Social norms and elections: How elected rules can make behavior (in)appropriate*

Arno Apffelstaedt Jana Freundt Christoph Oslislo

27th February 2021
(revised: 21st September 2021)

Abstract

Can elections change people's ideas about what is ethically right and what is wrong? A number of recent observations suggest that social norms can change rapidly as a result of election outcomes. We explore this conjecture using a controlled online experiment. In our experiment, participants rate the social appropriateness of sharing income with poorer individuals. We compare the ratings for situations in which a rule has been elected that asks people to share with ratings when the elected rule asks people not to share. We also compare both situations with ratings in a decision environment in which there is no official rule at all. In the absence of an elected rule, sharing is widely considered socially appropriate, while not sharing is considered socially inappropriate. We show that elections can change this social norm: They shift the modal appropriateness perception of actions and, depending on the elected rule, increase their dispersion, i.e. erode previously existing consensus. As a result, actions previously judged socially inappropriate (not sharing) can become socially appropriate. This power prevails, albeit in a weakened form, even if the election process is flawed (introducing a voting fee or "poll tax", bribing voters, disenfranchising poorer voters). An additional treatment suggests that both the social information contained in election results and the social appropriateness of following rules *per se* play a role in shifting social norms.

Keywords: social norms, elections, prosocial behavior, rule compliance
JEL Codes: D02, D91, C91

*Apffelstaedt: University of Cologne, Center for Social and Economic Behavior (C-SEB) and ECONtribute. Address: Albertus-Magnus-Platz, 50931 Cologne, Germany. Email: apffelstaedt@wiso.uni-koeln.de. Phone: +49 221 470 1842. Freundt: 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. Phone: +41 26 300 8267. Oslislo: University of Cologne, Institute for Economic Policy. Adress: Pohligstraße 1, 50969 Cologne, Germany. Email: oslislo@wiso.uni-koeln.de. Phone: +49 221 470 5351. We thank the Graduate School of Economics and Social Sciences of the University of Hamburg and the Chair of Industrial Economics at the University of Fribourg for funding. Apffelstaedt further thanks the German Research Foundation: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. We thank the editor and three anonymous referees for many excellent comments and suggestions that substantially improved the paper.

1 Introduction

Can elections shift people's ideas of what is ethically right and what is wrong? A number of recent observations suggest that social norms, often considered persistent long-standing social constructs, can change rapidly as a result of election outcomes. In 2016, shortly after the United Kingdom voted for Brexit, the country experienced a sharp rise in hate crime, which many observers attribute to a Brexit-induced increase in the social acceptability of xenophobic views and actions. As a result of the referendum, "anti-immigrant and anti-foreigner rhetoric had become 'normalised' ", making Britain effectively "a more racist country", the United Nations claim.[1] Similar claims were made after the election of Donald Trump as president of the United States that same year.[2] There are also examples where elections appear to have lead to more tolerant norms: Baskaran and Hessami (2018) and Kedia and Pareek (2020) observe that electoral successes of female candidates in Germany and the US, respectively, seem to have improved norms regarding the treatment of women and reduced bias against them in elections as well as in the workplace. Jung and Tavits (2021) document an increase in the social acceptability of abortion following a pro-abortion vote in the Irish referendum of 2018. If the general conjecture that election outcomes can influence norms is true, it most likely holds for other domains as well, be it norms regarding same-sex relationships (Aksoy et al., 2020), compliance with Covid-19 regulations (Galbiati et al., 2020), or everyday behavior such as alcohol consumption (Lane et al., 2021).

A fundamental challenge in interpreting the relationship between norms and election outcomes is that elections are generally not exogenous to the society in which social norms develop. Observed differences in behavior can therefore also be attributed to other (possibly unobservable) factors that correlate with election outcomes, making it difficult to prove causality. Moreover, the fact that in the field, norms can typically only be observed indirectly from behavior obstructs the separation of presumed norm changes from changes in personal moral views or behavioral preferences. In this paper, we address these challenges and provide clean causal evidence on the effect of elections on social norms using a simple experiment.

Our experimental approach is based on the Krupka-Weber method (Krupka and Weber, 2013), which elicits social norms by asking subjects to estimate the degree of social approval of different actions in a hypothetical choice situation. Subjects are incentivized to provide a rating that is identical with the most common rating in their session, making the elicited social approval a direct elicitation of social norms or "normative expectations". We apply this method to investigate how social norms concerning prosocial vs. selfish behavior are affected by the outcome of majority

[1] See Brown, D. and Coates, S., "UK 'more racist after Brexit"', *The Times*, May 12, 2018. Available at: `https://www.thetimes.co.uk/article/uk-more-racist-after-brexit-qb7hd7xl7` (accessed February 18, 2021).

[2] Many popular press articles reported an increase in racially motivated violence and sexism after Trump's election, the cause of which was attributed to a change in social norms. The election of a person who is "openly hostile to women [...] normalizes abusive behavior and gives implicit permission for others to perpetuate it", the Huffington Post (Nov 16, 2016), for instance, writes. See Jeltsen, M., "Trump's Election Raises Fears Of Increased Violence Against Women", *The Huffington Post*, Nov 16, 2016. Available at: `https://www.huffpost.com/entry/trump-women-rights-violence-fears_n_582a0f63e4b02d21bbc9f186` (accessed February 18, 2021).

elections. We speculate that focal points in the coordination game might be created via "cultural values" such as prosociality, obedience to societal rules or the importance of democratic principles. A major benefit of using the Krupka-Weber method is that we elicit social norms directly using social approval ratings instead of indirectly using behavior. We thereby complement existing studies that also target election-induced norm changes but due to their indirect elicitation method cannot disentangle them from, say, changes in personal moral views or behavioral preferences.

The specific setting for our norms experiment derives from a behavior experiment which we introduce in detail in another paper (Apffelstaedt and Freundt, 2020). In the behavior experiment, participants must choose whether to act prosocially and give away a fixed part of their experimental endowment (3 out of 10 lottery tickets) to a lesser endowed individual ("Give"), or to act selfishly and keep the entire endowment to themselves ("Don't Give"). Participants make this decision twice, once in a "rule-free" environment (stage 1) and once again after a "code of conduct" has been elected by a referendum among the participants (stage 2). Participants in our norms experiment rate the social appropriateness of (not) giving away income in this setting. Depending on the treatment, they are either asked to make ratings in a situation in which there is no rule (stage 1) or to condition their ratings on a rule having been elected and put into force (stage 2). This setting allows us to exogenously manipulate the election outcome as well as whether there has been an election at all. Holding everything else constant, we can thus directly measure changes in social norms as a result of election outcomes by comparing norm ratings across the different choice environments. In this way, our lab design provides what is, as far as we know, the first clean test of the effects of elected rules on social norms and complements existing field research on this topic.

In our main treatment (StdMajority), we elicit social approval ratings conditional on participants having elected a code of conduct by a simple majority vote. The elected code tells subjects either that "everybody should choose *Give*" (Rule:Give) or that "everybody should choose *Don't Give*" (Rule:Don't). To identify the effect of this election on social norms, we compare the ratings elicited in StdMajority to each other as well as to the ratings in a benchmark treatment NoRule in which subjects rate actions Give and Don't Give in the absence of an election. We find that the election indeed has a strong impact on social norms. Most impressively, and much in line with the anecdotal evidence reported above, we show that majority-elected rules can cause actions previously judged socially inappropriate (Don't Give) to become socially appropriate. This is the case, specifically, if the elected rule asks subjects to not give (Rule:Don't). We also find a statistically significant effect on social norms of Rule:Give being elected into power, but the size of the effect is much smaller.

Having established this main result, we answer two additional questions. First, we ask whether norm shifts require "free and fair" majority votes, where by by "free" we relate to the notion that all subjects are able to vote for the rule of their choice and by "fair" to the notion that all votes have equal power and are counted accurately (see the *Encyclopedia of American Civil Rights and Liberties*, Stooksbury et al., 2017). To answer this question, we run three additional treatments

describing a situation in which the electoral process is subject to salient deficiencies, namely a voting fee or "poll tax," a bribe that induces subjects to change their vote, or the disenfranchisement of poor voters. We find that in the case of such deficiencies, elected rules still considerably shift social norms about prosocial actions, but their power to do so is weakened. Second, we ask the question of whether the norm shifts we observe in our experiment can predict changes in behavior. Drawing on data from the behavior experiment in Apffelstaedt and Freundt (2020), we show that social norms indeed predict giving decisions under each elected rule in a different sample of participants.

We discuss and analyze the role of two possible mechanisms that might underlie the election-induced shifts in normative expectations we observe. As one mechanism, we discuss the possibility that elections carry informative value about the underlying preferences and values in the society. As a second mechanism, we discuss the possibility that showing compliance with rules per se (irrespective of their specific content) may be a source of social approval. To shed light on these two mechanisms, we present the results of an additional treatment variation in which we aim to disentangle the role of the informational value of election outcomes and that of mere rule compliance. Our results suggest that both mechanisms play a role and are important for understanding the impact of elections on social norms.

We conclude by discussing our main findings as well as other interesting patterns in our data in light of the above mechanisms. In our discussion, we point to an important open question: Do elections necessarily lead to more agreement on social norms? In the main body of the paper, we focus on modal, mean, and median ratings to analyze the impact of elections on social norms. However, our data also show that elections can lead to a greater dispersion of ratings. In other words, instead of leading to more agreement on what is ethically right and what is wrong, elections may also undermine pre-existing norm consensus. In principle, we think it is possible that changes to the perceived norm consensus may play at least as large a role in shaping observed behavior after elections as changes to the norm itself.

Related Literature. Social norms—defined as "shared understandings about actions that are obligatory, permitted, or forbidden" (Ostrom, 2000, p.144)—govern many parts of our everyday lives, ranging from economic and political decisions to cultural practices and are thus an important element of any social group. In this paper, we focus on so-called *injunctive norms* or normative expectations in a population.[3] Building on work by Cialdini and Trost (1998) and Ostrom (2000), Krupka and Weber (2013) define injunctive social norms as collective perceptions or judgments regarding the appropriateness of actions. This requires that norms are "jointly recognized, or collectively perceived, by members of a population" (Krupka and Weber, 2013, pp.498-499).

In this context, our results contribute to four different literatures. First, our work complements

[3]We do not consider descriptive norms or empirical expectations, i.e. expectations about what others actually do. For a discussion of both concepts see also Bicchieri (2010, 2016); Bicchieri and Dimant (2019). Several studies have shown that both, injunctive and descriptive norms, can influence behavior (e.g., Bicchieri and Chavez (2010); Bicchieri et al. (2020a); Bursztyn et al. (2020b); Cialdini et al. (1990); Krupka and Weber (2009)).

a handful of previous papers that examine factors in the institutional environment that can cause long-standing social norms to change. For instance, recent research has shown that social norms can change quickly as a result of policy interventions. One example are studies of so-called "norm-nudges" (for overviews see Bicchieri and Dimant, 2019; Hauser et al., 2018). In particular, our work adds to a recent literature that examines how public decision-making processes (such as elections, initiatives, or referenda) can lead to very rapid changes in social norms, which in turn can lead to changes in behavior. Jung and Tavits (2021) argue, based on results of a panel survey, that the outcome of the 2018 Irish abortion referendum changed Irish citizens' perceptions of the social norm regarding abortion. Baskaran and Hessami (2018) and Kedia and Pareek (2020) show that elections have the potential to affect norms regarding gender as well as gender-related behavioral outcomes. Using an experimental setup, Bursztyn et al. (2020a) show that Donald Trump's victory in the 2016 Presidential election increased individuals' willingness to publicly express xenophobic views as well as accept related expressions by others.[4] In a similar vein, Albornoz et al. (2020) argue that the increase in hate crime following the Brexit referendum should be attributed to a change in social norms: They show that hate crime spiked especially in regions in which the outcome of the election came as a surprise and thus, can be explained by an update of beliefs about whether xenophobic views are extreme or mainstream (Albornoz et al., 2020). In the existing studies, the effect of elections on social norms is inferred indirectly from observed behavior or revealed preferences. Our experimental design instead allows us to *directly* elicit social norms and thus examine the immediate response of norms to elections. Using a well-established norm elicitation method (Krupka and Weber, 2013), we highlight the role of shifts in the social appropriateness of actions in bringing about behavioral change. To our knowledge, our paper is the first to directly measure changes in social norms in response to the election of a behavioral rule.

Second, our paper contributes to a growing experimental literature on the effect of social norms on behavior. This literature assumes that most individuals tend to learn and follow social norms, leading, for instance, to a willingness to constrain selfish behavior (Ostrom, 2000, p.143,149). Following this conjecture, a number of recent experiments show that many people do indeed have an intrinsic preference to conform to what is collectively perceived as socially appropriate, and that norm conformity can explain behavior in a variety of social contexts (Kimbrough and Vostroknutov, 2016; Krupka et al., 2017; Gaechter et al., 2017). Krupka and Weber (2013) find that social norms vary with different framings in dictator games and argue that this variation can provide a plausible explanation for observed differences in behavior. By providing information about the moral views of others, Bursztyn et al. (2020b) manage to directly manipulate perceived social norms and show that this in turn changes behavior. Importantly, these studies suggest that there is a fairly stable preference for following social norms across different settings. Our results are consistent with these findings: In our experiment, we observe systematic shifts in individual perceptions of social norms and present evidence that—by assuming a stable preference for following social norms exists—

[4]For related results see Crandall et al. (2018) and Huang and Low (2017).

these shifts are well suited to explain behavioral changes after elections. Note that we are primarily interested in *changes* in perceived social norms in a society, which is conceptually distinct from studies that examine changes in individual adherence to existing *stable* norms (e.g. Bicchieri et al., 2020b).

Third, our work adds to the literature on determinants of prosocial behavior and the analysis of the broader set of motives that shape people's social conduct. Norms and social pressure have been found to be important driving factors of altruistic behavior by attaching honor to good deeds and shame to selfish behavior (Bénabou and Tirole, 2006). A number of experimental studies has since then investigated the crucial role of social norms for prosocial behavior in various contexts (Krupka and Weber, 2009; Gaechter et al., 2012; Kimbrough and Vostroknutov, 2016). Our results confirm the view that prosocial behavior can be highly context-dependent and that a shift in social norms can lead to large shifts in prosocial outcomes. We add to the literature by showing that norms regarding prosocial behavior can be influenced by elections.

Fourth, we link to an interdisciplinary literature that examines how rules and laws can change attitudes and social norms and thus influence behavior beyond the imposition of explicit sanctions. There are a number of different theoretical approaches to explain the expressive power of rules (see, e.g., Cooter, 1998; Benabou and Tirole, 2011; McAdams, 2015). Using survey data, Galbiati et al. (2020) show that social distancing rules during the Covid-19 pandemic causally affected social norms regarding social interactions. Several other studies empirically investigate how laws shape attitudes on morally controversial issues, e.g., by studying the efficacy of anti-discrimination laws (Aksoy et al., 2020; Barron and Hebl, 2010, 2013). However, these studies are usually not able to disentangle precise channels to explain where changes in behavior or personal opinions stem from. An exception is Lane et al. (2021), who provide direct evidence that the legal status of an action causally affects its normative appropriateness. We investigate how *democratically elected* rules, in particular, affect social norms and thus behavior.

The paper proceeds as follows. In the next section, we explain the experimental setup in detail. In section 3 we present our results. In subsection 4.4, we discuss our findings with a focus on possible mechanisms and ways in which elections can influence and change norms. Section 5 concludes.

