Online Appendix for: Corrupted Votes and Rule Compliance

Arno Apffelstaedt and Jana Freundt

B.1 Randomization on observables

	StdMajority	Pay4Vote	MoneyOffer	ExcludePoor	All
Female	0.38	0.42	0.42	0.45	0.42
	(100)	(100)	(100)	(100)	(400)
Age	30.22	29.27	33.06	31.66	31.06
	(97)	(98)	(99)	(100)	(394)
Western	0.70	0.64	0.66	0.78	0.69
	(100)	(100)	(100)	(100)	(400)

Number of observations in parentheses.

Table B.1: Share of female participants, mean age and share of participants from western countries in each treatment and in total.

B.2 Treatment effects: Type-level and type-weighted estimations

In the main text, we estimate treatment effects on rule compliance using regressions that control for type fixed effects, where $Type_i = (Give_i|NoRule) \times Vote_i$. While this approach effectively controls for unbalanced type distributions across treatments, this appendix presents the results of an alternative approach to estimating treatment effects, which explicitly accounts for the possibility of treatment effects varying across types. As we will demonstrate, results remain unchanged. The approach relies on a weights-based identification strategy proposed by Dal Bó, Foster and Kamei (2019). It entails estimating individual treatment effects for each type and then computing the average treatment effect by assigning weights to each type's treatment effect based on its prevalence in the entire population.

It is worth noting that Dal Bó, Foster and Kamei (2019) introduce this method primarily to address issues related to endogenous selection into treatment when a subject's type is not consistently observable across all conditions. These issues are not present in our setting since the type of each subject $Type_i = (Give_i|NoRule) \times Vote_i$ is observed in all

treatments. Instead, we employ their weights-based approach to account for potential heterogeneous treatment effects across types, offering insights into how various types respond to malpractice.[1]

The detailed results of the type-weighting approach are presented in Table B.2. Panel (a) shows the number of subjects per type across all four treatments: There are

- 92 subjects of type $((Give_i|NoRule), Vote_i) = (0, Rule{:}Don't)$,
- 17 subjects of type $((Give_i|NoRule), Vote_i) = (1, Rule{:}Don't)$,
- 63 subjects of type $((Give_i|NoRule), Vote_i) = (0, Rule{:}Give)$, and
- 228 subjects of type $((Give_i|NoRule), Vote_i) = (1, Rule{:}Give)$.

To calculate type-weighted average treatment effects, we first obtain baseline compliance rates and treatment effects for each of the four types separately. We then weigh types according to the relative frequency with which they occur in our sample. In Panel (b) of Table B.2, we present the proportion of compliant subjects per type in the baseline treatment StdMajority. In Panel (c) of Table B.2 we report treatment effects: White cells display the individual treatment effects for each type, estimated through simple (univariate) OLS regressions while conditioning on the type. Using these estimates, we compute population average treatment effects (reported in gray cells) by weighing types according to their prevalence in the population. For example, to compute the average treatment effect of offering money to voters (MoneyOffer) on compliance with Rule:Give, we use the formula:

$$(92/400)(.01) + (63/400)(-.57) + (17/400)(-.17) + (228/400)(-.04) = -.12.$$

Standard errors for weighted averages are calculated using the Delta method. For example, the standard error for the average treatment effect we just calculated can be determined from

$$\sqrt{(92/400)^2(.16)^2 + (63/400)^2(.18)^2 + (17/400)^2(.36)^2 + (228/400)^2(.03)^2} = .05.$$

The results obtained using the type-weighting approach closely align with the estimates derived from regressions with type fixed effects, as presented in the main text. To ensure better comparability, we reproduce Table 3 and Figures 3 and 4 below using the results from our type-weighting approach. Table B.3 replicates Table 3. Figures B.1 and B.2 replicate Figures 3 and 4. As both the table and figures illustrate, compliance rates and treatment effects, at both the population and subgroup levels, are nearly identical across the two estimation methods.

[1]With types being perfectly observed, there are several methods available to accommodate type-dependent treatment effects. The type-weighted approach we employ is essentially equivalent to using a matching estimator with exact matching on (discrete) type covariates. Identically sized effects (usually with higher levels of significance) are found with similar methods such as inverse probability weighting or regression adjustment.

(a) All treatments: n by $Type_i$

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

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

	...**Rule:Give**			...**Rule:Don't**		
	$Give_i \vert NoRule$			$Give_i \vert NoRule$		
$Vote_i$	0	1	w.avg.	0	1	w.avg.
Rule:Don't	0.57	0.50	0.56	0.96	0.62	0.91
Rule:Give	0.80	1.00	0.96	1.00	0.51	0.62
w.avg.	0.66	0.97	**0.85**	0.98	0.52	**0.70**

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

		...**Rule:Give**			...**Rule:Don't**		
		$Give_i \vert NoRule$			$Give_i \vert NoRule$		
	$Vote_i$	0	1	w.avg.	0	1	w.avg.
Pay4Vote	*Rule:Don't*	-0.18	0.50	-0.07	-0.05	-0.63	-0.14
		(0.14)	(0.42)	(0.14)	(0.07)	(0.30)	(0.08)
	Rule:Give	-0.35	-0.04	-0.10	-0.10	-0.07	-0.08
		(0.16)	(0.03)	(0.04)	(0.08)	(0.10)	(0.08)
	w.avg.	-0.25	0.00	**-0.10**	-0.07	-0.11	**-0.10**
		(0.11)	(0.04)	(0.05)	(0.05)	(0.09)	(0.06)
MoneyOffer	*Rule:Don't*	-0.01	-0.17	-0.03	-0.09	0.37	-0.02
		(0.16)	(0.36)	(0.14)	(0.08)	(0.26)	(0.08)
	Rule:Give	-0.57	-0.04	-0.16	-0.15	0.02	-0.02
		(0.18)	(0.03)	(0.05)	(0.09)	(0.09)	(0.08)
	w.avg.	-0.24	-0.05	**-0.12**	-0.12	0.04	**-0.02**
		(0.12)	(0.04)	(0.05)	(0.06)	(0.09)	(0.06)
ExcludePoor	*Rule:Don't*	-0.13	-0.00	-0.11	0.00	0.37	0.06
		(0.14)	(0.33)	(0.13)	(0.07)	(0.23)	(0.07)
	Rule:Give	-0.33	-0.02	-0.09	0.00	0.08	0.06
		(0.17)	(0.03)	(0.04)	(0.09)	(0.10)	(0.08)
	w.avg.	-0.21	-0.02	**-0.09**	0.00	0.10	**0.05**
		(0.11)	(0.04)	(0.05)	(0.05)	(0.09)	(0.06)
Pooled	*Rule:Don't*	-.12	.06	-0.09	-0.04	0.15	-0.01
		(0.11)	(0.26)	(0.10)	(0.06)	(0.23)	(0.06)
	Rule:Give	-0.40	-0.03	-0.11	-0.08	0.01	-0.01
		(0.14)	(0.03)	(0.04)	(0.07)	(0.08)	(0.07)
	w.avg.	-0.23	-0.03	**-0.11**	-0.06	0.02	**-0.01**
		(0.09)	(0.03)	(0.04)	(0.04)	(0.08)	(0.05)

Standard errors in parentheses.

Table B.2: Number of subjects (a), baseline compliance rates (b) and treatment effects (c) by $Type_i = (Give_i|NoRule) \times Vote_i$. White cells in panel (b) show compliance rates (share of subjects for whom $Comply_i|Rule = 1$) for each $Type_i$ in StdMajority. White cells in panel (c) show treatment effects for each $Type_i$ from OLS regressions. Gray cells show weighted-average baseline compliance rates (panel b) and weighted-average treatment effects (panel c). Weighted standard errors calculated assuming normally distributed standard errors (Delta method).

