

CESifo AREA CONFERENCES 2017

Behavioural Economics

Munich, 27–28 October 2017

Guilt in Voting and Public Good Games

Dominik Rothenhäusler, Nikolaus Schweizer and Nora Szech



Guilt in Voting and Public Good Games*

Dominik Rothenhäusler, Nikolaus Schweizer and Nora Szech

October 2017

Abstract

This paper analyzes how moral costs affect individual support of morally difficult group decisions. We study a threshold public good game with moral costs. Motivated by recent empirical findings, we assume that these costs are heterogeneous and consist of three parts. The first one is a standard cost term. The second, shared guilt, decreases in the number of supporters. The third hinges on the notion of being pivotal. We analyze equilibrium predictions, isolate the causal effects of guilt sharing, and compare results to standard utilitarian and non-consequentialist approaches. As interventions, we study information release, feedback, and fostering individual moral standards.

JEL Classification: D02, D03, D23, D63, D82.

Keywords: Moral Decision Making, Division of Labor, Shared Guilt, Diffusion of Responsibility, Institutions and Morals, Committee Decisions, Moral Transgression.

*This manuscript previously circulated under the title “Institutions, Shared Guilt, and Moral Transgression.” We would like to thank Peter Dürsch, Armin Falk, Werner Hildenbrand, Steffen Huck, Philipp Lergetporer, Clemens Puppe, Frank Rosar, Patrick Schmitz, Avner Shaked, Joel Sobel, Justin Valasek, and Christoph Vanberg. Dominik Rothenhäusler, Seminar for Statistics, ETH Zürich, Rämistr. 101, CH-8092 Zürich, email: rothenhaeusler@stat.math.ethz.ch, phone: +41 44 6325319. Nikolaus Schweizer, Department of Econometrics and OR, Tilburg University, P.O.box 90153, NL-5000 LE Tilburg, email: n.f.f.schweizer@uvt.nl. Nora Szech, Chair of Political Economy, ECON Institute, Karlsruhe Institute of Technology, Fritz-Erler-Str. 1-3, D-76128 Karlsruhe, email: nora.szech@kit.edu, phone: +49 721 60843809.

1 Introduction

Many morally problematic acts and decisions require the support of several people to become implementable, and often, morally difficult tasks are delegated to groups instead of one individual alone. Extreme examples are so-called “execution teams,” juries deciding about death penalties, and group military activity such as shooting squads. There may be different reasons for these arrangements. Certainly, they causally affect moral responsibility in people involved. An expert in military psychology, Dave Grossman, stresses that individual barriers towards morally problematic activity often break when people become part of teams or groups.¹ In a related vein, studies from social psychology (compare the overviews in Bandura 1999 and 2016) document that diffusion of responsibility and shared guilt reduce moral conscience in people. In addition, a diffused notion of being pivotal can decrease moral feelings (compare Falk and Szech 2014). Various studies from economics and related fields further document that people are heterogeneous in their moral behavior. Some personalities adhere to higher moral standards than others, even across different institutional contexts.²

Motivated by these findings, we analyze how moral costs affect individual support and outcomes in morally difficult group decisions. For this purpose, we study a threshold public good game with moral costs. We assume that agents are heterogeneous, and that moral costs consist of three terms. The first one is a standard cost term. The second, shared guilt, decreases in the number of supporters. The third hinges on the notion of being pivotal. We analyze equilibrium predictions, isolate the causal effects of guilt sharing, and compare results to standard utilitarian and non-consequentialist approaches. As interventions, we study information release, feedback, and fostering individual moral standards.

We first prove existence and uniqueness of a symmetric Bayesian Nash equilibrium in which transgression can happen with positive probability. We focus on this equilibrium throughout. Its outcome “Pareto dominates” the outcome

¹Compare Grossman (1996): “The individual is not a killer, but the group is” (p. 149).

²See O’Fallon and Butterfield (2005) and Loe, Ferrell and Mansfield (2000) for reviews of questionnaire and hypothetical scenario studies. Economic evidence comes from Albrecht, Krämer and Szech (2017) and from Deckers, Falk, Kosse and Szech (2016). Another related economic study focusing on social behavior is Bruhin, Fehr and Schunk (2016).

of the trivial equilibrium in which all agents remain passive.³ We find that this Bayesian Nash equilibrium takes the form of a threshold equilibrium. Above a certain moral cost type, agents stay out, while agents below this threshold support moral transgression actively.

An increase in individual moral standards, e.g. through training moral virtues, unambiguously reduces moral transgression. Yet releasing additional information to the agents, such that moral views become more spread out, can have different effects. If selfish benefits from transgression are high, information release can fight transgression. The opposite is true if selfish benefits are low. The reason behind is that information release shifts probability mass both into the upper and into the lower tail of the distribution of the agents' moral types. Morally "good" agents tend to become better, "bad" agents tend to become worse. A transgression that is only appealing to the morally worst agents can thus materialize more likely if information is released, while a transgression that requires the support of high moral types can become prevented more easily.⁴

Information on the exact number of fellow supporters affects feelings of moral responsibility as guilt can be shared with these members. Yet in many situations, this precise information may not become viable. For example, in a shooting squad, it may be that only the outcome itself can be seen. Therefore, we also analyze the case in which feedback on the number of supporters is not provided. Again, non-trivial, symmetric Bayesian equilibria exist. These are all in threshold strategies. We find that these thresholds are all larger than the threshold in the corresponding setting with feedback. Transgression therefore becomes more likely if feedback is lacking. The technical reason behind is Jensen's inequality. Moral costs diffuse better if agents can only build a conditional expectation about how many other agents supported the transgression. This prediction could be directly tested in an economic experiment.

We further contrast our findings with the case in which guilt does not diffuse in

³Pareto domination appears here in the sense that all agents strictly prefer the expected outcome of the symmetric Bayesian Nash equilibrium to the outcome of the trivial equilibrium. Pareto improvements here go together with a larger likelihood of moral transgression.

⁴These results consider information release that tends to foster more extreme opinions in agents who already hold different views. In contrast, information that fosters consensus in the population of agents leads to the reversed effects.

the number of agents. The quantitative predictions differ drastically. The impact of guilt diffusion is roughly comparable to multiplying the selfish benefit by the number of required supporters: We observe a kind of ‘strategic equivalence’ between the game in which guilt diffuses in the number of supporters, and the comparison game with a much smaller selfish benefit. Precisely, the selfish benefit is divided by the number of supporters needed, one agent less is required, and the total group size is one agent less.

Finally, we compare our predictions to those of the companion model in which agents’ moral reasoning follows a purely non-consequentialist approach. What matters for feelings of guilt is now the intention, not the outcome. If agents act as supporters, they face moral costs – regardless of whether transgression materializes or not. In this model, symmetric equilibria in which transgression occurs with positive probability only exist if the selfish benefit from transgression is sufficiently high. If the benefit is large, there exist multiple equilibria, all in threshold strategies. The equilibrium with the highest transgression probability Pareto dominates.

1.1 Related Literature

Many previous applications of discrete public good games to morally relevant action have studied so-called bystanding. Bystanders are people who observe a crime or accident. Often they do not help the person in distress. A rich empirical literature has documented that helping becomes less likely if others could help as well. Helping drastically decreases in the number of bystanders, see e.g. Darley and Latané (1968), Latané and Nida (1981), and Fischer et al. (2011) for an overview. For a victim, in order to receive help, it is often much better to have just one or few observers than many. This finding has been coined as the “bystander effect.”

In social psychology and in sociology, guilt diffusion and reduced feelings of breaking a norm have been put forward as explanations for the bystander effect (see also Zimbardo 2007). Diffusion of guilt has also been found to operate in committee situations (compare Bandura, Underwood and Fromson 1975, Bandura 1999, 2016). In a related vein, empirical studies have shown that a

reduced notion of being pivotal decreases feelings of moral responsibility, in individual decision contexts and in committees (Falk and Szech 2014).⁵

In contrast, the game-theoretic literature on bystanding, such as Harrington (2001), Osborne (2004), and Crettez and Deloche (2011), has argued via mixed equilibrium effects without diffusion of (moral) costs. The argument goes as follows. In mixed equilibrium, coordination is difficult. A high individual likelihood of helping could result in very many bystanders helping. This would be an unneeded over-provision of costly help. Therefore, avoiding this over-provision of help to large extents, in mixed equilibrium, individual willingness to help is small, and drastically decreases in the number of bystanders. This explanation works well for situations in which observers cannot coordinate. Yet the empirical literature on the bystanding effect also reports many cases in which observers could coordinate such that over-provision of help could have been prevented.⁶ In these cases, the mixed equilibrium explanation from standard game theory does not apply. It is thus likely that the aforementioned effects of reduced notions of moral responsibility and guilt diffusion play a critical role for the bystanding effect, specifically in cases where coordination of potential helpers is possible.

This motivates us to enrich the standard game-theoretic approach. We assume that costs of behaving in morally problematic ways can be shared and thereby diffuse, and notions of being pivotal affect feelings of guilt. Our model builds on the benchmark model due to Palfrey and Rosenthal (1984). We extend this model by including incomplete information⁷ about other agents' types and a richer model of how both, own behavior and the behavior of others affect moral costs. This enables us to study the effects of information provision and feedback. Moreover, we can analyze the impact and interplay of different types of moral costs. The complete information game possesses pure equilibria with perfect coordination as well as mixed equilibria. In the incomplete information case, the pure equilibria break away while the mixed equilibria survive. This renders the mixed equilibria specifically relevant.

⁵Sobel (2010) analyzes perceptions of being pivotal in market trading.

⁶See the overview in Fischer et al. (2011).

⁷There exists a literature on public good games with incomplete information, e.g. Nitzan and Romano (1990). Yet we are not aware of any contribution which addresses moral costs or cost sharing in these settings.