2 Design

Our experimental approach is based on the Krupka-Weber method (Krupka and Weber, 2013), which identifies social norms by asking subjects to estimate the degree of social approval for different actions in a hypothetical decision situation. We apply this method to examine how social norms concerning prosocial vs. selfish behavior are affected by the outcome of majority elections in which a "code of conduct" is elected.

2.1 The decision context

The specific context for our study of norms is an earlier behavior experiment (Apffelstaedt and Freundt, 2020) in which we investigate how elected rules can change behavior. For the present paper, we use the decision environment of the behavior experiment to investigate the extent to which such elections can also shift norms. To provide an understanding of the actions for which we elicit social norms and the different situations in which these actions were evaluated, we now briefly introduce the behavior experiment on which the norms experiment is based. After this, we describe in detail how we elicit the social norms for this setting.

Figure 1, panel a) shows the timeline for the behavior experiment. The experiment revolves around a simple paradigm:

Action: Give or Don't Give. Among the 100 subjects of each treatment, income is distributed unequally. Before learning whether one is rich or poor, each subject has to decide privately whether to *Give* or *Don't Give*, where "Give" means that, conditional on being rich, the subject shares her income with another poorer subject and "Don't Give" means that the subject does not share her income. We operationalize this paradigm using a lottery: In each treatment, we raffle a cash prize of £100 among the 100 participating subjects. At the beginning of the experiment, subjects learn that lottery tickets for the raffle will be distributed unequally: While 50 subjects receive 10 lottery tickets each, the remaining 50 subjects receive no (zero) lottery tickets. Actions Give and Don't Give are then introduced as follows: "*If you happen to be a receiver of lottery tickets, 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?*"[5]

Stages 1 and 2: No rule vs. elected rule. Each subject takes the decision to Give/Don't Give twice. In a first stage, right after being informed about the unequal distribution of income, and in a second stage, after taking part in a referendum in which a "code of conduct" is elected. The code of conduct can either ask that "*everyone should choose GIVE*" (Rule:Give) or, contrarily, that "*everyone should choose DON'T GIVE*" (Rule:Don't). Subjects can cast a vote for either of the two rules. Subsequently, they decide, for each of the two potential election outcomes (strategy method) whether they want to Give or Don't Give. Subjects are informed that the final rule is non-binding: "*Once a rule has been set, each individual can decide privately and anonymously whether he/she wants to follow the rule or not.*"

[5]Technically, this is a binary dictator game with role uncertainty which uses lottery tickets as the experimental currency. For possible problems with role uncertainty, see Iriberri and Rey-Biel (2011). We took great care in explaining the game to subjects, emphasizing that "*if you happen to be a nonreceiver (50% chance), your choice whether to GIVE or DON'T GIVE does <u>not</u> play a role*", and employed multiple control questions to make sure that they understood the role uncertainty correctly.

Elections: StdMajority, Pay4Vote, MoneyOffer, ExcludePoor. Between treatments, we vary the exact procedure with which the code of conduct is elected. The baseline treatment StdMajority implements a simple majority vote among the 100 subjects of the treatment ("*all 100 individuals who take part in the lottery are asked to vote for the rule they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.*"). The other three treatments introduce salient deficiencies to the standard majority vote in order to analyze the extent to which behavioral changes hinge on the election being "free and fair": Treatment Pay4Vote introduces a voting fee or "poll tax" (subjects need to pay £0.20 to make their vote count). Treatment MoneyOffer involves a bribe (subjects are offered a bonus payment of £0.20 if they reverse their vote). Finally, treatment ExcludePoor excludes all subjects from the election who have a household income below £40,000 (about half of our subjects). Subjects are informed that these conditions apply to all participants in the treatment. They are not informed, however, about the share of votes that are ultimately uncounted or manipulated, nor about the extent to which the intervention affects the final vote share.

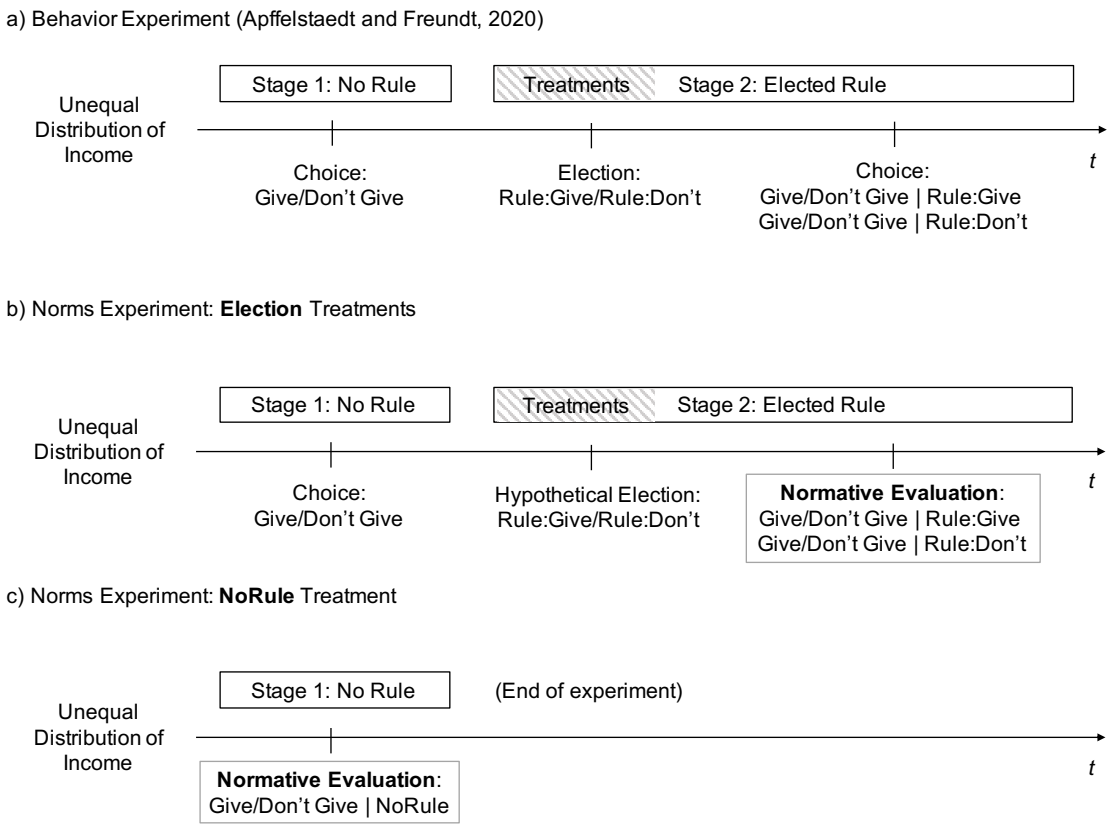


Figure 1: Timelines for behavior (panel a) and norms experiment (panels b and c). The behavior experiment is introduced and analyzed in detail in Apffelstaedt and Freundt (2020). The norms experiment (this paper) uses the decision context of the behavior experiment to investigate the extent to which elections can change social norms.

2.2 The norms experiment

We aim to understand how the election outcome (stage 2) affects the normative evaluation of actions Give and Don't Give and how the introduction of an elected rule changes social norms compared to the case of NoRule (stage 1). Toward this end, we invited 500 subjects into five treatments. 400 subjects (100 per treatment) provided normative evaluations for the situation in which a rule had been elected (election treatments), while 100 additional subjects provided normative evaluations for the case of no rule (NoRule treatment). Table 1 provides an overview of the treatments in the norms experiment.

Figure 1, panels b) and c) show the timeline for the norms experiment. We let subjects progress through the decision process as far as possible as active decision makers (until the point at which they are asked to provide normative evaluations). In both, the election treatments and the NoRule treatment, subjects participate in a raffle for a cash prize worth £100, identical to the subjects in the behavior experiment. We begin the experiment by distributing lottery tickets (i.e. income) unequally, meaning that 50% of subjects will be "poor" and 50% will be "rich" prior to redistribution. Before learning about their own position, subjects in the NoRule treatment (Figure 1, panel c) then provide social norm ratings for a "hypothetical decision situation" in which subjects can redistribute income by choosing to Give/Don't Give (stage 1). Subjects in the election treatments (Figure 1, panel b) move through stage 1 as active decision makers, choosing themselves whether to Give or Don't Give. After this, they provide social norm ratings for a "hypothetical decision situation" in which a code of conduct has been elected (stage 2).[6]

Social norms: Elicitation. We elicit social norms using the coordination game method suggested by Krupka and Weber (2013). We let subjects progress through the behavior experiment until the relevant point in the timeline. Following Krupka and Weber (2013), the respective decision for which they have to provide evaluations is then presented to subjects as a "hypothetical choice situation".[7] For this situation, we ask subjects to evaluate the "social appropriateness" of actions Give and Don't Give on a 6-point scale.[8] The scale allows subjects to evaluate the ac-

[6]Letting subjects progress through the experiment as active decision makers until the point at which they are asked to provide normative evaluations helps us measure social norms at the different natural stages of a society's development. The norms so obtained thus correspond exactly to the norms which arguably regulate individual behavior at the respective stages of this process. Letting subjects experience the unequal distribution of income themselves arguably helps them understand the hypothetical situation and the stakes involved better, meaning that again, the elicited norms should correspond better to those norms that also regulate behavior. In Appendix A.2, we study the role of stage 1 behavior for our results. We show that stage 1 behavior cannot explain our main findings about norm shifts with respect to the election outcome.

[7]According to the definition by Krupka and Weber (2013), a social norm is an empirically measurable collective judgment that assigns to each action a degree of appropriateness or inappropriateness. The idea is to elicit this "collective judgment" as the focal point of a coordination game in which each agent is incentivized to guess the modal normative assessment in society. Framing the situation as "hypothetical" helps to ensure that the focal point is indeed the modal *normative* assessment and nothing else, such as, for instance, the answer to the question what most people do (behavioral expectation) rather than what they ought to do (normative expectation).

[8]Note that the definition and measurement of social norms suggested by Krupka and Weber (2013) differs from previous binary conceptions of social norms by allowing for actions to vary in the degree to which they are regarded

tion negatively as "very socially inappropriate", "socially inappropriate", or "somewhat socially inappropriate", or positively as "somewhat socially appropriate", "socially appropriate", or "very socially appropriate". Two measures are taken to ensure that the elicited rating reveals a social norm—that is, a coordinated belief about what is wrong and what is right: First, we tell subjects that "*by socially appropriate, we mean behavior that most people agree is the 'correct' or 'ethical' thing to do*". Second, subjects are incentivized to provide a rating that is identical with how most of the other subjects evaluate the action. Specifically, we pay the subject a bonus payment of £2.00 if, for a randomly selected rating, the subject's rating matches the modal rating among the 99 other subjects in her treatment.

Description	Treatment	Elicits social norms for:
Social norms in the absence of election (benchmark rating)	NoRule	Give/Don't Give | NoRule
Social norms after standard majority vote (baseline election)	StdMajority	Give/Don't Give | Rule:Give Give/Don't Give | Rule:Don't
Social norms after simply majority vote in which...		
Voters have to pay £0.20 to make vote count	Pay4Vote	——— " ———
Voters are offered £0.20 to vote for the opposite rule	MoneyOffer	——— " ———
Voters with household inc. < GBP 40K excluded from ballot	ExcludePoor	——— " ———

Table 1: Overview of treatments in the norms experiment

Social norms after elections: Election treatments. We run four election treatments corresponding to the four different election procedures implemented in the behavior experiment. The four election treatments elicit social norms in stage 2, that is, for the situation in which a code of conduct has been elected asking people to Give (Rule:Give) or Don't Give (Rule:Don't). Subjects are presented with a "hypothetical decision situation" which describes the election procedure and are then asked, using the strategy method, to provide social approval ratings for actions Give and Don't Give conditional on Rule:Give/Rule:Don't being elected into power.[9] Our baseline election

as socially (in)appropriate. This is especially important for our purpose of studying changes in these perceptions dependent on the election procedure.

[9]Using the strategy method is in line with how choices conditional on Rule:Give and Rule:Don't, respectively, are elicited in the behavior experiment. In a survey of the literature, Brandts and Charness (2011) find little evidence that the strategy method generates systematically different behavioral responses than a direct elicitation. Most importantly, they do not find any case in which a treatment effect found with the strategy method is not observed with the direct-response method. Rauhut and Winter (2010) argue that when measuring social norms, the strategy method is preferable because the conditionality of responses can help make counterfactual states of the world more salient and thus improve the measurement of complex elements, such as the conditionality or the level of consensus of social norms. In our particular case, we do not believe that the results of a direct method would be very different from the results obtained by the strategy method. Even when not asking for a conditional response, the simple binary policy space (Rule:Give/Rule:Don't), together with the information that one of the two rules is chosen by election, will lead subjects to be aware of the counterfactual rule when making a normative judgment. We believe this "counterfactual thinking" is also consistent with the way people reason after a real-world election in which two

treatment is StdMajority, which provides us with the social norms after a standard simple majority vote. Treatments Pay4Vote, MoneyOffer, and ExcludePoor resemble the respective election treatments in the behavior experiment described above and allow us to investigate in how far salient deficiencies in the voting process (a voting fee or "poll tax", introducing bribes, excluding voters) affect the power of elected rules to shift norms.

Social norms before/in the absence of elections: NoRule treatment. To elicit social norms before/in the absence of an election (stage 1), we implement a fifth treatment (NoRule). Subjects in this treatment are presented with a "hypothetical decision situation" describing the basic choice between Give and Don't Give, and are then asked to provide a social approval rating for each action. The choice situation resembles stage 1 in the election treatments and in the behavior experiment. Consistent with stage 1 of the behavior experiment, no mention is made of an election (or rule) when describing the decision situation.

Implementation. Detailed instructions and screenshots can be found in appendix A.3. The experiment was conducted on the online survey platform Prolific using a randomly drawn sample of international participants. Prolific automatically provides us with basic (self-declared) demographic information about individual subjects. Additional to this basic information, which includes gender, age, and student status, we required that participants had filled in information about their nationality and country of residence. The four election treatments were conducted over a period of two weeks in September 2018. On average, subjects spent about 15 minutes to go through the experiment. In addition to the chance to win a cash prize of £100 and a possible bonus payment of £2.00 in the social norm task, subjects received a base payment of £1.60 for completing the experiment. The benchmark treatment (NoRule) was conducted as a separate treatment in November 2020. Since this experiment took only 10 minutes to complete, we reduced the base payment to £1.10. Data collection for each of the five treatments was preset to stop when the number of subjects reached 100.

The entire sample of 500 participants has a mean age of 28.73 years (SD 9.59), 46.60 percent of participants are female, and 38.87 percent are students. The largest share of participants have a British nationality (38.08 percent), followed by 11.62 percent US Americans.[10] The total share of "Western" subjects is 78.16 percent.[11] In treatments StdMajority, Pay4Vote, MoneyOffer and ExcludePoor, the share of subjects choosing action Give in stage 1 is .63, .67, .62, and .65, respectively. These shares are near-identical and not statistically different from each other.

conflicting policies (or politicians) were up for election.

[10] Not every participant had filled out all questions about demographics. Of 500 subjects, 494 subjects filled in information on their student status and 499 provided their nationality.

