Moral transgressions by groups: Guilt sharing and its underlying motives

Eberhard Feess[*], Florian Kerzenmacher[†], Gerd Muehlheusser[‡]

September 21, 2021

Abstract

We conduct an experiment where subjects are matched in groups of three and vote on a moral transgression. We find that the frequency of votes for the moral transgression increases with the number of votes required for it. This effect persists when considering pivotal votes only, thereby eliminating opportunities to rely on sufficiently many votes for the transgression by other group members. Our design enables us to cleanly identify guilt sharing as a driver of voting behavior. A series of novel treatments then allows separating shared responsibility and a preference for group consensus as independent motives underlying guilt sharing.

Keywords: group decisions, unethical behavior, experiment, voting, diffusion of responsibility, guilt sharing, donations

JEL Classification: C92, D02, D63, D71

[*]Victoria University of Wellington, School of Economics and Finance, Pipitea Campus, Lambton Quay, Wellington, New Zealand; eberhard.feess@vuw.ac.nz

[†]University of Innsbruck, Department of Economics, Universitätsstr. 15, 6020 Innsbruck, Austria; florian.kerzenmacher@uibk.ac.at

[‡]University of Hamburg, Department of Economics, Von-Melle-Park 5, 20146 Hamburg, Germany, and CESifo and IZA; gerd.muehlheusser@uni-hamburg.de

Acknowledgements This paper is a substantial revision of a previous version, circulated as CESifo DP No. 13383 and IZA DP No. 8384. We are grateful to Daniel Chen, Martin Dufwenberg, Paul Heidhues, Bernd Irlenbusch, Oliver Kirchkamp, Aniol Llorente-Saguer, Andreas Nicklisch, Petra Nieken, Ronald Peeters, Timo Promann, Anna Ressi, Simeon Schudy, Nora Szech, Wendelin Schnedler, Georg von Weizsäcker, Florian Zimmermann, and Daniel Zizzo for very valuable comments. We also benefited from comments at the ESA World Meeting (Vancouver, 2019), SABE 2020 Annual Congress, EEA Virtual 2020, the 2020 CESifo Behavioral Economics Conference, EALE 2021, and at seminar presentations at the Universities of Auckland, Dunedin, Hamburg, Hamilton, Paderborn, and Swansea. Financial support from the German Research Foundation (DFG, Grant 658143) and the Austrian Science Fund (FWF, SFB F63) is gratefully acknowledged.

1 Introduction

A large body of experimental research has shown that the frequency of morally questionable decisions is higher when people decide as members of a group instead of taking such decisions individually (e.g. Dana, Weber, and Kuang, 2007; Muehlheusser, Roider, and Wallmeier, 2015; Kocher, Schudy, and Spantig, 2018; Irlenbusch and Saxler, 2019; Falk, Neuber, and Szech, 2020). This phenomenon is often referred to as *guilt sharing*, and attributed to a perceived reduction in moral costs when people can share the responsibility for an immoral outcome with others (Bartling and Fischbacher 2012; Rothenhäusler, Schweizer, and Szech, 2018).

This paper makes two main contributions: First, we argue that the literature does often not cleanly identify guilt sharing as a relevant motive in the context of morally questionable group decisions, which is due to different financial incentives between the individual and group settings (see the discussion in Section 2). To see this, consider a setting where, by taking such a decision, single individuals can increase their payoff by 1$ at the expense of others. Compare this to a group setting, where the same decision taken by the group yields 1$ to each member of the group. If we observe that people are

more inclined towards moral transgressions in the group setting, then this could either be due to guilt sharing and/or the positive payoff externality on other group members. This payoff externality is an example where the expected marginal financial benefit of a decision differs across the two settings, thereby potentially constituting a confounding factor. Our first contribution is therefore to design an experiment where such confounding factors are excluded so that guilt sharing can be cleanly identified.

Our second contribution is to disentangle and quantify two related yet different motives underlying guilt sharing. We refer to these as a *preference for group consensus* and *shared responsibility*, respectively. To see the difference, consider a group with three members: A, B and C. Assume a situation in which A first collects the opinions of B and C, and then decides unilaterally for or against the transgression. The preferences of B and C may matter for A's decision, as she has a preference for group consensus, and also because she might take those as a signal on the social norm. Now compare this to a situation where the group decision is determined by a vote that requires unanimity. In this case, A knows that she is pivotal for the transgression if and only if both B and C also vote in favor of it. Again, a preference for group consensus may reduce A's moral concerns if B and C support the transgression, but this reduction may be reinforced when B and C could each have prevented it. We define this latter effect as shared responsibility, and hence distinguish between the cases where other group members cannot only share their preferences with respect to the transgression, but can also decide on it.[1] Making such a distinction is clearly supported by our experimental findings.

In our pre-registered experiment, subjects are matched in groups of three and decide on taking money designated for donation to a charity. The group decision is determined by a vote in which group members vote independently for or against taking the money (YES or NO). Such a well-structured voting procedure is vital for our research objective, as we can vary the relative importance of peoples' underlying motives by varying the voting

[1] Of course, our labelling of the different motives is somewhat arbitrary. The point we aim to make is that what is often referred to as guilt sharing may be induced by different motives, and that it is sensible to distinguish between them. Accordingly, sharing responsibility with someone who cannot influence the result is excluded by our definition of this motive.

rule. As discussed below, it also enables us to keep the financial incentives constant across treatments and conditions.[2] We consider three voting rules which differ in the number of affirmative votes required for the transgression (the *voting threshold* k), dictatorial[3] ($k = 1$), majority ($k = 2$), and unanimity ($k = 3$) voting.

In our baseline treatment *SIM* subjects vote simultaneously. In line with predictions from a recent game-theoretic model by Rothenhäusler, Schweizer, and Szech (2018), we find that the number of subjects voting YES increases in the threshold k, which we refer to as *threshold effect*. This finding is interesting as it shows that setting a high bar in terms of required support for a morally questionable outcome might potentially backfire, as it increases the willingness to vote for the moral transgression. However, treatment *SIM* does not yet allow us to cleanly identify guilt sharing as a motive, which is again due to differing expected marginal financial benefits across conditions as in the example above. In our setting, an individual might benefit from the transgression even when voting NO, thereby *free-riding* on sufficiently many YES votes by other group members. This free-riding possibility decreases in the voting threshold k: it is largest for $k = 1$ (only one other YES required), smaller for $k = 2$ (both other group members must vote YES), and fully eliminated for $k = 3$ as every group member needs to vote YES for the transgression.

To eliminate free-riding possibilities (and hence keep the expected marginal financial benefits constant), all of our subsequent treatments focus exclusively on pivotal votes. Each group consists of two *unconditional* voters and one *conditional* voter. Unconditional voters decide on their votes in the same way as in *SIM*. For conditional voters, we apply the strategy method and elicit their vote for each possible number of YES votes by the unconditional voters in their group. We then consider decisions of conditional voters when they are pivotal for the transgression, i.e. in case of $k-1$ unconditional YES votes.[4]

[2] In reality, the process of how group decisions are reached will typically be less stringently structured and also contain deliberative elements.

[3] We use "dictatorial voting" to capture the case where one YES vote suffices to implement the transgression, no matter from which voter. Hence, none of the group members is a "dictator" in the sense that (only) this member's vote (YES or NO) determines the group decision.

[4] Focusing on pivotal votes also rules out pivotality aversion as a further potential driver of the threshold effect.

Treatment differences in the number of YES votes can then no longer be attributed to different incentive structures. In our first treatment without free-riding, *P-BASE*, we indeed find that the frequency of pivotal conditional YES votes still increases with the voting threshold k, i.e. guilt sharing matters.

After having established the presence of guilt sharing, we turn to our second contribution, the identification of shared responsibility and group consensus as underlying motives. To this end, our treatment *P-RAND* eliminates shared responsibility by replacing the two unconditional votes in each group by random votes. The decision to take the money is now determined by the conditional vote and the two random votes. Comparing this treatment to *P-BASE*, one might be concerned that subjects may form treatment-specific beliefs about what constitutes a social norm, which may then also influence their voting behavior. We therefore introduce another treatment *P-INFO*, which is a replication of *P-BASE* except conditional voters are informed about the overall frequency of unconditional YES votes in treatment *P-BASE*. This information is supposed to provide a signal about the prevailing social norm, and is also given to conditional voters in *P-RAND*. This keeps the impact of social norm constant among treatments.

We find that the information on the social norm does not change the behavior of conditional voters in *P-INFO* compared to *P-BASE*, so that the threshold effect still persists. The elimination of shared responsibility in *P-RAND* then leads to a lower impact of the voting threshold k on the number of pivotal YES votes by conditional voters compared to *P-INFO*. Thus, it indeed makes a difference for a conditional voter's behavior if her fellow group members' votes just express their preference for the transgression, or whether these votes can also directly influence the outcome.

Treatment *P-RAND* also allows us to establish the preference for group consensus as an independent motive[5]: While the unconditional voters in each group are no longer

[5]Note that preferences for social conformity and group consensus may be aligned, but they may also go in opposite directions when the opinions of the own group members do not coincide with what is seen as the social norm. For example, consider a situation where subjects are assigned to 100 groups with three members each. Now suppose an individual knows that overall just 20 percent, but both own group members are in favor of the moral transgression. In this case, preferences for social conformity and for

relevant for the decision to take the money, we nevertheless elicit the conditional votes for each number of unconditional YES votes. We can thereby analyze, separately for each voting threshold k, if these unconditional YES votes have an impact on the decisions of conditional voters. We find that this is indeed the case: For each voting rule, the number of pivotal YES votes by conditional voters increases with the number of (irrelevant) unconditional votes for it. This provides strong evidence that a preference for group consensus is a relevant motive in our context.

Finally, our framework allows us to analyze whether our list of relevant motives (free-riding, social conformity, shared responsibility, and group consensus) is close to exhaustive. This is accomplished by resorting again to treatment *P-RAND* (which neutralizes the first three motives) and fixing in addition the number of unconditional YES votes in a group (which neutralizes the fourth one). Under the hypothesis that our design indeed captures all relevant motives, a necessary condition is that the threshold effect vanishes, that is, the number of pivotal conditional YES votes should no longer vary in the voting threshold k. We find that this is indeed the case.

The remainder of the paper is organized as follows: Section 2 discusses the related literature. Section 3 introduces the experimental design and the treatments from which the hypotheses are derived. The results are presented in Section 4. Section 5 concludes and points to further research. Appendix A develops a formal framework underlying the hypotheses. Appendix B contains further robustness checks. Appendix C contains more information about the experimental procedures.