We first analyze the consequentialist moral perspective. Costs for immoral activity only arise if transgression materializes. In Section 5.2, we contrast this with a purely principle-based (Kantian) moral perspective under which moral costs do not hinge on the actual outcome.⁸ This part of the analysis is related to Huck and Konrad (2005) who consider mixed equilibria of a similar voting game with complete information. They study a committee situation in which agents vote about confiscating foreign investment. Confiscation leads to a selfish benefit to the group yet voting for confiscation comes along with Kantian costs of guilt. Guilt is thus independent of the voting outcome, and they assume that it cannot be shared with others. They find that transgression, i.e. confiscation, becomes less likely if the group size becomes larger.

Dufwenberg and Patel (2015) also assume a complete information environment and focus on pure equilibria. They add reciprocity preferences and cost-sharing (similar to our guilt diffusion) to Palfrey and Rosenthal’s model. With this combination (and in the absence of private information), they show that there can be efficient equilibria which do not require coordination. An earlier application of “diffusive” costs in complete information public good games is found in Weesie and Franzen (1998) who study the “volunteer’s dilemma,” i.e. the special case in which only one single agent needs to contribute.

Diffusion of moral costs further plays a crucial role in the analysis of Lindbeck, Nyberg and Weibull (1999). They consider a continuum situation such that coordination is always feasible. Agents need to decide if they want to work and thereby support others, or if they want to free-ride on others’ contributions. Opting for the latter comes with social stigma and thus with a morally relevant cost, yet this cost decreases (and thus diffuses) the larger the fraction of free-riders is. Agents are heterogeneous in their types such that the appeal of working varies across them. Lindbeck et al. analyze the equilibria of this model and demonstrate that sorting is feasible such that agents for whom working is less costly work, while others free-ride. In a sense, these equilibria have parallels to our threshold equilibrium in which agents with low moral cost types support transgression, while others stay out.

⁸The study of principle-based moral approaches within game-theoretic models goes back to Harsanyi, see e.g. Harsanyi (1977). Kranz (2010) applies this approach in various economic settings. This includes non-binary public good games in which abstaining causes moral costs for some agents.

2 Model and Basic Properties

This section introduces the morally relevant threshold public good game we are going to study. Motivated by recent empirical findings, we assume that moral costs are heterogeneous and consist of three parts. The first one is a standard cost term. The second one is diffusive in the number of people sharing in. The third one hinges on the notion of being pivotal.

We prove existence and uniqueness of a non-trivial, symmetric equilibrium. We will rely on this equilibrium for all predictions.

2.1 The Model

We consider $n \geq 1$ agents who each face the decision to support a morally difficult action. One may want to think about voting Yes for going to war instead of No, or for taking some morally shaky business endeavor instead of playing fair.⁹ The action (which we will also call transgression) requires the support of at least k of these agents in order to materialize. We are thus in a threshold public good game with moral costs.

We assume that if transgression happens, all agents receive a benefit $V > 0$ each (e.g. the country is defended, the group becomes richer etc.). Otherwise each agent receives 0. Transgression is associated with moral costs. These only accrue in the active supporters, i.e. in those who voted Yes instead of No. Motivated by recent empirical research from economics, business ethics, and social and military psychology, such as Bandura (1999, 2016), Deckers et al. (2016), Falk and Szech (2013, 2014), Grossman (1996), Loe et al. (2000), and Zimbardo (2007), these moral costs are heterogeneous and consist of three terms. The first term is a standard cost from the transgression happening. The second term captures guilt sharing and decreases in the number of supporters. The third term hinges on the notion of being pivotal.

Each agent has a private moral type x_i . Types are drawn independently from a commonly-known, continuous distribution function F with $F(0) = 0$ whose density function f is strictly positive over the support $(0, a)$, $a \in (0, \infty]$ of

⁹The activity may thus be either morally difficult, or morally completely wrong. In the latter case, it is clear from a designer's point of view that transgression should be prevented. This is the perspective we will take when it comes to political interventions, see Section 4.

F. The **three components of moral costs** are parametrized by the non-negative constants α , β and γ . The first one, $\alpha \cdot x_i$, accrues if the agents successfully supports the transgression. We will sometimes call it the non-diffusive part in order to differentiate it from the second cost term. The second moral cost term captures the concept of guilt sharing, $\beta \cdot x_i / Y$. Y denotes the total number of agents who support the transgression. Moral costs become smaller the more supporters guilt can be shared with. Third, the agent faces a cost of $\gamma \cdot x_i$ which materializes if he is pivotal i.e. if the transgression would not have occurred without his support. If the transgression does not materialize, we assume that agents do not feel guilty. We thus follow a consequentialist approach.¹⁰ Agents who do not support the transgression do not face any moral costs either.

The collective decision process is modeled as a **threshold public good game**. Agents simultaneously opt either for Yes or No. If at least k agents choose Yes, transgression happens. $k \in \{1, \dots, n\}$ is commonly known. This game can be interpreted as the result of a decision rule that was prescribed or agreed upon beforehand. Alternatively, it can be thought of as a game of volunteering to participate in an immoral action, and k as the minimum number of volunteers needed to carry it out successfully.

To sum up, the realized utility of agent i from supporting immoral activity, i.e. opting for **Yes**, is

$$\left(V - x_i \left(\alpha + \frac{\beta}{1 + Y_{-i}} \right) \right) 1_{\{Y_{-i} \geq k-1\}} - \gamma x_i 1_{\{Y_{-i} = k-1\}} \quad (1)$$

where Y_{-i} denotes the number of agents other than i who act as supporters. Realized utility from choosing **No** is

$$V 1_{\{Y_{-i} \geq k\}}. \quad (2)$$

We assume throughout that at least one of the guilt parameters α , β and γ is positive.

¹⁰See Section 5.2 for a discussion of the corresponding purely non-consequentialist approach.

2.2 Equilibrium Analysis

The solution concept we employ is Bayesian Nash equilibrium. We focus on equilibria which are symmetric in the sense that agents with the same type take the same moral decision.

For $k > 1$ there exists a pooling equilibrium in which all agents choose No independent of their type. All agents receive a utility of 0 in this equilibrium. As is common e.g. in the voting-games and matching literatures, we ignore this equilibrium in the following and focus on the Pareto-superior equilibria in which the morally difficult action is taken with positive probability.¹¹ As a first main result, we find that there always exists exactly one equilibrium of this type.

Theorem 1. *There exists a unique symmetric Bayesian equilibrium in which transgression happens with positive probability. In this equilibrium, agent i opts for Yes if $x_i \leq \theta_{k,n}$, and for No if $x_i > \theta_{k,n}$. $\theta_{k,n}$ is the unique solution of*

$$\begin{aligned} & V b(n-1, k-1, F(\theta_{k,n})) \\ & = \theta_{k,n} \left(\sum_{j=k-1}^{n-1} \left(\alpha + \frac{\beta}{1+j} \right) b(n-1, j, F(\theta_{k,n})) + \gamma b(n-1, k-1, F(\theta_{k,n})) \right) \end{aligned} \quad (3)$$

with

$$b(n, j, p) = \binom{n}{j} p^j (1-p)^{n-j}.$$

If $k < n$, $\theta_{k,n}$ lies in the interior of the support of F , $F(\theta_{k,n}) \in (0, 1)$ such that the equilibrium is a separating equilibrium.¹²

Equation (3) has a straightforward interpretation in terms of the costs and benefits of a pivotal agent. The left hand side consists of the gains in utility if the agent is pivotal, i.e. if exactly $k-1$ other agents are willing to take

¹¹As agents can guarantee themselves a non-negative payoff by choosing No, all agents must earn a non-negative payoff in any equilibrium. Thus any agent strictly preferring Yes in an equilibrium in which the morally difficult action is taken with positive probability, must earn a strictly positive expected payoff.

¹²We refer to equilibria in which agents' actions depend on their types as "separating equilibria." Strictly speaking, they are only "semi-separating" as we cannot infer an agent's type from his action because there are infinitely many types and only two actions.

action. On the right hand side we find the expected costs of an agent with threshold-type $\theta_{k,n}$.¹³

Throughout the paper, we focus on **three measures of moral transgression**. $\theta_{k,n}$ captures individual willingness to support morally difficult activity. Morally difficult behavior can increase on an individual level if institutional design lowers moral costs, e.g. via introducing better possibilities for guilt sharing. We measure such institutional effects on the individual by the quantity $\theta_{k,n}$, i.e. by the threshold-type who is indifferent between Yes and No. $P_{k,n}$ refers to the likelihood of transgression materializing. $N_{k,n}$ is the number of agents supporting transgression. $P_{k,n}$ and $N_{k,n}$ are thus outcome-based measures, derived from the individual thresholds. The equilibrium probability $P_{k,n}$ with which transgression occurs is given by

$$P_{k,n} = \sum_{j=k}^n b(n, j, F(\theta_{k,n})),$$

and the according equilibrium expected number of supporters is $N_{k,n} = n F(\theta_{k,n})$.

3 Comparative Statics and Limit Behavior

In this section, we establish the comparative statics for the game, as well as the convergence behavior as the number of agents grows large.

We analyze the comparative statics in the benefit V , the moral cost parameters α , β , and γ , the number of supporters necessary k , and the overall group size n . To this end, it proves useful to rewrite the equilibrium condition (3) into the form

$$\frac{V}{\theta_{k,n}} = H(F(\theta_{k,n}), \alpha, \beta, \gamma, k, n) \quad (4)$$

with

$$H(F, \alpha, \beta, \gamma, k, n) = \gamma + \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{b(n-1, k-1, F)} \left(\alpha + \frac{\beta}{1+j} \right). \quad (5)$$

¹³Uniqueness holds up to the decision of agents with type $\theta_{k,n}$. These are indifferent between Yes and No. We ignore this technicality in the following as it only becomes relevant with a probability of zero.

