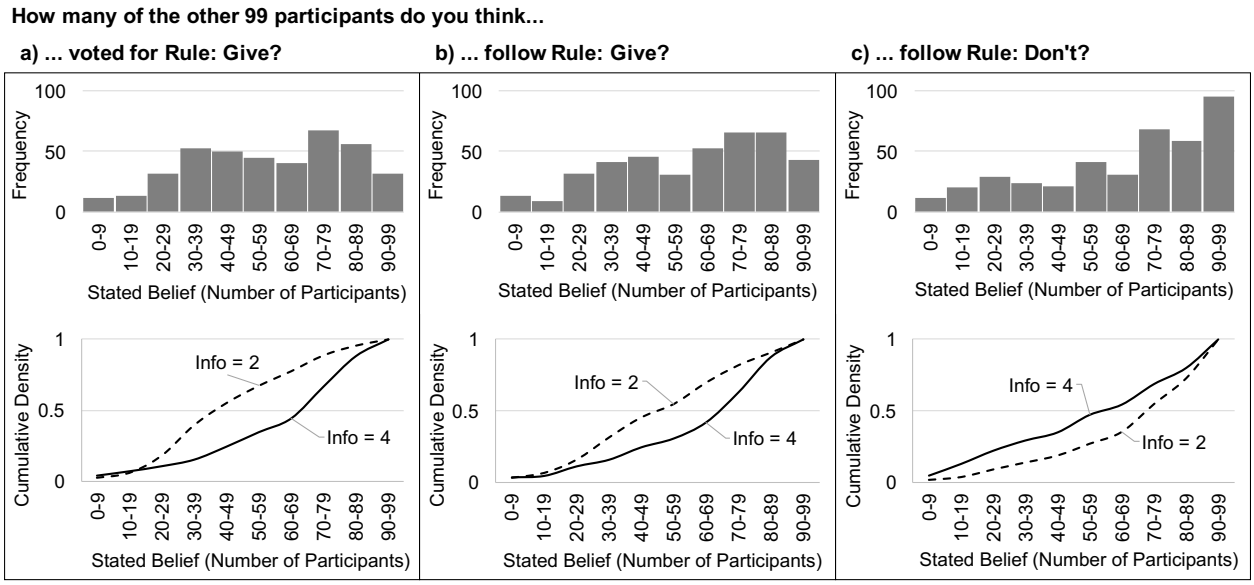


Figure 6: Beliefs about the choices of other participants (data from all treatments pooled, N=400). Top: Frequency of beliefs by answer bracket. Bottom: Cumulative density of answers among subjects having received *info*= 2 and *info*= 4, respectively.

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.[24] 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 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-

[24] Beliefs in each treatment follow very much the same distribution as the pooled data shown in Figure 6.

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*.

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.[25] 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 compliance with Rule:Give. Moreover, there is no evidence that*

[25] 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 $E_i(Comply_{-i}) = 44.5/99 \approx 0.45$.

	(a) *Rule:Give*					(b) *Rule:Don't*				
	$E_i(Comply_{-i})$	$Comply_i = 1$				$E_i(Comply_{-i})$	$Comply_i = 1$			
	(1) OLS	(2) OLS	(3) 2SLS	(4) OLS	(5) OLS	(1) OLS	(2) OLS	(3) 2SLS	(4) OLS	(5) OLS
$info_i$	.13***			-.04	-.04	-.13***			-.11***	-.09**
	(.02)			(.04)	(.04)	(.03)			(.04)	(.04)
$E_i(Comply_{-i})$		.46***	-.32				.51***	.87***		
		(.07)	(.30)				(.07)	(.33)		
Malpractice	-.02	-.10***	-.11**	-.11***	-.10**	.01	-.02	-.02	-.01	-.05
	(.03)	(.04)	(.05)	(.04)	(.04)	(.03)	(.05)	(.05)	(.05)	(.05)
Constant	.51***	.31***	.74***	.58***	.30*	.79***	.57***	.30	.99***	.59***
	(.03)	(.06)	(.18)	(.05)	(.16)	(.04)	(.08)	(.25)	(.06)	(.19)
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

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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.

beliefs about others' compliance causally affect baseline compliance with Rule:Give. A subject's compliance with Rule:Don't, on the other hand, is strongly and positively affected by 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).

3.3.2 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.

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_MoneyOffer*) 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 many votes were lost

