

# Institutions, Shared Guilt, and Moral Transgression\*

Dominik Rothenhäusler, Nikolaus Schweizer and Nora Szech

December 2015

## Abstract

Many tasks can only be completed if several people contribute. Likewise, many institutions, e.g. voting rules, require the support of several people to implement specific decisions. In such situations, individual costs from supporting may decrease in the number of supporters. This holds true for physical costs in division of labor. Psychological, sociological and economic evidence suggests that also moral costs due to feelings of guilt or having a bad conscience diffuse. This paper analyzes the influence of diffusion of costs in threshold public good games under incomplete information, and compares predictions to those when costs do not diffuse.

**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.

---

\*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, Mercator School of Management, University of Duisburg-Essen, Lotharstr. 65, D-47057 Duisburg, email: nikolaus.schweizer@uni-due.de, phone: +49 203 3792234. Nora Szech, Department of Economics, Karlsruhe Institute of Technology, Fritz-Erler-Str. 1-3, D-76128 Karlsruhe, email: nora.szech@kit.edu, phone: +49 721 60843809.

# 1 Introduction

This paper studies the critical role of institutions in affecting moral behavior. Specifically, we aim at understanding how several people acting together may promote or prevent moral transgression, depending on how easily feelings of responsibility and guilt diffuse. Many immoral acts require the support of several people to become implementable. On the one hand, this may put a natural barrier on moral transgression: If not just the “worst” people in a population, but also a considerable fraction of “ordinary” people have to participate, transgression may lack support and hence be prevented. On the other hand, for the individual, knowing that he or she is just one out of many people, may facilitate turning evil: Acting together with others allows for sharing guilt, for feeling less responsible for immoral outcomes, and for lowering the impression of breaking a social norm (compare Latané and Nida (1981)). Thus, if many have to be supportive for transgression to happen, transgression may occur exactly *because* of this necessity. It is this tension our paper aims to explore. Depending on how many people need to participate by institutional design, moral transgression may occur or be prevented. Likewise, situations of cost diffusion arise in division of labor. Our analysis applies to these situations as well, and compares predictions to those when costs do not diffuse.

Morally difficult tasks are often delegated to several people instead of one individual alone. An example are executions by shooting, where several people act together in a shooting squad. Similarly, death penalties are often executed by so-called execution teams who inject lethal doses of toxic drugs together. In his book “On Killing”, Lieutenant Colonel Dave Grossman (1996) points out that individual barriers towards morally problematic activity often break when people become part of teams or groups: “The individual is not a killer, but the group is” (p. 149).

Additional evidence that the diffusion of responsibility is an important catalyst for reducing moral behavior (Bandura 1999) comes from social psychology. Bandura et al. (1975) show how the diffusion of guilt lowers moral concerns in a (staged) scenario. Subjects can give electric shocks of various intensities to other participants. If subjects are individually responsible, the shock intensities are much lower than in a diffused-responsibility set-up. In the latter, the shock intensity a subject chooses is averaged with the shock-intensities cho-

sen by two other subjects. Subjects find it considerably easier to administer shocks of high intensity in the latter set-up. Likewise, in his book “The Lucifer Effect” (2007), social psychologist Philip Zimbardo analyzes how good people turn evil. He classifies the diffusion of responsibility as an important driver of moral transgression and suggests to make people individually responsible in order to prevent immoral activity (compare Ch. 12 and 16).

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. With regard to the diffusion of responsibility, the amount of supporters of an evil act may be crucial for moral versus immoral behavior. Likewise, depending on how easily guilt diffuses, organizations may choose different set-ups depending on the outcome they would like to establish. Our study aims to shed light on such scenarios. Even for the most drastic atrocity, some historians stress the role of institutional design facilitating moral transgression (compare Browning (2004) on the organization of the Holocaust). We focus on the role of diffusion of guilt and responsibility, and how institutional design affects moral behavior depending on how easily guilt diffuses.

We analyze two measures of immorality: Firstly, moral transgression can happen on an individual level. If a person would be willing to support transgression in one institution, but not in the other, institutions affect this person’s moral disposition. Secondly, institutions affect outcomes. If transgression requires the collaboration of several people, feelings of guilt are reduced in those who collaborate. Yet for the transgression to materialize, enough supporters have to share in. Therefore, we also study the likelihood of transgression to materialize as a second measure of immorality within a population. This second measure is focused on social rather than direct effects of institutions promoting moral transgression.

Our game-theoretic model can be understood as an extension of a discrete public goods game as in Palfrey and Rosenthal (1984). In our model, a group of agents faces the opportunity to commit a morally problematic act (this is, ironically, the “public good” in our case). Agents would enjoy the outcome of the act, focusing on some naïvely selfish benefit. Yet committing the act

is associated with feelings of guilt for those who supported it actively. If a sufficiently large group of agents acts as supporters, moral transgression happens – and supporters experience guilt.<sup>1</sup> We assume that the associated individual feelings of guilt are decreasing in the number of supporters.

We analyze the following factors influencing immoral outcomes: Firstly, we vary the degree to which guilt diffuses. Secondly, we study the impact of the number of supporters necessary for moral transgression to happen. Thirdly, we analyze the effect of the overall population size. Varying the number of required supporters could be interpreted as changing the decision rule in an organization, e.g., from unanimity voting to a simple majority vote, or to looking for a volunteer. Varying group size can be interpreted as shifting responsibility for a particular decision upwards or downwards in the hierarchy of a larger organization. Likewise, our results can be understood as a comparison of existing institutions.

One might argue that in real-life organizations, different mechanisms can simultaneously influence moral behavior. Besides the diffusion of guilt, one may think about authority, status concerns, reward systems, or the attribution of guilt to victims. Our approach to understand the effects of institutions on moral behavior better is to focus on one important mechanism. Further research is needed to explore the effects of other mechanisms – and their interplay – and how they affect moral behavior. Ultimately, we hope that the identification and detailed analysis of facets of organizational design will lead to a fundamental understanding of moral decision making and how institutions affect it.

## Main Results

We find that if transgression requires *unanimity* within the population, the results are clear-cut for sufficiently large groups: The probability of a transgression occurring approaches either zero or one. It depends on the distribution of individual levels of guilt within the population and on how easily guilt diffuses among many. While, individually, people find it easier to participate

---

<sup>1</sup>In the terminology coined by Admati and Perry (1991), the game we study is a “subscription game” since costs only accrue if there are sufficient contributions to the “public good”.

in transgression within a large population, transgression may be completely prevented if there are enough people having substantial moral concerns.

The probability for transgression to happen is *U*-shaped in the population size for many common distributions of individual levels of guilt. The intuition is as follows. The initial decrease is driven by the same effect which drives the transgression probability to zero in a model without sharing of guilt: In a larger group, more have to agree and thus the probability that someone will object is high. As the group gets even larger, this effect is overruled by another one. If very many have to agree guilt is shared so effectively that it becomes negligible.

We then move away from unanimity and study the effects of varying the number of required supporters on individual and outcome-based immorality. Individual willingness to support transgression increases monotonically in the number of required supporters since guilt can be shared. Yet the probability that transgression materializes first decreases in the number of required supporters, but increases in this number if the overall number of required supporters is large. Thus transgression can best be prevented if an intermediate number of supporters is necessary. In consequence, simple majority voting tends to be more successful at preventing transgression than both unanimity voting and looking for a volunteer.

Furthermore, we study effects of population size if the number of supporters required is fixed. We observe an effect of “immoral overshooting”: As the population grows large, many people find it easy to support transgression even though their support is not required at all.

Besides shared guilt, one may think of other, closely related factors lowering levels of guilt in people. For example, guilt may be larger if people know that their support was pivotal in the sense that if they had opted out, the moral transgression would not have materialized. In an extension, we therefore study such aspects of pivotality in isolation and in combination with effects of shared guilt. Our techniques and most of our results extend to both settings.

In the main part of the paper, we assume that guilt is driven by outcomes: People do not feel guilty if they supported an immoral action that never materialized. This is in line with the basic concept of guilt in western law systems,

and it corresponds to a utilitarian view of morality. Motivated by a rule-based, e.g. Kantian, moral conception, we also study an extension in which guilt is driven by supporting immoral actions no matter whether the transgression realizes or not. In the latter model, we observe the following transition: An equilibrium in which moral transgression occurs with a positive probability exists only if selfish “benefits” from transgression are sufficiently large. In particular, there typically exist several equilibria of this type, and the one with the largest transgression probability Pareto-dominates all other equilibria.

Finally, we show that there is a non-trivial strategic equivalence between our model and a more standard public goods model without any diffusion of guilt. Hence, our results inform us about standard public goods games with heterogeneous, private voting costs as well, and generate new insights for these models, too.

## Related Literature

We add to the literature in several ways: Previous applications of discrete public good games to morals and collective action focus mostly on the bystander effect.<sup>2</sup> Our model differs from this line of research in that the “public good” is an immoral outcome and not a moral one. Accordingly, we incorporate the psychological costs of guilt instead of the physical costs of doing a good deed.

Our work is more closely related to Huck and Konrad (2005) who also consider a trade-off between the costs and benefits of a moral transgression. The key difference between our model and theirs is the inclusion of diffusion of costs/guilt as the number of supporters increases. This assumption is very much in line with findings from social psychology<sup>3</sup> and leads to significant changes in the model’s qualitative behavior. Moreover, while Huck and Konrad mostly focus on varying the population size (or, equivalently, the size of the deciding committee), our analysis considers a broader set of questions. For instance,

---

<sup>2</sup>The bystander effect describes the phenomenon that helping a person in distress becomes less likely the more people are around and could in principle help the victim, see e.g. Latané and Nida (1981). The game-theoretic literature on the bystander effect includes Harrington (2001), Osborne (2004) and Crettez and Deloche (2011), see the last paper for more references.

<sup>3</sup>Compare Bandura (1999), Zimbardo (2007), Fischer et al. (2011), and the references therein.

varying the size of required supporters has – to our knowledge – only been studied in Palfrey and Rosenthal (1984) while most of the literature considers the more tractable cases of one required supporter or unanimity.<sup>4</sup> In addition, our model differs by assuming incomplete information over heterogeneous costs instead of the symmetric complete information case considered by Palfrey and Rosenthal and the subsequent literature cited above.<sup>5</sup>

In a recent paper, Dufwenberg and Patel (2015) investigate the effects of cost-reducing division of labor due to reciprocity. They study a public goods context in which in a network agents may want to reciprocate, i.e. share in, if other agents do so as well. While they focus on the complete information case and coordination equilibria in pure strategies, our analysis is related to the study of symmetric mixed strategies in this context, and thus to the corresponding incomplete information case.

On a technical level, our model is also related to models of strategic voting and committee decision making.<sup>6</sup> What sets our model apart from this literature are the costs of voting, as in our study, individual costs of an immoral “vote” decrease in the number of supporters. Moreover, our study addresses a different topic: While that literature centers around questions of efficiency and information aggregation, our main interest lies in characterizing conditions that promote or limit immoral dispositions and outcomes.

In some contributions to the literature on social norms, (e.g. Lindbeck, Nyberg and Weibull (1999)) agents’ preferences are also assumed to depend on how their actions differ from those of others. The focus is on large (continuous) populations. In contrast, our main interest is on how institutions affect decisions in medium-sized groups in which every single agent may have some impact on outcomes. Yet we also study results for the population growing large and compare outcomes for small versus large populations.

---

<sup>4</sup>See, for instance, the literature on the volunteer’s dilemma, e.g. Weesie and Franzen (1998), who also consider cost sharing but focus on the case of one supporter.

<sup>5</sup>There is, of course, also a literature on public good games with incomplete information, e.g., Nitzan and Romano (1990), but we are not aware of contributions which address questions similar to those we study, such as comparative statics with regard to levels of guilt diffusion and decision mechanisms.