2 Relation to the Literature

Our paper contributes to the recent literature on moral transgressions in groups by a clean identification of guilt sharing, and by separating shared responsibility and group consensus as independent underlying motives. Falk, Neuber, and Szech (2020) compare decision making by individuals and groups in a setting similar to Falk and Szech (2013),

group consensus go in opposite directions.

where subjects decide whether or not to forgo 10 Euros each in order to save mice from being killed. Groups consist of eight subjects and eight mice are killed if at least one subject chooses to take the money (in our terminology, they consider the dictatorial rule). They find that more individuals opt for the transgression in the group setting compared to the setting with individual decision making, and this result is also confirmed in an alternative design with donations. This suggests that guilt sharing plays a role, but the incentive structure in the two treatments differs: In the individual setting, subjects certainly trade 10 Euros against the life of a mouse. In the group setting, there are two cases: If at least one group member voted YES (i.e. for sacrificing the mice), then voting NO has no material benefit. If all group members voted NO, then voting YES effectively kills eight mice. The subjectively perceived expected negative externality when voting YES is thus identical to the individual setting if one assumes with a probability of 1/8 that all other members vote NO.

In Irlenbusch and Saxler (2019), a seller and a buyer can gain a monetary benefit at the expense of a charity if both agree on it. This unanimity rule for the moral transgression yields a higher willingness for the moral transgression compared to individual decision making. Their approach builds on Dana, Weber, and Kuang (2007) who compare the behavior in different variants of the dictator game. In one of their treatments, two dictators decide between sharing the money equally with a recipient and keeping the lion's share to themselves. With unanimity required for the moral transgression, the unfair outcome is chosen more often than in a standard dictator game. Both papers clearly demonstrate that groups are more prone to selfish behavior.[6] However, the incentive structure differs from individual decision making in two important respects: First, voting for the moral transgression yields a positive externality on the partner. In line with this point, Wiltermuth (2011) finds that subjects lie more often when the benefit is shared

[6]While we focus on individual behavior in groups, a number of studies have also investigated the erosion of moral values in market settings as compared to individual decision making, see e.g. , Bartling, Weber, and Yao (2015). However, Bartling, Fehr, and Özdemir (2021) show that the difference between the individual and the market setting disappears when the number of rounds is the same. Hence, the effect identified in Falk and Szech (2013) might also be driven by repeated interaction rather than markets.

with dummy players, which is related to preferences for white lies as in Erat and Gneezy (2012). Second, the partner's vote may serve as a signal on the social norm.

Conrads et al. (2013) find that lying in teams is more frequent compared to individual lying; even when the own marginal monetary benefit is lower. This, however, might not (only) be triggered by guilt sharing but by providing a positive externality for the partner. Kocher, Schudy, and Spantig (2018) also find a higher lying frequency in groups. In one of their settings, this can be attributed to the necessity of coordination, and hence to a different incentive structure compared to purely individual decision making. In the other setting, the effect is driven by communication, as each individual's report determines only their own payoff. Our setting differs as votes determine the outcome for the whole group.

A related strand of literature focuses on the impact of social conformity, and finds mixed results: Monroe et al. (2018) find that the assessment of moral behavior depends to a large degree on what the majority does and less on strict moral rules. Nook et al. (2016) find that subjects donate more after learning that other participants donated high amounts.[7] By contrast, recent studies suggest that social conformity and updates on the behavior of subjects one is matched with have little impact in lying experiments: In Peeters, Vorsatz, and Walzl (2015), the own behavior is not significantly influenced by the belief about the lying frequency of subjects one is matched with. Dato, Feess, and Nieken (2019) find that this holds even in highly competitive situations, as lying by competing contestants has no impact on the own lying frequency. We do not analyze the impact of social conformity but keep it constant among treatments to enable a clean identification of shared responsibility and group consensus.[8]

[7]Somewhat contradictory, Gächter, Gerhards, and Nosenzo (2017) find that, while the behavior of other subjects influences what is viewed as morally appropriate in a dictator game, it has only a limited impact on the actual own behavior.

[8]Both our notions of shared responsibility and group consensus can be considered *narratives* in the theoretical framework of Bénabou, Falk and Tirole (2020). In this case, a group member's YES vote is a message that allows a pivotal decision maker to downplay the own responsibility ("someone else could have prevented the transgression") or to consider the group member's payoff ("someone else needs the money") when taking the selfish decision.

We consider only pivotal votes to ensure that pivotality aversion cannot contaminate our analysis of threshold effects. The role of pivotality is studied in detail in Bartling, Fischbacher, and Schudy (2015), who find that those who are pivotal for a moral transgression are more severely punished compared to other subjects. Brütt, Schram, and Sonnemans (2020) find that a reduction in pivotality reduces moral concerns. Huck and Konrad (2005) show theoretically that, with non-consequentialist moral costs, the probability of reaching a predetermined threshold decreases in the committee size due to a decline in the probability of being pivotal.

Last but not least, our treatment *SIM*, where all group members are symmetric, tests one of the main propositions developed by Rothenhäusler, Schweizer, and Szech (2018). Their game theoretic model shows that, with consequentialist moral preferences and shared responsibility, the frequency of YES votes increases in k. However, they also point out that the same result can be derived without shared responsibility as free-riding opportunities decrease with k.

3 Experimental design, treatments, and hypotheses

In this section, we explain the basic setup of the experiment and the treatments we have conducted. The treatments are designed to identify guilt sharing as a driver of group decisions on moral transgression, and to disentangle shared responsibility and group consensus as underlying motives. This leads to a total of seven hypotheses to be tested (see Appendix A for a formal framework in which these hypotheses can be derived).

3.1 Experimental design

Subjects were randomly assigned to groups of three. They voted (YES or NO) on whether or not the group should take an amount of \$1.50 originally designated for donation to the charity *Make-A-Wish*, which engages in fulfilling the wishes of children with a critical

illness.[9] We adopt a donation setting, as we want to consider a decision that harms subjects in need outside of the laboratory, and because taking money already designated for donation can clearly be seen as a moral transgression (see e.g. Ariely, Bracha, and Meier, 2009; Bénabou and Tirole, 2010; Kirchler et al., 2016; Casal, Fallucchi, and Quercia, 2019; Feess et al., 2018). If sufficiently many group members voted for taking the money, the \$1.50 were split equally among all three of them. We consider three voting rules that differ with respect to the threshold k of YES votes required for taking the money. Throughout, we refer to $k = 1$, $k = 2$, and $k = 3$ as the *dictatorial*, *majority*, and *unanimity* rule, respectively, and to a *threshold effect* when the number of YES votes increases in k.

3.2 Treatments and hypotheses

We consider a total of four treatments (see Table 1 for an overview). As explained in more detail below, we have employed a between-subject design. Each subject was randomly allocated to one treatment and then played under one of the three voting rules.

Treatment *SIM* In this treatment all three group members are *unconditional* voters in the sense that none of them can condition her own vote on the votes of the other two group members. This treatment hence captures the case of simultaneous voting as considered in the theoretical framework by Rothenhäusler, Schweizer, and Szech (2018). They consider a group of size $n \geq 1$ where each member votes on a moral transgression that would yield the same monetary payoff to each group member, and which is implemented if there are at least k votes in favor of it. In the main part of their paper, they assume that subjects face moral costs if and only if they vote for the moral transgression *and* the transgression is actually implemented. Such types of moral costs are usually referred to as *consequentialist*

[9] See e.g. `www.wish.org`. Prior to the actual experiment, we ran an online survey (with a different subject pool) and asked subjects to rank several well-known charities such as the American Red Cross, the Children's Scholarship Fund, Doctors without borders, Donate Life and UNICEF with respect to their reputation. Make-A-Wish was ranked highest among these.

(Sinnott-Armstrong, 1988).[10]

Moral costs are divided into two components, *non-diffusive moral costs* and *diffusive moral costs*, where the latter decrease in the number of votes in favor of the moral transgression.[11] Rothenhäusler, Schweizer, and Szech (2018) analyze (symmetric Bayesian Nash) equilibria, where each subject votes for the moral transgression if and only if her (privately known) moral costs are below a critical value. This yields their Proposition 1, part (ii) (see Appendix), which translates directly into our first hypothesis.

Hypothesis 1. *In treatment* SIM*, the frequency of YES votes for the moral transgression increases in the voting threshold k.*

Our experiment provides a first empirical test of this basic prediction. It is important to note that the presence of non-diffusive moral costs alone is sufficient to derive Hypothesis 1 as the three voting rules give rise to different free-riding opportunities: The higher the likelihood that sufficiently many other group members vote for the moral transgression, the higher is the incentive to save the own moral costs by voting against it. In equilibrium, this likelihood decreases in the threshold k and is zero under unanimity. Hence, while testing Hypothesis 1 is adequate for analyzing the overall effect of voting thresholds, it is not informative about the underlying motives. Investigating those motives in more detail is the aim of our subsequent treatments.

Treatment *P-BASE* In order to identify guilt sharing as a relevant motive in our context, treatment *P-BASE* eliminates the differences in free-riding opportunities between the three voting thresholds. In contrast to *SIM*, one of the three members in each group is randomly chosen to become a *conditional* voter. For these conditional voters, we use the strategy method and elicit their decision for all possible numbers of YES votes by their

[10] Conversely, individuals incur *non-consequentialist* moral costs whenever they vote for the moral transgression, irrespective of whether or not it is actually carried out (Alexander and Moore, 2016).

[11] They use the term guilt sharing for diffusive moral costs. In their setting, there is no difference between shared responsibility and group consensus. Their model also includes an aversion against being pivotal. We neglect this cost type as it does not influence the impact of the voting threshold in their model. Furthermore, we consider only pivotal votes anyway in all of our subsequent treatments.

two fellow group members (zero, one, or two) who are unconditional voters as in *SIM*. We then analyze the voting behavior of conditional voters for those vote constellations in which they are pivotal, i.e. where the number of unconditional YES votes in her group equals $k-1$.[12] Throughout, we refer to YES votes by conditional voters when being pivotal as *pivotal conditional YES votes*. As pivotal voters know that the money is taken if and only if they vote YES, the consequences of their votes on their own monetary payoff and on the payoff of their two fellow group members are the same in all voting rules If the threshold effect continues to exist, then this would cleanly identify guilt sharing as one driver of group decisions on moral transgression. Note that considering pivotal votes only also has the merit that it does not matter whether moral costs are consequentialist or non-consequentialist.