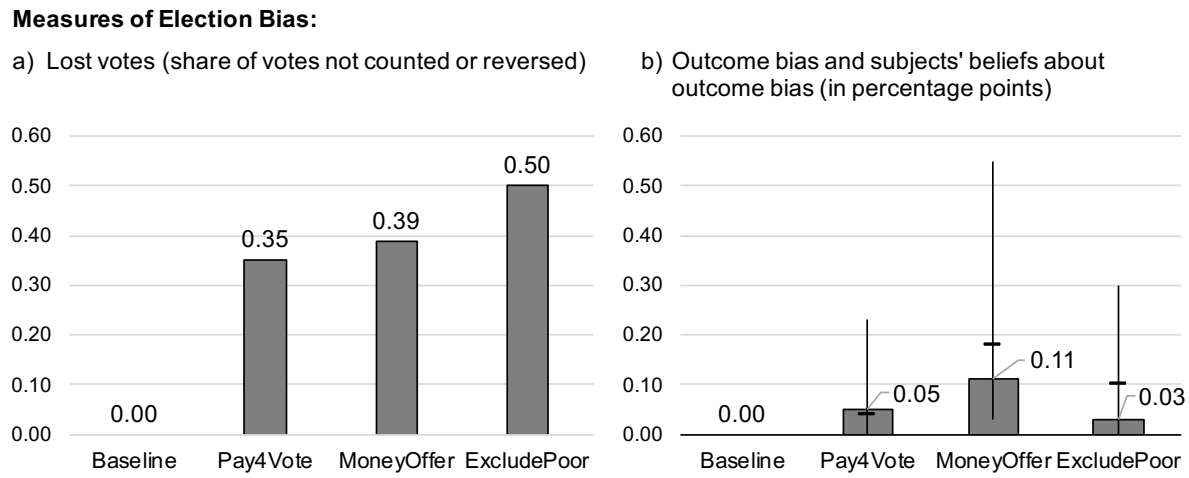


Figure 7: Measures of Election Bias. Panel a): Share of votes not counted or reversed due to interventions *Pay4Vote*, *MoneyOffer*, *ExcludePoor*. Panel b): Outcome bias (percentage point change in vote shares due to intervention, bar plot) and subjects' beliefs about outcome bias (10th to 90th percentile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

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.[26] 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 the variance in these two measures explain the variance in compliance with *Rule:Give* between treatments? 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

[26]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 ($T_MoneyOffer$), or the voting behavior among "poor" and "rich" subjects ($T_ExcludePoor$). In particular, we calculate individual i's belief about the outcome bias as $E_i[Outcome_Bias]$

$$:= \begin{cases} 0 & \text{if } i \text{ is in } T_Baseline, \\ \left| \dfrac{E_i[Accept_Pay_j|Vote_j = 1]E_i[Vote_j]}{E_i[Accept_Pay_j]} \right| & \text{if } i \text{ is in } T_Pay4Vote, \\ \big| E_i[Accept_MoneyOffer_j|Vote_j = 1]E_i[Vote_j] & \\ +E_i[Accept_MoneyOffer_j|Vote_j = 0](1 - E_i[Vote_j]) \big| & \text{if } i \text{ is in } T_MoneyOffer, \\ |E_i[Vote_j|Income_j > 40\text{K}] - E_i[Vote_j]| & \text{if } i \text{ is in } T_ExcludePoor. \end{cases}$$

by variance in $Lost_Vote_i$ and $E_i[Outcome_Bias]$.

	$Comply_i$|**Rule:Give**			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
$Lost_Vote_i$		-.11**		-.10**
		(.04)		(.04)
$E_i[Outcome_Bias]$			-.34***	-.33***
			(.12)	(.12)
T_Pay4Vote	-.11**	-.07	-.08*	-.05
	(.05)	(.05)	(.05)	(.05)
T_MoneyOffer	-.12**	-.08	-.04	.00
	(.05)	(.05)	(.06)	(.06)
T_ExcludePoor	-.09*	-.04	-.06	-.01
	(.05)	(.05)	(.05)	(.05)
Constant	.56***	.57***	.56***	.57***
	(.05)	(.05)	(.05)	(.05)
Add. Controls	Yes	Yes	Yes	Yes
Observations	400	400	400	400
Standard errors in parentheses	* $p < 0.1$,	** $p < 0.05$,	*** $p < 0.01$	

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 $info_i$.

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 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%. In a general sense, the effect sizes close to zero of the treatment variables in column (4) imply that participants who are *not* excluded and who do *not* hold the belief that the voting outcome looses its representativeness show the same compliance behavior as the average participant in *T_Baseline*.

Table 5 thus confirms our expectation that both *Lost_Vote*$_i$ and E_i[*Outcome_Bias*] are associated with significantly lower rates of rule compliance.[27] 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. Indeed, while E_i[*Outcome_Bias*] is not independent from treatment exposure, the results show that reductions in compliance are associated with holding the belief that the voting outcome is not representative of voting preferences in the population (see columns (3) and (4)).

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*.[28] 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*"

[27] 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".

[28] The treatment was run with 100 new participants in summer 2018 on *Prolific.ac*. Instructions and implementation were identical to the main treatments except for the description of the vote aggregation procedure as described here. The mean age of participants is 29 years, 53% are female, and 37% are students.

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 $\pm.00$ (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 a 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 -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 that may corrupt these institutions (such as bribing and lobbying). 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.

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 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).[29] 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

[29] 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*.

	$Comply_i \mid$ ***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
T_Pay4Vote	-.14**	-.08	-.15*	-.09	-.14**	-.06	-.12	-.15*	-.16**	-.05	-.18**	-.09	-.14*	-.03
	(.06)	(.08)	(.09)	(.08)	(.08)	(.09)	(.08)	(.08)	(.07)	(.08)	(.08)	(.07)	(.07)	(.08)
T_MoneyOffer	-.12*	-.12	-.21**	-.10	-.19**	-.00	-.19**	-.11	-.22***	-.01	-.19**	-.07	-.19***	-.07
	(.06)	(.09)	(.09)	(.08)	(.08)	(.08)	(.08)	(.08)	(.07)	(.08)	(.08)	(.07)	(.07)	(.09)
T_ExcludePoor	-.11*	-.01	-.18**	-.11	-.09	-.09	-.18**	.-04	-.13*	-.05	-.12	-.07	-.11	-.07
	(.06)	(.09)	(.09)	(.08)	(.08)	(.08)	(.08)	(.08)	(.07)	(.07)	(.08)	(.07)	(.07)	(.09)
Constant	.57***	.63***	.58***	.52***	.52***	.51***	.52***	.55***	.59***	.49***	.60***	.50***	.60***	.53***
	(.06)	(.09)	(.09)	(.07)	(.07)	(.07)	(.08)	(.07)	(.06)	(.07)	(.07)	(.07)	(.07)	(.08)
Add. 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

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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 $info_i$.