<sup>6</sup>For seminal contributions to that literature, see, e.g., Austen-Smith and Banks (1996) or Feddersen and Pesendorfer (1997). Li and Suen (2009) provide a recent survey.

A word is in order on what motivates our conception of guilt caused by agreeing to an immoral activity and how this conception of guilt differs from other conceptions in the economics literature: We study a context in which it is not debatable that the immoral act violates basic ethical standards. Participants decide about supporting moral transgression at a cost of bad conscience. Such moral transgression may for example refer to harming (or refusing to help) a third party. Motivated by results from social psychologists such as Bandura, Latané and Nida, or Zimbardo, we assume that it is salient that such harm and damage is immoral, and focus on the mere impacts of shared guilt. In other words, it is not determined by the expectations of the victim what behavior is morally right or wrong. Yet motivated by findings on the diffusion of guilt, we assume that when several people commit a moral transgression together, people have less difficulty in agreeing to harm (or not help) a victim.

A literature in economics has considered feelings of guilt that are caused by shortcoming the expectations of other individuals. For example, in the dictator game, if the game partner expects to receive a certain amount, behaving selfishly causes feelings of guilt due to not fulfilling the partner's expectations.<sup>7</sup> This definition of guilt is motivated by psychological research looking at a social relationship to a partner (compare, e.g., Baumeister, Stillwell and Heatherton, 1994), but not to a third party (such as a victim). For seminal work on incorporating the fulfillment of other's beliefs into a social game, see Geanakoplos, Pearce, and Stachetti, 1989. Such work gives a micro basis for the assumption that guilt diffuses within a group of activists. Studies such as Bénabou and Tirole (2011) take this microscopic approach one step further, focusing on a single agent's internal conflict - including guilt. Our approach is more mesoscopic in nature: We take diffusion of guilt as a given and study its consequences on the decisions of medium-size groups, i.e., groups which are small enough to still let the individual count.

In a recent study, Falk and Szech (2013a) find that deciding individually about harming a third party prevents immoral behavior compared to two people deciding together. People decide between saving the life of a mouse and foregoing some monetary amount versus agreeing to kill a mouse and receiving money. In the bilateral market situation, people find it much easier not to save the

---

<sup>7</sup>Compare, e.g., Battigalli and Dufwenberg (2007), Battigalli, Charness, and Dufwenberg (2013) and the references therein



life of a mouse than in the individual decision context. Knowing that another person has to support the killing, people possibly care less about the mouse life and opt for the money, too. This happens even though people know they remain fully pivotal for the death of the mouse.<sup>8</sup>

Sobel (2010) addresses the effects of diffused pivotality in markets in a theoretical study. Market participants act less pro-socially, knowing that their effect is small in a group of many traders. In an experimental study, Falk and Szech (2013b) show that diffusion of pivotality can indeed drive people to support immoral acts. Other mechanisms leading to “unfair” or less social outcomes are delegation or moral “wriggle rooms”. These are discussed, e.g., in Bartling and Fischbacher (2012), Hamman, Loewenstein and Weber (2010) and Dana, Weber and Kuang (2007).

## Outline

The paper is organized as follows: In Section 2, we introduce our model, prove existence and uniqueness of equilibrium, and define the individual and the outcome-based measures of morality used throughout. Section 3 focuses on effects of different levels of diffusion of guilt. In Section 4, we analyze the effects of institutions on individual moral thresholds and aggregate moral outcomes. Section 4.1 focuses on effects of the number of supporters necessary for transgression to happen, while Section 4.2 analyzes the effects of population size. Section 5 discusses extensions: direct costs of pivotality, heterogeneous benefits from the transgression and outcome-independent guilt. Section 6 compares our model to a standard model without sharing of guilt. Section 7 concludes.

## 2 Framework

This section introduces our model. We show existence of a single (non-trivial) symmetric equilibrium. Furthermore, we introduce our measures of morality: Individual thresholds  $\theta$  characterize at which personal level of moral concerns

---

<sup>8</sup>Another difference between the individual and the bilateral context was that the bilateral context used a market frame. People negotiated in the roles of “seller” and “buyer”. While the study cannot disentangle what drove immoral activity exactly, it provides a strong indication that sharing guilt lowers moral behavior.

people get tempted by institutions to agree to transgression. The transgression probability  $P$ , instead, focuses on outcomes. It describes the likelihood with which moral transgression materializes within a population.

## 2.1 The Model

We consider a group of  $n \geq 1$  agents who face the decision to take an action. We assume that taking the action is morally difficult. We thus also refer to it as the *moral transgression*. If transgression happens, each agent receives  $V > 0$ , otherwise each agent receives 0. Transgression is associated with costs of feeling guilty afterwards, or with some general “moral costs”, which are modeled as follows: Each agent has a private 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  $F$ . If agent  $i$  supported the transgression and the transgression indeed takes place, his or her moral costs are  $x_i$  divided by  $s(y)$  where  $y$  is the number of agents who supported the transgression. The division by  $s$  captures sharing, or diffusion, of guilt. For example,  $s(y) \equiv 1$  corresponds to a standard public goods game without sharing of guilt.  $s(y) = y$  captures a *proportional* diffusion of guilt. We assume that  $s$  is weakly increasing and that  $s(1) = 1$ . The latter assumption implies that the benchmark costs of being the only transgressor are the same under all sharing functions we consider. If transgression does not materialize, we assume that agents do not feel guilty.<sup>9</sup> Agents who did not support the transgression do not feel guilty either.

The collective decision process is modeled as a voting-type 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.

To sum up, the realized utility of agent  $i$  from supporting immoral activity

---

<sup>9</sup>See Sections 5 and 6 for extensions.

(opting for “Yes”) is given by

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

where  $Y_{-i}$  denotes the number of agents other than  $i$  who collaborated as well. Realized utility from opting against transgression (choosing “No”) is given by

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

## 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 where all agents choose “No” regardless of their type and where all agents have a utility of 0. 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 where the immoral action is taken with positive probability.<sup>10</sup> Our first main result shows that there always exists exactly one equilibrium of this type.

**Proposition 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*

$$V b(n-1, k-1, F(\theta_{k,n})) = \sum_{j=k-1}^{n-1} \frac{\theta_{k,n}}{s(1+j)} b(n-1, j, F(\theta_{k,n})), \quad (3)$$

where

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

If  $k < n$ ,  $\theta_{k,n}$  lies in the interior of the support of  $F$ ,  $F(\theta_{k,n}) \in (0, 1)$  and thus the equilibrium is a separating equilibrium.<sup>11</sup>

---

<sup>10</sup>Since agents can guarantee themselves a non-negative payoff by choosing “No”, all agents earn a non-negative payoff in any equilibrium. Any agent strictly preferring “Yes” in an equilibrium where the immoral action is taken with positive probability must earn a strictly positive expected payoff.

<sup>11</sup>We refer to equilibria in which agents’ actions depend on their type as “separating

Equation (3) has a straightforward interpretation in terms of the costs and benefits of a marginal agent: On the left hand side we have the gains in utility if the agent is indeed marginal, i.e., if exactly  $k - 1$  other agents are willing to take action. On the right hand side we find the expected costs of an agent with type  $\theta_{k,n}$ .

Obviously, uniqueness only holds up to the decision of agents with type  $\theta_{k,n}$  which are indifferent between “Yes” and “No”. We ignore this technicality in the following, since it only concerns zero-probability events. For  $k = n$ , (3) becomes  $\theta_{n,n} = s(n)V$ . Thus, for  $F$  with finite support, the equilibrium degenerates to a pooling equilibrium where all agents opt for “Yes” if  $n$  or  $V$  are sufficiently large.

### 2.3 Measures of Moral Transgression

We focus on two measures of moral transgression:  $\theta_{k,n}$  captures individual levels of immorality.  $P_{k,n}$  refers to the likelihood of immoral outcomes within the population.

Immorality can increase on an individual level, in the sense that an individual with given moral standards may get tempted by possibilities to share guilt via institutional design. We measure this institutional effect on the individual level by the quantity  $\theta_{k,n}$ , i.e., by the marginal type who supports the transgression. For an individual decision maker, we have  $\theta_{1,1} = V$ , i.e., the decision maker agrees whenever the benefits  $V$  are larger than his or her individual level of guilt  $x$ . In general,  $\theta_{k,n}$  may be both, smaller or larger than  $V$ . In some cases, agents who would have supported the transgression individually, opt against it in a different institution, hoping that others will do the dirty work for them. Thus, a free-riding-type effect may dominate. In other cases, agents who would not have supported the transgression individually may opt in its favor in a different institutional context, tempted by effects of shared guilt.

In addition, as an outcome-based measure of morality, we consider the equilibria. Strictly speaking, they are only “semi-separating” since we cannot infer an agent’s type from his action with infinitely many types and only two actions.

librium probability  $P_{k,n}$  with which transgression occurs within a population

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

For an individual decision maker, we have  $P_{1,1} = F(V)$ , i.e., the transgression probability equals the proportion of *individual* supporters in the population. As we will see in the following, the dependence of  $P_{k,n}$  on  $k$  and  $n$  is rather intricate. For instance,  $P_{n,n}$ , the transgression probability under unanimous decisions converges either to 0 or to 1, depending on the interplay of the distribution  $F$  and the sharing rule  $s$ .

### 3 Diffusion of Guilt

In this section, we study transgression probabilities in decisions made under unanimity. This allows to focus on the interplay between the distribution of moral standards within a population and the degree to which guilt can be shared. We show that except for knife-edge cases the transgression probability  $P_{n,n}$  converges either to 0 or 1 and give explicit conditions for both cases. We also show that for several common distribution functions,  $P_{n,n}$  is *U-shaped* in  $n$ .

We assume that for moral transgression to materialize, it has to be unanimously supported among the agents, i.e.,  $k = n$ . Hence agents have no incentive to free-ride on each other. We assume that the sharing function is of the form  $s(y) = y^\alpha$  with  $\alpha > 0$ .  $\alpha$  captures the degree to which guilt can be shared. We focus on the interplay between diffusion of guilt  $\alpha$ , population size  $n$  and distribution of moral concerns  $F$ .

If guilt diffuses easily, i.e., if  $\alpha$  is large, agents can more easily bear the burden of agreeing to the transgression. Accordingly, both of our measures of immorality increase in  $\alpha$ .

**Lemma 1.**

$$\theta_{n,n} = n^\alpha V \quad \text{and} \quad P_{n,n} = F(n^\alpha V)^n \tag{4}$$

are increasing in  $\alpha$ .

From the lemma, we see that  $\theta_{n,n}$  increases also in  $n$ . The dependence of  $P_{n,n}$  on  $n$  is less obvious in general. We begin with some simple observations:

**Corollary 1.**

- (i) If  $\alpha = 0$  and  $F(V) < 1$ ,  $P_{n,n}$  converges exponentially quickly to 0 in  $n$ .
- (ii) If  $\alpha > 0$  and  $F$  has finite support,  $P_{n,n} = 1$  for sufficiently large  $n$ .

Thus, for  $\alpha = 0$ , i.e. without sharing of guilt,  $P_{n,n}$  vanishes exponentially quickly in  $n$ , whenever some types reject the moral transgression on an individual level,  $F(V) < 1$ .<sup>12</sup> In contrast, under sharing of guilt,  $P_{n,n}$  will take the value 1 for sufficiently large  $n$  if the support of  $F$  is finite.

A finite support means that there exists a commonly-known maximum-possible level of individual guilt. If such a maximum-possible level of guilt is unknown or infinite, the support of  $F$  should be assumed as infinite. Proposition 2 and 3 characterize the behavior of  $P_{n,n}$  in this more general case. We show that while  $P_{n,n} \rightarrow 0$  is the rule without sharing of guilt, it only arises for very heavy-tailed distributions  $F$  if guilt can be shared easily. Thus, under sharing of guilt, in a large population moral transgression happens very likely.