Hypothesis 2. *In treatment* P-BASE *the frequency of pivotal conditional YES votes increases in the voting threshold k.*

To permanently eliminate free-riding incentives from the subsequent analysis, all further treatments also focus on pivotal conditional YES votes. Moreover, apart from the introduction of conditional voters, there is also a second difference between treatment *SIM* on the one hand and *P-BASE* (and the two subsequent treatments *P-INFO* and *P-RAND*) on the other: Before asking for their vote, unconditional voters were informed that it might be the case that their votes do not count towards the decision to take the money. We did so to ensure that unconditional voters are exactly in the same situation in all of these three treatments where the focus is on conditional voters. The reason is that, if unconditional voters were in different situations, and if conditional voters anticipated this, then they might interpret the same behavior of unconditional voters differently. In turn, this might influence their own behavior.[13]

[12]Note that whether the money is taken still depends on the actual choices of all three group members. For example, consider the case k = 3 with (i) a conditional voter who votes YES for any number of unconditional YES votes, and (ii) zero unconditional YES votes. For our analysis, we use only the decision where the conditional voter is pivotal, i.e. her decision for the case of two unconditional YES votes. However, the moral transgression is not implemented as there is only one actual YES vote in the group.

[13]In the empirical analysis, we do not utilize the unconditional votes from these treatments. Uncondi-

Treatment *P-INFO* This treatment is identical to *P-BASE* with the only difference that, before asking conditional voters for their decision strategy, they were informed about the average voting behavior of the unconditional voters in *P-BASE* (which was played beforehand) under the respective voting rule. For instance, under majority voting in *P-BASE*, 65% of the unconditional voters voted YES, and this information was then provided to all conditional voters assigned to the majority rule in *P-INFO*. Such information was meant to provide conditional voters with a signal about the prevailing social norm in the population. The reason why we provide this information is that the test of Hypothesis 5 below, which is key for the identification of shared responsibility, will be based on a comparison of the voting behavior in treatments *P-INFO* and *P-RAND*. To cleanly identify this motive, we need to make sure that the perception on what constitutes the social norm is the same in both treatments.

For *P-INFO* itself, we presume that providing the percentages of unconditional votes for the moral transgression does not erase the threshold-effect already predicted in *P-BASE*:

Hypothesis 3. *In treatment* P-INFO *the frequency of pivotal conditional YES votes increases in the voting threshold k.*

Treatment *P-RAND* This final treatment is identical to *P-INFO* except for one feature: The decision for the moral transgression is not determined by the three votes in a given group, but by the group's conditional voter and two random votes. The probability of the random vote being YES is taken from the frequency of YES votes in *P-BASE* for the respective voting threshold.

Whether a conditional voter is pivotal is hence no longer determined by the unconditional votes in her group, but by the outcome of the random votes. In our design, unconditional votes do not count for the group decision to take the money. But when the money is taken, it is still shared among all three group members. For example, suppose

tional votes are only used in the analysis of treatment *SIM*, where all voters are unconditional and where they are certain that their votes do count with respect to the decision to carry out the transgression.

Table 1: Overview over treatments and hypotheses

	SIM	*P-BASE*	*P-INFO*	*P-RAND*
Includes conditional voters	No	Yes	Yes	Yes
Unconditional voters know vote might not count	n/a	Yes	Yes	Yes
Signal about social norm	No	No	Yes	Yes
Random votes	No	No	No	Yes
Motive eliminated	None	Free-riding	Free-riding	Free-riding, Shared responsibility
Hypotheses	1	2	3 & 5	4 – 7

that $k = 3$ and that there are zero unconditional YES votes, two random YES votes and that the conditional voter votes YES for any number of random YES votes. Then the money is taken by the group and all three members receive their according payoffs, albeit the two unconditional voters actually voted against it.

The purpose of substituting the actual votes of unconditional voters by random votes is to eliminate shared responsibility as a potential motive for conditional voters. Recall that according to our definition, conditional pivotal voters can no longer share responsibility for an immoral outcome with the other group members when these votes do not matter for the decision. The effect of eliminating shared responsibility, and hence the difference in the number of pivotal conditional YES votes between *P-INFO* and *P-RAND*, should be strongest under the unanimity rule, followed by the majority rule. Under the dictatorial rule, responsibility cannot be shared as a conditional voter can decide alone on the moral transgression, irrespective of other votes (actual or random).

When eliciting the decisions of conditional voters via the strategy method, we take into account both unconditional and random votes, leading to $3^2 = 9$ decisions for conditional voters (one decision for each of zero, one, or two unconditional and random YES votes, respectively). We first investigate the behavior of conditional voters when we substitute

the unconditional votes by random votes but keep everything else constant. We then consider solely the case where the number of random YES *and* the number of unconditional YES votes in a group is $k-1$. We thereby eliminate shared responsibility because the actual votes of unconditional voters do not count for the final decision. At the same time, however, we keep the potential motive of group consensus in place as the number of $k-1$ unconditional YES votes that do not count for the final decision is zero in the dictatorial rule, one in the majority rule, and two in the unanimity rule. This allows us to test our next prediction:

Hypothesis 4. *In treatment* P-RAND*, when also the number of (irrelevant) unconditional votes in a group is $k-1$, the frequency of pivotal conditional YES votes increases in the voting threshold k.*

The presence of shared responsibility can be tested by comparing the conditional pivotal voters' behavior in *P-RAND* and in *P-INFO*. This motive is present in *P-INFO* but switched off in *P-RAND*, we expect that the threshold effect is smaller in *P-RAND* compared to *P-INFO*:

Hypothesis 5. *The difference in the frequencies of pivotal conditional YES votes between treatments* P-INFO *and* P-RAND *is increasing in the voting threshold k.*

Treatment *P-RAND* also allows us to examine the relevance the group consensus motive. For any given voting rule, the percentage of pivotal conditional YES votes should increase in the number of unconditional YES votes in the respective group. Thus, for a given number of $k-1$ random YES votes, we now consider all three possible values of unconditional YES votes (i.e. three out of the elicited nine decisions from each conditional voter). These unconditional votes matter neither for the transgression nor for determining pivotality of conditional voters, but they should still matter under the group consensus motive:

Hypothesis 6. *In treatment* P-RAND*, for each voting threshold k, the frequency of pivotal conditional YES votes increases in the number of unconditional YES votes in their group.*

Conversely to all previous hypotheses, the test of Hypothesis 6 is not based on comparisons of YES votes *between* different voting thresholds k, but on the comparison of YES votes *within* each threshold.

A final question is whether, in addition to shared responsibility and group consensus, there exist other motives underlying guilt sharing that we cannot capture with our design. This issue can also be addressed with treatment *P-RAND*. In *P-RAND*, out of the motives considered, only group consensus remains. Fixing the number of unconditional YES votes in a given group enables us to also eliminate group consensus. The absence of a threshold effect would then constitute a necessary (but of course not sufficient) condition that our framework indeed captures all relevant motives. This leads to our final prediction:

Hypothesis 7. *In treatment* P-RAND*, when fixing the number of unconditional YES votes in each group, the frequency of pivotal conditional YES votes is independent of the voting threshold k.*

3.3 Implementation of the experiment

The experiment was pre-registered at the RCT Registry of the American Economic Association (AEARCTR-0004317) and conducted in January and February 2020. A total of 4,680 subjects (52% female) were recruited through Amazon Mechanical Turk (MTurk).[14] On average, it took participants about four minutes to participate and they earned 1.14$ (including a show-up fee of 0.80$).

We employed a between-subjects design. Each subject participated only in one treatment and played under only one of the three voting rules. We conducted three sessions with subsets of participants for treatment *SIM* and four sessions for each of all other treatments at different points in time. Subjects were randomly assigned to voting rules and groups, and to roles within their group (unconditional or conditional voter). They then received instructions depending on their assigned role and voting rule. Instructions

[14]The requirements for participants were that they (i) were based in the US, (ii) had an approval rate for past work on MTurk greater than 98%, and (iii) had already completed more than 500 tasks. See Appendix C.1 for more information on the data collection process.

were identical for all subjects in treatment *SIM* playing under the same voting rule (see Appendix C).

Unconditional voters in all other treatments received the same (voting-rule specific) instructions, while the instructions for conditional voters varied across treatments. After they had read the instructions, subjects were asked some control questions. They could continue to the next page only after they had answered correctly. They were then asked for their vote. Unconditional voters were simply asked to vote for or against taking the money (YES or NO). For conditional voters, we used the strategy method to elicit their votes conditional on other voting decisions: In treatments *P-BASE* and *P-INFO*, there were three conditional votes (one for each possible number of YES votes by the unconditional voters in their group). In treatment *P-RAND*, there were nine conditional votes (one for each combination of unconditional and random YES votes).

After the voting stage, we elicited subjects' age and gender. In addition, we asked them about their beliefs concerning the voting behavior of the other participants and how trustworthy they considered the charity Make-A-Wish.[15] Finally, we elicited risk preferences by asking subjects for their willingness to take risks (measured on an ordinal scale from zero to ten; similar to Dohmen at al., 2011).

In treatment *SIM*, where all subjects are symmetric (all unconditional voters), we have 120 subjects for each of the three voting rules. In *P-BASE*, *P-INFO*, and *P-RAND*, where the focus is on the behavior of conditional voters, the number of subjects was tripled, which gives us 120 conditional voters for each voting rule.

4 Results

In this section, we report the results for all within- and between-treatment predictions from Section 3 (Hypotheses 1 – 7). For each treatment, Figure 1 shows the percentage of votes for taking the money, separated by the three voting thresholds. We test the

[15] As these two variables were elicited ex post, we only use them as a robustness check (see Appendix B).

hypotheses by estimating linear probability models. The regression results are shown in Tables 2 and 3, and most of our predictions are supported by the data. Our findings suggest that shared responsibility and group consensus are both independent drivers of voting behavior. As shown in Appendix B, the results remain robust when instead using a non-parametric approach and when considering a richer set of controls. Our results are also robust when estimating probit or logit models instead of linear probability models.

The impact of voting thresholds: Testing Hypothesis 1 We first consider the impact of the threshold k on voting behavior in treatment *SIM*. In this treatment, voters are fully symmetric and all three votes are taken into account. The percentage of votes in favor of taking the money is depicted in Figure 1(a) and increases with the voting threshold k. The respective regression results are shown in Table 2, column (1). The percentage of YES votes is significantly higher at the 5%-level when moving from dictatorial ($k = 1$) to unanimity ($k = 3$) voting. Moreover, we reject the hypothesis of no joint impact of majority and unanimity voting compared to dictatorial voting. These findings provide support for Hypothesis 1 and the theory of Rothenhäusler, Schweizer, and Szech (2018), but that does not allow us to identify guilt sharing and to investigate shared responsibility and group consensus as possible underlying motives in more detail.