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. The findings in table 6 also support our interpretation of the results in section 3.3. Together, they suggest that procedural concerns about the inclusiveness and unbiasedness of the election procedure might drive the decline in compliance observed for *Rule:Give*.This resonates with theories of "legitimate authority" (e.g., Weber, 1978; Tyler, 2006; Dickson, Gordon and Huber, 2015; Akerlof, 2017) and with empirical findings suggesting that people care about the " fairness" of decision making processes (see, e.g., Tyler, 1990; Frey, Benz and Stutzer, 2004; Cappelen et al., 2013). 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.

The additional treatment *T_Exo* sheds a new light on our surprising finding that malpractice seems to have an asymmetric effect: we find a strong and systematic malpractice effect for *Rule:Give* but not for *Rule:Don't*. Interestingly, the same asymmetric pattern can be found for the existence of a democracy effect. In other words, in our setting a malpractice effect can *always* be found in cases where a democracy effect exists.

Compliance with *Rule:Don't* is strongly driven by beliefs about what others do and since beliefs about others' behavior are not affected by the corrupted voting procedures, no differences in average compliance can be found. We can thus speculate that rules that are being complied with due to peer effects are one type of rule where procedural aspects do not play a role for compliance. In contrast, in the case of *Rule:Give*, compliance seems to rather occur due to a preference for following the rule and we find no evidence for beliefs about the rule compliance of others playing a role for own decisions.[30] This type of intrinsically motivated rule compliance seems to be sensitive to procedural aspects. Indeed, our experiment establishes a democracy effect for a fair majority vote as in Dal Bó, Foster and Putterman (2010) and at the same time shows how the same sensitivity to the procedure leads to a complete erosion of this effect if the majority vote has been corrupted. We thus speculate that democracy effects as well as malpractice effects might not be effective in all domains. Whether this speculation holds true in a more general sense and outside of our experimental setup will need to be uncovered by future research.

[30]For detecting a significant effect under *Rule:Give* with a power of 80% (which means that the effect will be significant 80% of the time with $\alpha = 0.05$), the minimum detectable effect size of the coefficient of $info_i$ (Table 4, column (4)) is $2.8 * 0.04 = 0.11$, where 0.04 is the standard error of the estimated coefficient. This is an effect we are able to find for *Rule:Don't* where the standard error is very similar. While the effect of beliefs on behavior is not zero under *Rule:Give*, the relationship is both, statistically insignificant and smaller in magnitude than for the case of *Rule:Don't*.

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. We find a strong and systematic reduction in voluntary compliance with *Rule:Give* but not with *Rule:Don't*. We demonstrate that the effects we observe under *Rule:Give* correspond to a complete erosion of a democracy effect on the same rule. Compliance with *Rule:Don't*, however, is driven by peer-effects and is not sensitive to procedural aspects. A sensitivity of rule compliance to the implementation procedure is mainly found among subjects who are themselves excluded from the ballot and those who believe the voting outcome to no longer be representative due to the corruption of the vote.

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) are 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 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.

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 Communities in 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.

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, Ernesto, Pedro Dal Bó, and Erik Eyster. 2017. "The demand for bad policy when voters underappreciate equilibrium effects." *The Review of Economic Studies*, 85(2): 964–998.

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 Kenju Kamei. 2019. "The Democracy Effect: a weights-based identification strategy." National Bureau of Economic Research.

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.

Dickson, Eric S, Sanford C Gordon, and Gregory A Huber. 2015. "Institutional Sources of Legitimate Authority: An Experimental Investigation." *American Journal of Political Science*, 59(1): 109–127.

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy*, 8(3): 203–232.

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.

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.

Fehr, Ernst, Holger Herz, and Tom Wilkening. 2013. "The Lure of Authority: Motivation and Incentive Effects of Power." *American Economic Review*, 103(4): 1325–59.

Feld, Lars P, and Bruno S Frey. 2002. "Trust breeds trust: How taxpayers are treated." *Economics of Governance*, 3(2): 87–99.

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.

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 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.

LeDuc, Lawrence, Richard G. Niemi, and Pippa Norris, ed. 2014. *Comparing Democracies.* Vol. 4, London:SAGE Publications.

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.

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.* New Haven: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.[31] 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.

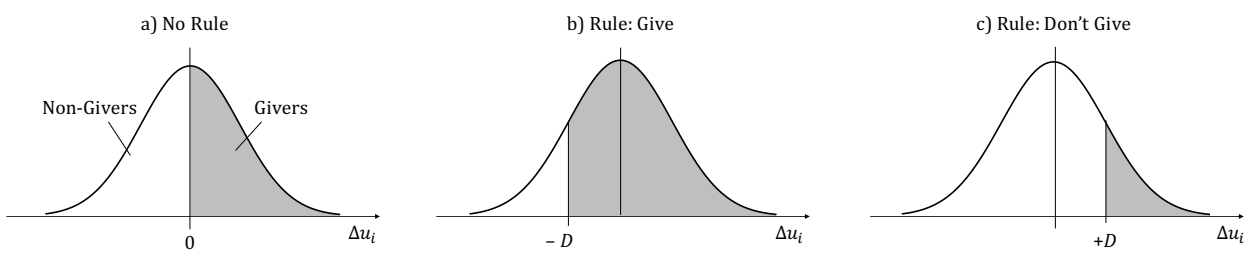


Figure A.1: Theory: Illustration of population shares choosing to give ($Give_i = 1$) and not to give ($Give_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with a standard majority vote (panels b and c).

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 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

$$\text{If } Malpractice = 0, \quad (Comply_i|Rule{:}Give) = 1 \Leftrightarrow \Delta u_i \geq -D,$$
$$\text{and} \quad (Comply_i|Rule{:}Don't) = 1 \Leftrightarrow \Delta u_i < +D.$$

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

[31] 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.

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.