Dep.Var.	a) Compliance with **Rule:Give**					b) Compliance with **Rule:Don't**				
		Subgroups					Subgroups			
		$Give_i$|*NoRule*		$Vote_i$			$Give_i$|*NoRule*		$Vote_i$	
	All Subjects	0	1	*Rule:Don't*	*Rule:Give*	All Subjects	0	1	*Rule:Don't*	*Rule:Give*
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	OLS (7)	OLS (8)	OLS (9)	OLS (10)
Pay4Vote	-0.10	-0.25	0.00	-0.07	-0.10	-0.10	-0.07	-0.11	-0.14	-0.08
	(0.04)	(0.11)	(0.04)	(0.14)	(0.04)	(0.06)	(0.05)	(0.09)	(0.08)	(0.08)
MoneyOffer	-0.12	-0.24	-0.05	-0.03	-0.16	-0.02	-0.12	0.04	-0.02	-0.02
	(0.05)	(0.12)	(0.04)	(0.14)	(0.05)	(0.06)	(0.06)	(0.09)	(0.08)	(0.08)
ExcludePoor	-0.09	-0.21	-0.02	-0.11	-0.09	0.06	-0.00	0.10	0.05	0.06
	(0.05)	(0.11)	(0.04)	(0.13)	(0.04)	(0.06)	(0.05)	(0.09)	(0.07)	(0.08)
Constant	0.85	0.66	0.97	0.56	0.96	0.70	0.98	0.52	0.91	0.62
(Compliance in StdMajority)	(0.03)	(0.08)	(0.03)	(0.09)	(0.03)	(0.04)	(0.04)	(0.07)	(0.05)	(0.06)
Malpractice	-0.11	-0.23	-0.03	-0.09	-0.11	-0.01	-0.06	0.02	-0.01	-0.01
(Pooled)	(0.04)	(0.09)	(0.03)	(0.10)	(0.04)	(0.05)	(0.04)	(0.08)	(0.06)	(0.07)
Observations	400	155	245	109	291	400	155	245	109	291

Standard errors in parentheses.

Table B.3: Treatment effects of Pay4Vote, MoneyOffer and ExcludePoor on rule compliance: Type-weighted estimates. The coefficient on Malpractice (shaded in grey) represents the pooled treatment effect estimated from conducting the same regression using a common indicator variable Malpractice = 1 if Treatment = Pay4Vote, MoneyOffer or ExcludePoor. Full results and details of the type-weighted analysis can be found in Table B.2. Replication of Table 3.

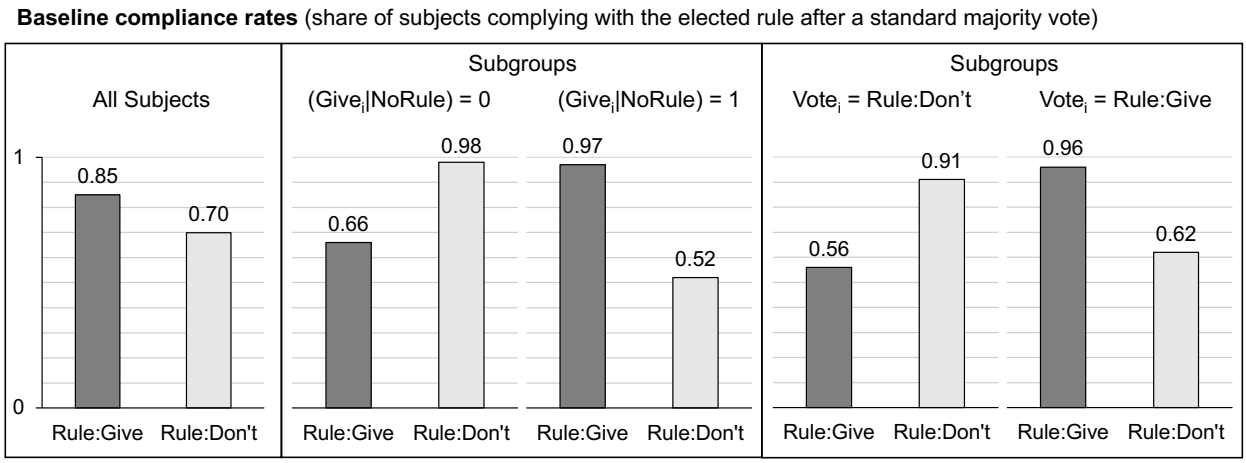


Figure B.1: Share of subjects complying with majority-elected rules in treatment StdMajority. Graphs show type-weighted shares. For details see Table B.2, Panel (b). Replication of Figure 3.

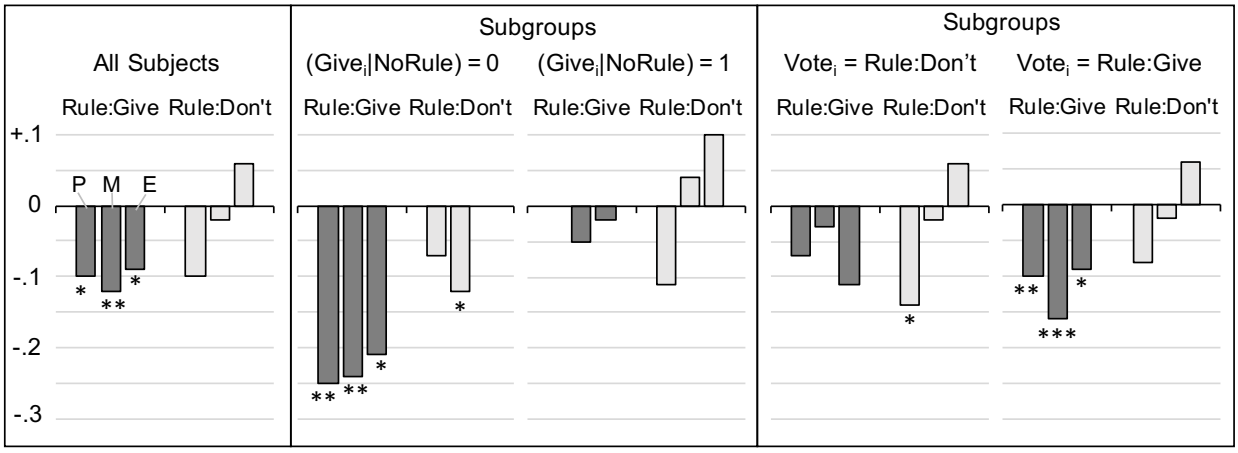


Figure B.2: Treatment effects of interventions Pay4Vote (P), MoneyOffer (M) and ExcludePoor (E) on rule compliance. Graphs show type-weighted estimates of treatment effects as reported in Table B.2 (panel c) and Table B.3. Replication of Figure 4. Stars denote statistically significant differences to compliance rates in StdMajority: * $p < .1$, ** $p < .05$, *** $p < .01$.