Eliminating free-riding: Testing Hypotheses 2 and 3 If results in treatment *SIM* were solely driven by free-riding, then the votes of subjects knowing that they are pivotal should not be influenced by the voting threshold. To test this, we next consider treatment *P-BASE* where groups consist of one conditional voter and two unconditional voters, and where the conditional voter is asked for her vote conditional on the number of unconditional YES votes (zero, one, or two). By considering only those vote constellations where the conditional voter is pivotal (i.e. where the moral transgression occurs if and only if the conditional voter votes YES), this removes free-riding incentives.

Figure 1(b) illustrates that in *P-BASE*, the percentage of pivotal conditional YES votes increases sharply in the voting threshold k: Compared to the dictatorial rule, the number of pivotal conditional YES votes is 14 and 29 percentage points higher under

Figure 1: Percentage of YES votes by voting thresholds (k) across treatments

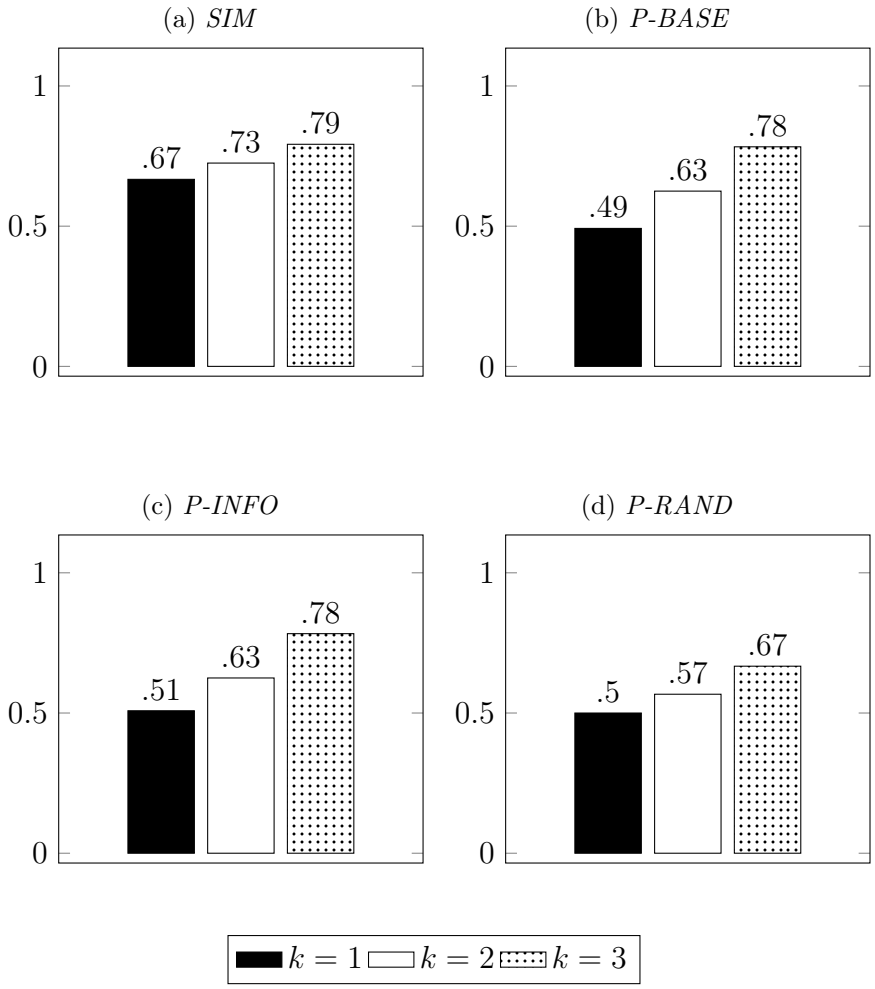

Table 2: Regressions results for the tests of Hypotheses 1 to 5

	(1) *SIM* (H1)	(2) *P-BASE* (H2)	(3) *P-INFO* (H3)	(4) *P-RAND* (H4)	(5) *P-INFO + P-RAND* (H5)
Majority	0.0567	0.139**	0.117*	0.0614	0.117*
	(0.350)	(0.028)	(0.067)	(0.339)	(0.066)
Unanimity	0.118**	0.273***	0.287***	0.138**	0.290***
	(0.040)	(0.000)	(0.000)	(0.031)	(0.000)
P-RAND					0.00259
					(0.968)
Majority x P-RAND					-0.0586
					(0.516)
Unanimity x P-RAND					-0.150*
					(0.086)
Observations	360	360	360	360	720
Adjusted R^2	0.018	0.082	0.050	0.045	0.050
T1: Una. vs. Majo.	0.258	0.021	0.004	0.218	
T2: (Una. + Majo.) vs. Dict.	0.096	0.000	0.000	0.075	
T3: Treatment effect Majo.					0.373
T4: Treatment effect Una.					0.011
Controls Pers. Charact.	Yes	Yes	Yes	Yes	Yes
Session Dummies	Yes	Yes	Yes	Yes	No

Notes: Each column refers to a linear probability model with YES votes as the dependent variable. Regression (1) uses the observations from all three voters in a group; all other regressions use only pivotal conditional YES votes. p-values are reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level (robust standard errors). The reference category is dictatorial voting. The bottom of the table shows p-values for the following Wald tests: (i) difference between unanimity and majority voting (T1), (ii) difference of unanimity and majority compared to dictatorial voting (T2), (iii) treatment effect for majority voting (T3) and unanimity voting (T4). All regressions contain controls for personal characteristics (gender, age, risk attitude). Session dummies are included in regressions (1) to (4), where within-treatment comparisons are conducted.

majority and unanimity, respectively. Regression results in Table 2, column (2), also show that, compared to dictatorial voting, the percentage of pivotal conditional YES votes is significantly higher for both majority and unanimity voting. Moreover, the Wald tests reveal at high levels of significance that (i) the impact of these two rules differs, and (ii) that there is a joint effect of these rules compared to dictatorial voting.[16] All in all, these results are supportive of Hypothesis 2, and hence provides evidence for guilt sharing as an empirically relevant driver of group decisions on moral transgressions.[17]

[16]Note that the behavior of pivotal voters may also be influenced by the fact that, ceteris paribus, they are least likely to actually be pivotal under the unanimity rule. In this case, however, we would rather under- than overestimate the impact of the motives our paper points to. To see this, note that consequentialists would not care about the probability of being pivotal at all as both monetary benefits *and* moral costs are multiplied with the probability, which hence cancels out. Non-consequentialists, however, would even be less inclined to vote YES if the probability that their vote actually counts decreases as it becomes more likely that they face moral costs at no benefit.

[17]While none of our hypotheses refer to a comparison of treatments *SIM* and *P-BASE*, it can be seen that, notwithstanding the elimination of free-riding, the impact of the threshold k is larger in *P-BASE*. A plausible explanation might be that, in *SIM*, subjects need to find out themselves that they will be pivotal only when there are $k-1$ YES votes in their group. By contrast, in *P-BASE* and *P-INFO*,

Very similar findings emerge in treatment *P-INFO* (see Figure 1(c)). Recall that this treatment is identical to *P-BASE* except that, before deciding on their voting strategy, conditional voters are informed about the percentage of unconditional YES votes in treatment *P-BASE*. This is meant as a signal about the prevailing social norm, as it refers to all 240 unconditional voters in *P-BASE* who played under the same voting rule. Regression results for *P-INFO* are reported in Table 2, column (3). They show that the differences to dictatorial voting are again statistically significant for both majority and unanimity voting. The same is true for the comparison of the effects of unanimity and majority voting, as well as for their joint effect compared to dictatorial voting. These findings support Hypothesis 3. All in all, the evidence for Hypotheses 2 and 3 strongly suggest that the observed threshold effect is not solely driven by free-riding, that guilt sharing matters, and that treatment-specific beliefs about the social norm to not seem to play a major role here.

Relevance of shared responsibility and group consensus: Testing Hypotheses 4 and 5 Treatment *P-RAND* plays a crucial role in the analysis of shared responsibility and group consensus as potential motives underlying guilt sharing, and hence as potential drivers of the threshold effect. Recall that *P-RAND* is identical to *P-INFO* with one exception: In *P-RAND*, whether a group takes the money is no longer determined by the three votes in a group, but by two random votes and the corresponding pivotal vote of the conditional voter. This removes shared responsibility as a potential motive for conditional voters, while keeping constant their information about the prevailing social norm (i.e. they still learn the overall behavior of unconditional voters in *P-BASE*). Hence, if the threshold effect persists in *P-RAND*, then this would hint at the empirical relevance of the group consensus motive (Hypothesis 4).[18]

The behavior in treatment *P-RAND* is illustrated in Figure 1(d) and regression results

conditional voters are directly asked about their vote for the case of $k-1$ unconditional YES votes.

[18]Recall from above, that for the test of Hypothesis 4 we use the votes of conditional voters only for the case where the number of unconditional YES votes also equals $k-1$ (i.e. they are the same as the number of random YES votes).

are shown in Table 2, column (4). As can be seen, the results are qualitatively similar to those for *P-BASE* and *P-INFO*. In particular, the threshold effect exists and the difference is statistically significant for unanimity voting. However, the coefficients in *P-RAND* become smaller compared to *P-INFO*, approximately half in size. Consequently, also the Wald tests become less significant or even insignificant. Arguably, these findings provide evidence for shared responsibility to matter in our context.

To investigate this motive in more detail, we compare treatments *P-INFO* and *P-RAND*, thereby exploiting the fact sharing responsibility is feasible only in the former, but not in the latter. Moreover, according to Hypothesis 5, responsibility can never be shared under dictatorial voting ($k = 1$), so that behavior under this rule should not differ between the two treatments. Conversely, for the majority and the unanimity rule, there should be fewer pivotal conditional YES votes in *P-RAND* than in *P-INFO* and the effect should be stronger under unanimity voting.

Moreover, compared to *P-INFO*, the percentage of pivotal conditional YES votes in *P-RAND* shrinks for both majority (by 6 percentage points) and unanimity (by 11 percentage points) voting, while there is no difference for the dictatorial rule. These findings are also largely supported in the regression analysis (see Table 2, column (5)): In line with Hypothesis 5, there is no treatment effect for dictatorial voting. Furthermore, the interaction term of the unanimity rule with the treatment dummy reveals that the difference between *P-RAND* and *P-INFO* is significantly larger for unanimity compared to dictatorial voting. However, the predicted difference for the comparison of majority and dictatorial voting does not show up in the data. Also in line with Hypothesis 5, a Wald test reveals that there are significantly less pivotal conditional YES votes in *P-RAND* for unanimity voting. Under majority voting, there are again fewer pivotal conditional YES votes in *P-RAND*, but the effect is not strong enough to be statistically significant. Overall, our findings support the view that shared responsibility is an empirically relevant motive in our context.