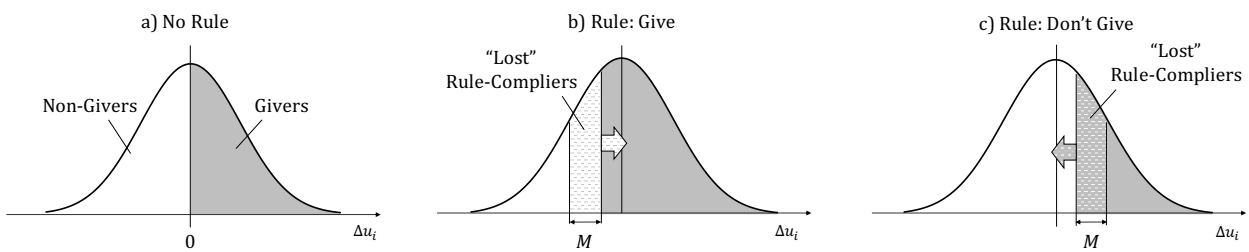


Figure A.2: Theory: Illustration of population shares choosing to give ($Give_i = 1$) and not to give ($Give_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with malpractice during the election (panels b and c).

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 $Rule \in \{\textit{Rule:Give}, \textit{Rule:Don't}\}$. 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*:

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

$$0.5 \cdot [u_i(Give_i = 1) + D] + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(Receive) \geq 0.5 \cdot u_i(Give_i = 1)$$
$$\Leftrightarrow \underbrace{\Delta F(D)}_{>0} \geq \underbrace{-\frac{D}{\Delta u(Receive)}}_{<0}.$$

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

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

$$0.5 \cdot u_i(Give_i = 0) + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(Receive) \geq 0.5 \cdot [u_i(Give_i = 0) + D]$$
$$\Leftrightarrow -D \geq -\Delta F(D) \cdot \Delta u(Receive)$$
$$\Leftrightarrow \Delta F(D) \geq \frac{D}{\Delta u(Receive)}$$

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

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

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

$$0.5 \cdot [u_i(Give_i = 1) + D] + 0.5 \cdot \Delta F[D] \cdot \Delta u_i(Receive) \geq 0.5 \cdot [u_i(Give_i = 0) + D]$$
$$\Leftrightarrow \Delta u_i \geq -\Delta F(D) \cdot \Delta u(Receive)$$
$$\Leftrightarrow \Delta F(D) \geq -\frac{\Delta u_i}{\Delta u(Receive)},$$

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).

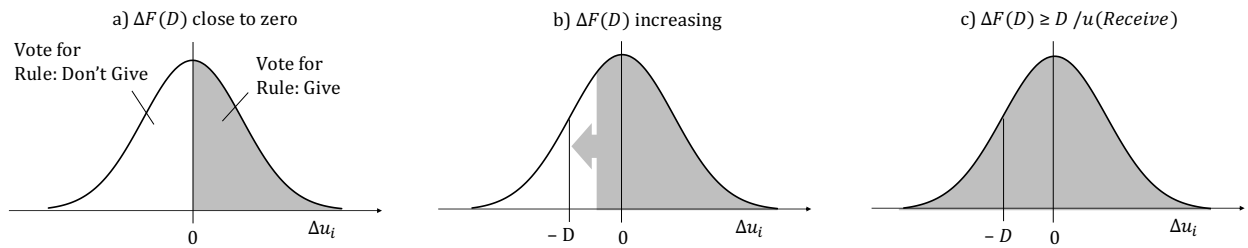


Figure A.3: Theory: Share of Population voting for Rule: Give

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)

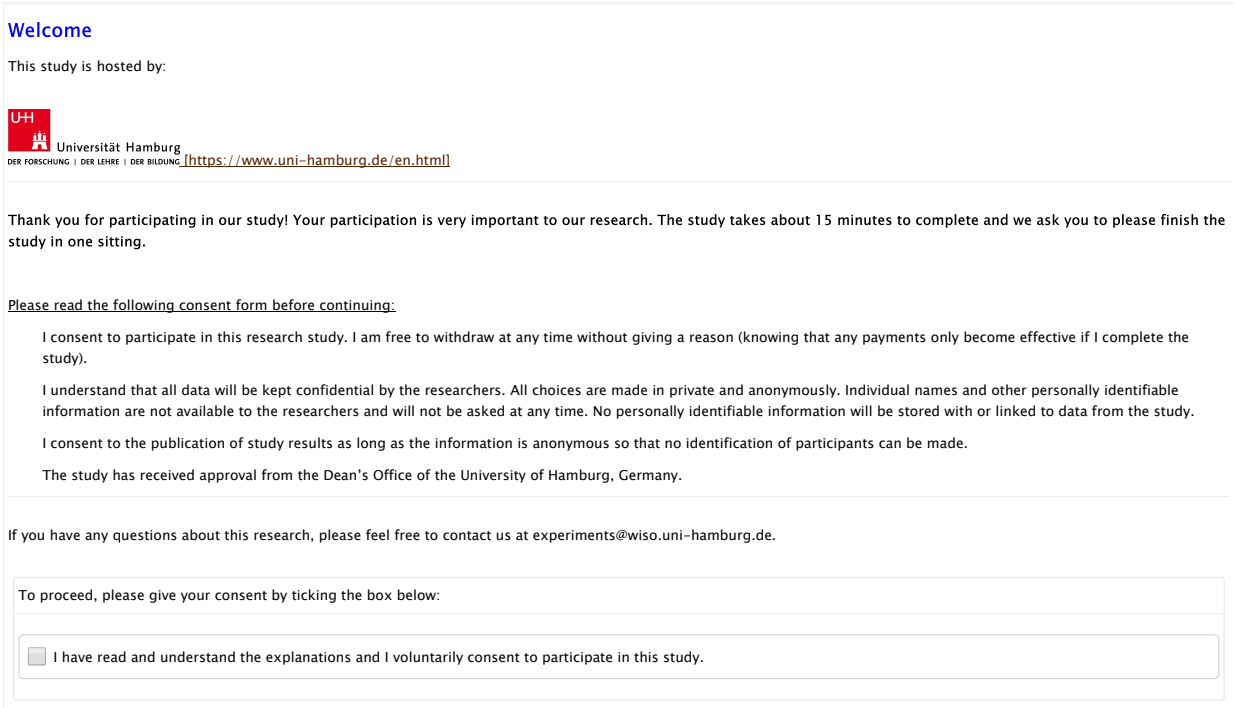
4. Assume that you had the opportunity to take part in the following gamble: There are 100 balls in an urn. Of these balls, 99 are black and 1 is red. One ball is randomly drawn from the urn. If it is red you win 1000 GBP. If it is black you win 0 GBP. What would be the maximal amount of money you would 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 you agree or disagree 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.