B.3 Replication of Table 4 with Individual Treatment Effects

	a) Compliance with **Rule:Give**					b) Compliance with **Rule:Don't**				
Dep. Var.	$E_i(Comply_{-i})$	$Comply_i = 1$		$Comply_i = 1$		$E_i(Comply_{-i})$	$Comply_i = 1$		$Comply_i = 1$	
	OLS	OLS	2SLS	OLS	OLS	OLS	OLS	2SLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
$info_i = 4$	0.13			-0.04	-0.05	-0.12			-0.11	-0.09
	(0.02)			(0.04)	(0.04)	(0.03)			(0.04)	(0.04)
$E_i(Comply_{-i})$		0.46	-0.32				0.51	0.89		
		(0.07)	(0.31)				(0.08)	(0.33)		
Pay4Vote	-0.04	-0.09	-0.12	-0.11	-0.11	0.02	-0.09	-0.09	-0.08	-0.12
	(0.03)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
MoneyOffer	-0.03	-0.11	-0.13	-0.12	-0.11	-0.02	-0.00	0.01	-0.01	-0.04
	(0.03)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
ExcludePoor	0.00	-0.10	-0.09	-0.09	-0.07	0.04	0.04	0.03	0.06	0.04
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Constant	0.49	0.31	0.74	0.58	0.25	0.78	0.57	0.29	0.99	0.52
	(0.03)	(0.06)	(0.18)	(0.05)	(0.17)	(0.04)	(0.08)	(0.24)	(0.06)	(0.21)
Type fixed eff.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Add. Controls					Yes					Yes
Observations	400	400	400	400	369	400	400	400	400	369

Standard errors in parentheses.

Table B.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. IV regressions are 2SLS with $E_i(Comply_{-i})$ being instrumented by $1.[info_i = 4]$. Additional controls in (5) are: $Female_i$, Age_i, $Western_i$, $Student_i$, $UGrad_i$, $Risk\_Seeking_i$, $Trust_i$, number of mistakes in control questions, factor variables measuring political and social values in questionnaire, as well as Big Five personality measures. $Female_i$ and $Risk\_Seeking_i$ (answer on 11-point 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 (0.08 ($p = 0.05$) and 0.02 ($p = 0.06$), respectively). $Trust_i$ (answer on 11-point 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 (0.04, $p = 0.001$). All other demographic and questionnaire controls are insignificant.

B.4 Calculation of subjective beliefs about *Outcome_Bias*

We define *Outcome_Bias* as the absolute value of the difference between the share of votes for Rule:Give before and after the intervention. We calculate subjective beliefs about the expected outcome bias ($E_i(Outcome\_Bias)$) from elicited beliefs about the share of subjects paying for their vote (Pay4Vote), the share of subjects accepting the bribe (MoneyOffer), or the voting behavior among "poor" and "rich" subjects (ExcludePoor).

In the Pay4Vote treatment, *Outcome_Bias* is equal to the absolute value of the difference between (1) the share of subjects who paid for the vote and voted for Rule:Give times the overall share of subjects who voted for Rule:Give, divided by the overall share of subjects who paid for the vote, and (2) the share of subjects who voted for Rule:Give. Expressed in form of beliefs directly elicited from subjects in the Pay4Vote treatment, the belief of subject i regarding *Outcome_Bias* is computed as $E_i(Outcome\_Bias|\, Treatment = Pay4Vote) :=$

$$\left| \frac{E_i\left(Pay_j = 1|\, Vote_j = Rule{:}Give\right) \cdot E_i\left(Vote_j = Rule{:}Give\right)}{\left[E_i\left(Pay_j = 1|\, Vote_j = Rule{:}Give\right) \cdot E_i\left(Vote_j = Rule{:}Give\right) + E_i\left(Pay_j = 1|\, Vote_j = Rule{:}Don't\right) \cdot \left(1 - E_i\left(Vote_j = Rule{:}Give\right)\right)\right]} - E_i\left(Vote_j = Rule{:}Give\right) \right|$$

where $E_i\left(Pay_j = 1|\, Vote_j = Rule{:}Give\right)$ ($E_i\left(Pay_j = 1|\, Vote_j = Rule{:}Don't\right)$) is the subject's belief regarding the share of individuals who pay for the vote, given that they voted for Rule:Give (Rule:Don't), and $E_i\left(Vote_j = Rule{:}Give\right)$ is the subject's belief regarding the total share of individuals voting for Rule:Give.

In the MoneyOffer treatment, *Outcome_Bias* is equal to the absolute value of the difference between (1) the share of subjects who accepted the bribe and voted for Rule:Don't times the share of subjects who voted for Rule:Don't plus the share of subjects who did not accept the bribe and voted for Rule:Give times the share of subject who voted for Rule:Give, and (2) the share of subjects who voted for Rule:Give. Expressed in form of beliefs directly elicited from subjects in the MoneyOffer treatment, the belief of subject i regarding *Outcome_Bias* is computed as $E_i(Outcome\_Bias|\, Treatment = MoneyOffer) :=$

$$\Big| E_i\left(Accept_j = 1|\, Vote_j = Rule{:}Don't\right) \cdot \left(1 - E_i\left(Vote_j = Rule{:}Give\right)\right) + \left(1 - E_i\left(Accept_j = 1|\, Vote_j = Rule{:}Give\right)\right) \cdot E_i\left(Vote_j = Rule{:}Give\right) - E_i\left(Vote_j = Rule{:}Give\right) \Big|$$

where $E_i\left(Accept_j = 1|\, Vote_j = Rule{:}Give\right)$ ($E_i\left(Accept_j = 1|\, Vote_j = Rule{:}Don't\right)$) is the subject's belief regarding the share of individuals who accept the bribe, given that they voted for Rule:Give (Rule:Don't), and $E_i\left(Vote_j = Rule{:}Give\right)$ is the subject's belief regarding the total share of individuals voting for Rule:Give.

In the ExcludePoor treatment, *Outcome_Bias* is equal to the absolute value of the difference between (1) the share of subjects with income larger than £40,000 who voted for Rule:Give, and (2) the share of subjects who voted for Rule:Give. Expressed in form of beliefs directly elicited from subjects in the ExcludePoor treatment, the belief of subject i regarding *Outcome_Bias* is computed as $E_i(Outcome\_Bias|\, Treatment = ExcludePoor) :=$

$$\left|E_i\left(Vote_j = Rule{:}Give | Income_j > £40,000\right) - E_i\left(Vote_j = Rule{:}Give\right)\right|$$

where $E_i\,(\,Vote_j = Rule{:}Give | Income_j > £40,000)$ is the subject's belief regarding the share of individuals with income larger than £40,000 who vote for Rule:Give, and $E_i\,(\,Vote_j = Rule{:}Give)$ is the subject's belief regarding the total share of individuals voting for Rule:Give.

By definition, $E_i(Outcome\_Bias | Treatment = StdMajority) := 0$ for all subjects in the baseline treatment StdMajority.

B.5 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.[2] 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 B.3, panel a), below.

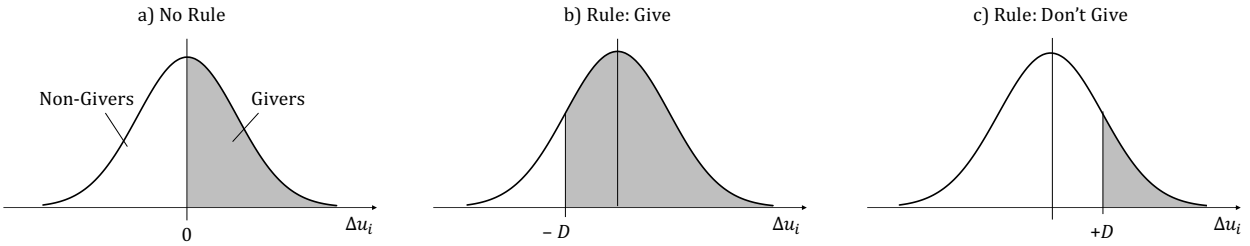


Figure B.3: Theory: Illustration of population shares choosing to give ($Give_i = 1$) and not to give ($Give_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with a standard majority vote (panels b and c).

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 (StdMajority) 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

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

[2]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.