The left hand side of (4) is strictly decreasing from $+\infty$ to 0 in θ while $H(F(\theta), \alpha, \beta, \gamma, k, n)$ is increasing in θ with

$$H(0, \alpha, \beta, \gamma, k, n) = \alpha + \frac{\beta}{k} + \gamma.$$

The two functions thus intersect once, at the equilibrium threshold $\theta_{k,n}$. Therefore, if we wish to determine the comparative statics, e.g. the effect of an increase in V on equilibrium thresholds, we can argue as follows. Increasing V shifts the function V/θ upwards. The point of intersection and thus $\theta_{k,n}$ accordingly move to the right. Increasing the benefit V therefore increases the equilibrium threshold and thus induces more agents to support the transgression. The next proposition collects further basic properties of the equilibrium threshold $\theta_{k,n}$ and the associated outcome-based measures $P_{k,n}$ and $N_{k,n}$.

Proposition 1.

- (i) $\theta_{k,n}$, $P_{k,n}$ and $N_{k,n}$ are increasing in the benefit V and decreasing in the cost parameters α , β and γ .
- (ii) If $\alpha + \beta > 0$, $\theta_{k,n}$ and $N_{k,n}$ are increasing in k .

The monotonicity behavior of $\theta_{k,n}$ and $N_{k,n}$ in k can be interpreted as follows. If k increases, i.e. if the number of supporters necessary increases, more agents become willing to support the transgression both individually and overall, in expectation. There are thus fewer agents who hope to free-ride on others with lower moral standards who could support the transgression instead.

In the case $\alpha + \beta = 0$, in which guilt purely depends on the notion of being pivotal, it holds that

$$\theta_{k,n} = \frac{V}{\gamma}. \tag{6}$$

Thus, thresholds are independent of the voting rule and of the size of the group. The reason behind is that agents compare their selfish benefit from transgression only to the cost that emerges if they turn out to be pivotal. Therefore, there are no incentives to free-ride on others in this case. Accordingly, the number of supporters $N_{k,n} = nF(V/\gamma)$ is also independent of k , while $P_{k,n}$, the probability of finding at least k supporters decreases with k .

The dependence of the transgression probability $P_{k,n}$ on k is in general more complex as demonstrated in Figure 1. The Figure shows that in a setting with

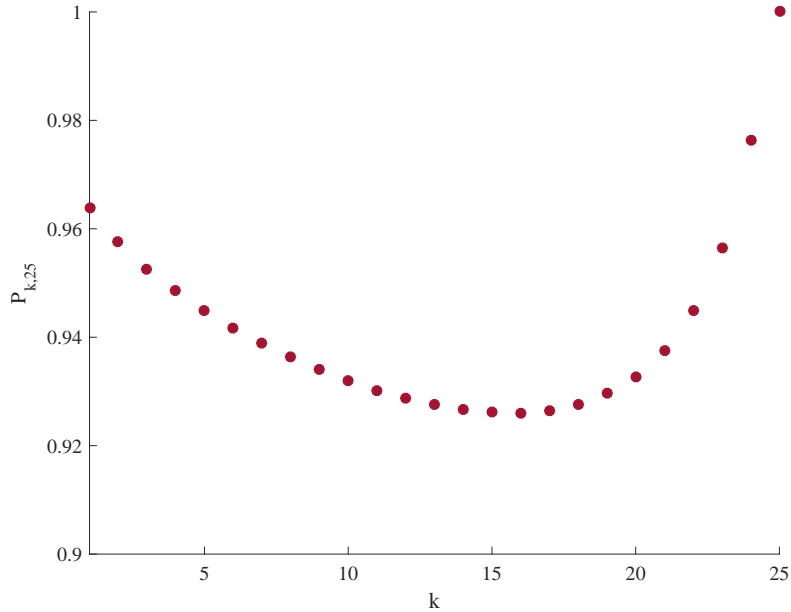


Figure 1: $P_{k,25}$ as function of k for $n = 25$, $V = 1$, $\beta = 1$, $\alpha = \gamma = 0$ and the exponential distribution $F(x) = 1 - \exp(-x)$.

purely diffusive guilt ($\alpha = \gamma = 0$), the probabilities form a U -shape. Transgression probability is smallest for an intermediate level of k . The intuition is as follows. If k is small, it is likely that there exist some agents with very low moral types. These will implement the transgression. For larger k , higher moral types would have to give their support. This only becomes an option if guilt diffuses well, i.e. if very many agents are required. For intermediate values of k neither of the two mechanisms helps agent much to overcome their moral concerns and, accordingly, immoral action materializes only with comparatively small probability.¹⁴

These effects still play a role in the general model in which guilt can only partly be shared, thereby precluding an easy comparative statics analysis of $P_{k,n}$. Yet there is one crucial difference. If guilt has either a non-diffusive component ($\alpha > 0$) or there are costs from being pivotal ($\gamma > 0$) (and when the support of F is unbounded), there will be some types who do not support transgression regardless of k and n . These very moral agents render transgression for k close to n less likely. The following lemma summarizes the results.

¹⁴See Rothenhäusler, Schweizer and Szech (2015) for further discussion and some analytical results on this U -shape.

Lemma 1.

(i) If $\alpha + \gamma > 0$, individual thresholds are bounded,

$$\theta_{k,n} \leq \frac{V}{\alpha + \gamma}.$$

In particular, if the support of F is unbounded, there exists a positive mass of agent types with high moral values who never support transgression, independently of k and n .

(ii) If $\alpha = \gamma = 0$, i.e. if guilt is fully diffusive as n grows large, for any type x in the support of F there exist k and n such that $x < \theta_{k,n}$. Thus any type x can become a supporter of transgression.

Finally, let us consider the limit results if the population size grows large.

Proposition 2. Suppose that $\alpha + \beta > 0$. Then the following claims hold.

(i) $\lim_{n \rightarrow \infty} \theta_{k,n} = 0$.

(ii) $\lim_{n \rightarrow \infty} N_{k,n} = \infty$.

(iii) $\lim_{n \rightarrow \infty} P_{k,n} = 1$.

While individual thresholds generally converge to 0, the expected number of supporters $N_{k,n}$ becomes arbitrarily large. We thus find an overshooting effect. In the limit, instead of the required k , infinitely many agents become supporters. This holds even if only one supporter is necessary, and also if only α is strictly positive and β is 0, i.e. if guilt does not diffuse. Instead of one, infinitely many agents join in. Accordingly, the transgression probability $P_{k,n}$ converges to 1. The reason behind is that the number of agents with very low moral standards grows that fast in n that it overrules the counteractive effect from the threshold which decreases in n . There is uncertainty about whether there are indeed many enough low types to free-ride on. This makes many more agents think about joining in than necessary. Coordination is impossible, and the gains from a transgression that is supported by too many are higher than the losses from the transgression not happening. In the case $\alpha + \beta = 0$, guilt only depends on notions of being pivotal. As shown in (6), the thresholds $\theta_{k,n}$ are positive and independent of k and n . Also in that case, (ii) and (iii) are satisfied.¹⁵

¹⁵Each agent participates with a probability $p = F(V/\theta)$ that does not depend on n . This

4 Interventions

In this section, we study the impact of three different interventions: training agents' individual moral virtues, increasing the precision of their information, and varying whether they receive feedback about others' support behavior after the decision (as in the basic model) or not.

4.1 Trained Virtues

Possibilities for training moral virtues have already been put forward by Aristoteles (see Barnes 1984). The next proposition demonstrates that an intervention which induces an upwards shift in the distribution of moral types indeed reduces all three measures of transgression. To this end, we compare two versions of our model with respective distributions F and G while we keep all other parameters constant. We indicate the distribution dependence through a superscript on the thresholds, $\theta_{k,n}^F$ and $\theta_{k,n}^G$, and analogously for all other quantities.

Proposition 3. *Suppose that G dominates F in terms of first order stochastic dominance, $G(x) \leq F(x)$ for all $x \geq 0$, so that G can be interpreted as a (non-linear) upwards shift of F . Then it holds that $\theta_{k,n}^G \geq \theta_{k,n}^F$, $N_{k,n}^G \leq N_{k,n}^F$, and $P_{k,n}^G \leq P_{k,n}^F$.*

An upwards shift in the distribution of moral types thus establishes higher moral standards in equilibrium. According to all three morally relevant measures, transgression decreases.

In a related vein, one may think about the impact of choice architectures. The idea of nudging agents towards more moral or socially responsible behavior has been discussed extensively in recent public debate, see e.g. Thaler and Sunstein (2008). Choice architectures may affect the moral costs that emerge from transgression directly. See e.g. the omission-commission bias as reported in Spranca, Misk and Baron (1991). These effects can be captured by a shift in the distribution of costs, as seen above. In other cases, nudging may affect the decision-making process, yet keep the underlying moral costs unchanged.

implies $N_{k,n} = np \rightarrow \infty$ and leads to $P_{k,n} \rightarrow 1$ far more directly than in the general case.

4.2 Information Release

We next investigate the effect of information release on moral outcomes. We here consider a situation in which information release leads to a more spread out distribution of moral types.¹⁶

We now interpret F as the distribution of (conditionally expected) moral costs, based on coarse information. Through releasing information, a designer can induce a more spread-out distribution G in which each agent has learned his moral costs more precisely.¹⁷ To formalize the notion of G being more spread out than F we rely on the following single-crossing criterion. G is more spread out in the sense of the single-crossing criterion, $G \succeq_{SC} F$, if there exists x_0 such that $F(x) \leq G(x)$ for $x \leq x_0$ and $G(x) \leq F(x)$ for $x \geq x_0$. If F and G share the same mean – as is the case if agents employ Bayesian updating – single-crossing is a sufficient condition for G being a mean-preserving spread of F .¹⁸

The next proposition demonstrates that the effect of information release crucially depends on the potential benefits of the moral transgression, i.e. on the level of V .