Preference for group consensus: Testing Hypothesis 6: Treatment *P-RAND* can also be used to test our remaining two hypotheses. Hypothesis 6 predicts a positive correlation between the pivotal conditional YES votes (where pivotality is determined by the outcome of the two random votes) and the number of unconditional YES votes in their group (which do not count towards whether or not the money is taken), measured by the variable $UncondYES \in \{0, 1, 2\}$. Such a positive relationship would provide evidence that the group consensus motive (i.e. the desire to take decisions in line with the other group members' preferences) matters in our context.

We test Hypothesis 6 separately for each of the three voting rules, utilizing the fact that we have elicited the conditional voters' decisions for all nine possible combinations of unconditional and random YES votes (see the treatment description in Section 3). Again, we focus on pivotal conditional YES votes, i.e. we still use only vote constellations with $k-1$ random YES votes. However, in contrast to the testing of Hypothesis 4, we here no longer confine attention to the case of $k-1$ unconditional YES votes. Instead, we use all three possible constellations of unconditional YES votes (zero, one, or two). As this yields three pivotal decisions per conditional voter, standard errors in the regressions are clustered at the subject level.

The results are illustrated in Figure 2 and the regression results are shown in Table 3, columns (1)–(3). For all three voting thresholds, the percentage of pivotal conditional YES votes increases in the number of unconditional YES votes in their group. For example, comparing the effect of zero and two unconditional YES votes leads to differences of 15, 23 and 14 percentage points for the respective thresholds. Five out of the six respective coefficients in the regression analysis are statistically significant at high levels, thereby lending strong support for Hypothesis 6 and the empirical relevance of preferences for group consensus in our context.

Relevance of further motives: Testing Hypothesis 7 The treatments discussed so far allowed us to address and isolate, in turn, several motives driving the observed threshold effect. In particular, we have *(i)* eliminated free-riding by considering only

Figure 2: Treatment *P-RAND*: Percentage of YES votes of conditional voters when pivotal, by voting thresholds (k) and by number of unconditional YES votes in group *(UncondYES)*

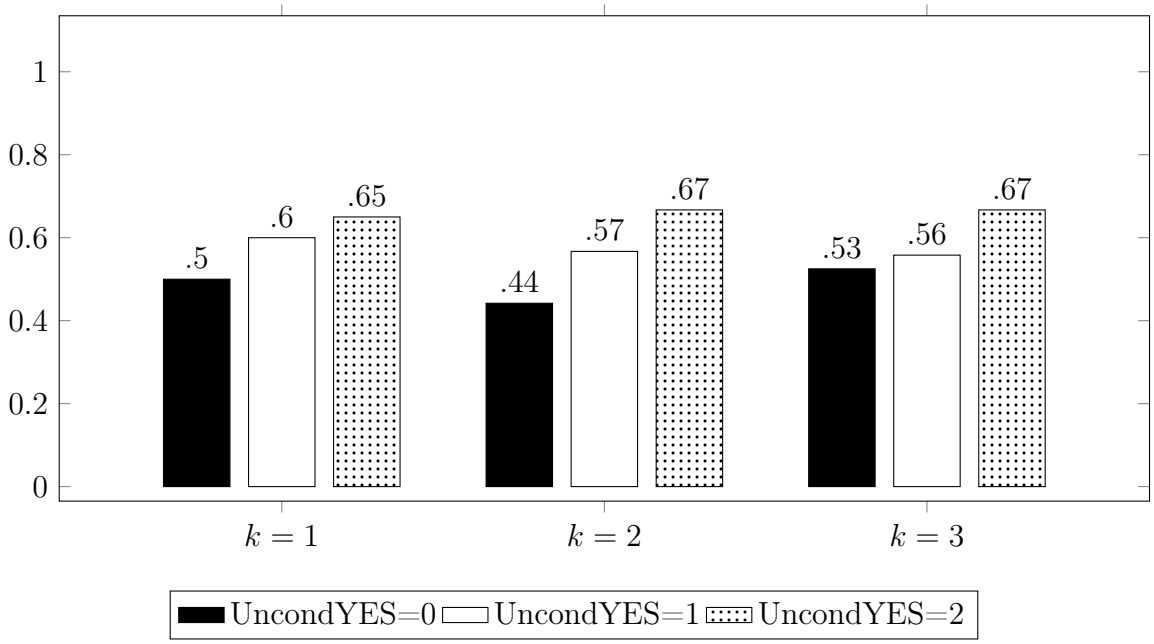

Table 3: Regressions results for the tests of Hypotheses 6 and 7

	(1)	(2)	(3)	(4)	(5)	(6)
	(H6)	(H6)	(H6)	(H7)	(H7)	(H7)
	Dictatorial	Majority	Unanimity	UncondYES=0	UncondYES=1	UncondYES=2
UncondYES=1	0.1000**	0.125***	0.0333			
	(0.024)	(0.003)	(0.163)			
UncondYES=2	0.150***	0.225***	0.142***			
	(0.001)	(0.000)	(0.001)			
Majority				-0.0589	-0.0406	0.0129
				(0.359)	(0.522)	(0.834)
Unanimity				0.000920	-0.0748	0.000926
				(0.989)	(0.242)	(0.988)
Observations	360	360	360	360	360	360
Adjusted R^2	0.036	0.091	0.109	0.031	0.029	0.011
T1: Both coeff. equal	0.113	0.003	0.005	0.349	0.588	0.847
T2: Sum both coeff. zero	0.004	0.000	0.003	0.606	0.297	0.898
Controls for Pers. Charact.	Yes	Yes	Yes	Yes	Yes	Yes
Session Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Clust. Subj. Level	Yes	Yes	Yes	No	No	No

Notes: All regressions are based on treatment *P-RAND*. Each column refers to a linear probability model with pivotal conditional YES votes as the dependent variable. p-values are reported in parentheses, and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level (robust standard errors). The reference category is dictatorial voting in regressions (1)-(3) and GroupYes=0 in regressions (4)-(6). The bottom of the table shows p-values for two Wald tests: one for the equality of the two coefficients in the respective columns (Test 1), and one for joint significance of both coefficients compared to the reference category (Test 2). All regressions include controls for personal characteristics (gender, age, risk attitude) as well as session dummies. Since each conditional voter is observed with three different decisions (one for each possible value of unconditional YES votes), standard errors are clustered at the subject level.

conditional voters in cases where they are pivotal (*P-BASE*, *P-INFO*, *P-RAND*), *(ii)* provided (and kept constant) information about the prevailing social norm (*P-INFO*, *P-RAND*), and *(iii)* isolated the effects of shared responsibility and group consensus (*P-RAND*).

This raises the question whether our design indeed captures all relevant motives. A straightforward way of addressing this issue is examining if the threshold effect as observed in *P-RAND* vanishes once we also neutralize the group consensus motive by fixing the number of YES votes by the unconditional voters in each group (Hypothesis 7). Empirical support for Hypothesis 7 would then constitute a necessary condition for the broader claim that no other motives outside our framework are relevant in our context. In this respect, our analysis can be seen as a first step towards this issue.

Descriptive evidence for Hypothesis 7 can be observed from Figure 2 when comparing the numbers for a given level of unconditional YES votes (i.e. comparing the three black, white and dotted bars, respectively). As can be seen, the threshold effect indeed virtually disappears. This is also confirmed in the regression analysis as shown in Table 3, columns (4)–(6), where none of the six coefficients for the voting thresholds is even close to being statistically significant.[19] These findings strongly support Hypothesis 7, and they are also in line with the (stronger) claim that our design indeed captures the relevant motives in the context of morally difficult decisions by groups.

5 Conclusion

Experimental papers find that moral transgressions are more likely when subjects make group decisions than when they decide on their own (e.g. Dana, Weber, and Kuang,

[19]We have also conducted an additional treatment that is identical to *P-RAND* except that we do not inform conditional pivotal voters on the number of YES votes in their group. In this case, the voting threshold may still matter as, depending on the threshold, conditional voters may form different expectations on the number of YES votes by their group fellows. Consistent with such a reasoning, we find that the frequency of YES votes by pivotal voters is significantly lower under dictatorial voting (53.3%) compared to both the majority (65%) and unanimity (64.2%) voting.

2007; Kocher, Schudy, and Spantig, 2018; Falk, Neuber, and Szech, 2020). While this is quite intuitively often attributed to guilt sharing, we have argued that the literature has not yet cleanly identified the presence of guilt sharing. The reason is that group decisions differ from individual decision making in more than just one respect: First, the expected financial benefit from favoring a moral transgression may well differ between groups and individual decision making. Second, the own behavior does not only affect the own payoff, but inevitably also the payoff of other group members. Therefore, it is notoriously difficult to identify guilt sharing by comparing group decision making to individual decision making. In this paper, we overcome this issue by varying, instead of the group size, the number of affirmative votes (the voting threshold) required for a moral transgression.

We then consider pivotal votes for each voting rule. This means that (i) the own vote certainly decides on whether or not the moral transgression occurs, and (ii) the impact of a YES vote on both the own monetary benefit and on the other group members is the same for all thresholds. We find that the frequency of YES votes for the moral transgression increases in the threshold, which means that pivotal voters do care about the behavior of their fellow group members, even when differences in the impacts of the own behavior are eliminated. Our first contribution is hence the clear identification of guilt sharing. We emphasize that we could achieve this only by a variation in the voting rule, and not by a variation in the group size. The reason is that, when the group size increases and the payoff per subject stays constant, then the positive externality of voting YES increases, as more other group members get the money. In order to avoid this, one would need to reduce the payoff per subject which, of course, would also alter the incentive structure.

In our second contribution, we go one step further by disentangling shared responsibility and group consensus as two potential motives underlying guilt sharing. Both motives are present in treatment *P-BASE*, while only the latter remains in *P-RAND*. The threshold effect also exists in *P-RAND* where shared responsibility is muted, which means that a preference for group consensus does contribute to guilt sharing. However, the threshold effect declines significantly, which shows that shared responsibility also has

an impact. As just one vote is decisive for the outcome in *P-RAND*, this treatment can also be interpreted as a situation where one group member has the right to decide, but beforehand collects her fellows' opinions. In this sense, identifying preferences for group consensus is not confined to settings with voting.

For future research, we aim at examining the robustness of our results by considering, instead of taking money designated for donation, a context of lying about the (privately observed) outcome of a lottery (see the seminal paper by Fischbacher and Föllmi-Heusi, 2013, and the meta study in Abeler, Nosenzo, and Raymond, 2019). Such an extension is interesting as one might assume that lying entails a larger non-consequentialist moral cost component, as it is perceived as immoral even when the final outcome is unaffected.[20] While the difference between consequentialist and non-consequentialist moral costs disappears for pivotal voters, as their votes always determine the final outcome by definition of pivotality, it may well influence the behavior of non-pivotal voters. In fact, the theoretical prediction of Rothenhäusler, Schweizer, and Szech (2018) that the frequency of YES votes increases in the voting threshold does not extend to non-consequentialist moral costs as there is a countervailing effect: A YES vote now bears the risk of incurring the non-consequentialist moral costs without a benefit (i.e. when the overall number of YES votes is below the respective threshold). When this "risk" is increasing in the voting threshold, this could lead to fewer YES votes accordingly.