[11] Western = 1 if Nationality = United Kingdom (190 participants), United States (58), Austria (4), Australia (7), Belgium(6), Canada (19), Denmark (4), Finland (5), Germany (11), Greece (12), Ireland (4), Italy (30), Netherlands (8), Norway (1), New Zealand (1), Portugal (25), Sweden (4), Switzerland (1).

2.3 Predictions

There are (at least) two theoretical mechanisms through which majority-elected rules may influence public consensus about how socially appropriate an action is: (1) Rules that are elected by a majority vote provide information about which action has greater support in society (see, e.g., McAdams, 2015). (2) Rule compliance itself may be considered socially appropriate (see, e.g., Nadler, 2017). Both mechanisms should lead to an increase in social approval for the action that conforms with the rule, while at the same time decreasing social approval for the action that opposes the rule.[12] Our first prediction is therefore:

Prediction 1. *Majority-elected rules shift social norms. The election of Rule:Give (Rule:Don't) will shift upward (downward) the social approval rating of action Give and will shift downward (upward) the social approval rating of action Don't Give.*

Salient deficiencies in the voting process are likely to obstruct both of the above-mentioned channels. When a considerable share of voters does not participate in the election, or their votes are manipulated, the elected rule becomes less indicative of what action enjoys majority support in society.[13] Such deficiencies also make the rule itself seem less legitimate, which is why rule compliance may be perceived as less socially appropriate following a flawed election process (Tyler, 2006; Norris, 2014). Our second prediction follows:

Prediction 2. *Salient deficiencies in the voting process (introducing a voting fee, bribing voters, or disenfranchising poor voters) decrease the power of elections to change norms. The effect of elected rules on social approval ratings will be lower in Pay4Vote, MoneyOffer, and ExcludePoor than in StdMajority.*

3 Results

3.1 Do elections shift norms?

To what extent can elections change social norms? We begin our analysis by comparing the social approval ratings after a "free and fair" majority election of Rule:Give or Rule:Don't (StdMajority) to each other and to those obtained without an election (NoRule).

Figure 2 displays the mean and median of social approval ratings across treatments NoRule and StdMajority. The left-hand side of the figure shows how subjects rate action Give in the absence of an election (NoRule), when Rule:Give is elected by majority vote (StdMajority) and when

[12] The two mechanisms and a follow-up treatment designed to shed light on their relevance are discussed in more detail in section 4.

[13] In the behavior experiment, 35% of participants in Pay4Vote refused to pay a fee to make their vote count, 39% of participants in MoneyOffer were willing to reverse their vote in exchange for the small bonus payment, and 50% of voters were excluded due to a low household income in ExcludePoor (see Apffelstaedt and Freundt, 2020). Subjects in our norms experiment were not informed of these shares.

Rule:Don't is elected by majority vote (StdMajority). The right-hand side of the figure (grey bars) shows how subjects rate action Don't Give under the same conditions. Following the convention introduced by Krupka and Weber (2013), we have converted subjects' responses into numerical scores. A rating of "very socially inappropriate" received a score of -1, "socially inappropriate" a score of -2/3, "somewhat socially inappropriate" a score of -1/3, "somewhat socially appropriate" a score of 1/3, "socially appropriate" a score of 2/3, and "very social appropriate" a score of 1.

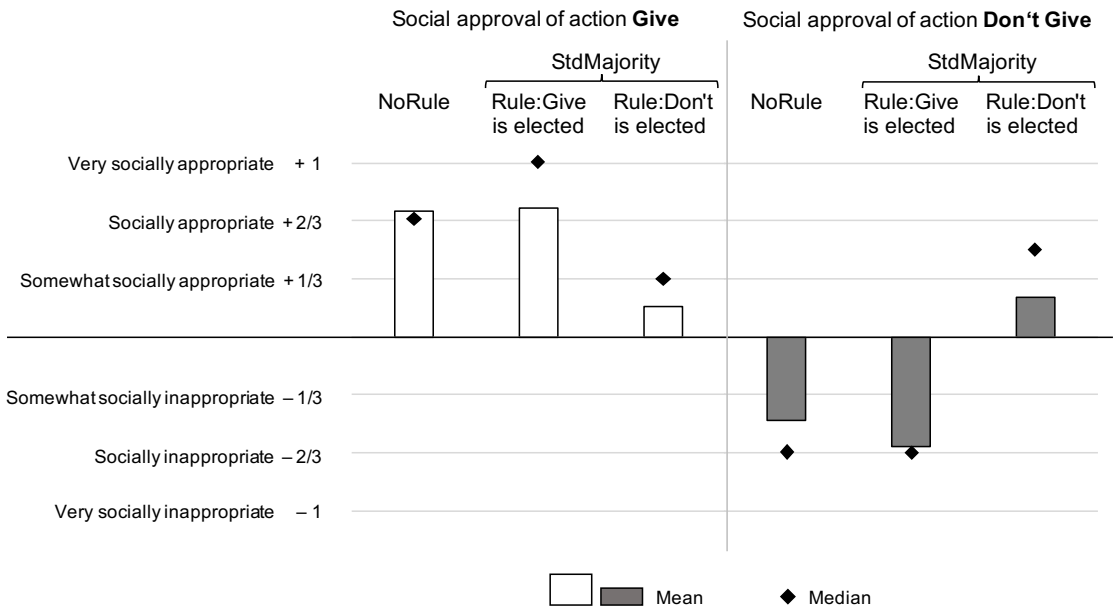


Figure 2: Elicited social approval (mean and median) of actions Give (left panel) and Don't Give (right panel) in the absence of an election (NoRule) and after a standard majority election of Rule:Give or Rule:Don't (StdMajority).

Recall our main prediction (Prediction 1): We predicted that majority-elected rules will shift upward the social approval of actions that comply with the rule and will shift downward the social approval of actions opposed to the rule. Specifically, we predicted that the election of Rule:Give (Rule:Don't) will shift upward (downward) the social approval of action Give and will shift downward (upward) the social approval of action Don't Give. This is exactly what we find. Consider first action Give (left-hand side of Figure 2): In the absence of an election (NoRule), the action is rated as "socially appropriate" (mean: .72). The election of Rule:Give shifts the rating of action Give moderately upward toward "very socially appropriate" (mean: .74), whereas the election of Rule:Don't leads to a strong downward shift to at or below "somewhat socially appropriate" (mean: .17). Although the mean and median ratings of action Give remain positive throughout, close to 40% of subjects rate action Give *negatively* under Rule:Don't, which is an increase of 35 and 32 percentage points, respectively, compared to NoRule and Rule:Give. We find a similar, flipped version of this pattern for action Don't Give (right-hand panel of Figure 2): In the absence of an election (NoRule), action Don't Give is rated as moderately socially *in*appropriate

	StdMajority				NoRule	
	Rule:Give		**Rule:Don't**		**NoRule**	
Rating	Give	Don't Give	Give	Don't Give	Give	Don't Give
− − −	5%	36%	8%	10%	0%	15%
− −	1%	41%	13%	11%	2%	41%
−	0%	15%	17%	13%	1%	32%
+	5%	4%	24%	16%	11%	8%
+ +	34%	1%	20%	32%	49%	4%
+ + +	55%	3%	18%	18%	37%	0%
Mean	.74	−.63	.17	.23	.72	−.48
Median	1.00	−.67	.33	.50	.67	−.67
Rating $\gtrless 0$ (Signed rank test (z))	7.64^{***}	-7.70^{***}	2.49^{**}	3.16^{***}	8.60^{***}	-7.40^{***}
vs. NoRule (Rank-sum test (z))	2.24^{**}	-3.68^{***}	-6.34^{***}	7.00^{***}		
vs. Rule:Give (Signed rank test (z))			-6.41^{***}	7.48^{***}		

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; all two-tailed.
Ratings are: "very socially inappropriate" (− − −), "socially inappropriate" (− −),"somewhat socially inappropriate" (−), "somewhat socially appropriate" (+), "socially appropriate" (+ +), "very socially appropriate" (+ + +); modal ratings are shaded. For means and medians, responses are converted into numerical scores −1 (− − −), −2/3 (− −), −1/3 (−), +1/3 (+), +2/3 (+ +), +1 (+ + +).

Table 2: Elicited social approval (full distribution and non-parametric tests) of actions *Give* and *Don't Give* in the absence of an election (NoRule) and after a standard majority election of Rule:Give or Rule:Don't (StdMajority).

(mean: -.48). The election of Rule:Give pushes this rating slightly further into the negative (mean: -.63), whereas the election of Rule:Don't produces a strong effect in the opposite direction. Under Rule:Don't, mean and median social approval of action Don't Give are *positive* (mean: .23), *and even higher* than that of the opposite action Give. Two thirds (66%) of subjects rate action Don't Give positively under Rule:Don't, an increase of 54 and 58 percentage points, respectively, compared to NoRule and Rule:Give. This finding shows that a majority election can cause actions previously judged socially inappropriate (Don't Give) to become socially appropriate.

Detailed information on the distributions of approval ratings can be found in Table 2. Similar shifts to those observed in the mean and median are also observed in the modal rating of actions Give and Don't Give (shaded values in Table 2). Non-parametric rank-sum and signed rank tests reported at the bottom of the table verify that social norms are significantly altered by the outcome of the majority election: Compared to NoRule, Rule:Give significantly shifts upward the social approval of action Give ($z = 2.24$, $p = .025$) and significantly shifts downward the social approval of action Don't Give ($z = -3.68$, $p < .001$). Analogously, Rule:Don't significantly shifts downward the social approval of action Give ($z = -6.34$, $p < .001$) and significantly shifts upward the social approval of action Don't Give ($z = 7.00$, $p < .001$). Under Rule:Don't, action Don't Give is evaluated positively ($z = 3.16$, $p = .002$).[14] We summarize our findings below:

[14] A notable additional observation (which we did not predict) is that social approval ratings become more dispersed following the election of Rule:Don't (see Table 2). That is, compared to the case of Rule:Give or NoRule, people seem to agree less on which action constitutes "the right thing to do" if Rule:Don't has been elected into power. One possible interpretation of this finding is that elections can sometimes lead to a fragmentation rather than to a

Result 1. *Majority-elected rules (StdMajority) shift social norms. The election of Rule:Give makes action Give (Don't Give) more (less) socially appropriate. The election of Rule:Don't makes action Don't Give (Give) more (less) socially appropriate. Majority-elected rules can cause actions previously judged socially inappropriate (Don't Give) to become socially appropriate.*

3.2 Do norm shifts require "free and fair" elections?

Having established that an inclusive and unbiased majority election has the power to shift social norms, we now ask whether this power is sensitive to salient deficiencies in the voting process that can make the election appear less democratic: Are norms affected less if there is a voting fee (Pay4Vote), voters are bribed (MoneyOffer), or parts of the electorate are excluded from the ballot (ExcludePoor)?

Figure 3 displays mean and median ratings of actions Give (left panel) and Don't Give (right panel) following the election of Rule:Give or Rule:Don't, respectively, across the four election procedures StdMajority, Pay4Vote, MoneyOffer, and ExcludePoor. In this figure, the benchmark average rating for the case where there exists no rule (Give: .72, Don't Give: −.48) is represented by a dashed line. Complementing the figure, in Table 3 we present OLS estimates of the effect of elected rules on mean approval ratings by treatment. Column (1) shows estimates for the effect of rules on the mean social approval of action Give. Column (3) shows estimates for the effect of rules on the mean social approval of action Don't Give. Here, the benchmark rating of treatment NoRule serves as the constant. Columns (2) and (4) replicate and add individual-specific controls for gender, age, student status and country of origin.[15]

Analyzing Figure 3 and Table 3, we first see that *all* four election procedures have the power to change norms: Relative to the baseline rating of NoRule, the election of Rule:Give shifts the social approval of action Give (Don't Give) slightly upward (slightly downward) relative no NoRule, while the election of Rule:Don't shifts ratings strongly into the opposite direction. Qualitatively, this is true for all four treatments, i.e, regardless of whether voting is costly (Pay4Vote), voters are bribed (MoneyOffer), or parts of the electorate are excluded from the ballot (ExcludePoor). As Table 3 shows, the effect of Rule:Don't on mean approval ratings is always highly significant, while the effect of Rule:Give is only sometimes weakly so. On average, the election of Rule:Give shifts the mean approval rating of action Give upward by .05 points and the mean approval rating of action Don't Give downward by −.12 points, which is about one-tenth and one-third, respectively, of a discrete step in the approval rating (where a discrete step means, e.g., going from "weakly socially appropriate" to "socially appropriate"). In comparison, the election of Rule:Don't leads to average shifts of −.49 points and .55 points, respectively, which converts to between one and two discrete

consolidation of normative expectations. We will elaborate on this additional finding and the particular aspects of our setting that may be driving it in subsection 4.4.

[15]The entire distribution of ratings for each of the four treatments StdMajority, Pay4Vote, MoneyOffer, and ExcludePoor, including non-parametric Rank-sum test vs. NoRule and StdMajority can be found in Table A.1 in the Appendix.

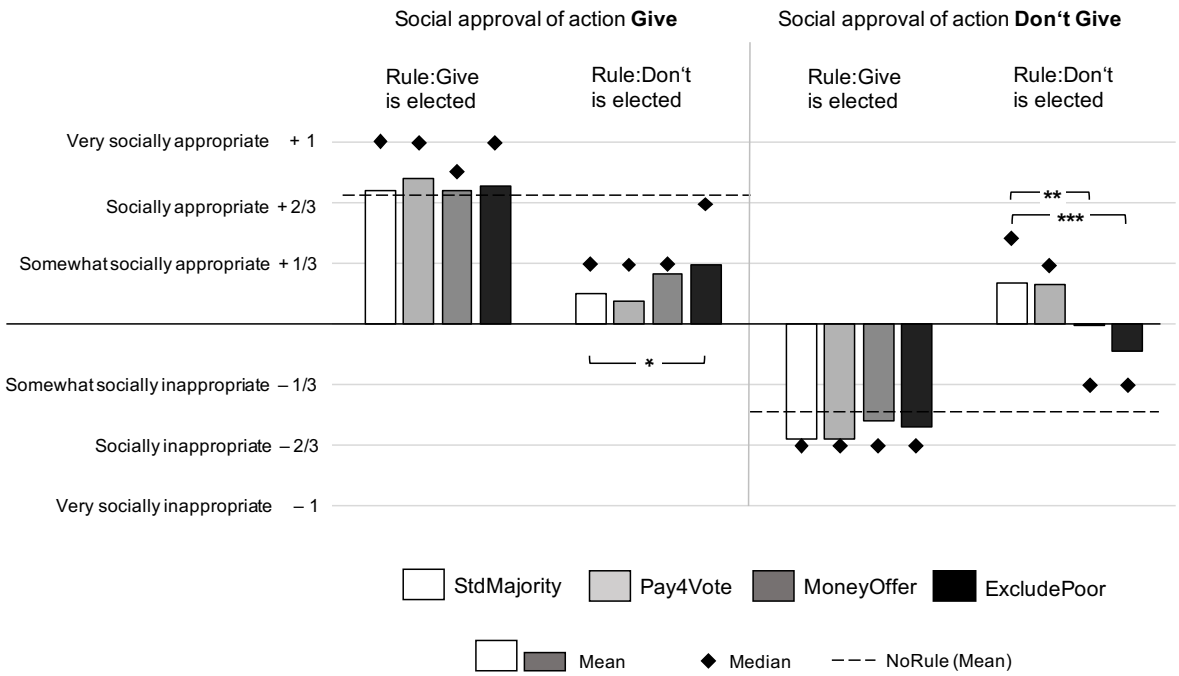


Figure 3: Elicited social approval (mean and median) of actions *Give* (left panel) and *Don't Give* (right panel) across different election procedures. Dashed line denotes mean social approval in the absence of an election (treatment NoRule; Give: .72, Don't Give: -.48). Stars denote significant results of Rank-sum tests comparing the social approval after a standard majority election (StdMajority) with the social approval after a non-standard majority election (Pay4Vote, MoneyOffer, ExcludePoor): $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

steps on the rating scale.