Proposition 3 shows that there is a sharp dichotomy between heavy-tailed distributions  $F$  for which  $P_{n,n}$  converges to 0, and light-tailed distributions for which  $P_{n,n}$  converges to 1. The intuition for this finding is as follows: As the population becomes larger, sharing guilt becomes easier. Yet the person with the strongest feelings of guilt out of the population is going to be a stricter and stricter moralist. If  $F$  has a sufficiently heavy tail, the latter effect dominates. In a first step, we characterize the boundary case of distributions for which  $P_{n,n}$  is constant, i.e., distributions for which the transgression probability is independent of population size  $n$ .

**Proposition 2.** Assume  $\alpha > 0$  and let  $F(x) = \exp(-\beta x^{-\frac{1}{\alpha}})$  where  $\beta = -V^{\frac{1}{\alpha}} \log(q)$  for some  $q \in (0, 1)$ . Then  $P_{n,n} = q$  for all  $n$ .

---

<sup>12</sup>If  $F(V) = 1$ , it seems safe to assert that we are not speaking of a moral transgression.

The distribution  $F$  in the proposition is the so-called Fréchet distribution<sup>13</sup> whose density is given by

$$f(x) = \frac{\beta}{\alpha} \exp(-\beta x^{-\frac{1}{\alpha}}) x^{-\frac{1+\alpha}{\alpha}}.$$

For large  $x$ , the behavior of this density function is dominated by the power decay  $x^{-\frac{1+\alpha}{\alpha}}$ . Proposition 3 shows that a power decay of order  $\frac{1+\alpha}{\alpha}$  marks indeed the boundary between distributions for which the transgression happens or does not happen in large populations.

**Proposition 3.**

- (i) If  $\lim_{x \rightarrow \infty} x^{\frac{1+\alpha}{\alpha}} f(x) = 0$ , then  $\lim_{n \rightarrow \infty} P_{n,n} = 1$
- (ii) If  $\lim_{x \rightarrow \infty} x^{\frac{1+\alpha}{\alpha}} f(x) = \infty$ , then  $\lim_{n \rightarrow \infty} P_{n,n} = 0$ .

The boundary case is thus a certain power decay and the critical power increases in the degree  $\alpha$  to which guilt can be shared. If the tail of  $F$  is lighter than  $x^{-\frac{1+\alpha}{\alpha}}$ , we are in case (i) of the proposition, if it is heavier we are in case (ii). If guilt is shared proportionally,  $\alpha = 1$ , the critical power is given by  $\frac{1+\alpha}{\alpha} = 2$ . Since distributions with heavier tails than  $x^{-2}$  are rarely observed in applications<sup>14</sup>, we can expect to be typically in case (i) of the proposition. In contrast, if we consider intermediate levels of guilt-sharing such as  $\alpha = 0.2$ , we find a critical power of  $\frac{1+\alpha}{\alpha} = 6$  and, accordingly, there are important classes of distributions  $F$  for which either  $P_{n,n} \rightarrow 1$  or  $P_{n,n} \rightarrow 0$ .

We close this section by showing that for many common distribution functions, exponential distributions and power distributions, the sequence  $P_{n,n}$  is  $U$ -shaped: Under unanimity voting and proportional sharing of guilt  $\alpha = 1$ , the smallest transgression probability occurs at some intermediate population size.

**Proposition 4.** *Assume that  $\alpha = 1$  and that  $F$  is either an exponential or a power distribution, i.e.,  $F(x) = 1 - e^{-cx}$  for some  $c > 0$  with density*

---

<sup>13</sup>The Fréchet distribution is best-known as a stable limiting distribution of rescaled first order statistics in extreme value theory, see, e.g., Chapter 22 of Johnson, Kotz and Balakrishnan (1994). Thus it is not surprising that this class of distributions arises here:  $P_{n,n} \equiv q$  can be understood as a stability property of first order statistics.

<sup>14</sup>For tails heavier than  $x^{-2}$ , expected values are infinite. Tails slightly lighter than  $x^{-2}$  do appear in applications. See Clauset, Shalizi and Newman (2009) for examples.

$f(x) = ce^{-cx}$ , or,  $F(x) = 1 - \frac{b^\gamma}{(x+b)^\gamma}$  for some  $b, c > 0$ ,  $\gamma > 1$  with density  $f(x) = \frac{\gamma b^\gamma}{(x+b)^{\gamma+1}}$ . Then, for sufficiently small  $V$ ,  $P_{n,n}$  is decreasing up to some unique and finite  $n_0 > 1$  and increasing from there on.

The (fairly technical) proof contains an explicit criterion which allows to check the result for further distribution functions. Using a transformation of the type  $G(nV) = F(n^\alpha V)$ , the result can be extended to  $\alpha \neq 1$ . The intuition for this result is as follows: The initial decrease is driven by the same effect which drives the transgression probability to zero in the case without sharing,  $\alpha = 0$ . As the group grows larger, more agents have to agree and thus the probability that someone will object is high. As  $n$  becomes even larger, this effect is overruled by the mechanisms of shared guilt. In a very large group, guilt can be shared so easily that its effect becomes negligible.

## 4 Institutions and Immoral Outcomes

In this section we compare various institutions: In Section 4.1 we analyze the influence of the number of required supporters for a transgression,  $k$ , within a population of fixed size. Section 4.2 focuses on effects of populations size  $n$  on immoral behavior and moral transgression.

Throughout Section 4, we assume a proportional sharing of guilt,  $s(y) = y$ . As seen in the previous section, in this case sharing of guilt can be expected to play a significant role for most natural distribution functions  $F$ .

As outlined in the introduction, we suggest to interpret these comparative statics results as an analysis of how the decision process within an institution affects moral outcomes.

### 4.1 Number of Required Supporters

In this subsection, we study how the number of supporters necessary for transgression to happen,  $k$ , affects moral outcomes. We find that individual thresholds  $\theta_{k,n}$  increase in  $k$ . This is intuitive as sharing guilt becomes easier the more supporters are needed. In contrast, the aggregate transgression probability  $P_{k,n}$  is decreasing in  $k$  for small  $k$  but increasing in  $k$  for large  $k$ : With very small  $k$  just the worst people in the population have to participate for



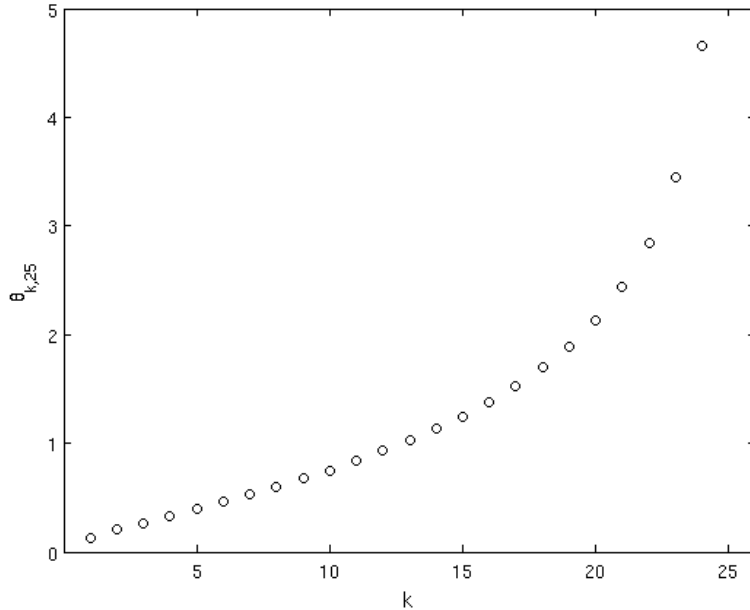


Figure 1:  $\theta_{k,n}$  as function of  $k$  for  $n = 25$ ,  $V = 1$  and the exponential distribution  $F(x) = 1 - \exp(-x)$ .

the transgression to happen. Yet as  $k$  increases, people with substantial moral concerns are needed as supporters. But these people have an incentive to free-ride on others. This makes transgression less likely overall. In contrast, for large  $k$  effects of shared guilt dominate. In consequence, the smallest values of  $P_{k,n}$  typically occur for intermediate values of  $k$ .

We start by rewriting the equilibrium condition (3) into a slightly more tractable form for further analysis.

**Lemma 2.** For  $s(y) = y$ , (3) is equivalent to

$$\frac{kV}{\theta_{k,n}} b(n, k, F(\theta_{k,n})) = \sum_{j=k}^n b(n, j, F(\theta_{k,n})). \quad (5)$$

We see that there is a simple relation between the equilibrium probability of a transgression, the right hand side of (5), and the equilibrium probability of a transgression which is supported by the minimal number of  $k$  agents,  $b(n, k, F(\theta_{k,n}))$ , on the left hand side.

We now turn to the impact of  $k$  for fixed  $n$ . Figures 1 and 2 display the behavior of  $\theta_{k,n}$  and  $P_{k,n}$  for  $F$  being the exponential distribution and  $V = 1$ .

As expected, the individual transgression thresholds  $\theta_{k,n}$  are increasing in  $k$ : If  $k$  is small, agents can free-ride, hoping that others with less scruples are around. For large  $k$ , the moral costs of supporting the transgression are small due to sharing of guilt. Accordingly, most types support the transgression. In particular,  $\theta_{k,n}$  is smaller than  $V$  for small  $k$  and larger than  $V$  for large  $k$ : For small  $k$ , agents who would support transgression if they were alone opt against it in order to free-ride. For large  $k$ , agents who would not behave immorally in an individual decision now decide to support it. This monotonicity behavior of  $\theta_{k,n}$  can easily be verified for general distribution functions:<sup>15</sup>

**Proposition 5.** *The sequence  $\theta_{k,n}$  is strictly increasing in  $k$ .*

We next consider the influence of  $k$  on the probability that the action is taken. In the example of Figure 2, we observe that the transgression probabilities  $P_{k,n}$  form a  $U$ -shape. The transgression probability is minimal at some intermediate value of  $k$  and larger if very few or very many agents have to support the transgression. The intuition for this  $U$ -shape is as follows. For small  $k$ , there will always be a small group of agents who really do not care about the moral dimension of the problem. For large  $k$ , it is certain that the action will not be taken unless guilt can be shared among many – resulting in small guilt associated with opting for “Yes”. For intermediate values of  $k$  neither of these two mechanisms helps agents as much to overcome their moral concerns and, accordingly, immoral action is taken with smaller probability. Thus, our model predicts that due to a “herd behavior”-like effect, a simple majority vote  $k = n/2$  is more likely to preserve moral standards than a vote made under unanimity,  $k = n$ .

The following proposition confirms that the basic logic behind this reasoning – a decrease at the left end and an increase at the right end – carries over to a large, non-parametric class of distributions, distributions with a decreasing density and a finite support.<sup>16</sup>

**Proposition 6.**

---

<sup>15</sup>A similar but slightly more involved proof generalizes this result to all weakly increasing sharing rules  $s$ .

<sup>16</sup>Backed by numerical investigations, we further conjecture the following: If  $f$  has a single peak, there still tend to be two maxima in the sequence  $(P_{k,n})_k$ , one at  $k = n$  and another for some small  $k \geq 1$ . If  $f$  has multiple peaks, the behavior of  $P_{k,n}$  can be more complex.