[20]Feess, Kerzenmacher, and Timofeyev (2021) consider a setting that yields opposing hypotheses for consequentialist and non-consequentialist moral preferences and find that the consequentialist (non-consequentialist) model performs better for donation (lying).

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for truth-telling. *Econometrica 87*(4), 1115–1153.

Alexander, L. and M. Moore (2016). Deontological ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford University.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–55.

Bartling, B., E. Fehr, and Y. Özdemir (2021). Does market interaction erode moral values? *The Review of Economics and Statistics, forthcoming*.

Bartling, B. and U. Fischbacher (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies 79*(1), 67–87.

Bartling, B., U. Fischbacher, and S. Schudy (2015). Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics 128*, 133–139.

Bartling, B., R. Weber, and L. Yao (2015). Do markets erode social responsibility? *Quarterly Journal of Economics 130*(1), 219–266.

Bénabou, R., A. Falk, and J. Tirole (2020). Narratives, imperatives, and moral persuasion. *University of Bonn, mimeo*.

Bénabou, R. and J. Tirole (2010). Individual and corporate social responsibility. *Economica 77*(305), 1–19.

Brütt, K., A. Schram, and J. Sonnemans (2020). Endogenous group formation and responsibility diffusion: An experimental study. *Games and Economic Behavior 121*, 1–31.

Casal, S., F. Fallucchi, and S. Quercia (2019). The role of morals in three-player ultimatum games. *Journal of Economic Psychology 70*, 67–79.

Conrads, J., B. Irlenbusch, R. Rilke, and G. Walkowitz (2013). Lying and team incentives. *Journal of Economic Psychology 34*, 1–7.

Dana, J., R. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Dato, S., E. Feess, and P. Nieken (2019). Lying and reciprocity. *Games and Economic Behavior 118*, 193–218.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Erat, S. and U. Gneezy (2012). White lies. *Management Science 58*(4), 723–733.

Falk, A., T. Neuber, and N. Szech (2020). Diffusion of being pivotal and immoral outcomes. *Review of Economic Studies 87*(5), 2205–2229.

Falk, A. and N. Szech (2013). Morals and markets. *Science 340*(6133), 707–711.

Feess, E., F. Kerzenmacher, and Y. Timofeyev (2021). Consequentialist or non-consequentialist models of moral behavior – what predicts moral transgressions? *Working Paper*.

Feess, E., H. Schildberg-Hörisch, M. Schramm, and A. Wohlschlegel (2018). The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory. *Journal of Economic Behavior & Organization 149*, 58–73.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in disguise – an experimental study on cheating. *Journal of the European Economic Association 11*(3), 525–547.

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

Huck, S. and K. A. Konrad (2005). Moral cost, commitment, and committee size. *Journal of Institutional and Theoretical Economics*, 575–588.

Irlenbusch, B. and D. J. Saxler (2019). The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior. *Journal of Behavioral and Experimental Economics 80*, 141–161.

Kirchler, M., J. Huber, M. Stefan, and M. Sutter (2016). Market design and moral behavior. *Management Science 62*(9), 2615–2625.

Kocher, M., S. Schudy, and L. Spantig (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science 64*(9), 3995–4008.

Monroe, A., K. Dillon, S. Guglielmo, and R. Baumeister (2018). It's not what you do, but what everyone else does: On the role of descriptive norms and subjectivism in moral judgment. *Journal of Experimental Social Psychology 77*, 1–10.

Muehlheusser, G., A. Roider, and N. Wallmeier (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters 128*, 25–29.

Nook, E., D. Ong, S. Morelli, J. Mitchell, and J. Zaki (2016). Prosocial conformity: Prosocial norms generalize across behavior and empathy. *Personality and Social Psychology Bulletin 42*(8), 1045–1062.

Peeters, R., M. Vorsatz, and M. Walzl (2015). Beliefs and truth-telling: A laboratory experiment. *Journal of Economic Behavior & Organization 113*, 1–12.

Rothenhäusler, D., N. Schweizer, and N. Szech (2018). Guilt in voting and public good games. *European Economic Review 101*, 664–681.

Sinnott-Armstrong, W. (1988). *Moral dilemmas*. Blackwell Publishers.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes 115*(2), 157–168.

Appendix

A Theory

A.1 Voting behavior when voters are unconditional

The basic model where all voters are unconditional and vote simultaneously follows Rothenhäusler, Schweizer, and Szech (2018) (RSS). Each member of a group of $n \geq 1$ agents votes on a moral transgression (YES or NO), which yields benefit $\pi > 0$ for each group member. The voting threshold k is defined as the minimum number of YES votes required for the transgression, where $1 \leq k \leq n$.

When the transgression occurs, each group member who has voted YES faces moral costs which consist of two components:[1] The first component is purely "internal" and hence independent of the number of YES votes in the group (sensitivity parameter $\alpha \geq 0$). The second component decreases in the overall number of YES votes in the group, thereby capturing the motive of shared responsibility.[2] In particular, we consider a function $g(Y_i)$ in combination with a sensitivity parameter $\beta \geq 0$, where Y_i is the number of YES votes in the group other than voter i, and where $g(0) = 1$ and $g'(\cdot) < 0$. Both of these two cost components are weighed with a voter's moral type $\theta_i \in [0, \theta^{\max}]$, which is voter i's private information. Types are distributed with continuous density $f(\theta)$ with full support on $[0, \theta^{\max}]$, where $f(\theta)$ is common knowledge. The utility of agent i from voting YES

[1] RSS also consider a cost of being pivotal (which accrues only in case the transgression would not have occurred without the respective voter's YES votes) as a third component. We drop this cost type as it has no impact on Proposition 1 and because, in all other propositions and hypotheses, we consider only pivotal voters anyway, so that preferences against being pivotal cannot influence the impact of voting thresholds.

[2] RSS call this the diffusive moral cost component or simply the part of guilt that can be shared. In their setting, there is no difference between shared responsibility and group consensus.

(u_Y^i) and NO (u_N^i), respectively, is then given by

$$u_Y^i := (\pi - \theta_i (\alpha + \beta g(Y_{-i}))) \cdot \Pr(Y_{-i} \geq k-1) \quad (1)$$

$$u_N^i := \pi \cdot \Pr(Y_{-i} \geq k). \quad (2)$$

Theoretical prediction for Hypothesis 1 RSS consider symmetric Bayesian Nash equilibria, and they prove existence of an equilibrium in which all types $\theta_i \leq \widehat{\theta}_{k,n}$ vote for the moral transgression and all types $\theta_i > \widehat{\theta}_{k,n}$ against it. Indifferent types $\widehat{\theta}_{k,n}$ are characterized by the set of solutions to

$$\pi b(k-1; n-1, p) = \widehat{\theta}_{k,n} \left(\sum_{j=k-1}^{n-1} (\alpha + \beta g(1+j))\, b(j; n-1, p) \right), \quad (3)$$

where

$$b(j; n-1, p) = \binom{n-1}{j} p^j (1-p)^{n-1-j}, \text{ and} \quad (4)$$

$$p = \Pr(u_y \geq u_n). \quad (5)$$

This yields the following result:

Proposition 1 (All voters unconditional)**.** *The indifferent type $\widehat{\theta}_{k,n}$, and hence the equilibrium frequency of YES votes, increases in the voting threshold k.*

Proof. See RSS.

The proposition directly leads to Hypothesis 1 as spelled out in Section 3 above, and which is tested in treatment *SIM*, where all votes are unconditional. For our purposes, it is crucial to note that the proposition also holds when there is no shared responsibility ($\beta = 0$), but only non-diffusive moral costs ($\alpha > 0$). The reason is that lower voting thresholds provide higher free-riding opportunities: The higher the likelihood that sufficiently many other group members vote YES, the higher is the incentive to save moral costs by voting NO. In equilibrium, this likelihood decreases in the threshold k and is zero for the unanimity rule ($k = n$).

A.2 The voting behavior of a pivotal voter

To eliminate such free-riding incentives, from now on we focus on pivotal votes. Consider a voter who knows that she is pivotal for the moral transgression (i.e. there are $k-1$ other YES votes). The utility of the pivotal voter when voting YES is then

$$u_Y^P := \pi - \theta_i \left(\alpha + \beta g(W_{-i})\right), \tag{6}$$

where W_{-i} is the number of YES votes by other voters in the group counting towards the moral transgression, so that the pivotal voter can share responsibility with them (and where, due to pivotality, $W_{-i} = k-1$ holds).

Now consider the case where the pivotal voter votes NO. Note first that in this case she has no chance of receiving the gain π, i.e. there are no free-riding opportunities for pivotal voters. Moreover, we introduce an additional cost when voting NO which is due to a preference for *group consensus*. As explained in the main text, the underlying idea is that individuals suffer a disutility when not acting in line with the (revealed) preferences of their group fellows. When voting NO, the pivotal voter incurs a disutility which increases with the preference of her group fellows in favor of the transgression. We capture this by a function $c(V_{-i})$ in combination with a sensitivity parameter $\phi \geq 0$, where V_{-i} denotes the number of YES votes by the group members other than i, and where $c(0) = 0$ and $c'(\cdot) > 0$.[3] In summary, when voting NO, the utility of a pivotal voter is given by

$$u_N^P := -\theta_i \phi c(V_{-i}), \tag{7}$$

where also $V_{-i} = k-1$ due to pivotality.

Note the subtle difference between W_{-i} and V_{-i}: In (6), W_{-i} is the number of group members other than i who vote YES *and* whose votes count for the moral transgression.

[3]The assumption that the costs $c(V_{-i}) > 0$ arise only when voting against the transgression is without loss of generality. An equivalent approach would be to introduce a cost when voting against the stated preferences of one's fellow group members. All we need is that the cost difference between a vote for and against the transgression increases in the number of YES votes in the group.

Hence, W_{-i} is the relevant measure with respect to shared responsibility. By contrast, in (7), V_{-i} is the number of group members other than i who vote YES, irrespective of whether their votes actually count. Hence, V_{-i} is the relevant measure with respect to group consensus. Recall that, in the experiment, in treatments *P-BASE* and *P-INFO* the decision concerning the moral transgression is based on the all three votes in the group, so that $W_{-i} = V_{-i}$ holds. By contrast, in treatment *P-RAND* the unconditional votes do not count for the decision concerning the transgression, i.e $W_{-i} \neq V_{-i}$.