A second observation we make is that not all of the elections shift norms to the same extent as the inclusive and unbiased majority election StdMajority. While Pay4Vote has virtually the same power as StdMajority, MoneyOffer and ExludePoor perform significantly worse in shifting social norms. Throughout Table 3, MoneyOffer and ExludePoor show systematically smaller coefficients than Majority. In Figure 3, stars denote statistically significant differences to StdMajority according to non-parametric Rank-sum tests (for detailed test results, see Table A.1 in the Appendix). We see that shifts of social approval ratings are significantly smaller following the election of Rule:Don't when this rule comes into force with an election in which voters received bribes (MoneyOffer) or in which voters with a low household income were excluded from the ballot (ExludePoor). The results are particularly pronounced and meaningful for the social approval ratings of action Don't Give (right panel in Figure 3): While in StdMajority and in Pay4Vote, the election of Rule:Don't pushes the evaluation of action Don't Give significantly into the positive (signed rank test $z = 3.16$, $p = .002$, and $z = 3.33$, $p < .001$, respectively), this is not the case following the election of Rule:Don't in MoneyOffer and ExludePoor ($z = -.06$, $p = .954$, and $z = -2.36$, $p = .019$, respectively). This comparison shows that the extent to which elections can change social norms depends on the election process: Elections that entail too large deficiencies in the voting process can

Dep. Var.	Mean social approval			
	Give		Don't Give	
	(1)	(2)	(3)	(4)
Election of **Rule:Give**				
StdMajority	.02	.03	-.16**	-.12*
	(.056)	(.059)	(.061)	(.064)
Pay4Vote	.09*	.09*	-.16**	-.12*
	(.045)	(.051)	(.064)	(.067)
MoneyOffer	.02	.02	-.06	-.02
	(.053)	(.056)	(.066)	(.067)
ExcludePoor	.04	.04	-.09	-.04
	(.049)	(.051)	(.068)	(.069)
Election of **Rule:Don't**				
StdMajority	-.55***	-.54***	.71***	.75***
	(.073)	(.075)	(.079)	(.082)
Pay4Vote	-.59***	-.59***	.70***	.74***
	(.071)	(.076)	(.076)	(.078)
MoneyOffer	-.44***	-.44***	.47***	.51***
	(.066)	(.071)	(.075)	(.077)
ExcludePoor	-.39***	-.39***	.33***	.38***
	(.070)	(.073)	(.076)	(.075)
Constant	.72	.63	-.48	-.30
	(.031)	(.080)	(.041)	(.087)
Mean rating NoRule	.72		-.48	
Controls	No	Yes	No	Yes
Observations	900	900	900	900
(Subjects)	(500)	(500)	(500)	(500)
R^2	.219	.224	.270	.279

Robust standard errors (clustered at subject level) in parentheses:
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Controls in columns (2) and (4) are: female (1/0), age (continuous), student (1/0), and Western (1/0).

Table 3: Marginal effects of elected rules (Rule:Give and Rule:Don't) on mean social approval of actions Give and Don't Give across different election procedures: OLS Regressions. Mean ratings in the absence of an election (NoRule; Give: .72, Don't Give: -.48) serve as the baseline.

still significantly shift norms, but may no longer be able to completely turn around prior normative evaluations.

Result 2. *Elected rules can shift social norms, but bribing voters (MoneyOffer) or excluding parts of the electorate (ExcludePoor) weaken this ability. While Pay4Vote has virtually the same power as StdMajority, elected rules shift social approval ratings significantly less in MoneyOffer and Exlude-Poor.*

3.3 Do election-induced norm shifts predict behavior change?

The literature on social norms typically argues that people have an intrinsic preference to conform to what is collectively perceived as socially appropriate and, in a variety of social contexts, refrain

from maximizing material profits in order to comply with social norms (Elster, 1989; Bicchieri, 2006; López-Pérez, 2008; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017; Gaechter et al., 2017). If elections can change norms, can these changes predict how people adapt their behavior to the election outcome? To answer this question, we draw on data from the behavior experiment (Apffelstaedt and Freundt, 2020) in which we elicited actual choices (Give or Don't Give) following the election instead of social approval ratings.[16]

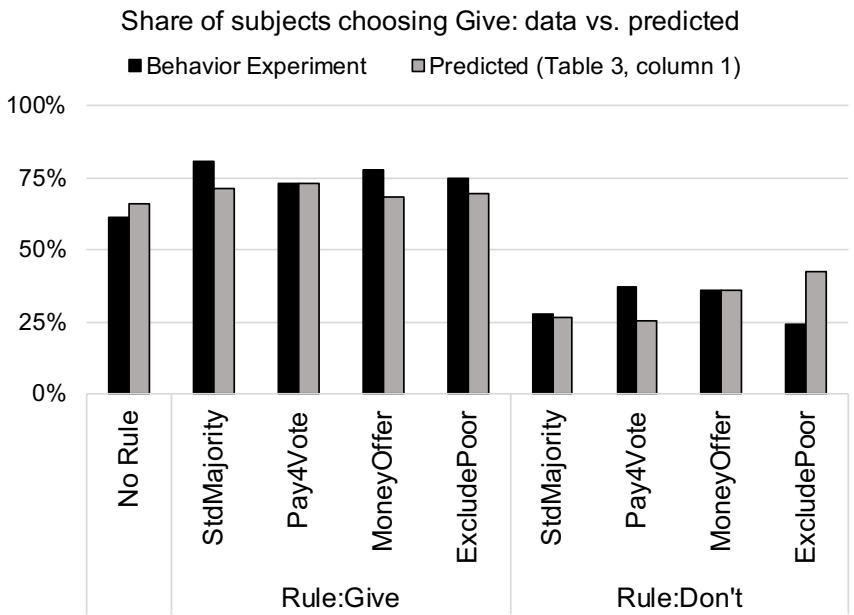


Figure 4: Share of subjects choosing action Give: data vs. predicted. Data from behavior experiment (Apffelstaedt and Freundt, 2020). Predictions using norms elicited through rating experiment (this paper), utility parameters according to specification Table 4, column (1).

How much of the change in behavior can be *directly* predicted by changes in social norms? Note that we use a setup without strategic incentives to follow a social norm. Rather, we assume that behavioral changes are driven by changes in the moral appeal associated with each action. Let us denote by $N(\text{Give}) \in [-1, 1]$ the elicited mean social approval of action Give in a given situation and by $N(\text{Don't Give}) \in [-1, 1]$ the elicited mean social approval of action Don't Give in the same situation. We seek to understand how much the propensity to choose action Give over Don't Give in the behavior experiment depends on the difference in social approval, $N(\text{Give}) - N(\text{Don't Give})$, elicited in the rating experiment. For this, assume that the utility from taking action Give takes the form

$$U_{\text{Give}} = const + \gamma \cdot [N(\text{Give}) - N(\text{Don't Give})], \tag{1}$$

[16]The behavior experiment, which was conducted in spring 2017 with a separate group of subjects on the same online platform as our rating experiment, forms the core of another paper (Apffelstaedt and Freundt, 2020). In that paper, we analyze the effects of different voting procedures on people's willingness to comply with elected rules. For a detailed description of the experiment and its results, see Apffelstaedt and Freundt (2020).

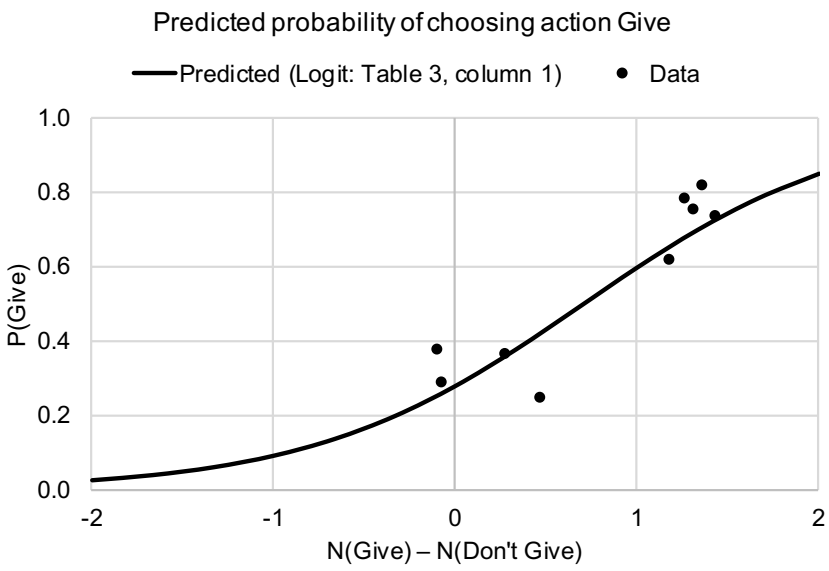


Figure 5: Predicted probability of choosing action Give as a function of $N(\text{Give}) - N(\text{Don't Give})$. Predictions using norms elicited through rating experiment (this paper) and actions from behavior experiment (Apffelstaedt and Freundt, 2020). Conditional Logit prediction with utility parameters according to specification Table 4, column (1).

and normalize the utility from taking action Don't Give to zero (i.e., $U_{\text{Don't Give}} = 0$).[17] In this simple utility framework, γ measures the weight that individuals attach to norms: A positive weight γ implies a utility gain from following that action (Give or Don't Give) which yields a higher social approval. The constant (*const*) captures the average utility individuals derive from choosing action Give over Don't Give that is independent of norms. Following the procedure in Krupka and Weber (2013), we combine the data from the behavior experiment with the data from the rating experiment to estimate the parameters of the utility function using conditional Logit. The results of this estimation are found in Table 4: In column (1), we estimate γ by fitting the utility function to the share of Givers in the behavior experiment using as only explanatory variable the elicited difference in social approval, $N(\text{Give}) - N(\text{Don't Give})$. We find a large, positive and highly significant estimate, $\gamma = 1.347$ ($p < .001$). This estimate tells us that, on average, the relative utility from taking action Give strongly increases with the difference in social approval between actions Give and Don't Give. Vice versa, if that difference in ratings becomes smaller or even turns negative as, for instance, when Rule:Don't is elected, the propensity to choose action Don't Give will become larger. Columns (2)-(4) show that the estimate of γ is robust to including demographic controls and does not vary significantly if we estimate it separately by treatment (column 3).

To get a better sense of the estimated relationship between norms and behavior, Figure 5 plots

[17] Because subjects can only choose between two actions, Give and Don't Give, only differences in utility matter for decisions. The normalization of $U_{\text{Don't Give}} = 0$ is thus without loss of generality.

Dep. Var.	Utility according to Eq. (1)			
	(1)	(2)	(3)	(4)
Appropriateness rating (γ)	1.347***	1.371***	1.403***	1.444***
	(.103)	(.106)	(.143)	(.148)
Appropriateness rating X				
Pay4Vote			-.194	-.233
			(.192)	(.198)
MoneyOffer			.248	.203
			(.216)	(.221)
ExcludePoor			-.154	-.140
			(.199)	(.199)
Constant (*const*)	-.944	-1.656	-.970	-1.671
	(.125)	(.393)	(.130)	(.393)
Controls	No	Yes	No	Yes
Observations	1200	1182	1200	1182
(Subjects)	(400)	(394)	(400)	(394)
Log-likelihood	-746.8	-725.9	-742.7	-722.4

Robust standard errors (clustered at subject level) in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.
Controls in columns (2) and (4) are: female (1/0), age (continuous), student (1/0), and Western (1/0).

Table 4: Conditional logit estimation of choice determinants (utility according to Eq. (1)). Choice data from behavior experiment (Apffelstaedt and Freundt, 2020). Includes mean appropriateness ratings N(Give) and N(Don't Give) from rating experiment (this paper) as explanatory variable.

the predicted probability of choosing action Give according to the model specification in Table 4, column (1): When there is no difference between the social approval of actions Give and Don't Give, $N(\text{Give}) - N(\text{Don't Give}) = 0$, $P(\text{Give})$ is predicted at 28%. That is, in the absence of clear guidance by a social norm, our model predicts that the majority of subjects will choose selfishly. From this position, increasing the social approval of action Give (or, equivalently, decreasing the social approval of action Don't Give) will lead on average to a 10 percentage point increase in the probability of taking action Give for every step on the rating scale (e.g., going from "weakly socially appropriate" to "socially appropriate"). This is a strong relationship: In the case of a standard majority election of Rule:Give, for instance, for which we elicit a difference in social approval of $N(\text{Give}) - N(\text{Don't Give}) = 1.37$, the probability of action Give is predicted at 71%—an increase of more than 40 percentage points relative to the case without normative guidance.

Finally, in Figure 4, we plot—next to the original data from the choice experiment—the share of Givers predicted by the model specification in Table 4, column (1), for each of our experimental settings. The figure demonstrates that, overall, the simple model of norm-dependent utility specified in Eq. (1) performs well in reproducing behavioral changes across different choice situations, i.e., going from NoRule to Rule:Give to Rule:Don't. Behavioral differences between treatments *within* a given rule (within Rule:Give/Rule:Don't) are not predicted as consistently as differences between NoRule, Rule:Give and Rule:Don't. In line with the analysis in Apffelstaedt and Freundt (2020), we suggest that these differences should be attributed to intrinsic preferences regarding the rule-

selection procedure rather than social norms.[18] We summarize our results on the relationship of norms and behavior as follows:

Result 3. *Election-induced norm shifts predict behavior change. Using choice data from the behavior experiment in Apffelstaedt and Freundt (2020), we predict a one-step increase in the mean approval rating of an action to increase the probability of taking that action by on average 10 percentage points.*

4 Mechanisms of Social Norms Change

Our results show that majority-elected rules can shift collective perceptions of what constitutes socially appropriate behavior. But what exactly is the cause for these shifts? In this section, we shed light on the role of two mechanisms, both of which are inherent to elections that select rules using a majority vote, but which are not easily separated in a natural setting (and in our experiment so far).[19] The first mechanism pertains to the informational value contained in majority elections, the second to the social appropriateness of following rules per se. Both mechanisms are extensively discussed in an interdisciplinary literature on the expressive function of law (see, e.g., Sunstein, 1996; Cooter, 1998; McAdams and Rasmusen, 2007). We present the results of an additional treatment that can shed light on the importance of the two mechanisms in the context of our experiment.

To fix ideas, consider the election of Rule:Don't in StdMajority. In line with our predictions, we showed in subsection 3.1 that such an election increases the social approval of action Don't Give and decreases the social approval of action Give. However, is this due to (1) the information that a majority of subjects prefer Rule:Don't over Rule:Give or (2) the fact that *regardless* of what the majority prefers, it is socially appropriate to take action Don't Give now that it is mandated by the rule?