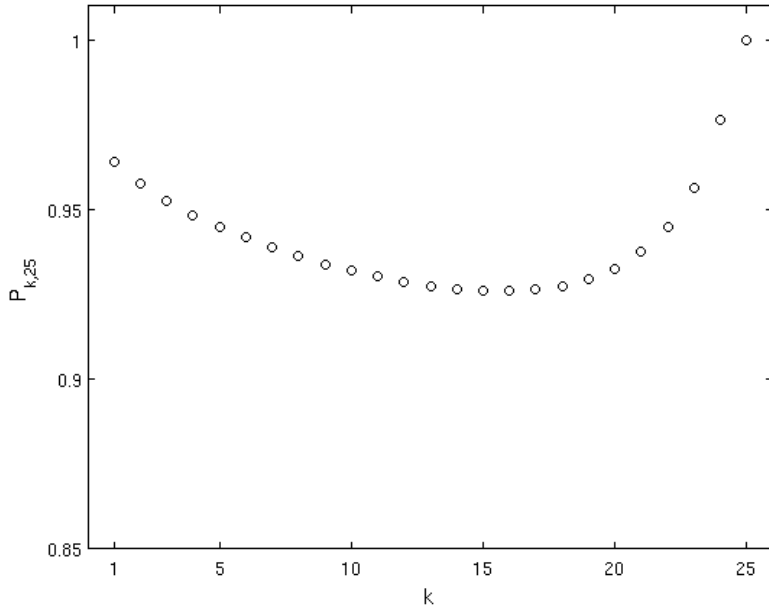


Figure 2:  $P_{k,25}$  as function of  $k$  for  $n = 25$ ,  $V = 1$  and the exponential distribution  $F(x) = 1 - \exp(-x)$ .

- (i) If the density function  $f$  of  $F$  is bounded and weakly decreasing over an interval  $[0, \varepsilon]$ ,  $\varepsilon > 0$ , then  $P_{1,n} > P_{2,n}$  for sufficiently large  $n$ .
- (ii) If the support of  $F$  is finite, we have  $P_{n-1,n} < P_{n,n}$  for sufficiently large  $n$ .

While an inspection of the proof of Proposition 6 (i) suggests that such a generalization may be hard to prove, we conjecture that the probabilities  $P_{k,n}$  are fully  $U$ -shaped for decreasing densities which are not too heavy-tailed.

## 4.2 Population Size

In this section, we study the influence of population size on moral transgression. We find that while individual transgression thresholds converge to 0 in  $n$ , the expected number of supporters becomes arbitrarily large as  $n$  increases. This is a strong overshooting effect: In the limit, instead of the required  $k$ , infinitely many people support moral transgression. This holds even if just one single supporter is required. Instead of one, infinitely many people become supporters. We furthermore study the behavior of the aggregate transgression as the population grows large. In the limit, transgression will happen independently of the number of supporters required. Yet this transgression probability

is typically not monotone in  $n$ . This is due to the interplay between competing effects of free-riding and sharing guilt.

We start with the behavior of the individual transgression threshold  $\theta_{k,n}$ :

**Proposition 7.**

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

The first part of the proposition shows that, not surprisingly, the individual willingness to participate vanishes as  $n$  grows large. The second part of the proposition provides an upper bound on the rate of this decrease. Moreover, it shows that the expected number of agents who support the decision,  $nF(\theta_{k,n})$ , gets arbitrarily large for large  $n$  – even though the necessary number of supporters  $k$  remains fixed.

Finally, we show that  $P_{k,n}$  converges to 1 as  $n$  gets large. With many agents, a sufficiently large group of supporters with sufficiently small moral concerns will most likely take immoral action.

**Proposition 8.**  $\lim_{n \rightarrow \infty} P_{k,n} = 1$ .

This result complements our findings on the (similar) limit behavior of  $P_{n,n}$  in Section 3. Interestingly, the convergence in Proposition 8 is typically not monotonic for  $k < n$ . To see this, note that for sufficiently large  $k$  and  $F$  with finite support,  $P_{k,k} = 1$ ,  $P_{k,n} < 1$  for  $n > k$  and  $\lim_{n \rightarrow \infty} P_{k,n} = 1$ . Figure 3 displays the effect for  $F$  being the exponential distribution. If we think of  $P_{k,n}$  as  $U$ -shaped in  $k$  for fixed  $n$ , the effect stems from the fact that for  $k = n$  we are at the right end of such a  $U$ -shape, for  $n \approx 2k$  we are somewhere in the middle of the  $U$ -shape for that  $n$ . For large  $n$ , we are at the left end of a  $U$ -shape.

## 5 Direct Costs of Pivotality and Outcome-Independent Guilt

In this section we discuss two natural variations of our basic model: In Section 5.1 we consider a model in which people face additional moral costs if they

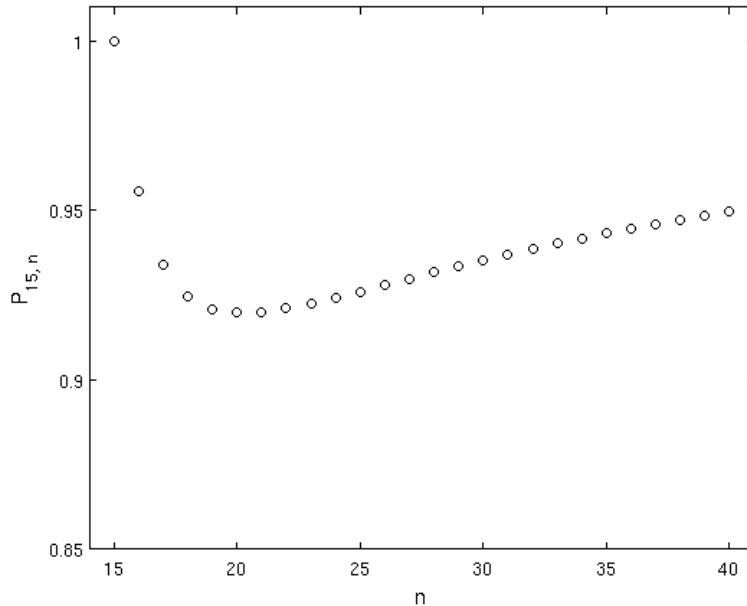


Figure 3:  $P_{15,n}$  as function of  $n$  for  $V = 1$  and the exponential distribution  $F(x) = 1 - \exp(-x)$ .

were pivotal, i.e., if the transgression would not have happened if they had chosen “No” instead of “Yes”. We find that most of our results carry over to this setting. The model with costs of pivotality is strategically equivalent to a model in which the benefit  $V$  from the transgression is heterogeneous across the population and decreasing in guilt. Our discussion thus covers this model as well.

In Section 5.2 we introduce a model in which the costs associated with shared guilt are paid regardless of whether the transgression occurs or not. In this model, we observe the following transition in equilibrium behavior depending on the transgression benefits  $V$ : For small  $V$ , the unique symmetric equilibrium is the equilibrium where all agents choose “No”. If  $V$  lies above some threshold there exist several symmetric equilibria in non-trivial threshold strategies, i.e., equilibria where the transgression probability  $P_{k,n}$  is strictly positive. Among these, the equilibrium with the highest transgression probability Pareto-dominates all other symmetric equilibria. Thus, in this model a critical mass of sufficiently willing agents is needed to make the transgression materialize.

## 5.1 Costs of Pivotality

We consider a model in which people experience particularly large levels of guilt if they could have stopped the transgression through their individual decision. We thus assume that there are additional, *direct costs of pivotality* for agents who chose “Yes” whenever exactly  $k$  agents opted “Yes” in total. For simplicity, we focus on a proportional sharing of guilt,  $s(y) = y$ . Let the pivotality cost function  $c$  be a non-negative, continuous, strictly increasing function with  $c(0) < V$ . The last condition ensures that at least some agents might selfishly benefit from the transgression. The realized utility of agent  $i$  with type  $x_i$  from opting for “Yes” is given by

$$V 1_{\{Y_{-i} \geq k-1\}} - c(x_i) 1_{\{Y_{-i} = k-1\}} - \frac{x_i}{1 + Y_{-i}} 1_{\{Y_{-i} \geq k-1\}} \quad (6)$$

where  $Y_{-i}$  denotes the number of agents other than  $i$  who chose “Yes”. Realized utility from choosing “No” is still given by

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

Guilt from pivotality is represented by the second term in (6), while the third term represents (diffused) guilt from supporting the transgression. We assume that these two facets of guilt are a function of an agent’s type  $x_i$  and thus perfectly correlated.

Before we discuss this generalized model, let us briefly address the case in which guilt is present only in the form of direct costs of pivotality, i.e., the case in which the third term in (6) is absent. Then it is easy to see that there exists a symmetric equilibrium in threshold strategies where agent  $i$  votes “Yes”, whenever  $V > c(x_i)$ . Equilibrium decisions are thus independent of  $k$  and  $n$ . In particular, agents stick to the decision they would have made when deciding alone. The reason is, of course, that in this model the cases where an agent is pivotal are the only cases where his decision matters.

In the full model with both costs of shared guilt and costs of pivotality, the counter-part of Proposition 1 is the following result:

**Proposition 9.** *There exists a unique symmetric Bayesian equilibrium in*

which transgression happens with positive probability. In this equilibrium, agent  $i$  votes “Yes” if  $x_i \leq \theta_{k,n}$  and “No” if  $x_i > \theta_{k,n}$ .  $\theta_{k,n}$  is the unique solution of

$$\frac{k(V - c(\theta_{k,n}))}{\theta_{k,n}} b(n, k, F(\theta_{k,n})) = \sum_{j=k}^n b(n, j, F(\theta_{k,n})). \quad (7)$$

Comparing (7) with the corresponding equation (5) in the basic model, it is easy to see the following strategic equivalence: (7) is also the equilibrium condition in a modified basic model where agents have heterogeneous benefits  $\tilde{V}_i = V - c(x_i)$  from the transgression. In that model, agents with higher costs of shared guilt also benefit less from the transgression. In particular, agents with strong moral concerns may feel bad about realized transgression even if they said “No” themselves. To sum up, this model with benefits  $\tilde{V}_i$  is strategically equivalent to the model with pivotality costs (even though the models are *not* payoff-equivalent).

By “sandwiching” the present model between two versions of the basic model, one with benefit  $V$  and one with benefit  $v < V - c(0)$ , one can easily see that the effect of increasing population size is similar to what we have found in Section 4.2:

**Proposition 10.** *For  $k \geq 1$  and  $n \rightarrow \infty$ , we have*

- (i)  $\theta_{k,n} \rightarrow 0$ .
- (ii)  $nF(\theta_{k,n}) \rightarrow \infty$ .
- (iii)  $P_{k,n} \rightarrow 1$ .

Whether there is a substantial qualitative difference between the present model and the basic model depends on the following distinction between two cases. In the case where  $V - c(x)$  is positive for all  $x$  in the support of  $F$ , one can show that the extended model behaves very similar to the basic one. In terms of the model with heterogeneous benefits, this is the case where all agents – regardless of their type  $x_i$  – have a positive benefit  $\tilde{V}_i$  from the transgression. If there exists some  $x^*$  such that  $V - c(x^*) = 0$ , equilibrium thresholds  $\theta_{k,n}$  are bounded by  $x^*$ :

**Corollary 2.** *If there exists  $x^*$  with  $V - c(x^*) = 0$  and  $F(x^*) < 0$ , then  $\theta_{k,n} \leq x^*$  for all  $k$  and  $n$ .*

This follows immediately from the fact that the right hand side of (7) is non-negative. In this case, there is a positive mass of “incorruptible” agents – those with  $x_i > x^*$  – who never vote in favor of the transgression. In consequence, we no longer have  $\lim_{n \rightarrow \infty} P_{n,n} = 1$ , unanimity does not lead to a transgression in the large-population limit.