Setting $u_Y^P = u_N^P$ then leads to the indifferent moral type of the pivotal voter, denoted by $\widehat{\theta}_k^P$, i.e.

$$\widehat{\theta}_k^P = \frac{\pi}{\alpha + \beta g(W_{-i}) - \phi c(V_{-i})}. \tag{8}$$

Theoretical prediction for Hypotheses 2 and 3 As explained above, treatments *P-BASE* and *P-INFO* differ only with respect to whether or not conditional voters receive information about the behavior of a population of unconditional voters outside their treatment which is not modeled here. Hence, the underlying proposition for the behavior of pivotal conditional voters is the same for both hypotheses. Taking into account that $W_{-i} = V_{-i} = k - 1$, this leads to

Proposition 2 (Pivotal votes in treatments *P-BASE* and *P-INFO*)**.** *The indifferent moral type of the pivotal voter $\widehat{\theta}_k^P$, and hence the equilibrium frequency of YES votes, increases in the voting threshold k.*

Proof. Equating u_Y^P and u_N^P, we get the indifferent type as:

$$\widehat{\theta}_k^P = \frac{\pi}{\alpha + \beta g(k-1) - \phi c(k-1)}, \tag{9}$$

where

$$\frac{\partial \widehat{\theta}_k^P}{\partial k} = \frac{\pi(\phi c'(k-1) - \beta g'(k-1))}{[\alpha + \beta g(k-1) - \phi c(k-1)]^2} > 0 \tag{10}$$

as $g'(k) < 0$ and $c'(k) > 0$. ■

Intuitively, in case of pivotality, a higher voting threshold k also implies a higher number of other YES votes $k - 1$. This reduces the moral costs of a YES vote for the

pivotal voter due to shared responsibility. In addition, a higher value of $k-1$ YES votes indicates a stronger preference for the moral transgression among the other group members, so that the cost of voting against it increases as well (group consensus motive). Both effects make a YES vote more attractive for the pivotal voter (guilt sharing).

Theoretical prediction for Hypothesis 4 This hypothesis applies to conditional voters in treatment *P-RAND* where both the number of random and unconditional YES votes is equal to $k-1$. Since only the former are relevant for the transgression, this mutes shared responsibility (i.e. $W_{-i}=0$) while the group consensus motive is still present (i.e. $V_{-i}=k-1$). This leads to

Proposition 3 (Pivotal votes in treatment *P-RAND*)**.** *Suppose $\phi>0$. Then the indifferent moral type of the pivotal voter $\widehat{\theta}_k^P$, and hence the equilibrium frequency of YES votes, increases in the voting threshold k.*

Proof. Without the possibility of sharing responsibility, we have $g(W_{-i}=0)=1$. The threshold type who is indifferent between voting YES or NO is given by:

$$\widehat{\theta}_k^P = \frac{\pi}{\alpha+\beta-\phi c(k-1)}, \tag{11}$$

where

$$\frac{\partial \widehat{\theta}_k^P}{\partial k} = \frac{\pi\phi c'(k-1)}{[\alpha+\beta-\phi c(k-1)]^2} > 0 \tag{12}$$

as $c'(V_{-i})>0$. ■

Intuitively, as $V_{-i}=k-1$ increases in k, so does the cost of voting NO due to the preference for group consensus. Therefore, the number of pivotal conditional YES votes increases in the voting threshold even when shared responsibility is not present.

Theoretical prediction for Hypothesis 5 The impact of shared responsibility can be gauged by comparing the behavior of pivotal voters in treatments *P-INFO* and *P-RAND*. Denoting the indifferent types under *P-INFO* and *P-RAND* by $\widehat{\theta}_{1,k}^P$ and $\widehat{\theta}_{2,k}^P$, respectively, and defining the difference $\Delta\widehat{\theta}_k^P := \widehat{\theta}_{1,k}^P - \widehat{\theta}_{2,k}^P$, this leads to

Proposition 4 (Impact of shared responsibility)**.** *Suppose $\beta > 0$. Then, the indifferent moral type of the pivotal voter (and thus the equilibrium frequency of YES votes) is higher in treatment* P-INFO *than in* P-RAND *(i.e. $\Delta\widehat{\theta}_k^P \geq 0$). For $k = 1$, there is no treatment difference ($\Delta\widehat{\theta}_1^P = 0$). For $k > 1$, $\Delta\widehat{\theta}_k^P$ is strictly positive and increasing in k.*

Proof. Without the possibility for sharing responsibility in treatment *P-RAND*, we have $Y_{-i} = 0$ and $g(Y_{-i}) = 1$. Similarly, in treatment *P-INFO*, we have $Y_{-i} = k - 1$ and $g(Y_{-i}) \leq 1$. The respective threshold types who are indifferent between voting YES or NO are then given by:

$$\widehat{\theta}_{1,k}^P = \frac{\pi}{\alpha + \beta g(k-1) - \phi c(k-1)} \tag{13}$$

and

$$\widehat{\theta}_{2,k}^P = \frac{\pi}{\alpha + \beta - \phi c(k-1)}, \tag{14}$$

where

$$\frac{\partial(\widehat{\theta}_{1,k}^P - \widehat{\theta}_{2,k}^P)}{\partial k} = \pi \frac{\phi c'(k-1) - \beta g'(k-1)}{[\alpha + \beta g(k-1) - \phi c(k-1)]^2} - \pi \frac{\phi c'(k-1)}{[\alpha + \beta - \phi c(k-1)]^2} > 0 \tag{15}$$

as the numerator of the first fraction exceeds the numerator of the second fraction (because $c'(k-1) > 0$ and $g'(k-1) < 0$) and the denominator of the first fraction is (weakly) smaller than the denominator of the second fraction (because $g(k-1) \leq 1$). Note that $\widehat{\theta}_{1,k}^P = \widehat{\theta}_{2,k}^P$ if and only if $k = 1$. ■

Theoretical prediction for Hypothesis 6 The hypothesis concerns the impact of the group consensus motive and is again based on treatment *P-RAND* where the sharhing of responsibility is impossible (i.e. $W_{-i} = 0$). However, we no longer fix $V_{-i} = k - 1$ but analyze, separately for each voting threshold k, how the behavior of pivotal voters varies in V_{-i}. This leads to

Proposition 5 (Impact of group consensus motive)**.** *Suppose $W_{-i} = 0$ and $\phi > 0$. Then, for a given voting threshold k, the indifferent moral type of the pivotal voter $\widehat{\theta}_k^P$, and hence the equilibrium frequency of YES votes, increases in the number of unconditional YES V_{-i} in her group.*

Proof. With $W_{-i} = 0$, we have $g(0) = 1$. The threshold type who is indifferent between voting YES or NO is given by:

$$\widehat{\theta}_k^P = \frac{\pi}{\alpha + \beta - \phi c(V_{-i})}, \tag{16}$$

where

$$\frac{\partial \widehat{\theta}_k^P}{\partial V_{-i}} = \frac{\pi \phi c'(V_{-i})}{[\alpha + \beta - \phi c(V_{-i})]^2} > 0 \tag{17}$$

as $c'(V_{-i}) > 0$. ■

Intuitively, although the unconditional YES votes in a group do not count towards the moral transgression, they still matter for a (conditional) pivotal voter because of the group consensus motive. This effect size increases with the number of unconditional YES votes in a group.

Theoretical prediction for Hypothesis 7 According to this final hypothesis, other than for the ones considered so far, there are no further motives giving rise to a threshold effect. The hypothesis is also tested in treatment *P-RAND* where, in addition to the absence of shared responsibility in this treatment, we also take out the group consensus motive by separately considering each value of V_{-i}. This leads to

Proposition 6 (No further motives)**.** *Suppose $W_{-i} = 0$ and $\phi > 0$. Then, for a given number of unconditional YES votes, $\overline{V}_{-i}$, the indifferent moral type of the pivotal voter $\widehat{\theta}_k^P$ is independent of the voting threshold k.*

Proof. Again, without the possibility of sharing responsibility, we have $g(W_{-i} = 0) = 1$. The threshold type who is indifferent between voting YES or NO is given by:

$$\widehat{\theta}_k^P = \frac{\pi}{\alpha + \beta - \phi c(\overline{V}_{-i})}, \tag{18}$$

where $\overline{V}_{-i}$ is constant. Thus, $\frac{\partial \widehat{\theta}_k^P}{\partial k} = 0$. ■

B Robustness

In this appendix, we provide additional robustness checks for our results: First, we analyze whether a threshold effect also arises in a non-parametric approach. Second, we check the robustness of the regressions results by considering a richer set of controls from the post-experimental questionnaire. Overall, these robustness checks do not lead to qualitatively different results than those presented in the main analysis.

Non-parametric approach Table 4 reports the p-values of the χ^2 tests for the comparison of the different thresholds in each treatment. This gives rise to results which are qualitatively very similar to those reported in Table 2. In particular, in each treatment there are significantly more YES votes under unanimity voting compared to dictatorial voting and also some of the other comparisons yield significant differences.

Table 4: Threshold effect in a non-parametric approach

	Dictatorial vs. Majority	Dictatorial vs. Unanimity	Majority vs. Unanimity
SIM	0.326	0.029^{**}	0.228
P-BASE	0.038^{**}	$< 0.001^{***}$	0.007^{***}
P-INFO	0.068^{*}	$< 0.001^{***}$	0.007^{***}
P-RAND	0.301	0.009^{***}	0.111

Notes: The entries in the table report the p-values for the respective χ^2 tests for differences of YES votes under the different thresholds in each treatment. As in the main text, in treatment *SIM* all votes are unconditional, and all of them are used in the analysis. In all other treatments, only pivotal conditional YES votes are used.

Additional controls Tables 5 and 6 replicate the regression analysis from the main text (Tables 2 and 3) but control in addition for the subjects' belief about the overall percentage of YES votes in the experiment and the degree of trust in the charity, measured on a ten-point scale. Both variables were elicited after subjects had taken their voting decisions.