Proposition 4. *Fix all parameters except for V and suppose that G is more spread out than F in single-crossing precision, $G \succeq_{SC} F$. Then there exists $v_0 \geq 0$ with the following properties.*

- (i) *For $V \geq v_0$, we have $\theta_{k,n}^G \geq \theta_{k,n}^F$, $N_{k,n}^G \leq N_{k,n}^F$, and $P_{k,n}^G \leq P_{k,n}^F$.*
- (ii) *For $V \leq v_0$, we have $\theta_{k,n}^G \leq \theta_{k,n}^F$, $N_{k,n}^G \geq N_{k,n}^F$, and $P_{k,n}^G \geq P_{k,n}^F$.*

Thus, additional information decreases transgression if stakes are high, but increases transgression if stakes are low. This happens because information release shifts probability mass both into the upper and into the lower tail. Morally “good” agents become better, and “bad” agents become worse. If stakes are low, only the worst agents in the population act as supporters. The

¹⁶The opposite case is, of course, also possible. Information could foster consensus in a population, which would lead to a less spread out distribution of moral types. In the latter case, the analysis thus applies in the reversed way.

¹⁷This is in line with Ganuza and Penalva (2010)’s model of information release.

¹⁸See Diamond and Stiglitz (1974) for further discussion of this relation. We do not need the assumption of F and G sharing the same mean and thus do not impose it.

moral costs of these agents tend to decrease as the distribution becomes more spread out. Therefore, transgression becomes more likely.

Conversely, if stakes are high, even good agents with high costs are near the threshold of becoming supporters. As the upper tail of the distribution becomes more spread out, these agents tend to end up above the threshold and thus do not become supporters. Hence transgression becomes less likely.

4.3 Feedback

In the baseline model, we assume that supporters of the transgression observe the precise number of other supporters and share their guilt with these. In the following, we contrast this baseline setting with the situation in which agents only observe whether the transgression has occurred. Thus, they do not learn the precise number of supporters. We call this counterpart model the game without feedback.

In the game without feedback, utility from voting Yes is given by

$$\left(V - x_i \left(\alpha + \frac{\beta}{1 + E[Y_{-i} | Y_{-i} \geq k-1]} + \gamma P(Y_{-i} = k-1 | Y_{-i} \geq k-1) \right) \right) \cdot \mathbf{1}_{\{Y_{-i} \geq k-1\}}$$

where the conditional expectation is taken under the distribution induced by equilibrium behavior. For instance, in an equilibrium in which agents vote Yes whenever their type is below the threshold $\tilde{\theta}_{k,n}$, this conditional expectation is given by

$$E[Y_{-i} | Y_{-i} \geq k-1] = \frac{\sum_{j=k-1}^{n-1} j b(n-1, j, F(\tilde{\theta}_{k,n}))}{\sum_{j=k-1}^{n-1} b(n-1, j, F(\tilde{\theta}_{k,n}))}.$$

Similarly, we have

$$P(Y_{-i} = k-1 | Y_{-i} \geq k-1) = \frac{b(n-1, k-1, F(\tilde{\theta}_{k,n}))}{\sum_{j=k-1}^{n-1} b(n-1, j, F(\tilde{\theta}_{k,n}))}.$$

The next proposition shows that symmetric equilibria of the game without feedback are equilibria in threshold strategies, just like in the baseline model. Such an equilibrium always exists, and the moral thresholds $\tilde{\theta}_{k,n}$ in any such equilibrium are higher than the thresholds $\theta_{k,n}$ in the corresponding baseline

model. Absence of feedback thus makes diffusion of guilt more effective, rendering moral transgression more likely.¹⁹

Proposition 5. *The game without feedback possesses a symmetric equilibrium in which transgression happens with positive probability. All such equilibria are equilibria in threshold strategies where the thresholds $\tilde{\theta}_{k,n}$ solve*

$$\frac{V}{\tilde{\theta}_{k,n}} = \tilde{H}(F(\tilde{\theta}_{k,n}), \alpha, \beta, \gamma, k, n) \quad (7)$$

with

$$\begin{aligned} \tilde{H}(F, \alpha, \beta, \gamma, k, n) &= \alpha \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} + \gamma \\ &+ \beta \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} \frac{1}{\sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{B(n-1, k-1, F)} (1+j)} \end{aligned}$$

and

$$B(n-1, k-1, F) = \sum_{j=k-1}^{n-1} b(n-1, j, F).$$

Moreover, any solution $\tilde{\theta}_{k,n}$ to (7) satisfies $\tilde{\theta}_{k,n} \geq \theta_{k,n}$ where $\theta_{k,n}$ is the unique equilibrium threshold of the baseline model with feedback.

Accordingly, the transgression probability $\tilde{P}_{k,n}$ and the expected number of supporters $\tilde{N}_{k,n}$ are larger than if feedback on the number of supporters is precise. The lack of feedback renders it easier for agents to share their guilt with others. Therefore, they become more willing to support moral transgression. Technically, this result stems from Jensen's inequality and the convexity of moral costs in the number of supporters. Without precise knowledge of the numbers of supporters, guilt tends to diffuse better.

5 Diffusion of Guilt

In this section, we further explore the effect of diffusion of guilt. Section 5.1 discusses the role of diffusion of guilt compared to the other components of

¹⁹We do not show uniqueness of equilibrium yet numerical results suggest that it holds. We prove existence and a comparison between *any* equilibrium and the unique equilibrium of the baseline model.

moral costs. Section 5.2 describes how predictions change if agents adopt a non-consequentialist perspective, thus experiencing guilt whenever they act as supporters – regardless of whether transgression actually happens.

5.1 Isolating the Effect of Diffusion

We will see that there is an almost substitutive relation between the diffusive and the non-diffusive components of moral costs. By and large, if guilt is completely diffusive ($\alpha = 0$) predictions are quite comparable to the case in which guilt is completely non-diffusive ($\beta = 0$) but where the benefit V from transgression is almost k times larger. Theorem 2 specifies this relationship in all precision.

Theorem 2. *Denote by $\theta_{k,n}(\alpha, \beta, \gamma, V)$ the equilibrium threshold of the game with parameters $\alpha, \beta, \gamma, V, k$ and n as defined in (4).*

(i) *Fix $w \in [0, 1]$ and $\alpha_0 > 0$ and define θ_1 and θ_2 as the equilibrium thresholds of the game without diffusion ($\beta = 0$) for (k, n) and for $(k + 1, n + 1)$,*

$$\theta_1 = \theta_{k,n}(\alpha_0, 0, \gamma, V) \quad \text{and} \quad \theta_2 = \theta_{k+1,n+1}(\alpha_0, 0, \gamma, V).$$

Then it holds that

$$\min(\theta_1, \theta_2) \leq \theta_{k,n}(w\alpha_0, k(1-w)\alpha_0, \gamma, V) \leq \max(\theta_1, \theta_2).$$

(ii) *For any $\delta > 0$, we have $\theta_{k,n}(0, \delta, 0, V) = \theta_{k+1,n+1}(\delta, 0, 0, k \cdot V)$.*

(iii) *We have the bound*

$$\theta_{k,n}(\alpha, \beta, \gamma, V) \leq \theta_{k,n}(\alpha, 0, \gamma, V) \leq \left(1 + \frac{\beta}{k\alpha}\right) \theta_{k,n}(\alpha, \beta, \gamma, V).$$

The theorem demonstrates a kind of strategic equivalence between the game with n agents out of whom k are required for transgression to happen and guilt diffuses in the number of supporters, and the comparison game in which guilt does not diffuse. The number of agents needs to be mildly adjusted by one, both for the group size and for the number of agents required. The benefit from transgression needs to be inflated by k .

Part (i) illustrates that up to adjusting between (k, n) and $(k+1, n+1)$, decreas-

ing the non-diffusive guilt term still leads to the same equilibrium thresholds if the diffusive term is increased by k -times that amount. Part (ii) focuses on the case in which notions of being pivotal do not affect guilt, $\gamma = 0$. Up to adjusting k and n by one, if guilt diffuses, $(\alpha, \beta) = (0, \delta)$, equilibrium thresholds are identical to those of the game with perfectly non-diffusive guilt, $(\alpha, \beta) = (\delta, 0)$ if the benefit V is inflated by k .²⁰ Part (iii) assesses quantitatively how the diffusive component of guilt affects equilibrium thresholds. Equilibrium thresholds of the two games with $\beta > 0$ and $\beta = 0$, lie within a factor $1 + \frac{\beta}{k\alpha}$ of each other. This holds independently of the distribution of moral types, F . For fixed α and β , the impact of the diffusive guilt component on equilibrium behavior vanishes quickly as k increases. Guilt sharing between agents is then so strong that this costs component has little impact on the agents' behavior.

5.2 A Purely Non-Consequentialist View

So far, guilt hinged on consequences of actions instead of intent. In this section, we focus on the latter. Under this purely non-consequentialist (or Kantian) moral perspective, agents face moral costs from choosing Yes, independent of the outcome.²¹ As before, we assume that agents share guilt with other supporters, while notions of being pivotal play not role anymore as consequences do not affect guilt.

The following dichotomy emerges. For small values of the selfish benefit V , the equilibrium in which all agents opt against transgression is the unique symmetric equilibrium. For larger V , there are several symmetric equilibria in which the transgression occurs with strictly positive probability. The equilibrium with the highest transgression probability Pareto-dominates.

Moral costs for agent i from opting for Yes are now given by

$$V1_{\{Y_{-i} \geq k-1\}} - x_i \left(\alpha + \frac{\beta}{1 + Y_{-i}} \right).$$

²⁰This is, of course, not a strategic equivalence in the strictest sense as the number of agents mildly varies across the two games.

²¹In the terminology of Admati and Perry (1991), we compare our basic model, a subscription game, to the associated contribution game.