## 5.2 Outcome-independent Guilt

In the basic model, we assume that agents feel guilt only when the transgression they supported occurs. In this section, we take a look at rule-based moral approaches: Agents feel guilt from voting “Yes” regardless of the outcome of the voting.<sup>17</sup> We assume that they still share guilt with all others who voted like them. We find the following dichotomy: For small values of the benefit  $V$  the equilibrium in which all agents opt against the transgression is the unique symmetric equilibrium. If  $V$  lies above some threshold, there are several symmetric equilibria in which the transgression occurs with strictly positive probability. The equilibrium with the highest transgression probability Pareto-dominates all others.<sup>18</sup>

We thus assume that the realized utility of agent  $i$  from opting for “Yes” is given by

$$V 1_{\{Y_{-i} \geq k-1\}} - \frac{x_i}{s(1 + Y_{-i})}$$

while 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 11.

---

<sup>17</sup>In the terminology of Admati and Perry (1991), we compare our basic model, a subscription game, to the associated contribution game.

<sup>18</sup>We leave a complete analysis of this model and, more importantly, intermediate cases between this and the basic model open for future research.



(i) *Symmetric equilibria in which the transgression occurs with positive probability are equilibria in symmetric threshold strategies where equilibrium thresholds solve*

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

*For sufficiently large  $V$  such an equilibrium exists.*

(ii) *Let  $f(0) < \infty$  and  $k > 2$ . Then, for sufficiently large  $V$ , there are at least two distinct symmetric equilibria where the transgression occurs with positive probability.*

(iii) *Let  $f(0) < \infty$  and  $k > 1$ . Then, for small  $V$ , the unique symmetric equilibrium is the one where all agents vote “No”.*

Thus, in equilibrium small temptations  $V$  do not lead to a moral transgression – but sufficiently large ones do. This is due to the fact that the present model contains a “two-sided” coordination problem: In addition to the possibility of free-riding on other agents’ “Yes”-votes, there is the risk – when voting “Yes” – of paying the costs of guilt without reaping the benefit  $V$ . The two-sided coordination problem also causes the multiplicity of equilibria in case (ii).

## 6 No Sharing Of Guilt

In this section, we discuss how our basic model with proportional sharing of guilt,  $s(y) = y$ , differs from the more standard model without any sharing of guilt,  $s(y) \equiv 1$ . If the number of supporters necessary,  $k$ , is fixed, the model without shared guilt is equivalent to a model with shared guilt in which the gain out of transgression is adjusted from  $V$  to  $V/(k-1)$ . This illustrates that sharing guilt makes transgression much more likely, causing the participation of people who would never like to become responsible for the immoral action to happen individually.

Furthermore, the models’ behavior in  $k$  differs drastically. While we observed a  $U$ -shape in the aggregate transgression probabilities under shared guilt, these probabilities tend to decrease in  $k$  if diffusion of guilt is ruled out. In the case of unanimity, i.e.  $k = n$ , the aggregate transgression probability decreases very quickly to 0 if guilt is not divisible.

For a more detailed comparison of the models with and without shared guilt, let us denote by  $\rho_{k,n}$  and  $Q_{k,n}$  the thresholds and transgression probabilities for the model without sharing,  $s(y) \equiv 1$ . The corresponding quantities under  $s(y) = y$  we continue to denote by  $\theta_{k,n}$  and  $P_{k,n}$ . Sometimes, we need to emphasize the dependence on  $V$  and write  $\rho_{k,n}(V)$  etc. We begin by inspecting the case in which only one supporter is needed for moral transgression to happen, i.e.  $k = 1$ , for the model without sharing.

**Proposition 12.**  *$\rho_{1,n}$  monotonically converges to 0 in  $n$ . Moreover,  $Q_{1,n} \rightarrow 1$  as  $n \rightarrow \infty$ .*

We have seen in Corollary 1 (i) that  $Q_{n,n} \rightarrow 0$ . In contrast,  $P_{n,n} \rightarrow 1$  for most distribution functions  $F$ . Accordingly, instead of the  $U$ -shape of the sequence  $(P_{k,n})_k$ , we find that without shared guilt, the transgression probability  $Q_{k,n}$  tends to decrease as more agents become necessary for transgression to take place.

If  $k > 1$ , there is a surprising strategic equivalence<sup>19</sup> between the models with and without shared guilt:

**Proposition 13.** *For  $k > 1$  and  $n > 1$ , we have*

$$\rho_{k,n}(V) = \theta_{k-1,n-1} \left( \frac{V}{k-1} \right).$$

Thus, up to small adjustments of  $k$  and  $n$ , thresholds are identical in a model without sharing and a model with sharing if the individual gain out of transgression is divided by  $k - 1$ . Hence, differences in the gains from transgression must be considerable to lead to comparable transgression probabilities in the models with and without sharing.

Applying Proposition 13, the results of Section 4.2 can easily be transferred to the model without sharing. As  $n$  gets large while  $k$  remains fixed, thresholds vanish, the expected number of supporters gets large, and the transgression probability converges to 1. Note that while limit behavior becomes similar, quantitative result should be very different for fixed  $k$  and  $n$  as illustrated by

---

<sup>19</sup>This is not a strategic equivalence in the strictest sense, since the number of players between the games differs.

the drastic adaption of the individual gain out of transgression which is necessary to make the models comparable.

**Proposition 14.** *For  $k > 1$  and  $n \rightarrow \infty$ , we have*

(i)  $\rho_{k,n} \rightarrow 0$ .

(ii)  $nF(\rho_{k,n}) \rightarrow \infty$ .

(iii)  $Q_{k,n} \rightarrow 1$ .

To sum up, in the free-riding regime of small, fixed  $k$  in a population that grows large, differences between the models are quantitative rather than qualitative – and they are substantial. For large  $k$  in a population of limited size, in addition to marked quantitative differences, there are qualitative differences as well.

## 7 Conclusion

We have studied the role of institutions in affecting individual moral behavior and moral outcomes within smaller and larger populations, focusing on the diffusion of guilt. In particular, we have analyzed the effects of shared guilt when institutions require the support of several people for moral transgression to happen. Our study follows a standard game-theoretic approach, but incorporates evidence from social psychology and experimental economics that people acting together with others tend to feel less responsibility and guilt (e.g. Bandura et al. (1975), Zimbardo (2007), Falk and Szech (2013b)).

When people share guilt, large groups might be willing to support a transgression – if institutions *require* such collaboration. Thus, barriers of individual moral concerns may be substantially lowered by institutional design – unless there is a considerable fraction of people with extremely high moral standards. Looking through human history, it becomes plausible that institutions asking for the collaboration of many people facilitate moral transgression, especially if individual moral standards are low anyway. Some historians even point out that most drastic atrocity was facilitated by such institutional design.<sup>20</sup>

---

<sup>20</sup>Compare Browning’s (2004) view on what made the Holocaust possible. For a discussion and conflicting views, compare Goldhagen, Browning and Wieseltier (1996).

It is hence important to understand institutional influences on moral behavior – ultimately, of course, to prevent moral transgression from taking place. We focused on the role of diffused guilt. Further research is necessary to identify and understand the influence of other morally relevant organizational design tools. One may think of authority, status concerns, attributions of blame to victims, arousal, or reward systems. 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 shape such social safeguards.

## A Proofs

*Proof of Proposition 1.* We first show that all equilibria, symmetric or not, in which the action is taken with positive probability must be equilibria in threshold strategies: For each agent 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  $i$ ’s opponents. In an equilibrium where the action is taken with positive probability, at least  $k - 1$  of the opponents vote “Yes” with positive probability. Assume agent  $i$  weakly prefers voting “Yes” over “No” at some value  $x_i$ . Comparing the expectations of (1) and (2) over the opponents’ strategies implies that this preference must be strict for types with smaller costs. Likewise, if agent  $i$  weakly prefers voting “No” over “Yes” at some value of  $x_i$ , this preference must be strict at all types with higher costs. Thus, all best-responses to the opponents’ strategies are threshold strategies. Therefore, all equilibria where the action is taken with positive probability are equilibria in threshold strategies.<sup>21</sup>

Thus, for  $k < n$  it remains to 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

---

<sup>21</sup>Since the equilibrium where all agents always vote “No” is also in threshold strategies, it follows that in fact all symmetric equilibria are in threshold strategies.

his opponents play a threshold strategy with  $\theta_{k,n} > 0$ . Expected payoff from voting “Yes” is then given by

$$\sum_{j=k-1}^{n-1} b(n-1, j, F(\theta_{k,n})) \left( V - \frac{x}{s(1+j)} \right) \quad (8)$$

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

$$\sum_{j=k}^{n-1} b(n-1, j, F(\theta_{k,n})) V. \quad (9)$$

Equilibria are characterized by values of  $\theta_{k,n}$  for which the two expressions coincide for  $x = \theta_{k,n}$ . Equating (8) and (9) for  $x = \theta_{k,n}$  yields the condition

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

which is (3). It remains to show existence of a unique solution. To this end, we write (10) as

$$\frac{V}{\theta_{k,n}} = \frac{\sum_{j=k-1}^{n-1} b(n-1, j, F(\theta_{k,n})) \frac{1}{s(1+j)}}{b(n-1, k-1, F(\theta_{k,n}))}. \quad (11)$$

The left hand side of (11) is strictly decreasing in  $\theta_{k,n}$ , diverging to  $\infty$  at 0. Writing the right hand side as

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

we see that it increases strictly from  $s(k)^{-1}$  to  $\infty$  as  $\theta_{k,n}$  moves through the support of  $F$ . This shows the existence of a unique solution  $\theta_{k,n}$  in the interior of the support. It remains to consider  $k = n$ . As in the case  $k < n$ , an interior separating equilibrium must be characterized by (11) which becomes

$$\frac{V}{\theta_{n,n}} = \frac{1}{s(n)}. \quad (12)$$

In the case where (12) does not possess an interior solution (and only then),

all agents voting “Yes” regardless of their type is an equilibrium: This holds whenever  $V \geq \frac{x}{s(n)}$  for all  $x$  in the support of  $F$ .  $\square$

*Proof of Proposition 2.* We have to solve the condition  $P_{n,n} = F(n^\alpha V)^n = q$  for  $F$ . Taking logarithms yields  $n \log(F(n^\alpha V)) = \log(q)$ . The monotone transformation  $x = n^\alpha V$ , i.e.,  $n = x^{1/\alpha} V^{-1/\alpha}$  yields

$$\log(F(x)) = x^{-\frac{1}{\alpha}} V^{\frac{1}{\alpha}} \log(q) = -x^{-\frac{1}{\alpha}} \beta.$$

Taking the exponential of both sides of the equation gives the desired expression for  $F$ . Since the resulting  $F$  is a Fréchet-distribution, we know in particular that it is indeed a distribution function on  $\mathbb{R}^+$ .  $\square$

*Proof of Proposition 3.* We prove (i) and (ii) in two steps: We first show that  $x^{\frac{1+\alpha}{\alpha}} f(x) \rightarrow 0$  and  $x^{\frac{1+\alpha}{\alpha}} f(x) \rightarrow \infty$  imply, respectively,  $n(1 - F(n^\alpha V)) \rightarrow 0$  and  $n(1 - F(n^\alpha V)) \rightarrow \infty$ . We then derive  $P_{n,n} \rightarrow 1$  and  $P_{n,n} \rightarrow 0$  from these conditions (which can in fact be shown to be equivalent). Fix  $\varepsilon > 0$  and assume  $x^{\frac{1+\alpha}{\alpha}} f(x) \rightarrow 0$ . Then for sufficiently large  $n$  we have  $x^{\frac{1+\alpha}{\alpha}} f(x) \leq \varepsilon$  for all  $x \geq n^\alpha V$ . It follows that

$$n(1 - F(n^\alpha V)) = n \int_{n^\alpha V}^{\infty} f(x) dx \leq \varepsilon n \int_{n^\alpha V}^{\infty} x^{-\frac{1+\alpha}{\alpha}} dx = \varepsilon \alpha V^{-\frac{1}{\alpha}}.$$