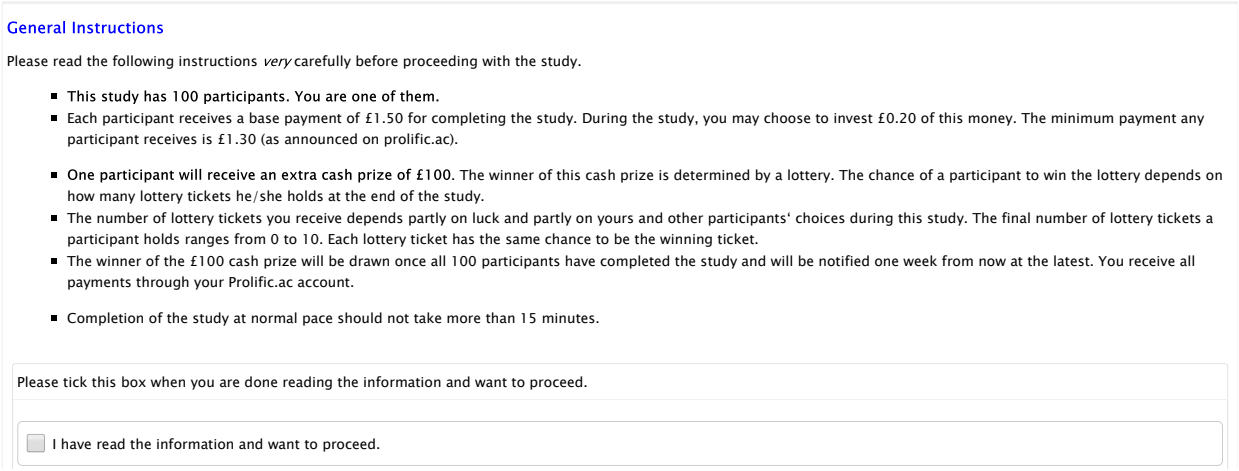
The study has received approval from the Dean's Office of the University of Hamburg, Germany.

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

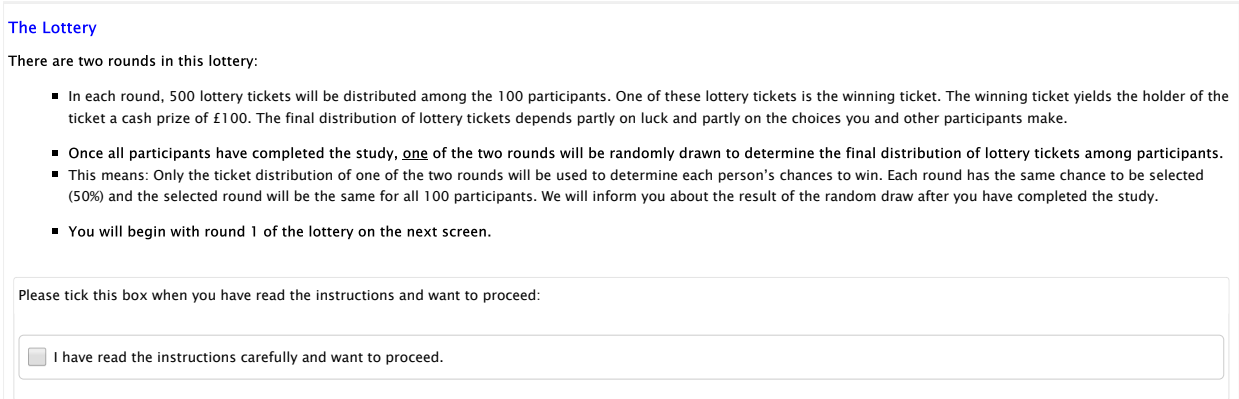
Please read the following instructions *very* carefully before proceeding with the study.

- 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.

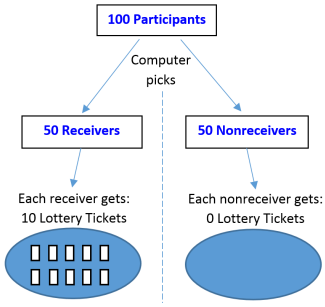
Figure A.6: Screenshot: Instructions about the Lottery

Distribution of lottery tickets

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

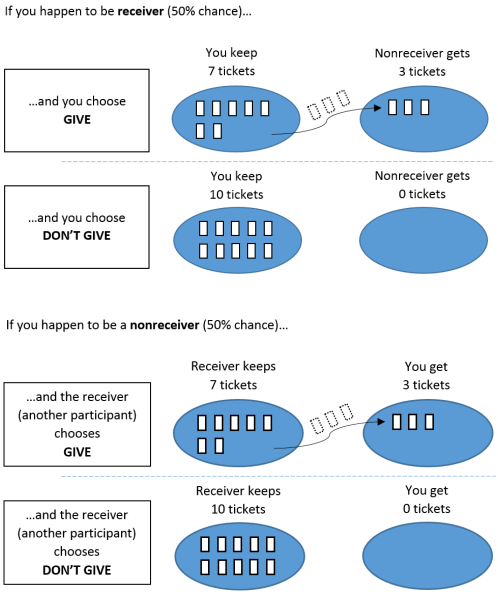
Step 1: The computer picks 50 receivers and 50 nonreceivers:

- 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.