We thus dropped the term that captured notions of being pivotal which was inherently consequentialist. Realized utility from voting No is given by

$$V 1_{\{Y_{-i} \geq k\}}.$$

The following proposition shows that – under mild conditions – transgression occurs in equilibrium only if V is sufficiently large.

Proposition 6.

(i) *For sufficiently large V , there exist symmetric equilibria in which transgression occurs with positive probability. These are in threshold strategies. Equilibrium thresholds solve*

$$\frac{V}{\bar{\theta}_{k,n}} = \bar{H}(F(\bar{\theta}_{k,n}), \alpha, \beta, k, n) \quad (8)$$

with

$$\bar{H}(F, \alpha, \beta, k, n) = \sum_{j=0}^{n-1} \frac{b(n-1, j, F)}{b(n-1, k-1, F)} \left(\alpha + \frac{\beta}{1+j} \right).$$

If two thresholds $\bar{\theta}_{k,n}^{(1)}$ and $\bar{\theta}_{k,n}^{(2)} > \bar{\theta}_{k,n}^{(1)}$ both solve (8), they both correspond to equilibria. The equilibrium associated with the larger threshold $\bar{\theta}_{k,n}^{(2)}$ Pareto dominates the other.

(ii) *Assume $f(0) < \infty$ and $k > 2$. Then, for sufficiently large V , there exist at least two symmetric equilibria in which transgression occurs with positive probability.*

(iii) *Assume $f(0) < \infty$ and $k > 1$. Then, for small V , the unique symmetric equilibrium is the one in which all agents choose No.*

Thus, in equilibrium, small temptations V do not cause moral transgression – but sufficiently large ones do. This is due to a “two-sided” coordination problem.²² For an agent, in addition to the possibility of free-riding on other agents’ Yes-votes, opting for No has another appeal. It avoids the risk of facing moral costs without reaping any benefits. For small V , the possibility to coordinate on the transgression breaks down. The worst agents in the population

²²The two-sided character of this coordination problem also causes the multiplicity of equilibria for large V .

now fear that the other agents with their “holier than thou” attitudes will let them down.

6 Conclusion

If people engage in “psychosocial manoeuvres – often aided by the institutions [...], which absolve them from moral responsibility for harmful acts” (Haidt and Kesebir, 2010, p. 812), further research is needed to understand and disentangle the mechanisms behind institutionally driven moral transgression. This paper contributes to this endeavor. So far, economic models of group activity in morally relevant contexts are rare. Notable exceptions are Huck and Konrad (2005) looking into committee decisions on confiscating foreign investments, Sobel (2010) analyzing moral responsibility in market activity, and Huesmann and Wambach (2015) studying mechanisms for rewarding dangerous services of moral relevance. These applications outline already that studying moral behavior can come with different goals.

The design of institutions can likely foster morally challenging behavior when societies need it, and limit unwanted transgression such as corruption and cruelty. Speaking with the social psychologist Albert Bandura (1999, p. 297), “...societies cannot rely entirely on individuals to deter human cruelty. Civilized life requires, in addition to human personal codes, effective social safeguards against the misuse of power for exploitative and destructive purposes.” We believe that economic research can and should contribute to shaping such social safeguards. It should contribute to a deeper understanding of how institutional design affects morally relevant decision-making, and shed light on effects that may be unwanted.

A Proofs

Proof of Theorem 1. We first show that all equilibria, symmetric or not, in which transgression happens with positive probability must be equilibria in threshold strategies. For each agent i , there is a threshold $\theta_{k,n,i} \in [0, \infty]$ such that i votes Yes if $x_i < \theta_{k,n,i}$ and No if $x_i > \theta_{k,n,i}$. To see this, fix the strategies of the other agents. In an equilibrium in which transgression happens with

positive probability, at least $k - 1$ of the other agents must vote Yes with positive probability. Assume agent i weakly prefers voting Yes over No for some type x_i . Comparing the expectations of (1) and (2) implies that this preference must be strict for all lower types. Likewise, if agent i weakly prefers voting No over Yes for some type x_i , this preference must be strict for all higher types. Thus, all best responses to the other agents' strategies are threshold strategies. Therefore, all equilibria in which transgression happens with positive probability are equilibria in threshold strategies.²³

For $k < n$, we now show that there exists a unique threshold $\theta_{k,n} > 0$ with the property that if all agents play a threshold strategy with $\theta_{k,n}$, an agent with type $\theta_{k,n}$ is indifferent between voting Yes and No. We treat the case $k = n$ separately at the end. Consider an agent with type x and assume that all other agents follow a threshold strategy with $\theta_{k,n} > 0$. Expected payoff from voting Yes is then given by

$$\sum_{j=k-1}^{n-1} b(n-1, j, F(\theta_{k,n})) \left(V - \left(\alpha + \frac{\beta}{1+j} \right) x \right) - b(n-1, j, F(\theta_{k,n})) \gamma x \quad (9)$$

where the Binomial distribution arises as the other agents independently vote Yes with probability $F(\theta_{k,n})$. Expected payoff from voting No is given by

$$\sum_{j=k}^{n-1} b(n-1, j, F(\theta_{k,n})) V. \quad (10)$$

Equilibria are characterized by thresholds $\theta_{k,n}$ for which the two expressions coincide for $x = \theta_{k,n}$. Equating (9) and (10) for $x = \theta_{k,n}$ yields condition (3) as stated in the proposition. It remains to show existence of a unique (positive) solution. To this end, we write (3) as

$$\frac{V}{\theta_{k,n}} = H(F(\theta_{k,n}), \alpha, \beta, \gamma, k, n) \quad (11)$$

²³As the equilibrium in which all agents always vote No is also in threshold strategies, it follows that in fact all symmetric equilibria are in threshold strategies.

with

$$H(F, \alpha, \beta, \gamma, k, n) = \gamma + \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{b(n-1, k-1, F)} \left(\alpha + \frac{\beta}{1+j} \right).$$

The left hand side of (11) is strictly decreasing in $\theta_{k,n}$, diverging to ∞ at 0 and becoming arbitrarily small for large $\theta_{k,n}$. Writing the right hand side as

$$\gamma + \sum_{j=k-1}^{n-1} \left(\alpha + \frac{\beta}{1+j} \right) \frac{\binom{n-1}{j}}{\binom{n-1}{k-1}} \left(\frac{F(\theta_{k,n})}{1-F(\theta_{k,n})} \right)^{j-k+1},$$

we see that it strictly increases from $\alpha + \frac{\beta}{k} + \gamma$ to ∞ as $\theta_{k,n}$ moves through the support of F . This proves the existence of a unique solution $\theta_{k,n}$ in the interior of the support.

It remains to study the case $k = n$. Like in the case of $k < n$, an interior separating equilibrium must be characterized by (11) which becomes

$$\frac{V}{\theta_{n,n}} = \alpha + \frac{\beta}{n} + \gamma \quad (12)$$

and thus

$$\theta_{n,n} = \frac{V}{\alpha + \frac{\beta}{n} + \gamma}. \quad (13)$$

In the case where this choice of $\theta_{n,n}$ lies above the support of F (and only then), all agents voting Yes regardless of their type is an equilibrium.²⁴ \square

Proof of Proposition 1. By (11), $\theta_{k,n}$ is characterized as the intersection between the decreasing function $V/\theta_{k,n}$ and the function H which is increasing in $\theta_{k,n}$ via F . Increasing the decreasing function $V/\theta_{k,n}$ by increasing V necessarily shifts the point of intersection to the right, thus increasing $\theta_{k,n}$. Similarly, $\theta_{k,n}$ decreases in α , β and γ : Increasing any of these parameters increases the function H , thus shifting the intersection to the left. The claims about $N_{k,n}$ and $P_{k,n}$ in (i) follow from the behavior of $\theta_{k,n}$: $N_{k,n}$ and $P_{k,n}$ can be seen as the mean and the probability of observing a realization above $k-1$ for a binomial distribution with “success probability” $F(\theta_{k,n})$. Both quantities increase in this probability. To prove (ii), we first show the following alternative

²⁴The separating equilibrium thus degenerates to a pooling equilibrium.

representation of the function H :

$$H(F, \alpha, \beta, \gamma, k, n) = \alpha \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{b(n-1, k-1, F)} + \frac{\beta}{k} \sum_{j=k}^n \frac{b(n, j, F)}{b(n, k, F)} + \gamma. \quad (14)$$

To see this, we rewrite the part of H which depends on β . Observe first the following connection between binomial distributions for $(k-1, n-1)$ and (k, n) :

$$\begin{aligned} & \sum_{j=k-1}^{n-1} b(n-1, j, F) \frac{1}{1+j} \\ &= \sum_{j=k-1}^{n-1} \binom{n-1}{j} F^j (1-F)^{n-j-1} \frac{1}{1+j} \\ &= \frac{1}{n} \sum_{j=k-1}^{n-1} \binom{n}{j+1} F^j (1-F)^{n-j-1} \\ &= \frac{1}{n} \sum_{l=k}^n \binom{n}{l} F^{l-1} (1-F)^{n-l} \\ &= \frac{1}{nF} \sum_{l=k}^n b(n, l, F). \end{aligned}$$

Combining this identity with $b(n-1, k-1, F) = \frac{k}{nF} b(n, k, F)$ yields

$$\beta \frac{\sum_{j=k-1}^{n-1} b(n-1, j, F) \frac{1}{1+j}}{b(n-1, k-1, F)} = \frac{\beta}{k} \frac{\sum_{j=k}^n b(n, j, F)}{b(n, k, F)}$$

and thus (14). From (14), we can conclude that H is decreasing in k . This follows immediately from the fact that both

$$\frac{\sum_{j=k-1}^{n-1} b(n-1, j, F)}{b(n-1, k-1, F)} \quad \text{and} \quad \frac{\sum_{j=k}^n b(n, j, F)}{b(n, k, F)}$$

are reciprocals of the *failure rate* of the Binomial distribution which is increasing in k , see Johnson, Kotz and Kemp (1992), Chapter 3. Consequently, $\theta_{k,n}$ is increasing in k as an increase in k decreases H and thus shifts the intersection of V/θ and H to the right. As $N_{k,n} = nF(\theta_{k,n})$ depends on k only through $\theta_{k,n}$, it increases in k as well. \square