Compared to the case without a rule, the share of subjects choosing to give increases or decreases, see Figure B.3, panel (b) and (c), respectively. Note, importantly, that rules only affect the behavior of those individuals who in the absence of a rule would have chosen the opposite action. While Rule:Give may convince a Non-Giver to give, it will leave the behavior of a Giver ($\Delta u_i \geq 0$) unaffected. Similarly, Rule:Don't may induce some Givers to stop giving, but will not affect the choice of Non-Givers ($\Delta u_i < 0$). We assume that electoral malpractice (in our experiment Pay4Vote, MoneyOffer, ExcludePoor) alters the value some people derive from obeying the elected rule. 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 B.4, panel b). Similarly, a lower share of Givers will be willing to follow Rule:Don't (Figure B.4, 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 rule, although such a strong reaction might be unlikely to be observed in the experiment.

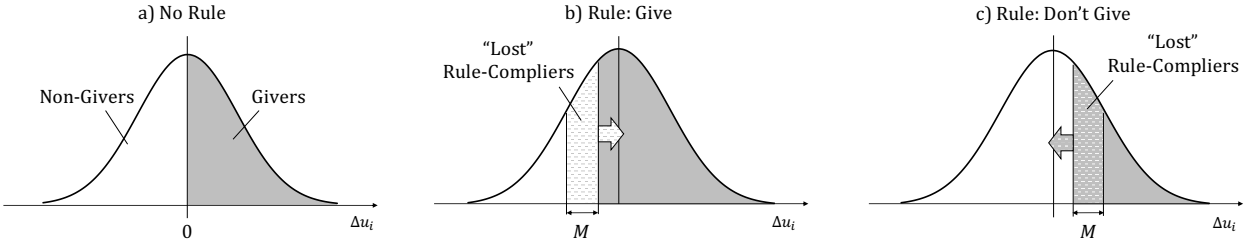


Figure B.4: Theory: Illustration of population shares choosing to give ($\textit{Give}_i = 1$) and not to give ($\textit{Give}_i = 0$) when there exists no code of conduct (panel a) and when there exists a code of conduct that came into force with malpractice during the election (panels b and c).

Voting behavior. We can extend the 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 $U_i[\textit{Rule}]$ denote i's expected utility given $\textit{Rule} \in \{\textit{Rule:Give}, \textit{Rule:Don't}\}$. 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 B.5, 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 B.5, 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 B.5, panel c).

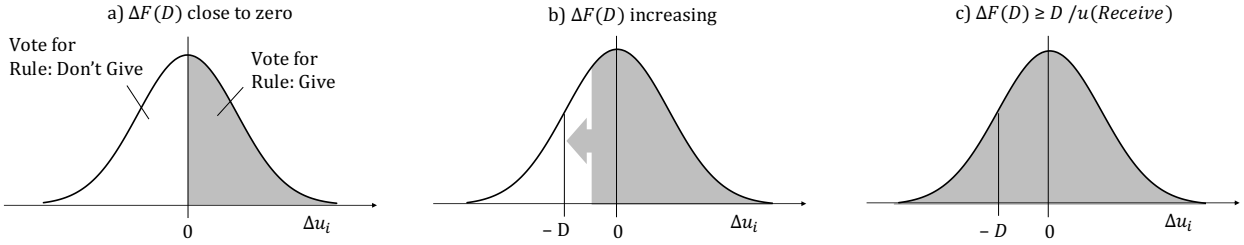


Figure B.5: Theory: Share of Population voting for Rule:Give

B.6 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)

B.7 Instructions and screenshots

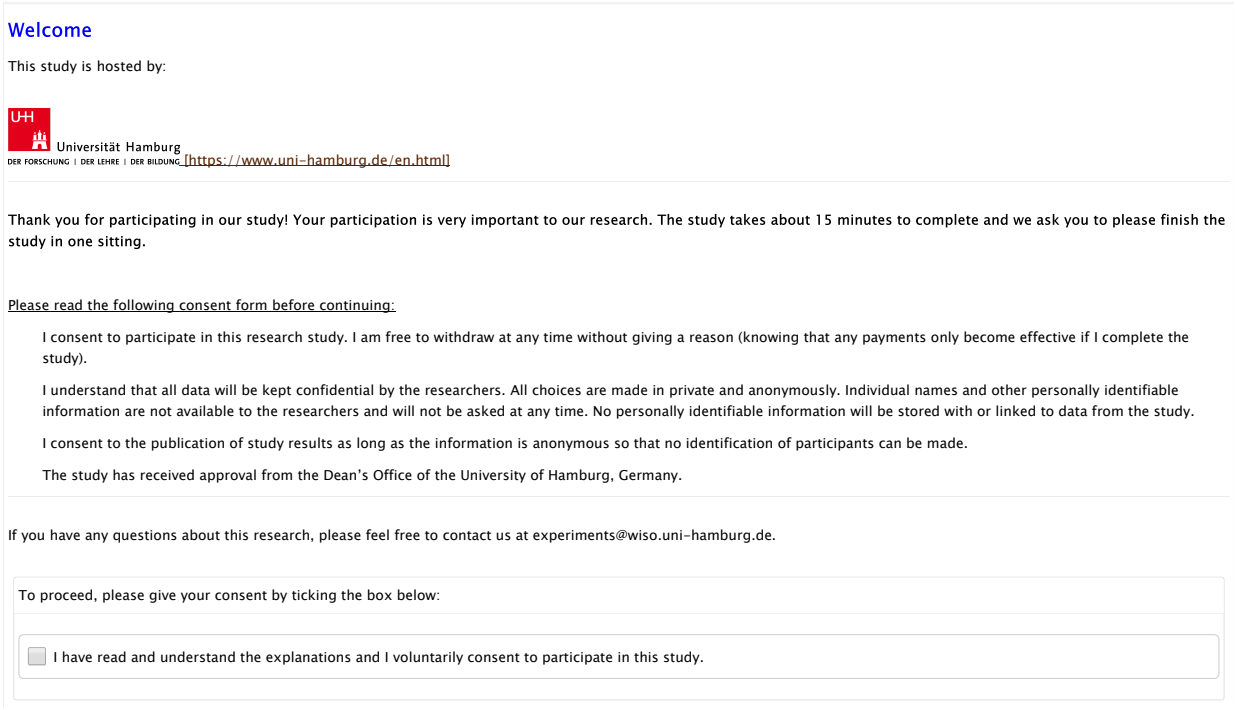


Figure B.6: Screenshot: Welcome and Consent Form

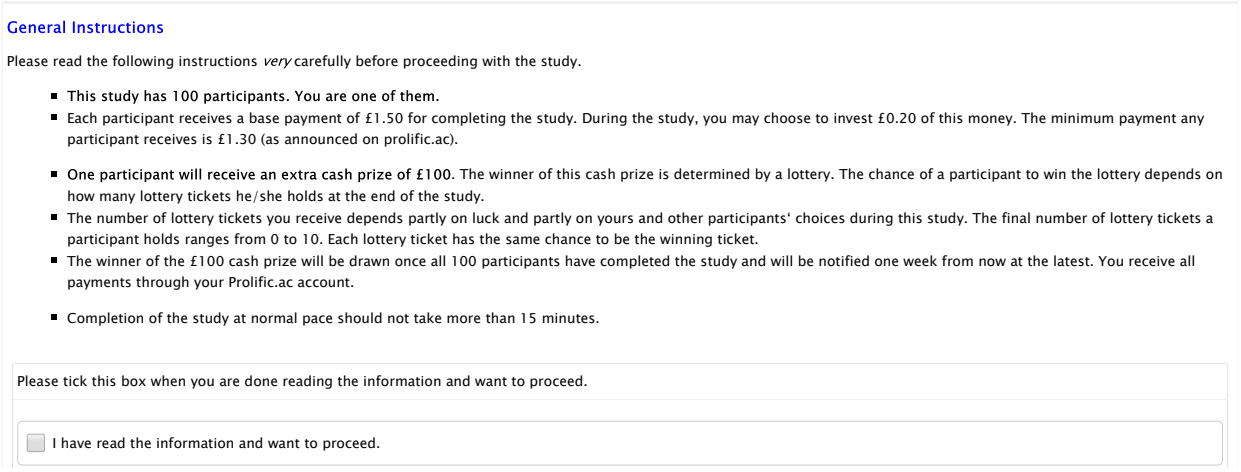


Figure B.7: Screenshot: General Instructions (Pay4Vote)