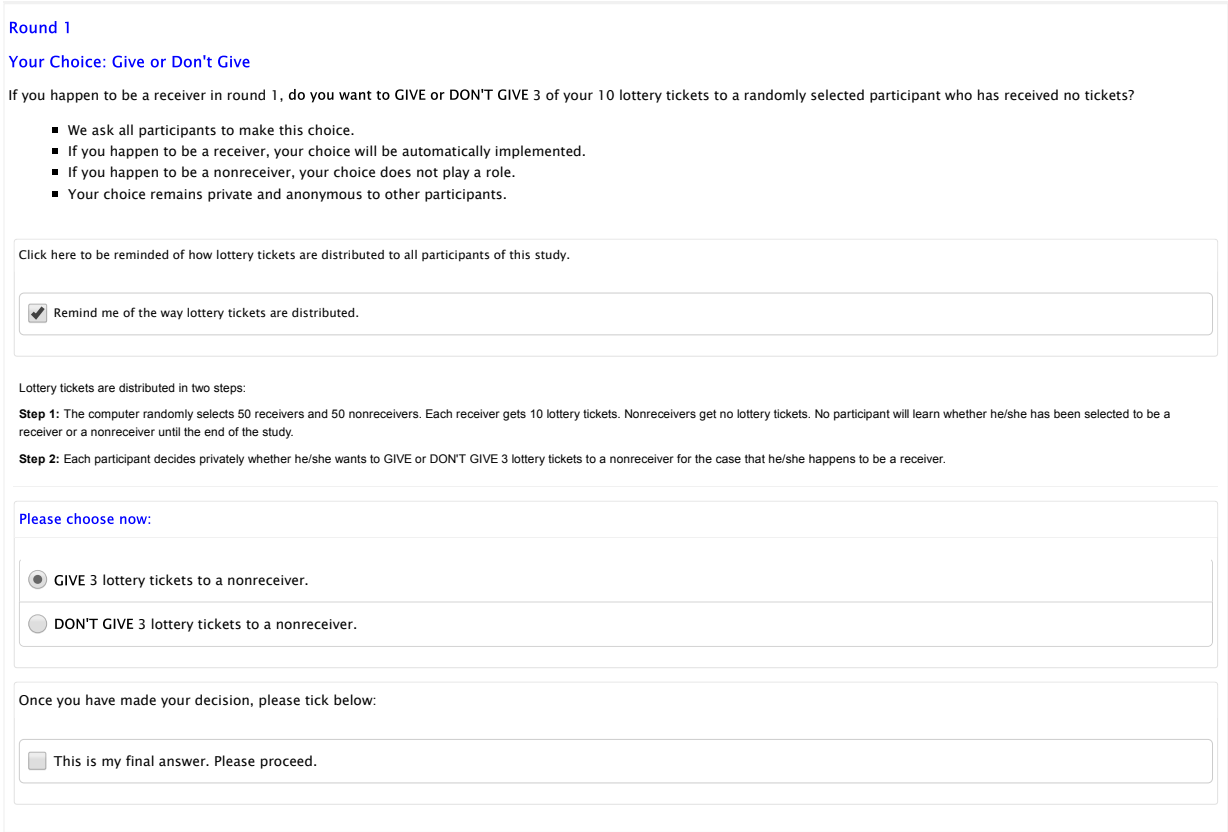
If you happen to be a nonreceiver (50 % chance), your choice whether to GIVE or DON'T GIVE does *not* play a role. In this case, the choice of *another* participant (who happens to be a receiver) determines the number of lottery tickets that you will receive.

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:

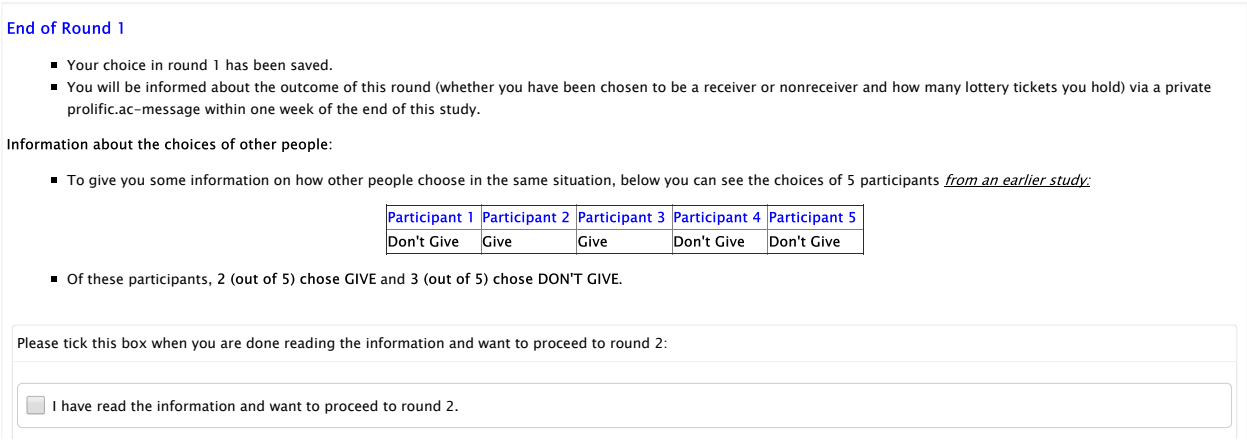
◉ GIVE 3 lottery tickets to a nonreceiver.