4.1 Mechanism 1: The informational content of majority elections

Outcomes of elections and referenda can serve as a public signal which contains information about the distribution of preferences and values in society (Bursztyn et al., 2020a; McAdams, 2015). In our setting, the election of Rule:Give signals broad public support for action Give, while the election of Rule:Don't signals broad public support for action Don't Give. If most people agree that whether an action is socially appropriate depends positively on the level of support for that action in society, then this information—*regardless* of whether or not the elected rule is actually implemented—will shift collective perceptions of what constitutes socially appropriate behavior. It follows that when being asked to estimate the social approval of actions Give and Don't Give

[18] For a detailed analysis of treatment effects on rule compliance (i.e. on behavior conditional on Rule:Give or Rule:Don't being elected) and possible mechanisms we refer the interested reader to Apffelstaedt and Freundt (2020).

[19] We thank an anonymous reviewer for pushing this argument further and motivating us to investigate the mechanisms behind our observation in more detail.

conditional on Rule:Don't being elected by a majority vote, subjects in our experiment will provide a different rating than when being asked to rate the same actions conditional on Rule:Give being elected.[20] Depending on their beliefs regarding public support for actions Give and Don't Give when *not* receiving information about a voting outcome, they will also give a different evaluation than in NoRule. Belief updating processes may play an important role in the coordination game of our experiment, as subjects have an incentive to consider any information they believe will change the modal answer.

4.2 Mechanism 2: The social appropriateness of following rules

An alternative explanation is based on the idea that the mere existence of a rule itself—irrespective of its formation process or underlying normative foundations—can influence collective opinion of what is ethically right and what is wrong. Rather than addressing information processes, this mechanism posits that subjects are morally obligated to obey any rule that is in place (Nadler, 2017). If such a general obligation is collectively acknowledged, this can lead to a social norm of *unconditional* rule compliance. There are various intuitive explanations for why such norms may exist. For instance, Benabou and Tirole (2011) propose a (self-)signaling model in which individuals reveal private information about themselves by (not) abiding by the law. If rule compliance is prime, subjects in our experiment will rate that action as socially appropriate that complies with the rule and will rate that action as socially inappropriate that opposes the rule—irrespective of the nature of actions and irrespective of the social norm that governs actions in the absence of a rule.

4.3 Uncoupling mechanism 1 and 2: Treatment ExoRule×MajVote

Our experiment so far cannot disentangle the two mechanisms.[21] To shed light on their role, we designed and implemented a follow-up treatment, ExoRule×MajVote. The new treatment was run in July 2021 with 100 new subjects on Prolific, using the same subject pool as our main experiment. Treatment ExoRule×MajVote inherits the structure and procedures of our election treatments (see section 2). In stage 2, subjects are asked to rate the social approval of actions Give and Don't Give conditional on a code of conduct (Rule:Give or Rule:Don't) being implemented. However, instead of presenting subjects with an election procedure, we present them with the following situation: "*To determine the rule, a coin is flipped by the computer. If the coin lands heads (50%chance), RULE: GIVE is implemented. If the coin lands tails (50% chance), RULE: DON'T is implemented.*

[20] While the strength of the behavioral response to conditional information might differ, the logic, and hence the potential relevance, of this information mechanism holds irrespective of whether we elicit norms using a direct method or the strategy method. Let $S(\text{Don't Give} \mid \text{Rule:Don't})$ be a subject's Bayesian posterior belief of the support for action Don't Give conditional on the information that Rule:Don't is elected. This posterior is the same whether we ask for conditional ratings (strategy method) or only present subjects with one (final) election outcome for which they give their rating (direct method).

[21] Real world elections, in contrast to simple polls, are typically followed by an adjustment of the institutional framework, i.e. by a change of the rules. Hence, in normal practice, elections and referenda combine the two aspects described above, as did our experiment.

Note: Whether RULE: GIVE or RULE: DON'T is implemented depends purely on chance. That is, the rule is random." When asking for their evaluations, we *additionally* ask them to condition their answer on whether the majority of individuals who take part in the lottery would vote for Rule:Give or Rule:Don't.[22] In total, we thus elicit the social norms for four different scenarios:

- Scenario 1: ExoRule (Coin flip) = Rule:Give; MajVote = Rule:Give
- Scenario 2: ExoRule (Coin flip) = Rule:Give; MajVote = Rule:Don't
- Scenario 3: ExoRule (Coin flip) = Rule:Don't; MajVote = Rule:Give
- Scenario 4: ExoRule (Coin flip) = Rule:Don't; MajVote = Rule:Don't

Hence, the information on the majority vote is now independent of the actual rule that is implemented. This design allows us to decouple the role of mere rule compliance (ExoRule) from the role of social information (MajVote).

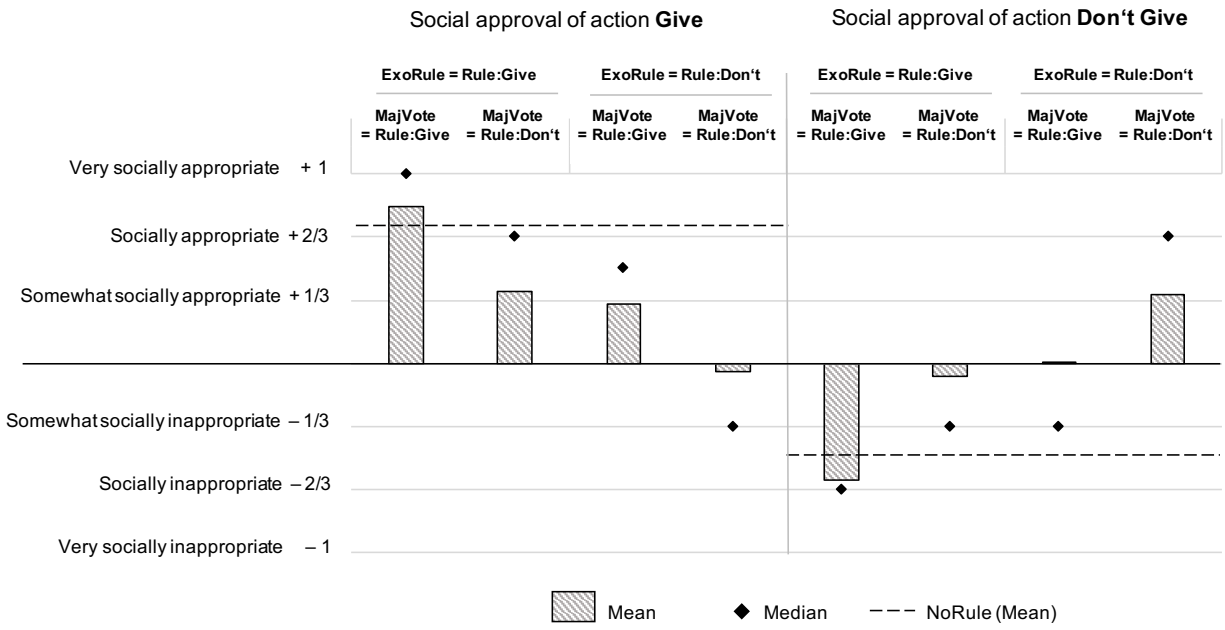


Figure 6: Elicited social approval (mean and median) of actions *Give* (left panel) and *Don't Give* (right panel) across different scenarios in ExoRule×MajVote. Dashed line denotes mean social approval in the absence of an election (treatment NoRule; Give: .72, Don't Give: -.48).

We present the main findings from treatment ExoRule×MajVote in Table 5, and Figures 6 and 7. Figure 6 displays elicited mean and median approval ratings of action Give (left panel) and action Don't Give (right panel) across the four scenarios. In Table 5 we estimate the effect of ExoRule and MajVote on mean approval ratings using OLS regressions. Figure 7 graphically illustrates the heterogeneity in individual reactions to shifts in ExoRule and MajVote.

[22] Detailed instructions can be found in Appendix A.X

Dep. Var.	Mean social approval	
	Give (1)	Don't Give (2)
ExoRule = Rule:Give	baseline	baseline
MajVote = Rule:Give	baseline	baseline
ExoRule = Rule:Don't	-.52***	.62***
	(.068)	(.069)
MajVote = Rule:Don't	-.45***	.55***
	(.065)	(.066)
ExoRule x MajVote = Rule:Don't x Rule:Don't	.09	-.19**
	(.093)	(.080)
Constant	.83	-.61
	(.034)	(.050)
Observations	400	400
(Subjects)	(100)	(100)
R^2	.236	.247

Robust standard errors (clustered at subject level) in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 5: Marginal effects of ExoRule and MajVote on mean social approval of actions Give and Don't Give in treatment ExoRule×MajVote: OLS Regressions.

Looking at Table 5, we see that both, a change in the exogenous rule *and* a change in the majority vote have a very significant, similar and sizeable effect on social norms. Going from ExoRule = Rule:Give to ExoRule = Rule:Don't, ceteris paribus, shifts mean social approval ratings by $-.52$ (action Give) and $+.62$ (action Don't Give) points, respectively. Similarly, going from MajVote = Rule:Give to MajVote = Rule:Don't, ceteris paribus, shifts mean social approval ratings by $-.45$ (action Give) and $+.55$ (action Don't Give) points, respectively.[23]

Looking at Figure 6, we see that in absolute terms, the social approval ratings for scenarios 1 (ExoRule = MajVote = Rule:Give) and 4 (ExoRule = MajVote = Rule:Don't) are very similar to the social approval ratings which we document in StdMajority following a majority election of Rule:Give and Rule:Don't, respectively. Going from scenario 1 (on the very left) to scenario 4 (on the very right), we find that—as in StdMajority, when going from Rule:Give to Rule:Don't—the approval ratings can flip, making action Don't Give (instead of action Give) the relatively more appropriate action. As in StdMajority, action Don't Give is rated significantly positively following such a change (signed rank test, $z = 4.87$, $p < .001$). However, for such a major shift in ratings to occur, both a change in the formal rule (ExoRule=Rule:Don't) and a majority vote confirming that rule (MajVote=Rule:Don't) are required. This observation suggests that both mechanisms also contribute significantly to the shift in social norms that we observe in our main experiment.

Last but not least, looking at Figure 7, we document that while subjects are quite heterogeneous in their reactions to ExoRule and MajVote, the majority of subjects show responses that are broadly

[23]The interaction term in Table 5 indicates that the marginal effects of changes in ExoRule and MajVote are somewhat smaller when the initial situation is one in which both variables are set to Rule:Don't (ExoRule = MajVote = Rule:Don't) rather than Rule:Give (ExoRule = MajVote = Rule:Give). This suggests that social norms associated with Rule:Don't are somewhat harder to change.

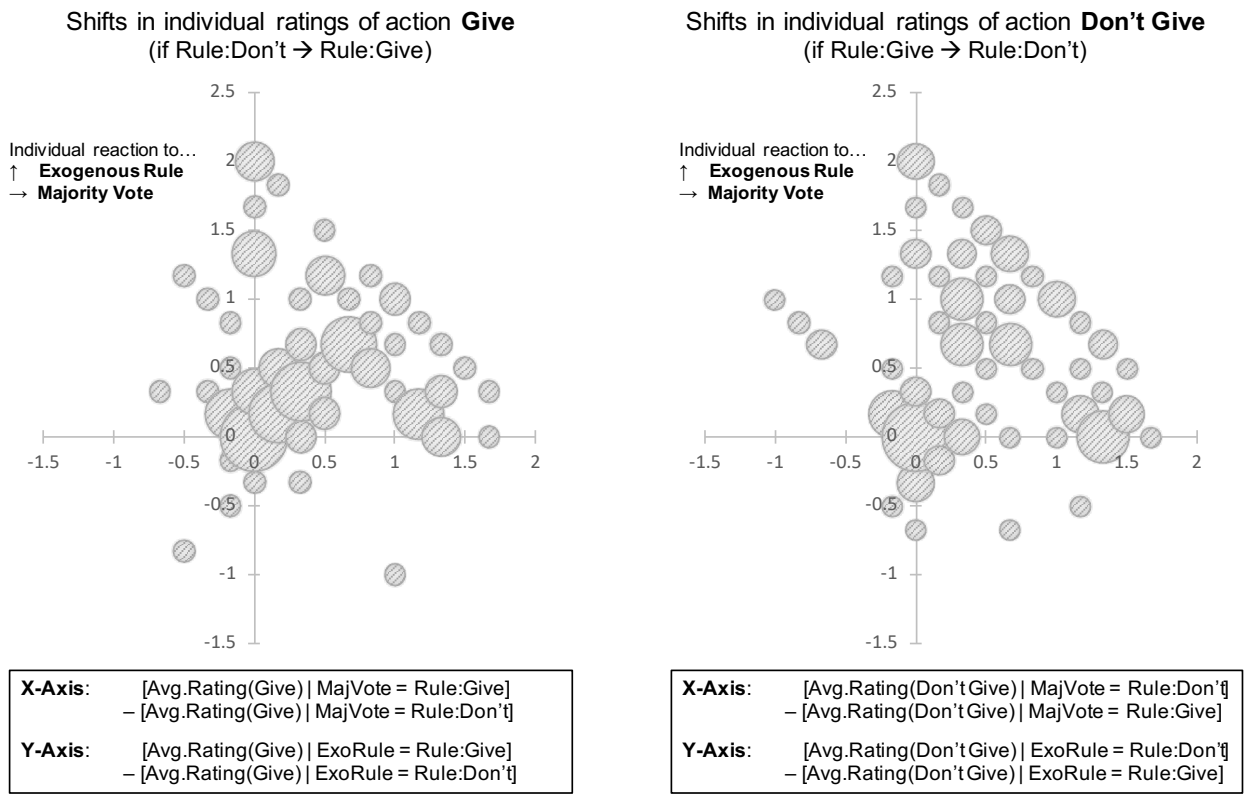


Figure 7: Graphical illustration of shifts in individual norm ratings for action Give (left) and action Don't Give (right) as a consequence of a change in the exogenous rule (y-axis) or the majority vote (x-axis): Treatment ExoRule×MajVote. Bubble sizes correspond to subject densities (n=100). The further away a bubble from the origin, the stronger the reaction. The strongest possible positive reaction is a change from "very socially inappropriate" (-1) to "very socially appropriate" (+1), which converts into a reaction of +2. The strongest possible negative reaction is a change from "very socially appropriate" (+1) to "very socially inappropriate" (-1), which converts into a reaction of −2.

in line with the aggregate effects reported above. Only 9 (10) out of 100 subjects do not adjust their rating of action Give (Don't Give) to either of the two aspects (bubbles at the origin). A minority of subjects (18 of 100 in both panels) only respond to either the rule or the majority vote (bubbles on the axes). The large majority of subjects (56 (50) out of 100) are found in the interior of the first quadrant: Consistent with the aggregate effects reported above, their ratings respond positively to both a change in the exogenous rule endorsing the action and a change in the majority decision endorsing the action. There seems to be some agreement among subjects, then, that both, the compliance with rules *and* the compliance with majority opinions, are relevant to the moral evaluation of behavior.

4.4 Interpreting the data in light of the two mechanisms

The results of our follow-up treatment ExoRule×MajVote suggest that both the informative value of elections as well as a general obligation to comply with rules are drivers of election-induced norm

shifts. Below, we discuss how these two mechanisms may relate to three interesting patterns that we observe in the data of our main experiment.