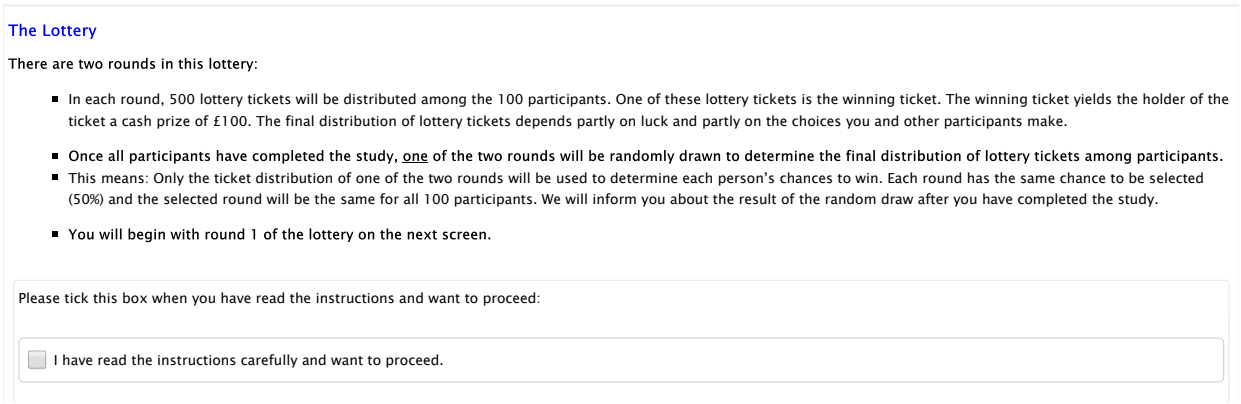


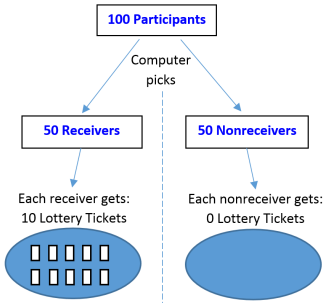
Figure B.8: Screenshot: Instructions about the Lottery

Distribution of lottery tickets

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

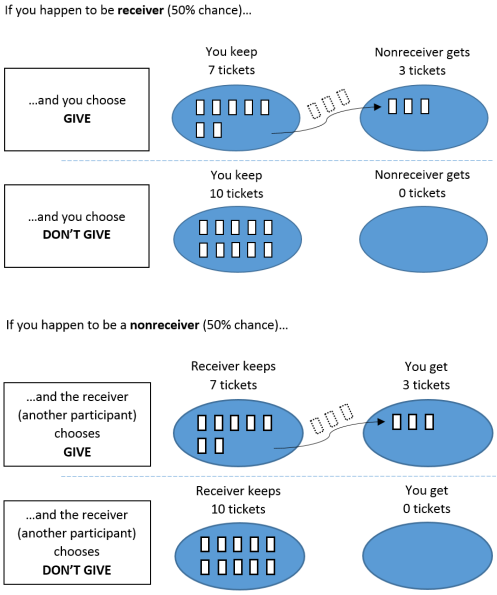
Step 1: The computer picks 50 receivers and 50 nonreceivers:

- 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 B.9: Screenshot: Instructions about the Distribution of Lottery Tickets

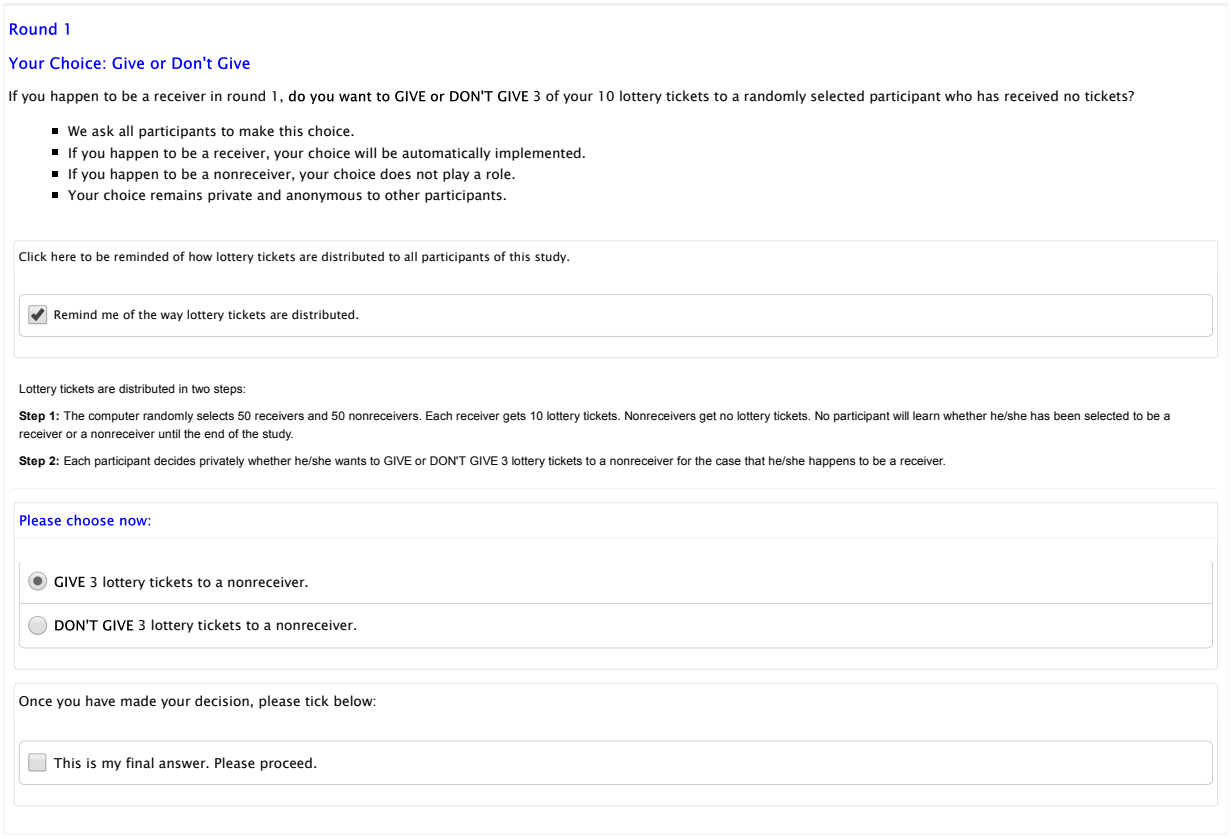


Figure B.10: Screenshot: Choice $Give_i|NoRule \in \{0, 1\}$ (Round 1)