○ 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:*

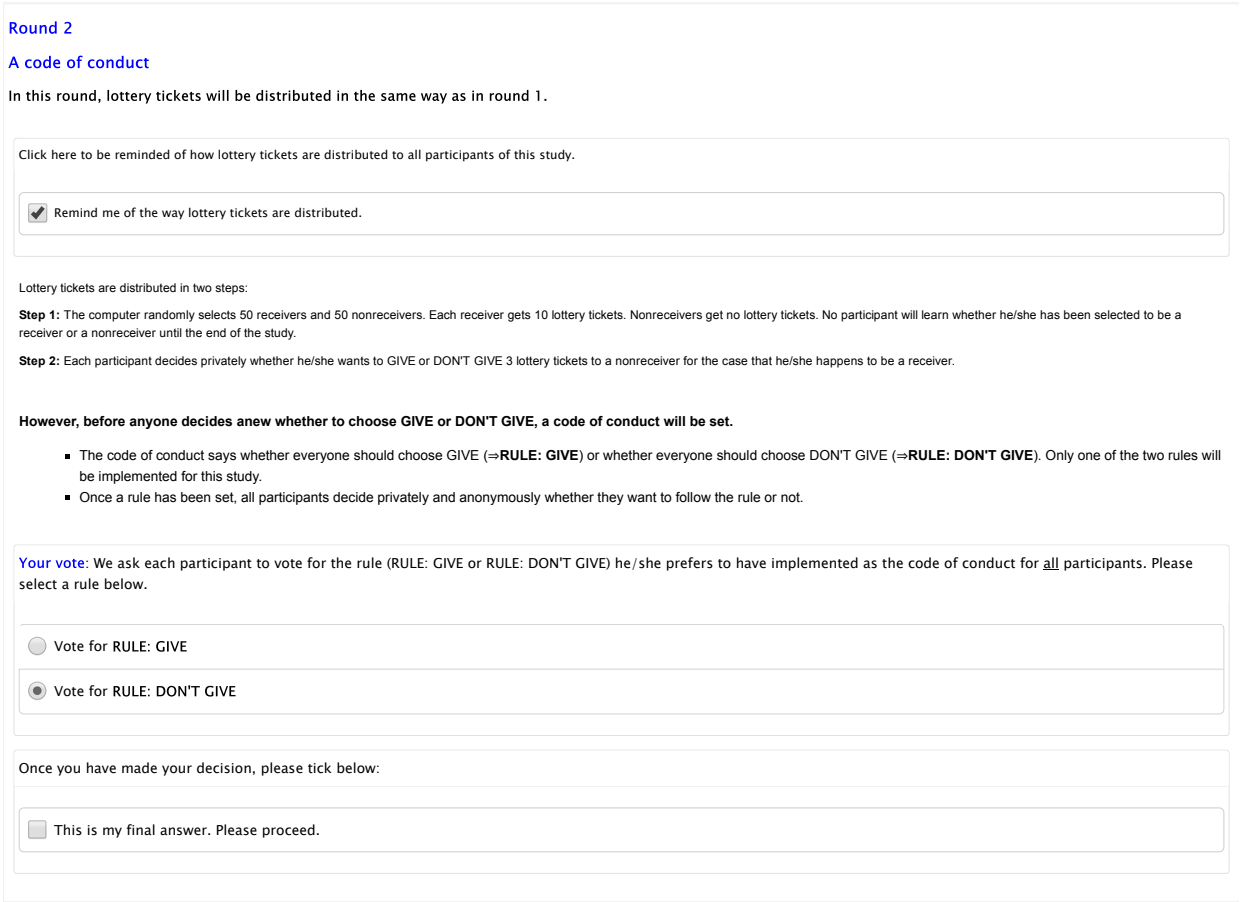
Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Don't Give	Give	Give	Don't Give	Don't Give

- 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)

Round 2

Pay £0.20 to make your vote count

- You just selected RULE: DON'T GIVE as the rule you want to vote for.
- You have to pay £0.20 to make your vote count.

The code of conduct will be determined as follows:

- The rule that receives more votes in total will be implemented as the code of conduct.*
- The votes of participants who pay £0.20 will be counted. Other votes will not be counted.

*Tie Breaker: In case there are exactly the same number of votes counted for RULE: GIVE as for RULE: DON'T GIVE, a coin-flip decides which of the two rules will be implemented.

- If you pay £0.20, your vote for RULE: DON'T GIVE will be counted. If you don't pay, your vote will not be counted.
- This payment is independent of which rule you have selected (and whether or not the rule you have selected will be implemented).
- If you choose to pay, £0.20 will be substracted from your base payment. All other payments are unaffected.
- We ask all 100 participants to make this choice. This means: Only the votes of those participants who pay £0.20 will be counted.

Please choose now:

- ○ Don't pay £0.20. Your vote will NOT be counted.
- ◉ Pay £0.20. Your vote will be counted.

Once you have made your decision, please tick below:

☐ This is my final answer. Please proceed.

Figure A.11: Screenshot: $Accept_Pay4Vote \in \{0, 1\}$ (Round 2, $T_Pay4Vote$)

Round 2

Receive £0.20 for changing your vote

You just selected RULE: DON'T GIVE as the rule you want to vote for.

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

*Tie Breaker: In case there are exactly the same number of votes counted for RULE: GIVE as for RULE: DON'T GIVE, a coin-flip decides which of the two rules will be implemented.