Proof of Lemma 1. To see (i), note that $H(F, \alpha, \beta, \gamma, k, n) \geq \alpha + \gamma$. Thus,

(11) implies $V/\theta_{k,n} \geq \alpha + \gamma$. To see (ii), note that by (13) the threshold $\theta_{n,n}$ becomes $\theta_{n,n} = nV/\beta$ for $\alpha = \gamma = 0$. $\theta_{n,n}$ thus becomes arbitrarily large as n increases. \square

Proof of Proposition 2. The **proof of (i)** proceeds in two steps. We first show that it suffices to prove the result for the case $\alpha = \gamma = 0$. Then we prove (i) for this case. Let $\theta'_{k,n}$ be the equilibrium threshold with parameters $\alpha' = 0$, $\gamma' = 0$ and $\beta' = \alpha + \beta$. By the definition of H , we have

$$H(F, \alpha, \beta, \gamma, k, n) \geq H(F, 0, \alpha + \beta, 0, k, n).$$

If $\theta'_{k,n} < \theta_{k,n}$, then by this inequality we would have

$$\frac{V}{\theta'_{k,n}} = H(F(\theta'_{k,n}), 0, \alpha + \beta, 0, k, n) \leq H(F(\theta_{k,n}), \alpha, \beta, \gamma, k, n) = \frac{V}{\theta_{k,n}},$$

which is a contradiction. Hence, we have the bound $\theta'_{k,n} \geq \theta_{k,n}$ and $\theta'_{k,n} \rightarrow 0$ implies $\theta_{k,n} \rightarrow 0$.

We prove $\theta'_{k,n} \rightarrow 0$ by contradiction. Suppose (i) is violated for $\theta'_{k,n}$, i.e. suppose that there exists a subsequence n_j of \mathbb{N} and $\varepsilon > 0$ such that $\theta'_{k,n_j} \geq \varepsilon > 0$ for all $j \in \mathbb{N}$. Note that the sum

$$\sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}$$

is monotonically increasing in p for $p \in [0, 1]$. Thus,

$$\begin{aligned} \sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta'_{k,n_j})^i (1 - F(\theta'_{k,n_j}))^{n_j-i} &\geq \sum_{i=k}^{n_j} \binom{n_j}{i} F(\varepsilon)^i (1 - F(\varepsilon))^{n_j-i} \quad (15) \\ &= 1 - \sum_{i=0}^{k-1} \binom{n_j}{i} F(\varepsilon)^i (1 - F(\varepsilon))^{n_j-i}. \end{aligned}$$

Since $\varepsilon > 0$ implies $F(\varepsilon) > 0$, the right hand side converges to 1: $\binom{n_j}{i}$ grows polynomially in n_j and $(1 - F(\varepsilon))^{n_j-i}$ decays exponentially in n_j . Now we rewrite the equilibrium condition (14) into

$$\beta' \frac{\sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta'_{k,n_j})^i (1 - F(\theta'_{k,n_j}))^{n_j-i}}{\binom{n_j}{k} F(\theta'_{k,n_j})^k (1 - F(\theta'_{k,n_j}))^{n_j-k}} = \frac{kV}{\theta'_{k,n_j}}$$

and then into

$$\begin{aligned}
& \beta' \sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta'_{k,n_j})^i (1 - F(\theta'_{k,n_j}))^{n_j-i} \\
&= \frac{kV}{\theta'_{k,n_j}} \binom{n_j}{k} F(\theta'_{k,n_j})^k (1 - F(\theta'_{k,n_j}))^{n_j-k} \\
&\leq \frac{kV}{\varepsilon} \binom{n_j}{k} (1 - F(\varepsilon))^{n_j-k}.
\end{aligned} \tag{16}$$

As $F(\varepsilon) > 0$, and as $\binom{n_j}{i}$ grows polynomially in n_j and $(1 - F(\varepsilon))^{n_j-i}$ decays exponentially in n_j , the final term converges to 0. This is a contradiction as we proved in (15) that the left hand side of (16) converges to 1. This completes the proof of (i).

The **proof of (ii)** proceeds in two similar steps. We first argue that we can focus on the case in which the diffusive guilt term is absent, and then prove the result for this case. Let $\theta'_{k,n}$ now be the equilibrium threshold for the case $\alpha' = \alpha + \beta$, $\gamma' = \gamma$ and $\beta' = 0$. By definition,

$$H(F, \alpha, \beta, \gamma, k, n) \leq H(F, \alpha + \beta, 0, \gamma, k, n).$$

If $\theta'_{k,n} > \theta_{k,n}$, then by this inequality

$$\frac{V}{\theta_{k,n}} = H(F(\theta_{k,n}), \alpha, \beta, \gamma, k, n) \leq H(F(\theta'_{k,n}), \alpha + \beta, 0, \gamma, k, n) = \frac{V}{\theta'_{k,n}}$$

which is a contradiction. Hence, we have $\theta'_{k,n} \leq \theta_{k,n}$, and it suffices to show that $nF(\theta'_{k,n}) \rightarrow \infty$. To this end, consider a subsequence n_l such that $n_l F(\theta'_{k,n_l})$ converges to a number in $c \in [0, \infty)$. Now we use that

$$\frac{V}{\theta'_{k,n}} = \gamma + \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F(\theta'_{k,n}))}{b(n-1, k-1, F(\theta'_{k,n}))} (\alpha + \beta). \tag{17}$$

Calculating the limits for $l \rightarrow \infty$ results in

$$\begin{aligned}
& \lim_l b(n_l - 1, k - 1, F(\theta'_{k,n_l})) \\
&= \lim_l \binom{n_l - 1}{k - 1} F(\theta'_{k,n_l})^{k-1} (1 - F(\theta'_{k,n_l}))^{n_l - k} \\
&= \lim_l \frac{\binom{n_l - 1}{k - 1}}{n_l^{k-1}} \lim_l n_l^{k-1} F(\theta'_{k,n_l})^{k-1} \lim_l (1 - F(\theta'_{k,n_l}))^{n_l - k} \\
&= \frac{1}{(k - 1)!} c^{k-1} \lim_l (1 - F(\theta'_{k,n_l}))^{n_l} \\
&= \frac{c^{k-1}}{(k - 1)!} \exp(-c)
\end{aligned}$$

and, analogously,

$$\begin{aligned}
\sum_{j=k-1}^{n-1} b(n - 1, j, F(\theta'_{k,n})) &= 1 - \sum_{j=0}^{k-2} b(n - 1, j, F(\theta'_{k,n})) \\
&\rightarrow 1 - \sum_{j=0}^{k-2} \frac{c^j}{j!} \exp(-c).
\end{aligned}$$

Hence, along the subsequence, the right hand side of (17) converges to

$$\gamma + (\alpha + \beta) \left(\exp(c) - \sum_{j=0}^{k-2} \frac{c^j}{j!} \right) \frac{(k - 1)!}{c^{k-1}}.$$

But, as shown in part (i) of the proposition, the left hand side of (17) diverges to ∞ . This is a contradiction. Hence there exists no such subsequence n_l . Thus $nF(\theta'_{k,n}) \rightarrow \infty$.

It remains to **prove (iii)**. For this purpose, fix k and V and let $\kappa > 0$ be arbitrary. By part (ii) of the proposition, $nF(\theta_{k,n}) \rightarrow \infty$. Thus, for n sufficiently large, $nF(\theta_{k,n}) \geq \kappa$. As

$$\sum_{i=k}^n \binom{n}{i} p^i (1 - p)^{n-i}$$

is increasing in p for $p \in [0, 1]$, we have for sufficiently large n ,

$$\begin{aligned} P_{k,n} &= \sum_{i=k}^n \binom{n}{i} F(\theta_{k,n})^i (1 - F(\theta_{k,n}))^{n-i} \geq \sum_{i=k}^n \binom{n}{i} \left(\frac{\kappa}{n}\right)^i \left(1 - \frac{\kappa}{n}\right)^{n-i} \\ &= 1 - \sum_{i=0}^{k-1} \binom{n}{i} \left(\frac{\kappa}{n}\right)^i \left(1 - \frac{\kappa}{n}\right)^{n-i}. \end{aligned} \quad (18)$$

We also know that for fixed i ,

$$\lim_{n \rightarrow \infty} \binom{n}{i} \frac{1}{n^i} = \lim_{n \rightarrow \infty} \frac{1}{i!} \prod_{j=0}^{i-1} \frac{n-j}{n} = \frac{1}{i!}.$$

Thus, for $n \rightarrow \infty$, (18) converges to

$$1 - \sum_{i=0}^{k-1} \frac{\kappa^i}{i!} e^{-\kappa}$$

which converges to 1 as $\kappa \rightarrow \infty$. □

Proof of Proposition 3. Throughout the proof, we drop the dependence of H on $(\alpha, \beta, \gamma, k, n)$, writing just $H(F)$. In the proof of Theorem 1, we saw that $H(F)$ is increasing. By assumption, we have $H(F(\theta)) \geq H(G(\theta))$ and thus

$$\frac{V}{\theta_{k,n}^F} = H(F(\theta_{k,n}^F)) \geq H(G(\theta_{k,n}^F)),$$

implying that $\theta_{k,n}^F$ lies to the left of the unique intersection of V/θ and $H(G(\theta))$. Accordingly, $\theta_{k,n}^F \leq \theta_{k,n}^G$. Moreover,

$$H(F(\theta_{k,n}^F)) = \frac{V}{\theta_{k,n}^F} \geq \frac{V}{\theta_{k,n}^G} = H(G(\theta_{k,n}^G))$$

implies $F(\theta_{k,n}^F) \geq G(\theta_{k,n}^G)$ by the monotonicity of H . The latter inequality implies the results for $N_{k,n}$ and $P_{k,n}$. □