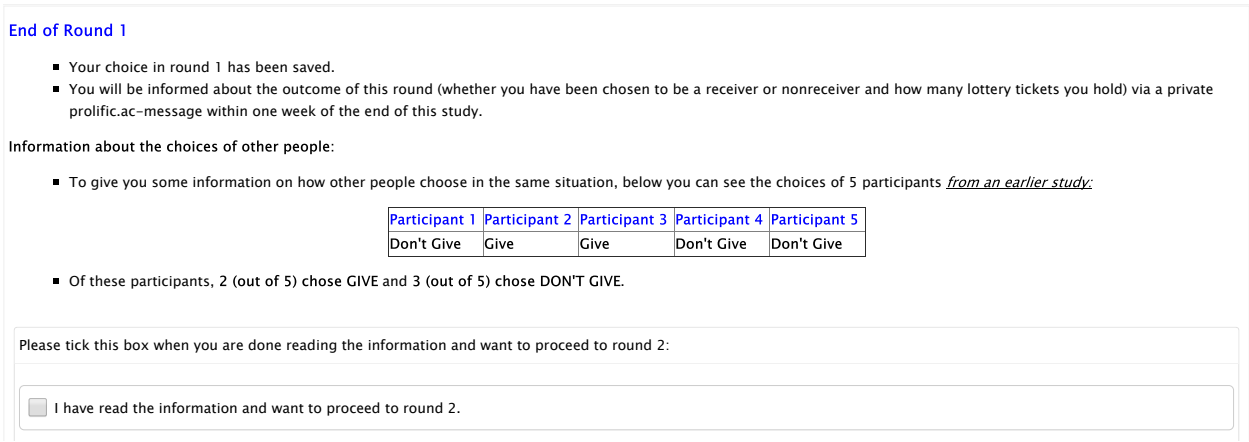


Figure B.11: Screenshot: Randomized social information $info_i \in \{2, 4\}$ (Round 1)

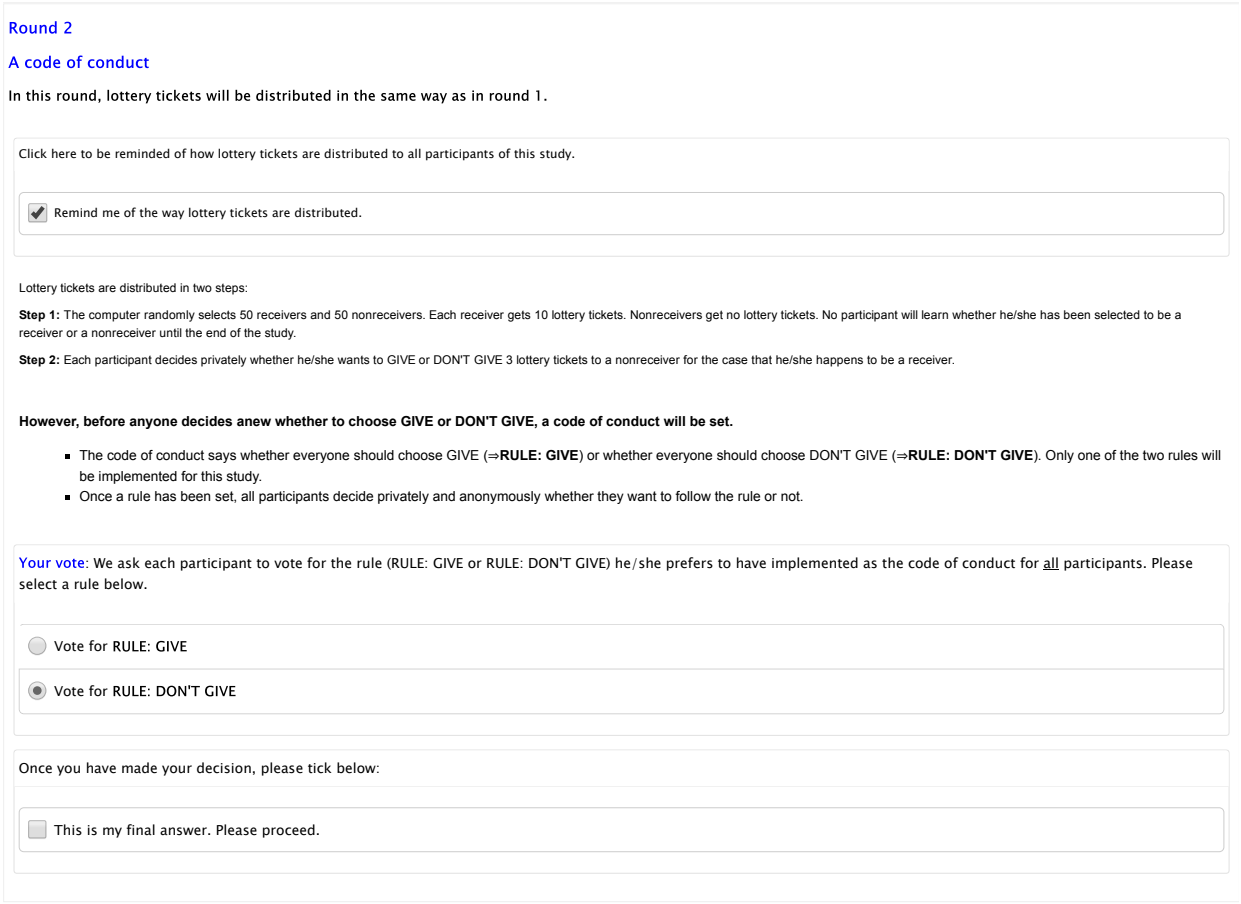


Figure B.12: Screenshot: $Vote_i \in \{Rule{:}Give, Rule{:}Don't\}$ (Round 2)