For an extra payment of £0.20: Are you willing to vote for the opposite rule instead?

- If you vote for the rule that is opposite to what you wanted to vote for (RULE: GIVE instead of RULE: DON'T GIVE), you will receive an extra payment of £0.20 on top of your base payment.
- This will be your final vote. Only the vote that you cast on this page will be counted.
- We ask all 100 participants to make the same choice. This means: All participants are offered an extra payment of £0.20 to vote for the rule that is *opposite to* what they originally wanted to vote for. Only the final vote of each participant will be counted.

Please choose now:

- ◉ Accept extra payment of £0.20 and change my vote to RULE: GIVE.
- ○ Reject extra payment of £0.20 and keep my vote for RULE: DON'T GIVE.

Once you have made your decision, please tick below:

☐ This is my final answer. Please proceed.

Figure A.12: Screenshot: $Accept_MoneyOffer \in \{0, 1\}$ (Round 2, $T_MoneyOffer$)

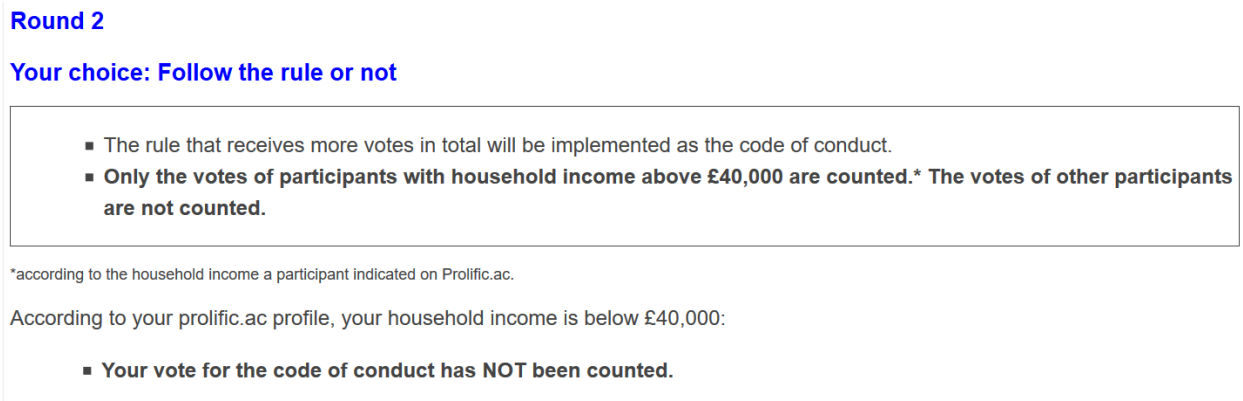


Figure A.13: Screenshot: Information about intervention *Exclude_Poor* (Round 2)

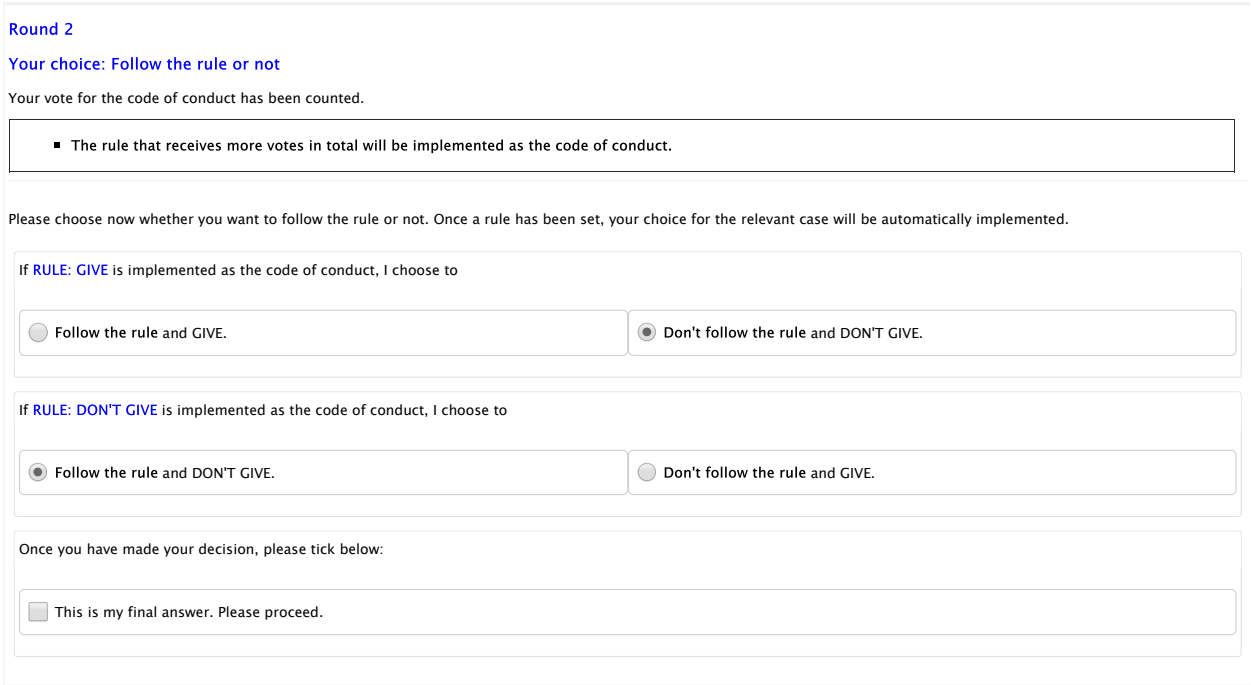


Figure A.14: Screenshot: $Give_i|Rule \in \{0, 1\}$ (Round 2, *T_Baseline*)

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.
- [x] 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:

- [x] 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*)