Proof of Proposition 4. The proof extends the one of Proposition 3. Note that the functions $H(F(\theta))$ and $H(G(\theta))$ inherit the single-crossing property of F and G . Denote by $x_0 > 0$ a point of intersection of F and G and define $v_0 = x_0 H(F(x_0))$ such that v_0/θ intersects both $H(F(\theta))$ and $H(G(\theta))$ for

$\theta = x_0$. Consequently, for $V = v_0$ we have $\theta_{k,n}^F = \theta_{k,n}^G = x_0$. For $V > v_0$, it follows that both, $\theta_{k,n}^F$ and $\theta_{k,n}^G$ lie in $[x_0, \infty]$. For $\theta \in [x_0, \infty]$, we know by the single-crossing condition that $H(F(\theta)) \geq H(G(\theta))$. For $V > v_0$, it thus follows as in the proof of Proposition 3 that $\theta_{k,n}^F \leq \theta_{k,n}^G$ and $F(\theta_{k,n}^F) \geq G(\theta_{k,n}^G)$, implying the results about $N_{k,n}$ and $P_{k,n}$. The argument for $V < v_0$ is analogous, using that the ranking of F and G is reversed on $[0, x_0]$. \square

Proof of Proposition 5. Arguing as in the proof of Theorem 1, best responses to any non-trivial strategies of the other agents must be threshold strategies. This implies that equilibria must be in threshold strategies. Consider agent i of type x_i who faces opponents who play a threshold strategy with threshold $\theta_{k,n}$. In determining a symmetric equilibrium in threshold strategies, the main difference to the proof of Theorem 1 lies in the diffusive part of the expected cost from voting Yes. We now obtain

$$\begin{aligned} P(Y_{-i} \geq k-1) & \frac{\beta x_i}{E[1 + Y_{-i} | Y_{-i} \geq k-1]} \\ & = B(n-1, k-1, F(\theta_{k,n})) \frac{\beta x_i}{\sum_{j=k-1}^{n-1} \frac{b(n-1, j, F(\theta_{k,n}))}{B(n-1, k-1, F(\theta_{k,n}))} (1+j)}. \end{aligned}$$

For the expected costs of pivotality, we obtain

$$\gamma P(Y_{-i} \geq k-1) P(Y_{-i} = k-1 | Y_{-i} \geq k-1) = \gamma b(n-1, k-1, F(\theta_{k,n}))$$

Modifying the equilibrium condition (11) accordingly shows that equilibria are now characterized by

$$\frac{V}{\theta_{k,n}} = \tilde{H}(F(\tilde{\theta}_{k,n}), \alpha, \beta, \gamma, k, n) \quad (19)$$

with

$$\tilde{H}(F, \alpha, \beta, \gamma, k, n) = \alpha \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} + \gamma \quad (20)$$

$$+ \beta \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} \frac{1}{\sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{B(n-1, k-1, F)} (1+j)}. \quad (21)$$

The function $\tilde{H}(F, \alpha, \beta, \gamma, k, n)$ is continuous in F and satisfies

$$\tilde{H}(0, \alpha, \beta, \gamma, k, n) = \alpha + \frac{\beta}{k} + \gamma.$$

Moreover, it is bounded from below by $\alpha + \frac{\beta}{n} + \gamma > 0$. The last claim follows from (20) because of the estimates $b(n-1, k-1, F) \leq B(n-1, k-1, F)$ and $j+1 \leq n$. As V/θ is arbitrarily large for small θ and arbitrarily small for large θ , it follows that (19) possesses at least one solution $\tilde{\theta}_{k,n} > 0$. To complete the proof of the proposition, we need to show that any solution $\tilde{\theta}_{k,n}$ satisfies $\tilde{\theta}_{k,n} \geq \theta_{k,n}$. To prove this, we show that $\tilde{H}(F, \alpha, \beta, \gamma, k, n) \leq H(F, \alpha, \beta, \gamma, k, n)$. The result then follows from the monotonicity of V/θ in θ . To compare \tilde{H} and H , it suffices to compare the β -dependent diffusive terms in (5) and (20). Applying Jensen's inequality with the convex function $h(x) = 1/x$ and the probability distribution with probabilities $b(n-1, j, F)/B(n-1, k-1, F)$ on k, \dots, n , we find that

$$\begin{aligned} & \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} \frac{1}{\sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{B(n-1, k-1, F)} (1+j)} \\ & \leq \frac{B(n-1, k-1, F)}{b(n-1, k-1, F)} \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{B(n-1, k-1, F)} \frac{1}{1+j} \\ & = \sum_{j=k-1}^{n-1} \frac{b(n-1, j, F)}{b(n-1, k-1, F)} \frac{1}{1+j} \end{aligned}$$

and thus $\tilde{H}(F, \alpha, \beta, \gamma, k, n) \leq H(F, \alpha, \beta, \gamma, k, n)$. □

Proof of Theorem 2. We begin with claim (i). To simplify the notation, we define functions $h_1(\theta)$ and $h_2(\theta)$ by

$$h_1(\theta) = \frac{B(n-1, k-1, F(\theta))}{b(n-1, k-1, F(\theta))} \quad \text{and} \quad h_2(\theta) = \frac{B(n, k, F(\theta))}{b(n, k, F(\theta))},$$

and notice that due to (14), the following identity holds:

$$H(F(\theta), \alpha, \beta, \gamma, k, n) = \alpha h_1(\theta) + \frac{\beta}{k} h_2(\theta) + \gamma. \quad (22)$$

The thresholds θ_i , $i = 1, 2$ defined in the proposition solve $L_i(\theta_i) = 0$ where

the strictly increasing functions L_i are given by

$$L_i(\theta) = \alpha_0 h_i(\theta) + \gamma - \frac{V}{\theta}.$$

It follows that for the strictly increasing function

$$L_w(\theta) = wL_1(\theta) + (1-w)L_2(\theta), \quad w \in (0, 1),$$

there exists a unique θ_w with the properties that $L_w(\theta_w) = 0$ and

$$\min(\theta_1, \theta_2) < \theta_w < \max(\theta_1, \theta_2),$$

i.e. the unique zero of a convex combination of the strictly increasing functions h_1 and h_2 lies between the unique zeros of these functions. It thus remains to show that $\theta_w = \theta_{k,n}(w\alpha_0, k(1-w)\alpha_0, \gamma, V)$. This follows from the observation that by (22),

$$\begin{aligned} L_w(\theta) &= wL_1(\theta) + (1-w)L_2(\theta) \\ &= w\alpha_0 h_1(\theta) + \frac{k(1-w)\alpha_0}{k} h_2(\theta) + \gamma - \frac{V}{\theta} \\ &= H(w\alpha_0, k(1-w)\alpha_0, \gamma, V, k, n) - \frac{V}{\theta}. \end{aligned}$$

To see (ii), observe that it follows from (22) that

$$H(F(\theta), 0, \delta, 0, k, n) = k \cdot H(F(\theta), \delta, 0, 0, k+1, n+1)$$

which immediately leads to the result. We finally prove (iii). Define the short notation $\theta = \theta_{k,n}(\alpha, \beta, \gamma, V)$ and $\theta' = \theta_{k,n}(\alpha, 0, \gamma, V)$. By Proposition 1, we have $\theta \leq \theta'$. Note further that

$$H(F, \alpha, \beta, \gamma, k, n) \leq H(F, \alpha + \beta/k, 0, \gamma, k, n). \quad (23)$$

Hence,

$$\begin{aligned}
\frac{V}{\theta} &= H(F(\theta), \alpha, \beta, \gamma, k, n) \\
&\leq H(F(\theta), \alpha + \beta/k, 0, \gamma, k, n) \\
&\leq \left(1 + \frac{\beta}{k\alpha}\right) H(F(\theta), \alpha, 0, \gamma, k, n) \\
&\leq \left(1 + \frac{\beta}{k\alpha}\right) H(F(\theta'), \alpha, 0, \gamma, k, n) \\
&\leq \left(1 + \frac{\beta}{k\alpha}\right) \frac{V}{\theta'}
\end{aligned}$$

which yields $\theta \leq \theta' \leq \left(1 + \frac{\beta}{k\alpha}\right) \theta$ as required. □

Proof of Proposition 6. Throughout the proof, we do not use the exact functional form of how guilt depends on the total number y of Yes votes. Instead we assume that shared guilt takes the form of $x/s(y)$ where x denotes an agent's type and s is an increasing, positive function in y . This shortens the notation and illustrates that the result holds also in this more general setting. In our case, we have

$$\frac{1}{s(y)} = \alpha + \frac{\beta}{y} \Leftrightarrow s(y) = \frac{1}{\alpha + \frac{\beta}{y}}.$$

Let us prove the first assertion. We look for $\theta > 0$ where the expected payoff of voting No is equal to the expected payoff of voting Yes, i.e. we look for solutions $\theta > 0$ of the equation

$$V \sum_{j=k}^{n-1} b(n-1, j, F(\theta)) = V \sum_{j=k-1}^{n-1} b(n-1, j, F(\theta)) - \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)), \quad (24)$$

i.e.

$$Vb(n-1, k-1, F(\theta)) = \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)). \quad (25)$$

The right hand side is nonnegative, 0 for $\theta = 0$, and diverges to ∞ for $\theta \rightarrow \infty$. The left hand side is bounded, continuous in θ and positive for $\theta > 0$. This implies that for large V there exists at least one nontrivial solution $\theta > 0$ to this equation. A solution θ of equation (24) is the threshold of a symmetric

Bayesian equilibrium in which each agent i votes Yes if $x_i \leq \theta$ and No if $x_i > \theta$ as the expected utility of voting Yes decreases in x_i .