Asymmetric effects of Rule:Don't vs. Rule:Give. We observe that the election of Rule:Don't seems to shift social norms relative to NoRule to a greater extent than the election of Rule:Give (see, for instance, Table 3). This observation lends itself to an interpretation using either the first or the second mechanism.[24] Assume, first, that subjects primarily care about the informative value of elections (mechanism 1). Compared to NoRule, the election of a rule then affects ratings to the extent that the elected rule changes subjects' beliefs about the degree of public support for actions Give and Don't Give. Subjects will react asymmetrically to rules if the election of one rule confirms their prior and the election of the other rule rejects their prior. It follows that if subjects have the prior (NoRule) belief that most participants would vote for Rule:Give, then the election of Rule:Give will shift social norms less than the election Rule:Don't. Assume, instead, that subjects primarily care about rule compliance (mechanism 2). Elected rules then shift norms to the extent to which rule compliance after the election contradicts any prior assessment of social appropriateness. In the case of Rule:Give, rule compliance and prosocial behavior (which is arguably the most important factor when evaluating actions Give and Don't Give in the absence of a rule) coincide. This may lead to a reinforcement of the previous (NoRule) norm, but should not greatly change the basic evaluation of the actions. The election of Rule:Don't, on the other hand, leads to a clash of the two rating dimensions. The more important rule compliance is to the subjects, the more strongly norms will shift in this case. As long as rule compliance is an important criterion, Rule:Don't will shift social norms more than Rule:Give.

Salient deficiencies in the voting process. In subsection 3.2, we document differences in the extent to which social norms are affected by elections in which participants are offered a bribe (MoneyOffer) or excluded from the election (ExcludePoor) relative to the case of a standard majority vote. While we can only speculate about this, it seems intuitive that when evaluating the social appropriateness of actions, the more democratic the election of the rule, the greater the weight that should be given to rule compliance (mechanism 2). Following this argumentation, controversial practices such as vote buying and the disenfranchisement of poor voters should decrease the power of elections to change social norms. It is also possible that the effect is due to a signal extraction problem (mechanism 1): If the election result is biased or no longer representative, it reveals less about the public support for a given policy. Again, social norms should react to a lesser extent to the rule that was elected. We also observe that the introduction of a voting fee or "poll tax"

[24] We do not mean to suggest that other mechanisms are not at work. An alternative explanation is that there is less room for a shift in ratings in response to Rule:Give than for a shift in ratings in response to Rule:Don't Give, because the social approval ratings for Give and Don't Give are already relatively close to the upper and lower limits, respectively, in the absence of a rule. Another explanation is that Rule:Don't and Rule:Give differ in the basic mechanisms by which they affect social norms, e.g., because Rule:Don't can also be used as an "excuse" for selfish behavior. For instance, Bénabou et al. (2018) model how such narratives can influence prosocial behavior.

(Pay4Vote) leaves the power of elected rules to shift norms largely unaffected. In line with our reasoning above, it could be that a voting fee is not seen as such a massive interference with democratic principles, rendering the obligation to comply with the elected rule more important than in treatments MoneyOffer and ExcludePoor. It also seems plausible that the informative value of the election is higher in this treatment, as the election result is potentially less biased and still (somewhat) representative.

Elections and norm consensus. In the results so far, we have focused on modal, mean, and median ratings to analyze effects of elections on social norms. Another noteworthy finding which we have only referred to in a footnote (see footnote 14) is an observation that may, prima facie, appear counterintuitive: We find that elections do not only have the potential to cause a shift in modal, median and mean social appropriateness ratings but also an increase in the variance of the distribution of individual ratings (see Table 2 in the main text and Table A.1 in the Appendix). We interpret this finding as a decrease in *norm consensus*, i.e. in the degree to which members of a society agree on which action constitutes "the right thing to do". How can we explain the fact that social norms may become *less* clear even though elections are thought to provide *better* information about the moral preferences in society? In our data, this finding is again particularly prevalent in the case of Rule:Don't.

We can make sense of this finding by acknowledging that people may be heterogeneous when it comes to weighting the social appropriateness of following elected rules (mechanism 2) against the social appropriateness of taking prosocial actions. In the case of Rule:Don't, these two dimensions disagree regarding the evaluation of actions Give (high on prosociality, low on rule compliance) and Don't Give (high on rule compliance, low on prosociality). A subject who weights rule compliance strongly (and believes other people to do so as well), will rate action Give socially inappropriate and action Don't Give socially appropriate. At the same time, however, a respondent who strongly weights prosociality over compliance will give the opposite rating. As people become unsure or begin to disagree about what the social norm is, rating decisions diverge as a consequence of the election of Rule:Don't. Ultimately, either the shift in modal, median, and mean social appropriateness ratings that we observe under Rule:Don't *or* the observed erosion of norm consensus could be responsible for why people's behavior is affected by the election.

5 Conclusion

We have investigated how elected rules can affect what is perceived as socially appropriate behavior. Participants in our online experiment rate the social appropriateness of sharing versus not sharing experimental income with other participants. We find that majority-elected rules that ask people to share or not to share, respectively, can change social norms: They shift the modal appropriateness perception of actions and, as a result, can cause actions previously judged socially inappropriate (not sharing) to become socially appropriate. Comparing different voting procedures, we show that

this power prevails, albeit in a weakened form, even if the election process is flawed (introducing a voting fee or "poll tax", bribing voters, disenfranchising poorer voters). Using behavioral data from a related experiment (Apffelstaedt and Freundt, 2020), we show that the norm shifts we observe are able to predict changes in behavior that result from the election of rules. A follow-up treatment suggests that both the informational value contained in election results and the social appropriateness of following rules *per se* play a role in shifting social norms.

We hope that our paper will stimulate future research on the importance of democratic procedures in general, and elections in particular, for the formation and dissolution of social norms in a society. Our data show that elected rules not only shift modal appropriateness ratings of behavior, but can also alter their distribution and lead to an erosion of a previously existing norm consensus. Psychological research underscores the importance of perceived social consensus or "norm clarity" in shaping one's opinions and the ability of norms to guide behavior (Lewandowsky et al., 2019; Zitek and Hebl, 2007). Thus, to fully understand how elections influence norms and how these norms in turn shape behavior, we believe it is an important task of future research to shed light on the role of norm consensus as well as the role of individual uncertainty about social norms in driving behavioral responses.

References

Aksoy, C.G, Carpenter, C.S, De Haas, R, Tran, K.D, 2020. Do laws shape attitudes? Evidence from same-sex relationship recognition policies in Europe. European Economic Review 124, 103399.

Albornoz, F, Bradley, J, Sonderegger, S, 2020. The Brexit Referendum and the Rise in Hate Crime; Conforming to the New Norm. Working Paper. This Version: July 31, 2020.

Apffelstaedt, A, Freundt, J, 2020. Corrupted Votes and Rule Compliance. Working Paper. This Version: February 4, 2020.

Barron, L.G, Hebl, M, 2010. Reducing "Acceptable" Stigmatization Through Legislation. Social Issues and Policy Review 4, 1–30.

Barron, L.G, Hebl, M, 2013. The force of law: The effects of sexual orientation antidiscrimination legislation on interpersonal discrimination in employment. Psychology, Public Policy, and Law 19, 191–205.

Baskaran, T, Hessami, Z, 2018. Does the election of a female leader clear the way for more women in politics? American Economic Journal: Economic Policy 10, 95–121.

Bénabou, R, Tirole, J, 2006. Incentives and prosocial behavior. American Economic Review 96, 1652–1678.

Benabou, R, Tirole, J, 2011. Laws and norms. Working Paper 17579, National Bureau of Economic Research.

Bicchieri, C, 2006. The Grammar of Society, Cambridge: Cambridge University Press.

Bicchieri, C, 2010. Norms, preferences, and conditional behavior. Politics, Philosophy & Economics 9, 297–313.

Bicchieri, C, 2016. Norms in the wild: How to diagnose, measure, and change social norms, Oxford: Oxford University Press.

Bicchieri, C, Chavez, A, 2010. Behaving as expected: Public information and fairness norms. Journal of Behavioral Decision Making 23, 161–178.

Bicchieri, C, Dimant, E, 2019. Nudging with care: The risks and benefits of social information. Public Choice 2019, 1–22.

Bicchieri, C, Dimant, E, Sonderegger, S, 2020a. It's Not A Lie if You Believe the Norm Does Not Apply: Conditional Norm-Following with Strategic Beliefs. Working Paper. This version: October 9, 2020.

Bicchieri, C, Dimant, E, Gächter, S, 2020b. Observability, social proximity, and the erosion of norm compliance. Working Paper. This version: March 24, 2020.

Brandts, J, Charness, G, 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. Experimental Economics 14, 375–398.

Bursztyn, L, Egorov, G, Fiorin, S, 2020a. From extreme to mainstream: The erosion of social norms. American Economic Review 110, 3522–3548.

Bursztyn, L, González, A.L, Yanagizawa-Drott, D, 2020b. Misperceived social norms: Women working outside the home in saudi arabia. American Economic Review 110, 2997–3029.

Bénabou, R, Falk, A, Tirole, J, 2018. Narratives, imperatives, and moral reasoning. Working Paper 24798, National Bureau of Economic Research.

Cialdini, R.B, Reno, R.R, Kallgren, C.A, 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology 58, 1015.

Cialdini, R.B, Trost, M.R, 1998. Social influence: Social norms, conformity and compliance. In: Gilbert, D.T, Fiske, S.T, Lindzey, G (Eds.), The Handbook of Social Psychology, New York. McGraw-Hill, 151–192.

Cooter, R, 1998. Expressive law and economics. The Journal of Legal Studies 27, 585–607.

Crandall, C.S, Miller, J.M, White, M.H, 2018. Changing norms following the 2016 US presidential election: The Trump effect on prejudice. Social Psychological and Personality Science 9, 186–192.

Elster, J, 1989. Social norms and economic theory. Journal of economic perspectives 3, 99–117.

Gaechter, S, Gerhards, L, Nosenzo, D, 2017. The importance of peers for compliance with norms of fair sharing. European Economic Review 97, 72–86.

Gaechter, S, Nosenzo, D, Sefton, M, 2012. Peer effects in pro-social behavior: Social norms or social preferences? Journal of the European Economic Association 11.

Galbiati, R, Henry, E, Jacquemet, N, Lobeck, M, 2020. How Laws Affect the Perception of Norms: Empirical Evidence from the Lockdown. Working Paper. This version: July 2020.

Hauser, O.P, Gino, F, Norton, M.I, 2018. Budging beliefs, nudging behaviour. Mind & Society 17, 15–26.

Huang, J, Low, C, 2017. Trumping norms: Lab evidence on aggressive communication before and after the 2016 US presidential election. American Economic Review 107, 120–124.

Iriberri, N, Rey-Biel, P, 2011. The role of role uncertainty in modified dictator games. Experimental Economics 14, 160–180.

Jung, J.h, Tavits, M, 2021. Do referendum results change norm perceptions and personal opinions? Electoral Studies 71, 102307.

Kedia, S, Pareek, A, 2020. Women in Politics: The Effect on Board Diversity. Working Paper. This version: December 2020.

Kimbrough, E, Vostroknutov, A, 2016. Norms make preferences social. Journal of the European Economic Association .

Krupka, E, Weber, R.A, 2009. The focusing and informational effects of norms on pro–social behavior. Journal of Economic Psychology 30, 307–320.

Krupka, E.L, Leider, S, Jiang, M, 2017. A meeting of the minds: Informal agreements and social norms. Management Science 63, 1708–1729.

Krupka, E.L, Weber, R.A, 2013. Identifying Social Norms Using Simple Coordination games: Why Does Dictator Game Sharing Vary? Journal of the European Economic Association 11, 495–524.

Lane, T, Nosenzo, D, Sonderegger, S, 2021. Law and Norms: Empirical Evidence. Working Paper. This version: June 30, 2021.

Lewandowsky, S, Cook, J, Fay, N, Gignac, G.E, 2019. Science by social media: Attitudes towards climate change are mediated by perceived social consensus. Memory and Cognition 47, 1445–1456.

López-Pérez, R, 2008. Aversion to norm-breaking: A model. Games and Economic behavior 64, 237–267.

McAdams, R, 2015. The Expressive Powers of Law: Theories and Limits.

McAdams, R.H, Rasmusen, E, 2007. Norms and the law. vol. 2, chap. 20. Elsevier, 1 edn., 1573–1618.

Nadler, J, 2017. Expressive law, social norms, and social groups. Law & Social Inquiry 42, 60–75.

Norris, P, 2014. Why Electoral Integrity Matters: Cambridge University Press.

Ostrom, E, 2000. Collective action and the evolution of social norms. Journal of Economic Perspectives 14, 137–158.

Rauhut, H, Winter, F, 2010. A sociological perspective on measuring social norms by means of strategy method experiments. Social Science Research 39, 1181–1194.

Stooksbury, K.E, Scheb II, J.M, Stephens Jr., O.H, 2017. Encyclopedia of American Civil Rights and Liberties: Revised and Expanded Edition, 2nd Edition, Volume 2.

Sunstein, C, 1996. On the expressive function of law. 144 U. Pa. L. Rev. 2021 .

Tyler, T.R, 2006. Psychological Perspectives on Legitimacy and Legitimation. Annual Review of Psychology 57, 375–400.

Zitek, E.M, Hebl, M.R, 2007. The role of social norm clarity in the influenced expression of prejudice over time. Journal of Experimental Social Psychology 43, 867–876.

Appendix

A.1 Distribution of ratings across treatments

Panel (a): Action **Give**	**Rule:Give**				**Rule:Don't**				**NoRule**
Rating	StdMaj	P4Vote	MOffer	ExPoor	StdMaj	P4Vote	MOffer	ExPoor	NoRule
– – –	5%	1%	3%	2%	8%	9%	6%	10%	0%
– –	1%	1%	2%	1%	13%	11%	4%	3%	2%
–	0%	1%	0%	1%	17%	21%	21%	13%	1%
+	5%	4%	10%	6%	24%	23%	25%	19%	11%
+ +	34%	36%	31%	40%	20%	23%	28%	37%	49%
+ + +	55%	57%	54%	50%	18%	13%	16%	18%	37%
Mean	.74	.80	.74	.76	.17	.13	.27	.33	.72
Median	1.00	1.00	1.00	.83	.33	.33	.33	.67	.67
Rating $\gtrless 0$ (Signed rank test (z))	7.64^{***}	8.61^{***}	7.98^{***}	8.28^{***}	2.49^{**}	2.00^{**}	4.36^{***}	4.50^{***}	8.60^{***}
vs. NoRule (Rank-sum test (z))	2.24^{**}	2.88^{***}	1.82^{*}	1.80^{*}	-6.34^{***}	-7.08^{***}	-5.99^{***}	-4.94^{***}	
vs. StdMaj (Rank-sum test (z))		0.52	−0.31	−0.51		−0.49	1.06	1.80^{*}	

Panel (b): Action **Don't Give**	**Rule:Give**				**Rule:Don't**				**NoRule**
Rating	StdMaj	P4Vote	MOffer	ExPoor	StdMaj	P4Vote	MOffer	ExPoor	NoRule
– – –	36%	43%	35%	37%	10%	5%	11%	12%	15%
– –	41%	32%	25%	35%	11%	14%	15%	30%	41%
–	15%	16%	27%	15%	13%	17%	25%	17%	32%
+	4%	2%	7%	4%	16%	20%	20%	19%	8%
+ +	1%	4%	3%	5%	32%	26%	22%	16%	4%
+ + +	3%	3%	3%	4%	18%	18%	7%	6%	0%
Mean	−.63	−.63	−.53	−.57	.23	.22	−.01	−.15	−.48
Median	−.67	−.67	−.67	−.67	.50	.33	−.33	−.33	−.67
Rating $\gtrless 0$ (Signed rank test (z))	-7.70^{***}	-7.45^{***}	-7.02^{***}	-6.88^{***}	3.16^{***}	3.33^{***}	−.06	-2.36^{**}	-7.40^{***}
vs. NoRule (Rank-sum test (z))	-3.68^{**}	-3.87^{***}	-1.70^{*}	-2.92^{***}	7.00^{***}	7.36^{***}	5.30^{***}	3.27^{***}	
vs. StdMaj (Rank-sum test (z))		−0.51	1.45	.42		−0.24	-2.73^{**}	-4.00^{***}	

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; all two-tailed.
Ratings are: "very socially inappropriate" (– – –), "socially inappropriate" (– –),"somewhat socially inappropriate" (–), "somewhat socially appropriate" (+), "socially appropriate" (+ +), "very socially appropriate" (+ + +); modal ratings are shaded. For means and medians, responses are converted into numerical scores −1 (– – –), −2/3 (– –), −1/3 (–), +1/3 (+), +2/3 (+ +), +1 (+ + +).