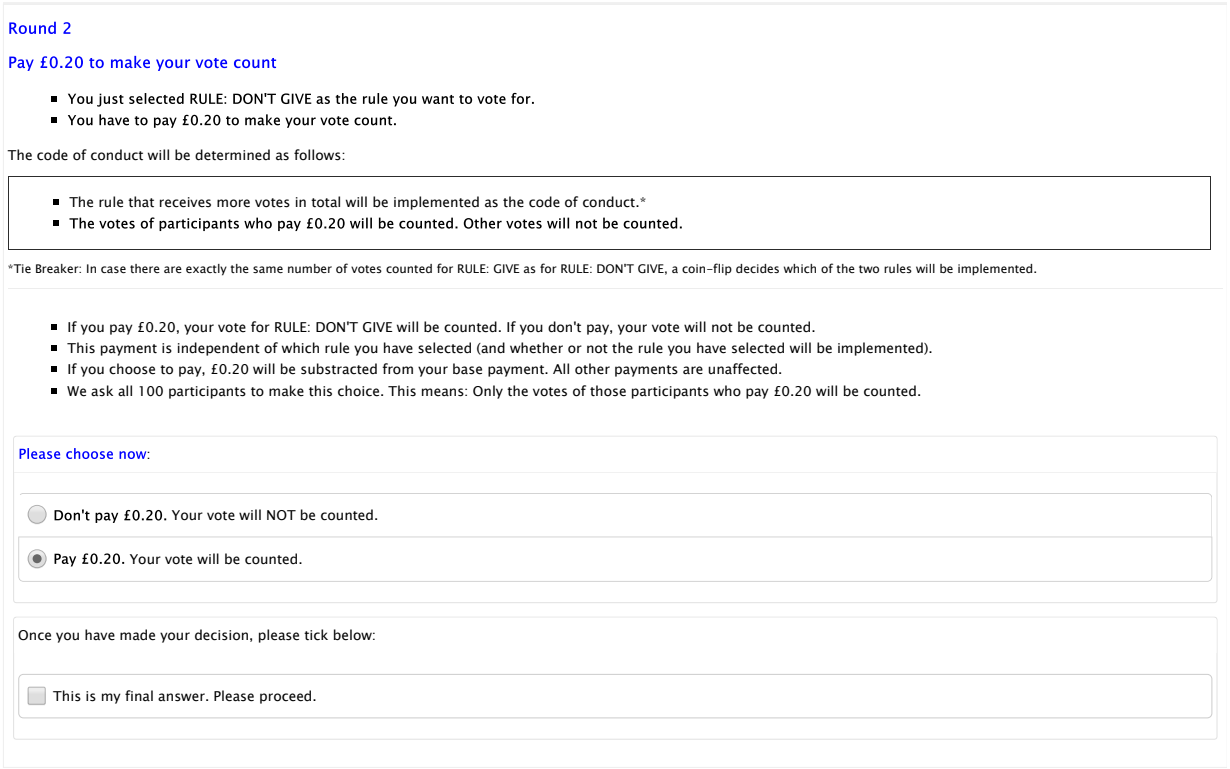


Figure B.13: Screenshot: $Accept\_Pay4Vote \in \{0, 1\}$ (Round 2, Pay4Vote)

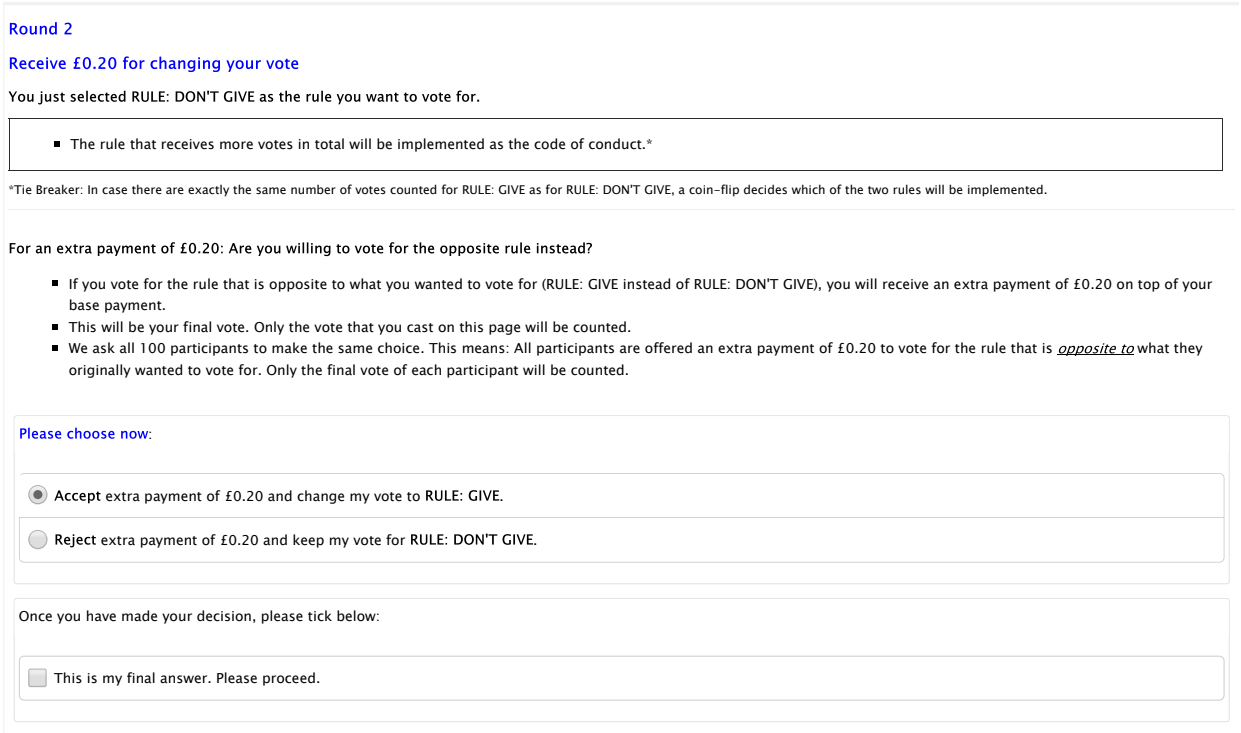


Figure B.14: Screenshot: $Accept\_MoneyOffer \in \{0, 1\}$ (Round 2, MoneyOffer)

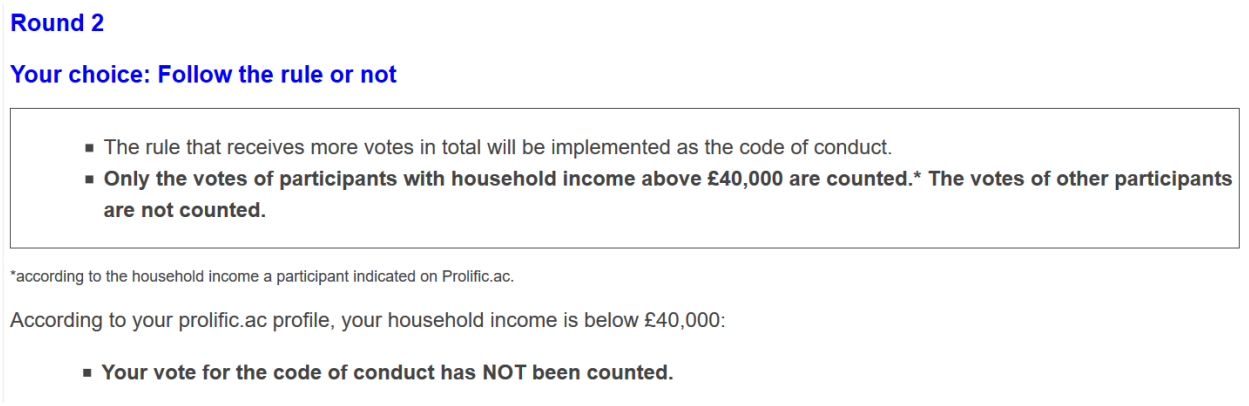


Figure B.15: Screenshot: Information about intervention Exclude_Poor (Round 2)