Comparing Tables 5 and 2 shows that the results for the threshold effect are quite similar. In total, there are three cases where coefficients become insignificant when using

Table 5: Robustness: Regressions results for the tests of Hypotheses 1 to 5

	(1) *SIM* (H1)	(2) *P-BASE* (H2)	(3) *P-INFO* (H3)	(4) *P-RAND* (H4)	(5) *P-INFO* + *P-RAND* (H5)
Majority	0.0561	0.159***	0.101*	0.0736	0.0975
	(0.285)	(0.004)	(0.092)	(0.199)	(0.102)
Unanimity	0.0410	0.312***	0.249***	0.142**	0.248***
	(0.387)	(0.000)	(0.000)	(0.016)	(0.000)
P-RAND					-0.0179
					(0.762)
Majority x P-RAND					-0.0265
					(0.748)
Unanimity x P-RAND					-0.110
					(0.171)
Belief	0.00866***	0.00820***	0.00783***	0.00801***	0.00797***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Trust	-0.0412***	-0.0273***	-0.0368***	-0.0516***	-0.0454***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
Observations	360	360	360	360	720
Adjusted R^2	0.282	0.307	0.167	0.211	0.195
T1: Una. vs. Majo.	0.750	0.002	0.007	0.240	
T2: (Una. + Majo.) vs. Dict.	0.270	0.000	0.001	0.032	
T3: Treatment effect Majo.					0.447
T4: Treatment effect Una.					0.019
Controls Pers. Charact.	Yes	Yes	Yes	Yes	Yes
Session Dummies	Yes	Yes	Yes	Yes	No

Notes: The structure of the regressions is the same as in Table 2 in the main text, so that also the explanation below Table 2 applies. The only difference is that the regressions reported here include additional controls for the reported beliefs about other participants' voting behavior (*Belief*) and the degree of trust in the charity (*Trust*).

the richer set of controls: (i) the comparison of unanimity and dictatorial voting in *SIM* (see column (1)), (ii) the comparison of majority and dictatorial voting in *P-INFO* (see column (5)), and (iii) the interaction term for the treatment comparison between *P-INFO* and *P-RAND* (see column (5)). The comparison of Tables 6 and 3 (i.e. the test of Hypothesis 6 and Hypothesis 7) reveals that the two additional controls do not lead to any qualitative differences. In all regressions reported in Tables 5 and 6, *Belief* and *Trust* are highly significant in plausible directions. As these variables were elicited after the voting decisions were taken, the answers may well be influenced by rationalizations of the own behavior. It is hence not surprising that some of the effects become insignificant after controlling for *Belief* and *Trust*.

C Experimental Procedures

This appendix contains more information on the experimental procedures, in particular data collection (C.1) and instructions (C.2).

Table 6: Robustness: Regressions results for the tests of Hypotheses 6 and 7

	(1)	(2)	(3)	(4)	(5)	(6)
	(H6)	(H6)	(H6)	(H7)	(H7)	(H7)
	Dictatorial	Majority	Unanimity	UncondYES=0	UncondYES=1	UncondYES=2
UncondYES=1	0.1000**	0.125***	0.0333			
	(0.025)	(0.003)	(0.164)			
UncondYES=2	0.150***	0.225***	0.142***			
	(0.001)	(0.000)	(0.001)			
Majority				-0.0485	-0.0293	0.0217
				(0.409)	(0.613)	(0.711)
Unanimity				0.0117	-0.0736	0.00460
				(0.846)	(0.219)	(0.938)
Belief	0.00720***	0.00419**	0.00716***	0.00573***	0.00778***	0.00570***
	(0.000)	(0.046)	(0.000)	(0.000)	(0.000)	(0.000)
Trust	-0.0490***	-0.0550***	-0.0316*	-0.0593***	-0.0433***	-0.0395***
	(0.000)	(0.000)	(0.054)	(0.000)	(0.000)	(0.000)
Observations	360	360	360	360	360	360
Adjusted R^2	0.187	0.197	0.193	0.169	0.167	0.107
T1: Both coeff. equal	0.114	0.003	0.005	0.321	0.453	0.772
T2: Sum both coeff. zero	0.004	0.000	0.003	0.718	0.312	0.796
Controls for Pers. Charact.	Yes	Yes	Yes	Yes	Yes	Yes
Session Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Clust. Subj. Level	Yes	Yes	Yes	No	No	No

Notes: The structure of the regressions is the same as in Table 3 in the main text, so that also the explanation below Table 3 applies. The only difference is that the regressions reported here include additional controls for the reported beliefs about other participants' voting behavior (*Belief*) and the degree of trust in the charity (*Trust*).

C.1 Experimental data

We recruited participants through MTurk. After accepting to take part in our study, they received a link to our own website. This website was written in PHP and randomized participants into the three different voting rules. Depending on the treatment, they were given the instructions as shown in Appendix C.2 below. After the experiment, we also asked participants about their belief on how many other participants would vote YES, their gender, age, risk attitude, and how much they trusted Make-A-Wish. Finally, they received a random code to enter in MTurk so we could match responses to participants. All this information was stored in a database connected to our website.

For the test of Hypothesis 1, we used the simple response of YES/NO of all participants in treatment *SIM*. For all other treatments except *P-RAND*, we only used the responses in case participants were pivotal. In treatment *P-RAND*, we used the responses in case the random votes were equal to the actual votes (Hypotheses 4 and 5), the responses in case of all actual votes given pivotality (Hypothesis 6), and the responses in case of all random votes for a given number of actual votes (Hypothesis 7).

C.2 Instructions

Here, we provide the instructions for the exemplary case of majority voting ($k = 2$).

C.2.1 Introduction: all treatments

The instructions for all treatments started with the following text.

> Thank you for participating in this study. Please read the instructions carefully!
>
> You get \$0.80 for your participation and have the possibility to earn additional money.
>
> You are randomly paired with two other workers to form a **group of three**. Each member of your group votes separately on whether to take \$1.50 originally designated by us for donation to the Make-A-Wish Foundation, a charity that grants wishes to children with a critical illness.
>
> If there **are at least two votes** in favor of taking the money, then your group gets the \$1.50 designated for donation (each of you gets \$0.50 as a bonus payment). The charity Make-A-Wish gets nothing.
>
> If there **are less than two votes** for taking the money, then your group does not get the \$1.50. The money will instead be donated to Make-A-Wish.

C.2.2 Treatment *SIM*

After answering some control questions, participants in treatment *SIM* were then simply asked for their vote.

> Control questions:
>
> - Suppose exactly one of you votes for taking the money. Who gets the money in this case?

  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50
- Suppose exactly two of you vote for taking the money. Who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50

Now it is your decision: What do you vote for?

- Take the money
- Do not take the money

C.2.3 Unconditional voters in treatments *P-BASE*, *P-INFO*, and *P-RAND*

All unconditional voters in treatments *P-BASE*, *P-INFO*, and *P-RAND* received the same instructions. After the introduction, their instructions continued as follows.

> The timing is as follows: First, we ask you and one of the two other workers you are grouped with to vote independently from each other for or against taking the money. After the two of you have voted, we are going to ask the third worker for their vote. You then get the bonus payment if there are **at least two votes** for taking the money.
>
> However, it may be that your actual votes do not matter for the transfer of the money. In some cases, your votes are instead substituted by the votes of two other randomly chosen workers from a similar study. The third worker will then be informed that these randomly chosen votes matter for the final outcome. The third worker's vote always counts and is never substituted.
>
> As you do not know whether your vote actually counts for your own bonus payment or not you should make your decision assuming that it does count.
>
> **Before you make your decision, please answer the following two questions:**

- Suppose exactly one of you votes for taking the money. Who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50
- Suppose exactly two of you vote for taking the money. Who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50

Now it is your decision: What do you vote for?

- Take the money
- Do not take the money

C.2.4 Conditional voters in treatments *P-BASE* and *P-INFO*

After the introduction, conditional voters in treatments *P-BASE* and *P-INFO* received the same instructions except for the first paragraph, which was only shown to conditional voters in treatment *P-INFO*. Instead of just voting on whether or not to take the money, conditional voters in these treatments were asked for their vote depending on the votes of their group members.

[Only treatment *P-INFO* also included the following sentence:] In a similar study, **65%** of all workers who voted first voted for taking the money.

We have already collected the votes of the two workers you are grouped with. No one, one, or two of your group members may have voted for taking the money. We will ask for your decision for each of these three possibilities.

Before you make your decision, please answer the following two questions:

- Suppose exactly one of you votes for taking the money. Who gets the money in this case?

  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50

- Suppose exactly two of you vote for taking the money. Who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50

Now it is your decision. Please vote for the following situations:

None of the workers in your group have voted for taking the money.

- Take the money
- Do not take the money

Exactly one of the workers in your group has voted for taking the money.

- Take the money
- Do not take the money

Both of the workers in your group have voted for taking the money.

- Take the money
- Do not take the money

C.2.5 Conditional voters in treatment *P-RAND*

After the introduction, conditional voters in treatment *P-RAND* received the following instructions. Instead of just voting on whether or not to take the money, conditional voters in this treatment were asked for their vote depending on the votes of their group members and the outcome of the random draws.

> We have already collected the votes of the two workers you are grouped with. No one, one, or both may have voted for taking the money. In a similar study, **65%** of all workers who voted first voted for taking the money.

However, you were randomly assigned to a situation where the votes of the two workers you are grouped with do **NOT** matter for whether your group gets the money. Instead, the votes of your two group members are substituted by random draws. Each random draw has a probability of 65% for taking the money (as in the similar study).

We will ask for your decision for all possible combinations of these random draws (which count for the outcome) and the actual votes of your group members (which do not count).

Before you make your decision, please answer the following two questions:

Recall that the actual votes of the two other workers in your group do not count for the outcome.

- Suppose exactly one of the two random draws was for taking the money. If you vote against taking it, who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50
- Suppose again exactly one of the two random draws was for taking the money. If you vote for taking it, who gets the money in this case?
  - My group gets \$1.50 (which means that I get a bonus of \$0.50)
  - The charity Make-A-Wish gets \$1.50

We now ask you for your vote in the following situations:

Suppose **none** of the two random draws were for taking the money. Please vote for the following situations:

- **None** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.
  - Take the money
  - Do not take the money

- **Exactly one** of the other two workers in your group (and whose votes do not count for the final outcome) has voted for taking the money.
  - Take the money
  - Do not take the money
- **Both** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.
  - Take the money
  - Do not take the money

Suppose **exactly one** of the two random draws was for taking the money. Thus, the money will be taken only if you vote for it. Please vote for the following situations:

- **None** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.
  - Take the money
  - Do not take the money
- **Exactly one** of the other two workers in your group (and whose votes do not count for the final outcome) has voted for taking the money.
  - Take the money
  - Do not take the money
- **Both** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.
  - Take the money
  - Do not take the money

Suppose **both** of the two random draws were for taking the money. Please vote for the following situations:

- **None** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.

  - Take the money
  - Do not take the money

- **Exactly one** of the other two workers in your group (and whose votes do not count for the final outcome) has voted for taking the money.
  - Take the money
  - Do not take the money

- **Both** of the other two workers in your group (and whose votes do not count for the final outcome) have voted for taking the money.
  - Take the money
  - Do not take the money