Table A.1: Elicited social approval of actions *Give* (panel a) and *Don't Give* (panel b) across all treatments.

	ExoRule×MajVote			
Panel (a): Action **Give**	**ExoRule = Rule:Give**		**ExoRule = Rule:Don't**	
Rating	MajVote = Rule:Give	MajVote = Rule:Don't	MajVote Rule:Give	MajVote Rule:Don't
– – –	1%	3%	6%	13%
– –	1%	11%	7%	19%
–	2%	11%	12%	21 %
+	3%	17%	25%	18%
+ +	26%	34%	37%	19%
+ + +	67%	24%	13%	10%
Mean	.83	.38	.31	−.04
Median	1.00	.67	.50	−.33
Rating ⋚ 0 (Signed rank test (z))	8.73***	5.40***	4.67***	−.57
vs. NoRule (Rank-sum text (z))	4.060***	−4.000***	−5.842***	−8.181***
Panel (a): Action **Don't Give**	**ExoRule = Rule:Give**		**ExoRule = Rule:Don't**	
Rating	MajVote = Rule:Give	MajVote = Rule:Don't	MajVote Rule:Give	MajVote Rule:Don't
– – –	42%	10%	7%	6%
– –	32%	24%	25 %	11%
–	14%	20%	20%	10%
+	5%	20%	17%	13%
+ +	6%	20%	18%	34 %
+ + +	1%	6%	13%	26%
Mean	−.61	−.07	.00	.36
Median	−.67	−.33	−.33	.67
Rating ⋚ 0 (Signed rank test (z))	−7.58***	−1.09	.13	4.87***
vs. NoRule (Rank-sum text (z))	−3.557***	4.449***	5.056***	8.181***

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; all two-tailed.
Ratings are: "very socially inappropriate" (– – –), "socially inappropriate" (– –),"somewhat socially inappropriate" (–), "somewhat socially appropriate" (+), "socially appropriate" (+ +), "very socially appropriate" (+ + +); modal ratings are shaded. For means and medians, responses are converted into numerical scores −1 (– – –), −2/3 (– –), −1/3 (–), +1/3 (+), +2/3 (+ +), +1 (+ + +).

Table A.2: Elicited social approval (full distribution and non-parametric tests) of actions *Give* (Panel a) and *Don't Give* (Panel b) across choice environments in ExoRule×MajVote.

A.2 Subgroup Analysis (stage 1 Behavior)

We analyze the role played by stage 1 choices (Give vs. Don't Give) for ratings elicited in the election treatments. We call a subject who chose action Give in stage 1 a "Giver" and a subject who chose action Don't Give in stage 1 a "Nongiver". Table A.3 reports the distribution of types (Giver and Nongiver) across the four election treatments. The distributions do not statistically differ from each other (Fisher's exact tests, smallest $p = .55$ (two-sided) for Pay4Vote vs. MoneyOffer).

	Number of Subjects				
stage 1 choice	StdMajority	Pay4Vote	MoneyOffer	ExcludePoor	Total
Give (Giver)	63	67	62	65	257
Don't Give (Nongiver)	37	33	38	35	143
Total	100	100	100	100	400

Table A.3: Distributions of choices (Give vs. Don't Give) in stage 1 of election treatments

Decisions in stage 1 cannot account for our main results. Below, we show that: First, differences in ratings for actions Give and Don't Give with respect to the election outcome in StdMajority are highly significant and go into the same direction in *both* subsamples. That is, both Givers and Nongivers agree that the election of Rule:Give makes action Give (Don't Give) more (less) socially appropriate, while the election of Rule:Don't makes action Give (Don't Give) less (more) socially appropriate. Second, norm shifts between StdMajority and NoRule cannot be explained by stage 1 behavior. Significant differences in ratings with respect to the existence of an elected rule are found in both subsamples. Third, differences in ratings between the election treatments cannot be explained by stage 1 behavior since the distribution of Givers and Nongivers is near identical across treatments. Controlling for stage 1 behavior in a regression on ratings across the four election treatment thus leaves treatment effects unaffected.

Table A.4 reports the results of OLS regressions on ratings across the four election treatments by type (Giver, NonGiver) and pooled. Ratings conditional on Rule:Give in treatment StdMajority serve as the baseline. We see that changing the election outcome (Rule = Rule:Don't) has highly significant effects on the social approval ratings of both actions (Give and Don't Give) in both the Giver and the Nongiver subsample. In both subsamples, the effects are large, highly significant and go into the same direction: The election of Rule:Don't makes action Give (Don't Give) less (more) socially appropriate. Effect sizes are higher for Nongivers than for Givers, meaning that Nongivers react somewhat stronger to a change in the election outcome than Givers. We also find that Givers, on average, rate action Don't Give worse than Nongivers (column (8), Giver = 1). In sum, while we find differences between types, these differences do not challenge our main finding that election outcomes (on average and across the entire population) shift norm ratings. Regarding treatment effects (Pay4Vote, MoneyOffer, ExcludePoor), we also do not see large differences across types. If anything, Nongivers react a bit stronger to salient deficiencies in the voting process, especially when interacted with a change of the election outcome (Rule = Rule:Don't). Since the distribution

of Givers and Nongivers is near identical across treatments, controlling for stage 1 behavior (Giver = 1) in the pooled regressions leaves average treatment effects unaffected (columns (4) and (8)).

Dep.Var.	Mean social approval							
	Give				Don't Give			
	Giver	Nongiver	Pooled	Pooled	Giver	Nongiver	Pooled	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pay4Vote	.14	-.07	.07	.06	-.08	.16	.00	.01
	(.077)	(.083)	(.057)	(.058)	(.080)	(.123)	(.067)	(.067)
MoneyOffer	.02	-.04	-.00	-.00	.08	.13	.10	.10
	(.090)	(.080)	(.064)	(.064)	(.090)	(.101)	(.068)	(.098)
ExcludePoor	.06	-.04	.02	.02	.03	.12	.07	.07
	(.087)	(.061)	(.060)	(.060)	(.090)	(.113)	(.070)	(.070)
Rule = Rule:Don't	-.46***	-.76***	-.57***	-.57***	.67***	1.20***	.86***	.86***
(StdMajority)	(.098)	(.129)	(.079)	(.079)	(.101)	(.118)	(.081)	(.081)
×Pay4Vote	-.16	-.05	-.11	-.11	.18	-.33*	-.01	-.01
	(.133)	(.168)	(.106)	(.106)	(.142)	(.197)	(.117)	(.117)
×MoneyOffer	.05	.20	.10	.10	-.23*	-.53***	-.34***	-.34***
	(.124)	(.173)	(.102)	(.102)	(.142)	(.177)	(.114)	(.114)
×ExcludePoor	.08	.22	.14	.14	-.32**	-.65***	-.44***	-.44***
	(.130)	(.174)	(.105)	(.105)	(.141)	(.209)	(.120)	(.120)
Giver = 1				.03				-.18***
				(.040)				(.040)
Constant	.68	.84	.74	.72	-.62	-.66	-.63	-.52
	(.068)	(.051)	(.047)	(.050)	(.064)	(.054)	(.045)	(.048)
Observations	514	286	800	800	514	286	800	800
(Subjects)	(257)	(143)	(400)	(400)	(257)	(143)	(400)	(400)
R^2	.169	.315	.214	.215	.232	.380	.274	.291

Robust standard errors (clustered at subject level) in parentheses:
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.4: Marginal effects on social approval ratings across election treatments by type (Giver vs. Nongiver) and pooled: OLS Regressions. Ratings conditional on Rule:Give in treatment StdMajority serve as the baseline.

In Table A.5 we report the overall distribution of ratings in StdMajority separately for Givers and Nongivers alongside the ratings in NoRule. We see that differences in ratings for actions Give and Don't Give in StdMajority compared to NoRule go into the same direction for Givers and Nongivers. Within StdMajority, the only significant difference in the rating distributions between Givers and Nongivers is found for action Don't Give in the case where Rule:Don't has been elected. Here, the social approval of action Don't Give is significantly higher in the Nongiver compared to the Giver subsample. Even here, however, the effect of the elected rule compared to NoRule goes into the same direction: The election of Rule:Don't makes action Don't Give significantly more appropriate compared to NoRule. We conclude that norm shifts between NoRule and Rule:Give/Rule:Don't are not attributable to one subgroup only, but are driven by a reaction of both types.

Panel (a): Action **Give**	StdMajority				NoRule
	Rule:Give		**Rule:Don't**		**NoRule**
Rating	Givers	Nongivers	Givers	Nongivers	all subjects
– – –	8%	0%	5%	14%	0%
– –	0%	3%	14%	11%	2%
–	0%	0%	17%	16%	1%
+	6%	3%	19%	32%	11%
+ +	37%	30%	25%	11%	49%
+ + +	49%	65%	19%	16%	37%
Mean	.68	.84	.22	.08	.72
Median	.67	1.00	.33	.33	.67
Rating $\gtrless 0$ (Signed rank test (z))	5.50^{***}	5.39^{***}	2.67^{***}	.68	8.60^{***}
vs. NoRule (Rank-sum test (z))	1.10	2.87^{***}	-5.08^{***}	-5.53^{***}	
vs. Givers (Rank-sum test (z))		1.70^{*}		-1.17	

Panel (b): Action **Don't Give**	StdMajority				NoRule
	Rule:Give		**Rule:Don't**		**NoRule**
Rating	Givers	Nongivers	Givers	Nongivers	all subjects
– – –	40%	30%	16%	0%	15%
– –	37%	49%	16%	3%	41%
–	14%	16%	11%	16%	32%
+	3%	5%	21%	8%	8%
+ +	2%	0%	25%	43%	4%
+ + +	5%	0%	11%	30%	0%
Mean	$-.62$	$-.66$	.05	.54	$-.48$
Median	$-.67$	$-.67$	.33	.67	$-.67$
Rating $\gtrless 0$ (Signed rank test (z))	-5.75^{***}	-5.26^{***}	.34	4.66^{***}	-7.40^{***}
vs. NoRule (Rank-sum test (z))	-3.28^{***}	-2.63^{***}	4.37^{***}	7.79^{***}	
vs. Givers (Rank-sum test (z))		.46		3.76^{***}	

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$; all two-tailed.
Ratings are: "very socially inappropriate" (– – –), "socially inappropriate" (– –),"somewhat socially inappropriate" (–), "somewhat socially appropriate" (+), "socially appropriate" (+ +), "very socially appropriate" (+ + +); modal ratings are shaded. For means and medians, responses are converted into numerical scores -1 (– – –), $-2/3$ (– –), $-1/3$ (–), $+1/3$ (+), $+2/3$ (+ +), $+1$ (+ + +).

Table A.5: Elicited social approval of actions *Give* (panel a) and *Don't Give* (panel b) by subgroup (Givers vs. Nongivers): StdMajority vs. NoRule.

A.3 Experimental Instructions

Welcome and Consent Form

This study is hosted by the University of Hamburg [Fribourg/Cologne].
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.

If you have any questions about this research, please feel free to contact us at [email hidden for manuscript].
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.

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 GBP 1.60 for completing the study.
- One participant will receive an extra cash prize of GBP 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 GBP 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.co 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.

Instructions about the Lottery

- 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 GBP 100. The final distribution of lottery tickets depends partly on luck and partly on the choices you and other participants make.
- You will begin with task 1 on the next screen.

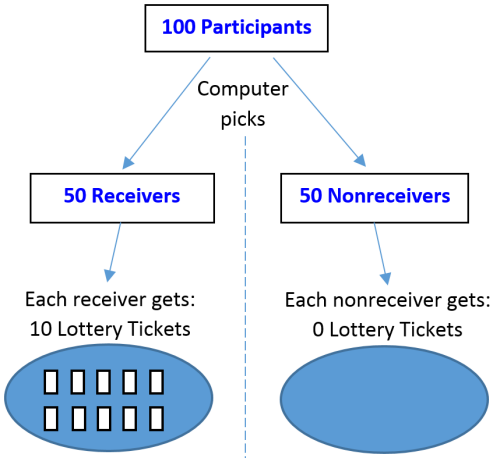
Please tick this box when you have read the instructions and want to proceed:

☐ I have read the information and want to proceed.

Instructions about the Distribution of Lottery Tickets

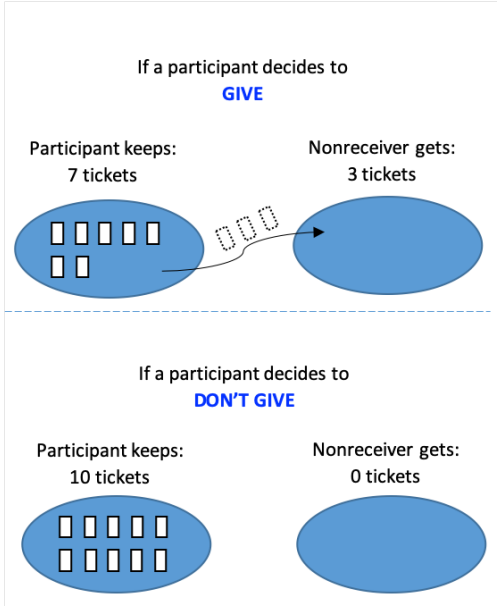
The lottery tickets are distributed in two steps.
Step 1: The computer picks 50 receivers and 50 nonreceivers:

- 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.

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.

To show that you have read and understood the instructions, please answer the following control questions:

If you happen to be a receiver and choose GIVE, you will hold: (numbers appear in random order)

◯ 3 tickets

◯ 10 tickets

◯ 7 tickets (correct)

◯ 0 tickets

If you happen to be a receiver and choose DON'T GIVE, you will hold:

◯ 7 tickets

◯ 10 tickets (correct)

◯ 0 tickets

◯ 3 tickets

If you happen to be a nonreceiver and the other participant chooses GIVE, you will hold:

◯ 0 tickets

◯ 10 tickets

◯ 3 tickets (correct)

◯ 7 tickets

If you happen to be a nonreceiver and the other participant chooses DON'T GIVE, you will hold:

◯ 0 tickets (correct)

◯ 3 tickets

◯ 7 tickets

◯ 10 tickets

Instructions for The Giving Decision (Stage 1)

Task 1

Your Choice: Give or Don't Give

If you happen to be a receiver, 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.

Instructions for the Evaluation Stage (Stage 2)

Task 2

Evaluate choices in a similar situation

On the following screen, you will read the description of a hypothetical choice situation that is very similar to the choice situation you just faced: 100 individuals take part in a lottery that has the exact same structure as the lottery you just took part in. Similar to your choice, each of these individuals has to decide whether to GIVE or DON'T GIVE 3 out of 10 lottery tickets to a nonreceiver.