Now let us turn to the second assertion. Analogously as in Proposition 1, all symmetric Bayesian equilibria must be in threshold strategies. The case in which the threshold is ∞ is trivial. Let us look again at (25). We want to show that this equation has no solution $\theta > 0$ for small V . For large θ ,

$$Vb(n-1, k-1, F(\theta)) < \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)).$$

Thus, by continuity in θ , it suffices to show that for small $\theta > 0$ and small V ,

$$Vb(n-1, k-1, F(\theta)) < \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)).$$

First note that

$$\frac{\theta}{s(1)}(1-F(\theta))^{n-1} < \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)) \quad (26)$$

with

$$Vb(n-1, k-1, F(\theta)) = V \binom{n-1}{k-1} F(\theta)^{k-1} (1-F(\theta))^{n-k} \leq V \binom{n-1}{k-1} F(\theta) \quad (27)$$

as $k > 1$. Now choose V sufficiently small such that $V \binom{n-1}{k-1} F'(0) < 1/s(1)$ which implies that

$$V \binom{n-1}{k-1} F(\theta) < \frac{\theta(1-F(\theta))^{n-1}}{s(1)}$$

for small θ as $F(\theta)/\theta \rightarrow F'(0)$ and $F(\theta) \rightarrow 0$ for $\theta \rightarrow 0$. This proves the claim.

Now let us turn to the third assertion. Similarly as in (26) and (27), one can show that

$$Vb(n-1, k-1, F(\theta)) < \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta))$$

for a fixed (arbitrary large) V if $\theta > 0$ is small enough. Hence, by continuity and positivity of both terms for $\theta > 0$, for large V there are at least two positive solutions to (25) with at least one solution satisfying $F(\theta) < 1$. Thus, the two solutions correspond to two distinct threshold strategies.

Note that

$$\sum_{j=0}^{n-1} \frac{1}{s(1+j)} b(n-1, j, F(\theta)) \quad (28)$$

is weakly decreasing in θ . By linearity, and as the function $s(1+j)$ is weakly increasing, this claim can be reduced to the well-known fact that

$$\sum_{j=0}^m b(n-1, j, F(\theta))$$

is weakly decreasing in θ for $m \leq n-1$.

Now let us denote by $\bar{\theta}_{k,n}^{(1)} < \bar{\theta}_{k,n}^{(2)}$ two solutions of (24) which correspond to two different threshold strategies, i.e. $F(\bar{\theta}_{k,n}^{(1)}) < F(\bar{\theta}_{k,n}^{(2)})$. We show that the equilibrium associated with the second threshold strategy strictly Pareto dominates the first. This argument also covers the trivial equilibrium $\bar{\theta}_{k,n}^{(1)} = 0$.

Consider an agent of type $x \leq \bar{\theta}_{k,n}^{(1)}$. In both equilibria, voting Yes leads to the the expected payoffs

$$V \sum_{j=k-1}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(i)})) - \sum_{j=0}^{n-1} \frac{x}{s(1+j)} b(n-1, j, F(\bar{\theta}_{k,n}^{(i)}))$$

for $i \in \{1, 2\}$.

The expected payoff is larger in the second equilibrium as the first term is increasing whereas the second term is weakly decreasing, see (28). The case $\bar{\theta}_{k,n}^{(2)} < x$ can be treated similarly. In the case $\bar{\theta}_{k,n}^{(2)} < x \leq \bar{\theta}_{k,n}^{(1)}$, we have to compare the expected payoffs

$$V \sum_{j=k}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(1)}))$$

and

$$V \sum_{j=k-1}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) - \sum_{j=0}^{n-1} \frac{x}{s(1+j)} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})).$$

We use $F(\bar{\theta}_{k,n}^{(1)}) < F(\bar{\theta}_{k,n}^{(2)})$ to show that

$$\begin{aligned} & V \sum_{j=k-1}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) - \sum_{j=0}^{n-1} \frac{x}{s(1+j)} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) \\ & \geq V \sum_{j=k-1}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) - \sum_{j=0}^{n-1} \frac{\bar{\theta}_{k,n}^{(2)}}{s(1+j)} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) \\ & = V \sum_{j=k}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(2)})) \\ & > V \sum_{j=k}^{n-1} b(n-1, j, F(\bar{\theta}_{k,n}^{(1)})) \end{aligned}$$

which proves the claim. □

References

- [1] Admati, Anat, and Motty Perry. Joint projects without commitment. *Review of Economic Studies*, 58, 259-276, 1991.
- [2] Albrecht, Konstanze, Krämer, Florentin, Szech, Nora. Ethics and Animal Welfare: A Personality Study. Working Paper, 2016.
- [3] Bandura, Albert. Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193-209, 1999.
- [4] Bandura, Albert. *Moral Disengagement: How People do Harm and Live with Themselves*. Worth Publishers, New York, 2016.
- [5] Bandura, Albert, Bill Underwood, and Michael E. Fromson. Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality*, 9, 253-269, 1975.
- [6] Barnes, Jonathan (Ed.) *The Complete Works of Aristotle: The Revised Oxford Translation*. Princeton University Press, Princeton, 1984.

- [7] Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. The Many Faces of Human Sociality - Uncovering the Distribution and Stability of Social Preferences. Working Paper, 2016.
- [8] Crettez, Bertrand, and Régis Deloche. On the optimality of a duty-to-rescue rule and the cost of wrongful intervention. *International Review of Law and Economics* 31 (4), 263-271, 2011.
- [9] Darley, John, and Bibb Latané. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383, 1968.
- [10] Deckers, Thomas, Armin Falk, Fabian Kosse, and Nora Szech. Homo Moralis: Personal Characteristics, Institutions, and Moral Decision-Making. Working Paper, 2016.
- [11] Diamond, Peter, and Joseph Stiglitz. Increases in risk and in risk aversion. *Journal of Economic Theory*, 8, 337-360, 1974.
- [12] Dufwenberg, Martin, and Amrish Patel. Reciprocity networks and the participation problem. *Games and Economic Behavior*, forthcoming, 2015.
- [13] Falk, Armin, and Nora Szech. Morals and Markets. *Science*, 340 (6133), 707-711, 2013.
- [14] Falk, Armin, and Nora Szech. Organizations, Diffused Pivotality and Immoral Outcomes. Working Paper, 2014.
- [15] Fischer, Peter, Joachim Krueger, Tobias Greitemeyer, Claudia Vogrinic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalene Wicher, and Martina Kainbacher. The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4): 517-537, 2011.
- [16] Ganuza, Juan-José and José Penalva. Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions. *Econometrica*, 78: 1007-1030, 2010.
- [17] Grossman, Dave. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Little Brown, Toronto, 1996

- [18] Haidt, Jonathan, and Selin Kesebir. Morality. Handbook of Social Psychology, 5th Edition. Chapter 22. Fiske, S., Gilbert, D., Lindzey, G. (Eds.), Wiley, New York, 2010.
- [19] Harrington, Joseph. A simple game-theoretic explanation for the relationship between group size and helping. Journal of Mathematical Psychology, 45, 389-392, 2001.
- [20] Harsanyi, John C. Rule utilitarianism and decision theory. Erkenntnis, 11, 25-53, 1977.
- [21] Huck, Steffen, and Kai Konrad. Moral cost, commitment and committee size. Journal of Institutional and Theoretical Economics, 161(4), 575-588, 2005.
- [22] Huesmann, Katharina, and Achim Wambach. Constraints on Matching Markets Based on Moral Concerns. Working Paper, 2015.
- [23] Johnson, Norman, Samuel Kotz, and Adrienne Kemp. Univariate discrete distributions. 2nd edition, Wiley, New York, 1992.
- [24] Kranz, Sebastian. Moral norms in a partly compliant society. Games and Economic Behavior, 68, 255-274, 2010.
- [25] Latané, Bibb, and Steve Nida. Ten years of research on group size and helping. Psychological Bulletin, 89, 308-324, 1981.
- [26] Lindbeck, Assar, Sten Nyberg, and Jörgen Weibull. Social norms and economic incentives in the welfare state. Quarterly Journal of Economics, 114, 1-35, 1999.
- [27] Loe, Terry, Linda Ferrell, and Phylis Mansfield. A Review of Empirical Studies Assessing Ethical Decision Making in Business. Journal of Business Ethics, 25, 185-204, 2000.
- [28] Nitzan, Shmuel, and Richard Romano. Private provision of a discrete public good with uncertain cost. Journal of Public Economics, 42, 357-370, 1990.

- [29] O'Fallon, Michael, and Kenneth Butterfield. A Review of the Empirical Ethical Decision-Making Literature: 1996-2003. *Journal of Business Ethics*, 59(4), 375-413, 2005.
- [30] Osborne, Martin. *An introduction to game theory*. Oxford University Press, Oxford, 2004.
- [31] Palfrey, Thomas, and Howard Rosenthal. Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics*, 24, 171-193, 1984.
- [32] Rothenhäusler, Dominik, Nikolaus Schweizer, and Nora Szech. *Institutions, Shared Guilt, and Moral Transgression*. Working Paper, 2015.
- [33] Sobel, Joel. *Markets and Other-Regarding Preferences*. Working Paper, 2010.
- [34] Spranca, Mark, Elisa Minsk, and Jonathan Baron. Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76-105, 1991.
- [35] Thaler, Richard, and Cass Sunstein. *Nudge. Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, 2008.
- [36] Weesie, Jeroen, and Axel Franzen. Cost sharing in a volunteer's dilemma. *Journal of Conflict Resolution*, 42, 600-618, 1998.
- [37] Zimbardo, Philip. *The Lucifer effect: Understanding how good people turn evil*. Random House, New York, 2007.