Since  $\varepsilon$  was arbitrary, this proves  $n(1 - F(n^\alpha V)) \rightarrow 0$ . Fix  $\kappa > 0$  and assume  $x^{\frac{1+\alpha}{\alpha}} f(x) \rightarrow \infty$ . Then for sufficiently large  $n$ , we have  $x^{\frac{1+\alpha}{\alpha}} f(x) \geq \kappa$  for all  $x \geq n^\alpha V$ . Arguing similarly to before, we have  $n(1 - F(n^\alpha V)) \geq \kappa \alpha V^{-\frac{1}{\alpha}}$ .  $n(1 - F(n^\alpha V)) \rightarrow \infty$  and the first step of the proof is complete. For the second step, write

$$F(n^\alpha V)^n = \left(1 + \frac{n(F(n^\alpha V) - 1)}{n}\right)^n \tag{13}$$

Assume  $n(1 - F(n^\alpha V)) \rightarrow 0$  and fix  $\varepsilon > 0$ . Then for sufficiently large  $n$  we have  $n(F(n^\alpha V) - 1) \geq -\varepsilon$ . Thus by (13)

$$F(n^\alpha V)^n \geq \left(1 - \frac{\varepsilon}{n}\right)^n \rightarrow \exp(-\varepsilon)$$

Since we also have  $F(n^\alpha V)^n \leq 1$  this implies  $F(n^\alpha V)^n \rightarrow 1$ . Now, assume  $n(1 - F(n^\alpha V)) \rightarrow \infty$  and fix  $\kappa > 0$ . For sufficiently large  $n$ , we have

$n(F(n^\alpha V) - 1) \leq -\kappa$ . This yields

$$F(n^\alpha V)^n \leq \left(1 - \frac{\kappa}{n}\right)^n \rightarrow \exp(-\kappa).$$

Since we can choose  $\kappa$  arbitrarily large and since we also have  $F(n^\alpha V)^n \geq 0$ , it follows that  $F(n^\alpha V)^n \rightarrow 0$  and the proof is complete.  $\square$

*Proof of Proposition 4.* We first prove a criterion on  $F$  which ensures that the result holds. Then we prove that exponential distributions and power distributions satisfy this criterion. The criterion runs as follows: Let  $F : [0, \infty) \rightarrow [0, \infty)$  be twice continuously differentiable with  $f(x) > 0$  for  $x \geq 0$  and  $F(xV)^x \rightarrow 1$  for  $x \rightarrow \infty$ . Furthermore, let there be exactly one  $x_0 \in (0, \infty)$  such that

$$2 = x_0 V \left( \frac{F'(x_0 V)}{F(x_0 V)} - \frac{F''(x_0 V)}{F'(x_0 V)} \right) \quad (14)$$

Then  $F(xV)^x$  has exactly one minimum on  $(0, \infty)$ . In particular, for sufficiently small  $V$  and  $\alpha = 1$ , the sequence  $P_{n,n}$  decreases up to some  $n_0 > 1$  and increases from  $n_0 + 1$  on. We first prove the criterion. We begin by showing that

$$\lim_{x \rightarrow 0} F(xV)^x = 1$$

As  $F(x)$  is differentiable in 0 with  $F'(0) > 0$ , there exists a  $\gamma > 0$  such that for all  $x$  small enough,  $F(xV) \geq x\gamma$ . Thus, for all  $x$  small enough,

$$1 \geq F(xV)^x \geq (\gamma x)^x.$$

For  $x \rightarrow 0$ ,  $x > 0$  the right hand side tends to 1 as  $x^x \rightarrow 1$ .

Now let us turn to the central part of the proof:

$$(F(xV)^x)' = (e^{x \log F(xV)})' = e^{x \log F(xV)} \left( \log F(xV) + x \frac{VF'(xV)}{F(xV)} \right) \quad (15)$$

We show that there exists exactly one  $x_0 \in (0, \infty)$  such that this term vanishes. Note that the whole term vanishes exactly when  $\log F(xV) + x \frac{VF'(xV)}{F(xV)}$  vanishes.

We first show that the derivative of  $\log F(xV) + x \frac{VF'(xV)}{F(xV)}$  is 0 exactly once in

$(0, \infty)$ :

$$\begin{aligned} & \left( \log F(xV) + x \frac{VF'(xV)}{F(xV)} \right)' \\ &= \frac{VF'(xV)}{F(xV)} + \frac{VF'(xV)}{F(xV)} + x \frac{V^2F''(xV)}{F(xV)} - x \frac{V^2F'(xV)^2}{F(xV)^2} \end{aligned}$$

which is 0 exactly when

$$2 = xV \left( \frac{F'(xV)}{F(xV)} - \frac{F''(xV)}{F'(xV)} \right)$$

We have assumed that this term is 0 exactly once.

The mean value theorem tells us now that the term  $g(x) := \log F(xV) + x \frac{VF'(xV)}{F(xV)}$  vanishes at most two times. We will now show that  $g(x)$  vanishes exactly once.

As  $F(xV)^x \rightarrow 1$  for  $x \rightarrow 0$  and  $F(xV)^x \rightarrow 1$  for  $x \rightarrow \infty$  and by equation (15),  $g(x)$  is  $< 0$  for small  $x$  and  $> 0$  for large  $x$ . Let us assume that  $g(x)$  has two different roots  $x_0 < x_1$ . As we have shown above these are the only roots of this function. Thus,  $g(t) < 0$  for  $t < x_0$  and  $g(t) > 0$  for  $t > x_1$ . As we assume that the derivative of  $g$  has exactly one root the mean value theorem tells us that this has to happen for a  $t_0 \in (x_0, x_1)$ . If  $g(t_0) = 0$ , we have a contradiction as the derivative of  $g$  vanishes at least two times (mean value theorem). This implies  $g(t_0) \neq 0$ . Assume without loss of generality that  $g(t_0) > 0$ . We have  $t_0 < x_1$ ,  $g(t_0) > g(x_1)$  and  $g(t) > g(x_1)$  for  $t > x_1$ , therefore the function  $g$  attains a local minimum in  $(t_0, x_1]$ . Therefore, the derivative of  $g$  vanishes at least two times. Contradiction.

Therefore we proved that there exists at most one  $x_0 \in (0, \infty)$  such that  $g(x_0) = 0$ . As  $g(x)$  is continuous, negative near 0 and positive for high  $x$ , there exists exactly one  $x_0$  with this property. Therefore,  $F(xV)^x$  has exactly one minimum in  $[0, \infty)$ . This shows, that our criterion is valid.

We now turn to the exponential distribution  $F(x) = 1 - e^{-cx}$  for a  $c > 0$ . By Proposition 3,  $F(nV)^n \rightarrow 1$  for  $n \rightarrow \infty$ ,  $n \in \mathbb{N}$ . We have



$$1 \geq F(xV)^x \geq (F([x]V)^{[x]})^{x/[x]} \quad (16)$$

and hence also  $F(xV)^x \rightarrow 1$  for  $x \rightarrow \infty$ ,  $x \in \mathbb{R}$ .

Hence we only have to show that the equation (14) has exactly one solution:

$$2 = xV \left( \frac{ce^{-cxV}}{1 - e^{-cxV}} + c \right) = xcV \left( \frac{1}{1 - e^{-cxV}} \right) \quad (17)$$

Without loss of generality let us assume  $c = V = 1$ . As the derivative of  $e^{-x}$  at 0 is  $-1$ , for  $x \rightarrow 0$  the right hand side of the equation converges to 1:

$$\lim_{x \rightarrow 0} \frac{e^{-x} - 1}{x} = -1$$

As the right hand side diverges to  $\infty$  for  $x \rightarrow \infty$  it suffices to show that the right hand side is increasing in  $x$ : We have

$$\left( \frac{x}{1 - e^{-x}} \right)' = \frac{(1 - e^{-x}) - xe^{-x}}{(1 - e^{-x})^2}$$

and, as  $e^x > 1 + x$  for  $x > 0$ ,  $1 - e^{-x} - xe^{-x} > 0$ . This proves the result for the exponential distribution.

We next turn to  $F(x) = 1 - (\frac{b}{x+b})^\gamma$  for  $b > 0$ ,  $\gamma > 1$ . By Proposition 3 and (16),  $F(xV)^x \rightarrow 1$  for  $x \rightarrow \infty$ . Hence we only have to show that equation (14) has exactly one solution:

$$2 = Vx \left( \frac{\gamma \frac{b^\gamma}{(Vx+b)^{\gamma+1}}}{1 - \frac{b^\gamma}{(Vx+b)^\gamma}} + \frac{(\gamma + 1)}{Vx + b} \right) = \frac{Vx}{Vx + b} \left( \gamma \frac{\frac{b^\gamma}{(Vx+b)^\gamma}}{1 - \frac{b^\gamma}{(Vx+b)^\gamma}} + (\gamma + 1) \right) \quad (18)$$

Without loss of generality, let us assume  $V = 1$ . The right hand side of the equation converges to  $\gamma + 1 > 2$  as  $x \rightarrow \infty$ . For  $x \rightarrow 0$  one can prove analogously to equation (17) that the right hand side converges to 1; note that the derivative of  $1 - \frac{b^\gamma}{(x+b)^\gamma}$  is  $\gamma \frac{b^\gamma}{(x+b)^{\gamma+1}}$ . We show now that the right hand side is strictly increasing in  $x$ :

$$\text{r.h.s.} = \frac{x}{x + b} \gamma \left( \frac{1}{1 - \frac{b^\gamma}{(x+b)^\gamma}} + \frac{1}{\gamma} \right)$$

It obviously suffices to show that

$$\frac{x+b}{x} \left( 1 - \frac{b^\gamma}{(x+b)^\gamma} \right)$$

is weakly decreasing in  $x$ .

$$\begin{aligned} \frac{x+b}{x} \left( 1 - \frac{b^\gamma}{(x+b)^\gamma} \right) &= \frac{x+b}{x} \left( 1 - \frac{b}{x+b} + \frac{b}{x+b} - \frac{b^\gamma}{(x+b)^\gamma} \right) \\ &= 1 + \frac{x+b}{x} \left( \frac{b}{x+b} - \frac{b^\gamma}{(x+b)^\gamma} \right) = 1 + \frac{b}{x} \left( 1 - \frac{b^{\gamma-1}}{(x+b)^{\gamma-1}} \right) \end{aligned} \quad (19)$$

Now note that  $\gamma > 1$  implies that  $x \mapsto 1 - \frac{b^{\gamma-1}}{(x+b)^{\gamma-1}}$  is a concave function on  $[0, \infty)$ . We now use the following fact: Let  $g$  be a concave function on  $\mathbb{R}$ ,  $x > x_0$ . Then

$$\frac{g(x) - g(x_0)}{x - x_0}$$

is weakly decreasing in  $x$ . This fact for  $x_0 = 0$  shows that (19) is weakly decreasing in  $x$ . We have shown that the right hand side of (18) converges to 1 for  $x \rightarrow 0$ , is strictly increasing and converges to a  $\gamma + 1 > 2$  for  $x \rightarrow \infty$ . Therefore this equation has exactly one solution in  $(0, \infty)$ .  $\square$

*Proof of Lemma 2.* Since we have  $\theta_{k,n} > 0$  and  $s(y) = y$ , we can rewrite the sum on the right hand side of (10) as follows:

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

Inserting  $b(n-1, k-1, F(\theta_{k,n})) = \frac{k}{nF(\theta_{k,n})} b(n, k, F(\theta_{k,n}))$  on the left hand side

of (10) and multiplying by  $nF(\theta_{k,n})$  we obtain

$$k b(n, k, F(\theta_{k,n})) V = \theta_{k,n} \sum_{j=k}^n b(n, j, F(\theta_{k,n})) \quad (20)$$

which is (5). □

*Proof of Proposition 5.* The equilibrium condition (5) can be written as

$$\frac{kV}{\theta_{k,n}} = \frac{\sum_{j=k}^n b(n, j, F(\theta_{k,n}))}{b(n, k, F(\theta_{k,n}))}. \quad (21)$$