In this new situation, however, you will NOT be asked to choose yourself. Instead, you will be asked to EVALUATE the different choices available to the other individuals. For each of the possible actions, you will have to decide whether taking that action would be

- ”socially appropriate” and ”consistent with moral or proper social behavior”, or
- ”socially inappropriate” and ”inconsistent with moral or proper social behavior.”

By socially appropriate, we mean behavior that most people agree is the ”correct” or ”ethical” thing to do.

All of the 99 other participants of today’s study will evaluate the same choices in the same hypothetical situation. We will compare your evaluation with the evaluation of the 99 other participants. **If your evaluation is the same as the evaluation most frequently given by the other 99 participants, then you will receive an additional payment of GBP 2.00!**

Note: You and the other 99 participants will evaluate several choices. For the extra payment of GBP 2.00 we will select one of these choices at random. If you evaluate this choice the same way as most of the other 99 participants do then you will receive an additional payment of GBP 2.00. Each of your evaluations has the same chance to be selected for your payment. That is, you maximize your chances to earn GBP 2.00 by trying to always match the most common evaluation in your group.

Note: Your evaluation in Task 2 does NOT influence your chances to win the lottery! The lottery tickets for your group have been distributed in Task 1.

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

To show that you have read and understood the instructions, please answer the following control questions:

	True	False
You will receive GBP 2.00 if your evaluation is the same as the evaluation provided by most of the 99 other participants.	○	○
By socially appropriate, we mean behavior that most people agree is the ”correct” or ”ethical” thing to do.	○	○

Description of the Hypothetical Situation

Consider the following hypothetical situation: 100 other individuals take part in a lottery that has exactly the same structure as the lottery you took part in a few minutes ago. However, in this new lottery, **before anyone of the 100 individuals decides 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.

In the choice situation you have to evaluate, the rule for the code of conduct will be determined in the following way:

Treatment StandardMajority:

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.

Treatment Pay4Vote:

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.
- However, only the votes of those participants who pay GBP 0.20 to make their vote count will be counted in the election.

Treatment MoneyOffer:

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.
- However, before the final votes are counted, 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.

Treatment ExcludePoor:

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.
- However, only the votes of participants with a household income above GBP 40,000 will be counted in the election.

Once a rule has been set, each individual can decide privately and anonymously whether he/she wants to follow the rule or not.

This is the situation you will be asked to evaluate on the next screen. Please make sure to remember it will. In particular, **on the next screen, you will be asked whether it is socially appropriate to follow the rule under the circumstances it has come into force.**

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

To show that you have read and understood the instructions, please answer the following control questions: In the hypothetical choice situation, the code of conduct will equal...
Please choose only one of the following: *[StdMajority, items appear in random order]*

◯ the rule that the majority of individuals voted for.

◯ the rule that was selected using a coin-flip by the computer.

◯ the rule that was selected by the researcher.

In the three malpractice treatments, the additional (correct) items were, respectively:

◯ the rule that the majority of those individuals who pay £0.20 voted for. *[Pay4Vote]*

◯ the rule that the majority of individuals finally voted for—after being offered GBP 0.20 to change their vote. *[MoneyOffer]*

◯ the rule that the majority of individuals with annual household income above GBP 40,000 voted for. *[ExcludePoor]*

The Evaluation Decision

Please evaluate: Choice in the presence of a code of conduct
Here is a reminder of how the rule for the code of conduct is determined: *[Example StdMajority]*

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.

StdMajority, Pay4Vote, MoneyOffer, ExcludePoor:

For each of the two possible rules (RULE: GIVE and RULE: DON'T GIVE), please indicate below how socially appropriate you believe it is to follow the rule and how socially appropriate you believe it is to not follow the rule. Remember that you will earn

money (GBP 2.00) if your evaluation is identical with the most common evaluation given by the other 99 participants of this study.

StdMajority, Pay4Vote, MoneyOffer, ExcludePoor:

RULE: GIVE. Below you see the choices available for each individual if RULE: GIVE is implemented as the code of conduct. Please indicate how socially appropriate you believe each choice to be.

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Follow the rule and GIVE.	○	○	○	○	○	○
Don't follow the rule and DON'T GIVE.	○	○	○	○	○	○

RULE: DON'T GIVE. Below you see the choices available for each individual if RULE: DON'T GIVE is implemented as the code of conduct. Please indicate how socially appropriate you believe each choice to be.

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Follow the rule and DON'T GIVE.	○	○	○	○	○	○
Don't follow the rule and GIVE.	○	○	○	○	○	○

All treatments:

Feedback Screen and End of the Study

Thank you very much for your participation. Your evaluations have been saved.

Did you feel that it makes a large difference whether to choose GIVE or DON'T GIVE or were you very much indifferent between choosing any of the two?

○ Was very much indifferent/did not care.

○ Felt that it makes a large difference.

Below, you have the opportunity to leave a more general personal comment. We greatly appreciate your feedback.

Please make sure to click ”Submit” and then the link on the next page in order to prove that you have completed the study! Thank you!

Treatment variation ExoRule×MajVote

The instructions in treatment variation ExoRule×MajVote were identical to the StdMajority treatment, except from the following parts:

Description of the Hypothetical Situation

- To determine the rule, a coin is flipped by the computer. If the coin lands heads (50% chance), RULE: GIVE is implemented. If the coin lands tails (50% chance), RULE: DON'T GIVE is implemented. Note: Whether RULE: GIVE or RULE: DON'T GIVE is implemented depends purely on chance. That is, the rule is random.

In ExoRule×MajVote, we asked the following two control questions:

If RULE: GIVE is implemented as the code of conduct, this is because...

◯ ... the majority of individuals who take part in the lottery prefer RULE: GIVE

◯ ... by chance, the coin flip landed heads.

◯ ... RULE: GIVE is the only rule that can be selected.

If RULE: DON'T GIVE is implemented as the code of conduct, this is because...

◯ ...the majority of individuals who take part in the lottery prefer RULE: DON'T GIVE

◯ ... by chance, the coin flip landed tails.

◯ ... RULE: DON'T GIVE is the only rule that can be selected.

In ExoRule×MajVote, the information given on the evaluation decision page differed from the other treatment variations:

The Evaluation Decision

Please evaluate now:
How socially appropriate is it to follow/not follow **RULE: GIVE** if it is implemented?
How socially appropriate is it to follow/not follow **RULE: DON'T GIVE** if it is implemented?
Below, we ask you to answer these questions **under two different scenarios**:

- **Scenario 1:** The majority of individuals who take part in the lottery prefer RULE: GIVE. That is, in a majority vote, RULE: GIVE would win against RULE: DON'T GIVE.

- **Scenario 2:** The majority of individuals who take part in the lottery prefer RULE: DON'T GIVE. That is, in a majority vote, RULE: DON'T GIVE would win against RULE: GIVE.

In total there are 4 possible situations that you have to evaluate.
Please evaluate each situation according to what you think most people find socially appropriate in that particular situation
Remember that you will earn money (£2.00) if your evaluation is identical with the most common evaluation given by the other 99 participants of this study.

ExoRule×MajVote: example of the rating decision table

Evaluation 1:
Please evaluate the following situation:

- The coin flip lands heads. **RULE: GIVE is implemented.**
- **The majority prefers RULE: GIVE.** That is, in a majority vote, RULE: GIVE would win against RULE: DON'T GIVE.

How socially appropriate is it to follow/not follow RULE: GIVE in this case?

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
To follow the rule and GIVE.	○	○	○	○	○	○
To not follow the rule and DON'T GIVE.	○	○	○	○	○	○

Example Screenshots: StdMajority

Task 2

Evaluate choices in a similar situation

On the following screen, you will read the description of a hypothetical choice situation that is very similar to the choice situation you just faced: 100 individuals take part in a lottery that has the exact same structure as the lottery you just took part in. Similar to your choice, each of these individuals has to decide whether to GIVE or DON'T GIVE 3 out of 10 lottery tickets to a nonreceiver.

In this new situation, however, you will NOT be asked to choose yourself. Instead, you will be asked to EVALUATE the different choices available to the other individuals. For each of the possible actions, you will have to decide whether taking that action would be

- "socially appropriate" and "consistent with moral or proper social behavior", or
- "socially *in*appropriate" and "*in*consistent with moral or proper social behavior."

By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do.

All of the 99 other participants of today's study will evaluate the same choices in the same hypothetical situation. We will compare your evaluation with the evaluation of the 99 other participants. **If your evaluation is the same as the evaluation most frequently given by the other 99 participants, then you will receive an additional payment of £2.00!**

Note: You and the other 99 participants will evaluate several choices. For the extra payment of £2.00 we will select one of these choices at random. If you evaluate this choice the same way as most of the other 99 participants do then you will receive an additional payment of £2.00. Each of your evaluations has the same chance to be selected for your payment. That is, you maximize your chances to earn £2.00 by trying to always match the the most common evaluation in your group.

Note: Your evaluation in Task 2 does NOT influence your chances to win the lottery! The lottery tickets for your group have been distributed in Task 1.

Figure A.1: Screenshot: Description of the evaluation decisions 1

Consider the following hypothetical situation:

100 other individuals take part in a lottery that has exactly the same structure as the lottery you took part in a few minutes ago. However, in this new lottery, **before anyone of the 100 individuals decides 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.

In the choice situation you have to evaluate, the rule for the code of conduct will be determined in the following way:

- All 100 individuals who take part in the lottery are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.

Once a rule has been set, each individual can decide privately and anonymously whether he/she wants to follow the rule or not.

This is the situation you will be asked to evaluate on the next screen. Please make sure to remember it well. In particular, **on the next screen, you will be asked whether it is socially appropriate to follow the rule under the circumstances it has come into force.**

Figure A.2: Screenshot: Description of the evaluation decisions 2 (*StdMajority*)

Please evaluate: Choice in the presence of a code of conduct

Here is a reminder of how the rule for the code of conduct is determined:

- All 100 individuals are asked to vote for the rule (RULE: GIVE or RULE: DON'T GIVE) they prefer to have implemented as the code of conduct. The rule that receives more votes in total will be implemented as the code of conduct.

For each of the two possible rules (RULE: GIVE and RULE: DON'T GIVE), please indicate below how socially appropriate you believe it is to follow the rule and how socially appropriate you believe it is to *not* follow the rule. Remember that you will earn money (£2.00) if your evaluation is identical with the most common evaluation given by the other 99 participants of this study.

RULE: GIVE. Below you see the choices available for each individual if RULE: GIVE is implemented as the code of conduct. Please indicate how socially appropriate you believe each choice to be.

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Follow the rule and GIVE.	○	○	○	○	○	○
Don't follow the rule and DON'T GIVE.	○	○	○	○	○	○

RULE: DON'T GIVE. Below you see the choices available for each individual if RULE: DON'T GIVE is implemented as the code of conduct. Please indicate how socially appropriate you believe each choice to be.

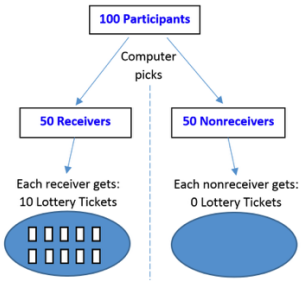
Figure A.3: Screenshot: Evaluation decisions (*StdMajority*)

Example Screenshots: NoRule

Instructions: Evaluate Choices in a Similar Situation

For the following questions we ask you to please consider the following hypothetical situation:

100 individuals in another group take part in a lottery that has the exact same structure as the lottery you are taking part in. That is, of the 100 participants, the computer selects 50 as receivers (who each get 10 lottery tickets) and 50 as nonreceivers (who get 0 lottery tickets):

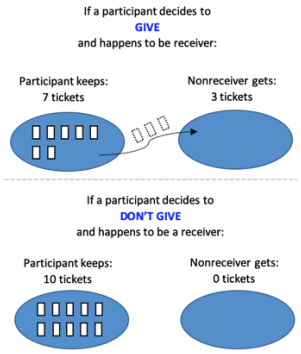


However, in this new hypothetical situation, the distribution of lottery tickets does not stop here.

Before the winning ticket is drawn, all participants get the chance to decide whether they want to share tickets with nonreceivers:

- Before participants learn whether they are a receiver or nonreceiver, each participant decides whether he/she wants to give away 3 of the 10 lottery tickets to a nonreceiver.
- This decision (**GIVE** or **DON'T GIVE**) has the following consequences:

Figure A.4: Screenshot: Description of the Evaluations 1



- **Note that everyone makes this decision without knowing whether he/she is a receiver or nonreceiver of lottery tickets.** In the case that the participant is a non-receiver, his/her decision to GIVE or DON'T GIVE does not play a role. The number of lottery tickets this participant holds then depends on the choice of another participant.

On the following screen, we will ask you to evaluate how socially appropriate you believe each of the two possible actions (GIVE and DON'T GIVE) to be:

You will have to decide whether you think that taking action GIVE would be

- "socially appropriate" and "consistent with moral or proper social behavior", or
- "socially inappropriate" and "inconsistent with moral or proper social behavior"

and whether you think that taking action DON'T GIVE would be

- "socially appropriate" and "consistent with moral or proper social behavior", or
- "socially inappropriate" and "inconsistent with moral or proper social behavior".

By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do.

Please note: How you evaluate actions GIVE and DON'T GIVE does NOT influence your chances to win the lottery. In your group, lottery tickets will only be distributed by the computer and not by anyone else.

Figure A.5: Screenshot: Description of the Evaluations 2

Your Evaluation

Your answers to the control questions were correct.

Below, we ask you to evaluate actions GIVE and DON'T GIVE in the hypothetical choice situation explained on the previous screen.

For each of those two actions, we ask you to decide whether you think that taking that action would be

- "socially appropriate" and "consistent with moral or proper social behavior", or
- "socially inappropriate" and "inconsistent with moral or proper social behavior".

By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do.

With your evaluation you can earn an additional £2.00 to your base payment!

Whether you receive the additional £2.00 is determined as follows:

- All of the 99 other participants of today's study evaluate the same actions in the same hypothetical situation.
- We will compare your evaluation with the evaluation of the 99 other participants. If your evaluation is the same as the evaluation most frequently given by the other 99 participants, then you will receive an additional payment of £2.00.*

*For the extra payment of £2.00 we will select one of the two evaluations (GIVE or DON'T GIVE) at random. Both of your evaluations have the same chance to be selected for the extra payment. You maximize your chances to earn an additional £2.00 by trying to always match the most common evaluation among the other 99 participants. Neither your base payment of £1.10, nor your chances to win the cash prize of £100.00 in the lottery is affected by whether you receive the extra payment.

Figure A.6: Screenshot: Evaluations of Giving Decisions 1

*Evaluation 1. Please indicate how socially appropriate you believe action GIVE to be:

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Action **GIVE** is	○	○	○	○	○	○

Remember that you will earn £2.00 if your evaluation is identical with the most common evaluation given by the other 99 participants of this study!

*Evaluation 2. Please indicate how socially appropriate you believe action DON'T GIVE to be:

	Very socially inappropriate	Socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Socially appropriate	Very socially appropriate
Action **DON'T GIVE** is	○	○	○	○	○	○

Remember that you will earn £2.00 if your evaluation is identical with the most common evaluation given by the other 99 participants of this study!

Figure A.7: Screenshot: Evaluations of Giving Decisions 2