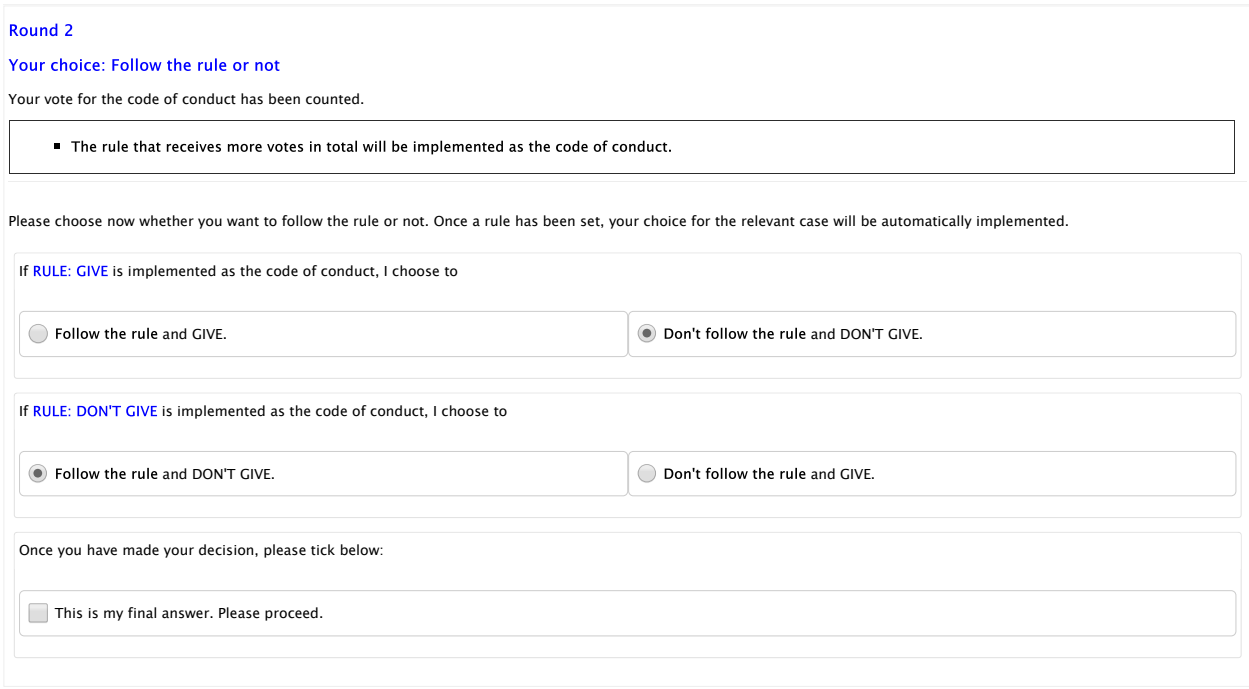


Figure B.16: Screenshot: $Give_i|Rule \in \{0,1\}$ (Round 2, StdMajority)

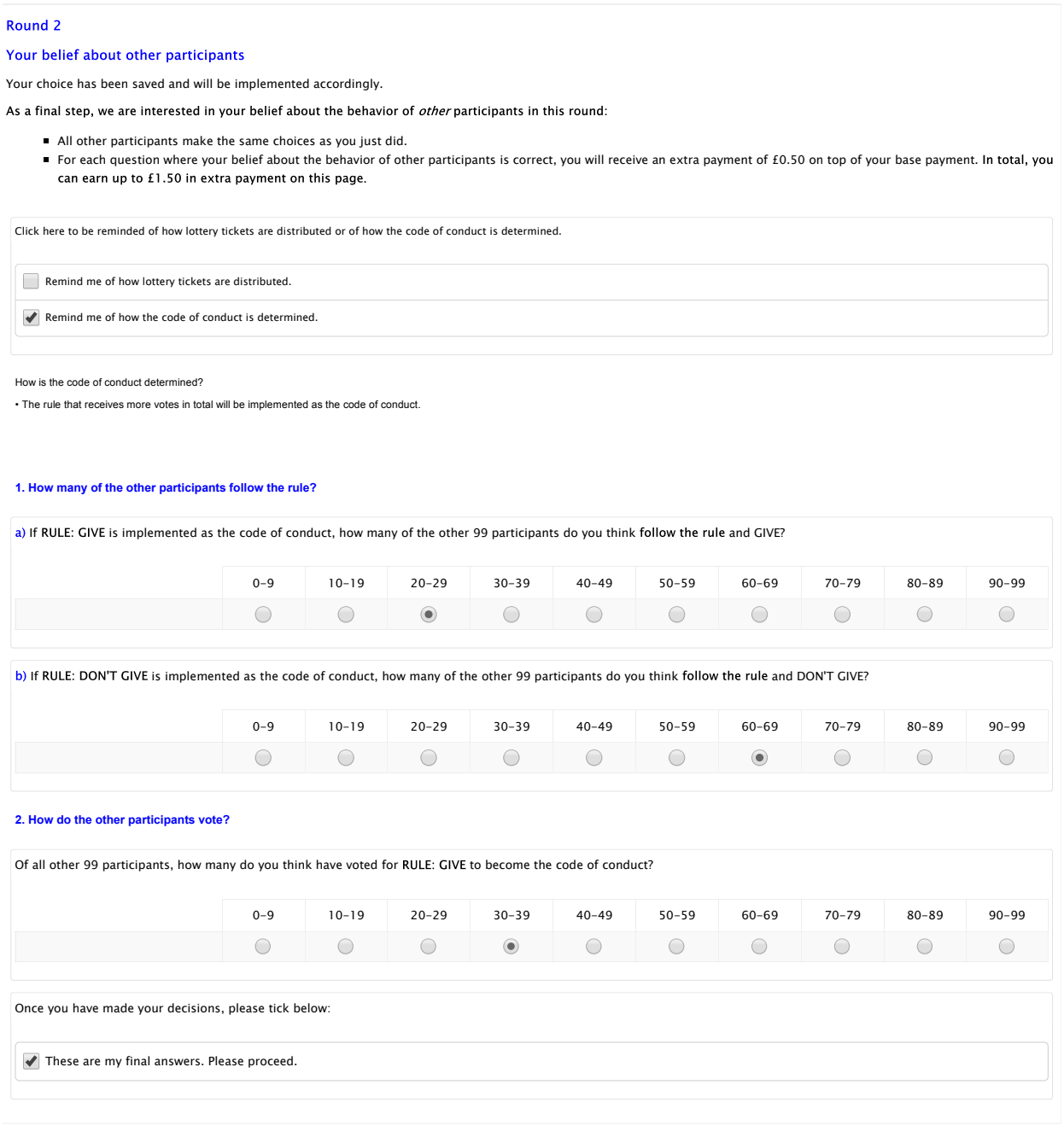


Figure B.17: Screenshot: Beliefs about Others (Round 2, StdMajority)

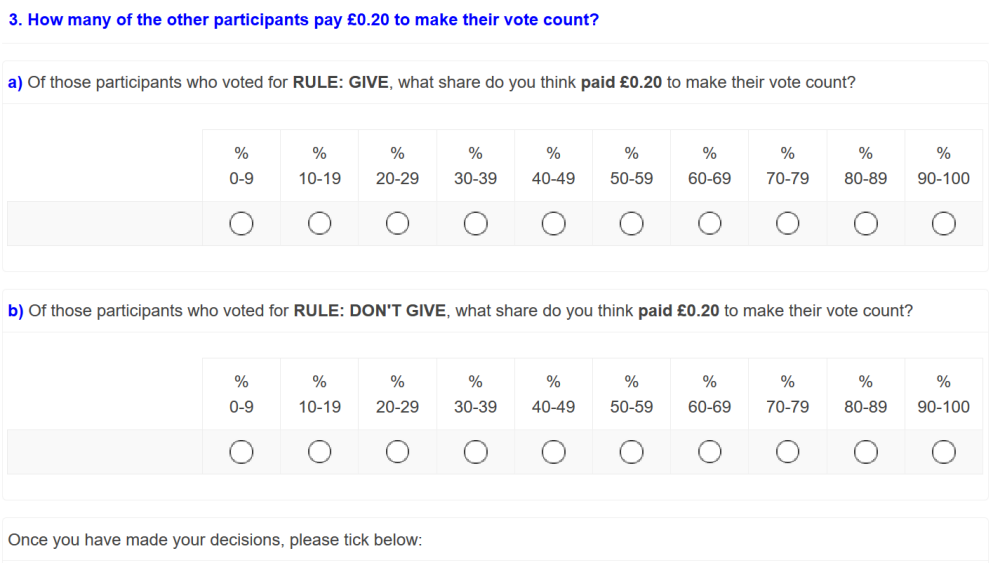


Figure B.18: Screenshot: Beliefs about Intervention (Round 2, Pay4Vote)

References (Online Appendix)

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.

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

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.

Dal Bó, Pedro, Andrew Foster, and Kenju Kamei. 2019. "The Democracy Effect: a weights-based identification strategy." NBER Working Paper No. 25724.

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

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

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