To see the result, we insert  $\theta_{k+1,n}$  into the left hand and right hand sides of (21) for  $k$ :

$$\begin{aligned} \frac{kV}{\theta_{k+1,n}} &< \frac{(k+1)V}{\theta_{k+1,n}} \\ &= \frac{\sum_{j=k+1}^n b(n, j, F(\theta_{k+1,n}))}{b(n, k+1, F(\theta_{k+1,n}))} \\ &\leq \frac{\sum_{j=k}^n b(n, j, F(\theta_{k+1,n}))}{b(n, k, F(\theta_{k+1,n}))}. \end{aligned} \quad (22)$$

The final inequality relies on the fact that the expression at hand is the reciprocal of the *failure rate* of the Binomial distribution which is increasing in  $k$ , see Johnson, Kotz and Kemp (1992), Chapter 3. The left hand side and right hand side of (11) cross only once in  $\theta$  for fixed  $k$  with the left hand side being larger to the left of the intersection. We thus conclude from (22) that  $\theta_{k+1,n}$  lies to the right of the intersection,  $\theta_{k+1,n} > \theta_{k,n}$ . □

*Proof of Proposition 6.* <sup>22</sup> Part (ii) of the proposition follows from the observations that  $P_{n-1,n} \in (0, 1)$  while  $P_{n,n} = 1$  for sufficiently large  $n$  if the support of  $F$  is finite. It remains to prove (i). By the definitions of  $\theta_{1,n}$  and  $\theta_{2,n}$ , we know that

$$P_{1,n} = Vn \frac{F(\theta_{1,n})}{\theta_{1,n}} (1 - F(\theta_{1,n}))^{n-1} = 1 - (1 - F(\theta_{1,n}))^n$$

---

<sup>22</sup>While we present the results in an order guided by economic considerations, the logical contingencies are slightly different: The proof of Proposition 6 depends on the results of Section 4.2 in addition to the results stated earlier.

and

$$\begin{aligned}
P_{2,n} &= 1 - (1 - F(\theta_{2,n}))^n - nF(\theta_{2,n})(1 - F(\theta_{2,n}))^{n-1} \\
&= Vn(n-1)\frac{F(\theta_{2,n})^2}{\theta_{2,n}}(1 - F(\theta_{2,n}))^{n-2}.
\end{aligned} \tag{23}$$

Let  $a_n = (1 - F(\theta_{1,n}))^{n-1}$  and  $b_n = (1 - F(\theta_{2,n}))^{n-2}$ . Then, by inserting this into the equations above:

$$a_n = \frac{1}{Vn\frac{F(\theta_{1,n})}{\theta_{1,n}} + (1 - F(\theta_{1,n}))}$$

and

$$b_n = \frac{1}{Vn(n-1)\frac{F(\theta_{2,n})^2}{\theta_{2,n}} + (1 - F(\theta_{2,n}))^2 + nF(\theta_{2,n})(1 - F(\theta_{2,n}))}.$$

It follows that

$$\begin{aligned}
P_{2,n} &= Vn(n-1)\frac{F(\theta_{2,n})^2}{\theta_{2,n}}(1 - F(\theta_{2,n}))^{n-2} = Vn(n-1)\frac{F(\theta_{2,n})^2}{\theta_{2,n}}b_n \\
&= \frac{1}{1 + \frac{(1-F(\theta_{2,n}))^2\theta_{2,n}}{Vn(n-1)F(\theta_{2,n})^2} + \frac{(1-F(\theta_{2,n}))\theta_{2,n}}{V(n-1)F(\theta_{2,n})}}
\end{aligned}$$

and

$$P_{1,n} = Vn\frac{F(\theta_{1,n})}{\theta_{1,n}}a_n = \frac{1}{1 + \frac{\theta_{1,n}}{F(\theta_{1,n})}\frac{1-F(\theta_{1,n})}{Vn}}.$$

Therefore,  $P_{1,n} > P_{2,n}$  iff

$$\frac{\theta_{1,n}}{F(\theta_{1,n})}\frac{1 - F(\theta_{1,n})}{Vn} < \frac{(1 - F(\theta_{2,n}))^2\theta_{2,n}}{Vn(n-1)F(\theta_{2,n})^2} + \frac{(1 - F(\theta_{2,n}))\theta_{2,n}}{V(n-1)F(\theta_{2,n})}$$

i.e., iff

$$\frac{\theta_{1,n}}{F(\theta_{1,n})}(1 - F(\theta_{1,n})) < \frac{\theta_{2,n}}{F(\theta_{2,n})}\left(\frac{(1 - F(\theta_{2,n}))^2}{(n-1)F(\theta_{2,n})} + \frac{n(1 - F(\theta_{2,n}))}{(n-1)}\right).$$

Note that for any concave function  $F$  and  $x < y < z$ ,

$$\frac{F(z) - F(x)}{z - x} \leq \frac{F(y) - F(x)}{y - x}$$

By our assumptions and by Propositions 5 and 7, we know that  $\theta_{2,n} \rightarrow 0$ ,  $F(0) = 0$ ,  $\theta_{1,n} < \theta_{2,n}$  and that  $F$  is indeed concave (weakly decreasing density) near 0. Hence, for sufficiently large  $n$  we have

$$\frac{F(\theta_{2,n})}{\theta_{2,n}} = \frac{F(\theta_{2,n}) - F(0)}{\theta_{2,n} - 0} \leq \frac{F(\theta_{1,n}) - F(0)}{\theta_{1,n} - 0} = \frac{F(\theta_{1,n})}{\theta_{1,n}}.$$

Thus, it suffices to show that for large  $n$ :

$$1 - F(\theta_{1,n}) < \frac{(1 - F(\theta_{2,n}))^2}{(n-1)F(\theta_{2,n})} + \frac{n(1 - F(\theta_{2,n}))}{(n-1)}. \quad (24)$$

As preliminary observations, note that  $P_{1,n} = Vn \frac{F(\theta_{1,n})}{\theta_{1,n}} (1 - F(\theta_{1,n}))^{n-1} \rightarrow 1$  by Proposition 8, and that

$$\frac{F(\theta_{1,n})}{\theta_{1,n}} = \frac{F(\theta_{1,n}) - F(0)}{\theta_{1,n} - 0} \rightarrow f(0) > 0 \quad (25)$$

for  $n \rightarrow \infty$  as  $\theta_{1,n} \rightarrow 0$ . Thus, both  $\frac{F(\theta_{1,n})}{\theta_{1,n}}$  and  $n(1 - F(\theta_{1,n}))^{n-1}$  are bounded from above and from below by strictly positive numbers.

In inequality (24), we now take both sides to the power of  $n-1$  and multiply them by  $n$ . We have just seen that the left hand term is bounded. Thus, if we show that

$$n \left( \frac{(1 - F(\theta_{2,n}))^2}{(n-1)F(\theta_{2,n})} + \frac{n(1 - F(\theta_{2,n}))}{(n-1)} \right)^{n-1} \rightarrow \infty$$

the proof is complete. We have the lower bound

$$\begin{aligned} & n \left( \frac{(1 - F(\theta_{2,n}))^2}{(n-1)F(\theta_{2,n})} + \frac{n(1 - F(\theta_{2,n}))}{(n-1)} \right)^{n-1} \\ &= n(1 - F(\theta_{2,n}))^{n-1} \left( 1 + \frac{1}{F(\theta_{2,n})(n-1)} \right)^{n-1} \geq \frac{n}{F(\theta_{2,n})} (1 - F(\theta_{2,n}))^{n-1}. \end{aligned} \quad (26)$$

In the last inequality we have used that by binomial expansion,

$$\left( 1 + \frac{1}{F(\theta_{2,n})(n-1)} \right)^{n-1} > (n-1) \frac{1}{(n-1)F(\theta_{2,n})} = \frac{1}{F(\theta_{2,n})}.$$

As  $P_{2,n} \rightarrow 1$  and  $\theta_{2,n} \rightarrow 0$ , we deduce from (23) that

$$(1 - F(\theta_{2,n}))^{n-1} \sim (1 - F(\theta_{2,n}))^{n-2} \sim \frac{\theta_{2,n}}{Vn(n-1)F(\theta_{2,n})^2}. \quad (27)$$

Arguing as in (25), we see that  $\frac{\theta_{2,n}}{F(\theta_{2,n})}$  is bounded from below by a strictly positive number. Combining (26) and (27), it thus remains to show that

$$(n-1)F(\theta_{2,n})^2 \rightarrow 0.$$

To this end, let  $n_j$  be a subsequence such that  $(n_j - 1)F(\theta_{2,n_j})^2 \geq \kappa > 0$  for a  $\kappa > 0$ . Then, as  $\frac{F(\theta_{2,n})}{\theta_{2,n}} \leq C$  for a  $C > 0$  and  $F(\theta_{2,n_j}) \leq 1$ ,

$$\begin{aligned} P_{2,n_j} &= Vn_j(n_j - 1) \frac{F(\theta_{2,n_j})^2}{\theta_{2,n_j}} (1 - F(\theta_{2,n_j}))^{n_j-2} \\ &\leq CVn_j(n_j - 1) \left(1 - \frac{\sqrt{\kappa}}{\sqrt{n_j - 1}}\right)^{n_j-2}. \end{aligned}$$

As  $\kappa > 0$ , this term converges to 0, which is a contradiction to  $P_{2,n_j} \rightarrow 1$ . To see this convergence to 0, write

$$\left(1 - \frac{\sqrt{\kappa}}{\sqrt{n_j - 1}}\right)^{n_j-2} = \left(\left(1 - \frac{\sqrt{\kappa}}{\sqrt{n_j - 1}}\right)^{\sqrt{n_j - 1}}\right)^{\frac{n_j-2}{\sqrt{n_j - 1}}}.$$

The inner term converges to  $\exp(-\sqrt{\kappa}) < 1$ . Therefore,  $(n-1)F(\theta_{2,n})^2 \rightarrow 0$ .  $\square$

*Proof of Proposition 7.* We first prove (i). Suppose (i) is violated, i.e., suppose 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,

$$\begin{aligned} \sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta_{k,n_j})^i (1 - F(\theta_{k,n_j}))^{n_j-i} &\geq \sum_{i=k}^{n_j} \binom{n_j}{i} F(\varepsilon)^i (1 - F(\varepsilon))^{n_j-i} \quad (28) \\ &= 1 - \sum_{i=0}^{k-1} \binom{n_j}{i} F(\varepsilon)^i (1 - F(\varepsilon))^{n_j-i}. \end{aligned}$$

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

$$\frac{\sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta_{k,n_j})^i (1 - F(\theta_{k,n_j}))^{n_j-i}}{\binom{n_j}{k} F(\theta_{k,n_j})^k (1 - F(\theta_{k,n_j}))^{n_j-k}} = \frac{kV}{\theta_{k,n_j}}$$

to

$$\begin{aligned} \sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta_{k,n_j})^i (1 - F(\theta_{k,n_j}))^{n_j-i} &= \frac{kV}{\theta_{k,n_j}} \binom{n_j}{k} F(\theta_{k,n_j})^k (1 - F(\theta_{k,n_j}))^{n_j-k} \\ &\leq \frac{kV}{\varepsilon} \binom{n_j}{k} (1 - F(\varepsilon))^{n_j-k}. \quad (29) \end{aligned}$$

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 (28) that the left hand side of (29) converges to 1.

Now we turn to (ii) which we prove by contradiction as well: Let  $n_j$  be a subsequence such that  $n_j F(\theta_{k,n_j})$  is bounded from above by  $\kappa > 0$ . Observe that

$$\frac{kV}{\theta_{k,n_j}} \binom{n_j}{k} F(\theta_{k,n_j})^k (1 - F(\theta_{k,n_j}))^{n_j-k} = \sum_{i=k}^{n_j} \binom{n_j}{i} F(\theta_{k,n_j})^i (1 - F(\theta_{k,n_j}))^{n_j-i} \leq 1 \quad (30)$$

Thus, we can conclude that the right hand side of the following inequality is bounded in  $j$ :

$$\frac{1}{\theta_{k,n_j}} \binom{n_j}{k} F(\theta_{k,n_j})^k (1 - F(\theta_{k,n_j}))^{n_j-k} \geq \binom{n_j}{k} \frac{1}{\theta_{k,n_j}} F(\theta_{k,n_j})^k \left(1 - \frac{\kappa}{n_j}\right)^{n_j-k}$$

As  $(1 - \frac{\kappa}{n_j})^{n_j-k}$  converges to  $\exp(-\kappa)$ , it follows from the above inequality that

$\binom{n_j}{k} \frac{1}{\theta_{k,n_j}} F(\theta_{k,n_j})^k$  is bounded. Since we have

$$\lim_{j \rightarrow \infty} \binom{n_j}{k} \frac{1}{n_j^k} = \lim_{j \rightarrow \infty} \frac{1}{k!} \prod_{i=0}^{k-1} \frac{n_j - i}{n_j} = \frac{1}{k!}, \quad (31)$$

we know that  $n_j^k \frac{1}{\theta_{k,n_j}} F(\theta_{k,n_j})^k$  is bounded. As  $\theta_{k,n_j} \rightarrow 0$ , it follows that  $F(\theta_{k,n_j}) n_j \rightarrow 0$ . It remains to show that this leads to a contradiction: We use the following identity (see, e.g., Chapter 1 of Johnson, Kotz and Kemp (1992)), which holds for all  $1 \leq k \leq n, p \in [0, 1]$ :

$$\sum_{i=0}^{k-1} p^i (1-p)^{n-i} = \frac{n!}{(n-k)!(k-1)!} \int_0^{1-p} u^{n-k} (1-u)^{k-1} du.$$

From this we deduce (for  $p = 0$ , both sides are equal to 1) that

$$\sum_{i=k}^n p^i (1-p)^{n-i} = \frac{n!}{(n-k)!(k-1)!} \int_{1-p}^1 u^{n-k} (1-u)^{k-1} du.$$

On the right hand side we substitute  $u \mapsto 1-u$ :

$$\sum_{i=k}^n p^i (1-p)^{n-i} = \frac{n!}{(n-k)!(k-1)!} \int_0^p (1-u)^{n-k} u^{k-1} du.$$

We can plug this into (30) to obtain:

$$\begin{aligned} V \frac{1}{\theta_{k,n_j}} F(\theta_{k,n_j})^k (1 - F(\theta_{k,n_j}))^{n_j-k} &= \int_0^{F(\theta_{k,n_j})} p^{k-1} (1-p)^{n_j-k} dp \\ &= \int_0^1 F(\theta_{k,n_j})^k p^{k-1} (1 - pF(\theta_{k,n_j}))^{n_j-k} dp \end{aligned} \quad (32)$$

where in the last equality we substituted  $p \mapsto F(\theta_{k,n_j})p$ . For any  $p \in [0, 1]$ ,

$$1 \geq (1 - pF(\theta_{k,n_j}))^{n_j-k} \geq \left(1 - \frac{F(\theta_{k,n_j})n_j}{n_j}\right)^{n_j-k} \rightarrow \exp(-0) = 1$$

as  $F(\theta_{k,n_j})n_j \rightarrow 0$ . Thus,  $(1 - pF(\theta_{k,n_j}))^{n_j-k}$  converges uniformly in  $p \in [0, 1]$  to  $e^{-0} = 1$  for  $j \rightarrow \infty$ . This implies that

$$\lim_{j \rightarrow \infty} \int_0^1 p^{k-1} (1 - pF(\theta_{k,n_j}))^{n_j-k} dp = \frac{1}{k} \quad (33)$$



and

$$\lim_{j \rightarrow \infty} (1 - F(\theta_{k,n_j}))^{n_j - k} = 1. \quad (34)$$

If we now look at (32), the left hand side is  $\sim V \frac{1}{\theta_{k,n_j}} F(\theta_{k,n_j})^k$  by (34), while by (33) the last term in the equation  $\sim \frac{1}{k} F(\theta_{k,n_j})^k$  for  $j \rightarrow \infty$ . As  $\theta_{k,n_j} \rightarrow 0$  for  $j \rightarrow \infty$ , this is a contradiction.  $\square$

*Proof of Proposition 8.* Fix  $k$  and  $V$ . Let  $\kappa > 0$  be arbitrary. By Proposition 7,  $nF(\theta_{k,n}) \rightarrow \infty$ . Thus, for  $n$  sufficiently large,  $nF(\theta_{k,n}) \geq \kappa$ . Since

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

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

$$\begin{aligned} P_{k,n} &= \sum_{i=k}^n \binom{n}{i} F(\theta_{k,n})^i (1 - F(\theta_{k,n}))^{n-i} \geq \sum_{i=k}^n \binom{n}{i} \left(\frac{\kappa}{n}\right)^i \left(1 - \frac{\kappa}{n}\right)^{n-i} \\ &= 1 - \sum_{i=0}^{k-1} \binom{n}{i} \left(\frac{\kappa}{n}\right)^i \left(1 - \frac{\kappa}{n}\right)^{n-i} \end{aligned} \quad (35)$$

Arguing as in (31), we also know that for fixed  $i$ ,

$$\lim_{n \rightarrow \infty} \binom{n}{i} \frac{1}{n^i} = \frac{1}{i!}$$

Thus, for  $n \rightarrow \infty$ , (35) converges to

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

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

*Proof of Proposition 9.* The proof is parallel to the one of Proposition 1: It suffices to note that  $\frac{(V-c(\theta))}{\theta}$  is decreasing in  $\theta$  and positive for small  $\theta$  since  $V - c(0) > 0$ .  $\square$

*Proof of Proposition 10.* We first prove  $\lim_{n \rightarrow \infty} \theta_{k,n} = 0$ . Denote by  $\theta'_{k,n}$  the

(unique) nonnegative solution of the equation

$$\frac{kV}{\theta'_{k,n}} b(n, k, F(\theta'_{k,n})) = \sum_{j=k}^n b(n, j, F(\theta'_{k,n})),$$

i.e., of the equilibrium condition of the basic model with benefit  $V$ . Let us assume  $\theta_{k,n} > \theta'_{k,n}$ . Then we have

$$\frac{k(V - c(\theta_{k,n}))}{\theta_{k,n}} < \frac{kV}{\theta'_{k,n}}$$

and as

$$\frac{\sum_{j=k}^n b(n, j, F(\theta))}{b(n, k, F(\theta))} = \sum_{j=k}^n \binom{n}{j} / \binom{n}{k} \left( \frac{F(\theta)}{1 - F(\theta)} \right)^{j-k}$$

is weakly increasing in  $\theta$ , it follows that

$$\frac{\sum_{j=k}^n b(n, j, F(\theta'_{k,n}))}{b(n, k, F(\theta'_{k,n}))} \leq \frac{\sum_{j=k}^n b(n, j, F(\theta_{k,n}))}{b(n, k, F(\theta_{k,n}))}$$

which yields a contradiction to the definition of  $\theta_{k,n}$ . So we have  $\theta_{k,n} \leq \theta'_{k,n}$ . We know from Proposition 7 (i) that  $\theta'_{k,n} \rightarrow 0$ , so it follows that  $\theta_{k,n} \rightarrow 0$ .

Now let us prove that  $\lim_{n \rightarrow \infty} nF(\theta_{k,n}) = \infty$ . Since  $\theta_{k,n} \rightarrow 0$ , and  $c(0) < V$  there exists an  $\varepsilon > 0$  such that  $V - c(\theta_{k,n}) > \varepsilon$  for large  $n$ . Denote by  $\theta''_{k,n}$  the (unique) nonnegative solution of

$$\frac{k\varepsilon}{\theta''_{k,n}} b(n, k, F(\theta''_{k,n})) = \sum_{j=k}^n b(n, j, F(\theta''_{k,n})),$$

i.e., of the equilibrium condition of the basic model with benefit  $\varepsilon$ .

We know from Proposition 7 (ii) that  $nF(\theta''_{k,n}) \rightarrow \infty$ . By similar reasoning as above,  $\theta''_{k,n} \leq \theta_{k,n}$  for large  $n$ , so it follows that  $nF(\theta_{k,n}) \rightarrow \infty$ . Finally, analogously to the proof of Proposition 8 we get  $\lim_{n \rightarrow \infty} P_{k,n} = 1$ .  $\square$

*Proof of Proposition 11.* 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

$$V \sum_{j=k}^{n-1} b(n-1, j, F(\theta)) = V \sum_{j=k-1}^{n-1} b(n-1, j, F(\theta)) - \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)) \quad (36)$$

i.e.

$$Vb(n-1, k-1, F(\theta)) = \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)). \quad (37)$$

The right hand side is nonnegative, 0 for  $\theta = 0$ , and diverges to  $\infty$  for  $\theta \rightarrow \infty$ . The left hand side is bounded, continuous in  $\theta$  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  $\theta$  of equation (36) is the threshold of a symmetric Bayesian equilibrium, where all agents vote “Yes”, if  $x_i \leq \theta$  and “No”, if  $x_i > \theta$ , as the expected utility of an agent voting “Yes” decreases in the private type  $x_i$ .

Now let us turn to the second assertion. Analogously as in Proposition 1 all symmetric Bayesian equilibria are equilibria in threshold strategies. The case threshold =  $\infty$  is trivial. Let us look again at (37). We want to show that this equation has no solution  $\theta > 0$  for small  $V$ . For large  $\theta$ ,

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

Thus, by continuity in  $\theta$ , 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

$$\frac{\theta}{s(1)} (1 - F(\theta))^{n-1} < \sum_{j=0}^{n-1} \frac{\theta}{s(1+j)} b(n-1, j, F(\theta)) \quad (38)$$

with

$$Vb(n-1, k-1, F(\theta)) = V \binom{n-1}{k-1} F(\theta)^{k-1} (1 - F(\theta))^{n-k} \leq V \binom{n-1}{k-1} F(\theta) \quad (39)$$

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

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

for small  $\theta$  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 (38) and (39), 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 (37) with at least one solution satisfying  $F(\theta) < 1$ . Thus, the two solutions correspond to two distinct threshold strategies.

Note that

$$\sum_{j=0}^{n-1} \frac{1}{s(1+j)} b(n-1, j, F(\theta)) \quad (40)$$

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

Now let us denote by  $\theta_{k,n}^{(1)} < \theta_{k,n}^{(2)}$  two solutions of (36) which correspond to two different threshold strategies, i.e.  $F(\theta_{k,n}^{(1)}) < F(\theta_{k,n}^{(2)})$ . We show that the equilibrium associated with the second threshold strategy is strongly Pareto superior to the first. Note that this argument covers the trivial equilibrium  $\theta_{k,n}^{(1)} = 0$ .

Let us consider a player of type  $x \leq \theta_{k,n}^{(1)}$ . In both equilibria, he votes “Yes”, the expected payoffs are

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

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 (40). The case  $\theta_{k,n}^{(2)} < x$  can be treated similarly. In the case  $\theta_{k,n}^{(2)} < x \leq \theta_{k,n}^{(2)}$  we have to compare the expected payoffs

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

and

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

We use  $F(\theta_{k,n}^{(1)}) < F(\theta_{k,n}^{(2)})$ , to show

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

which proves the claim. □

*Proof of Proposition 12.* For  $s(y) \equiv 1$  and  $k = 1$  the equilibrium condition (3) reduces to

$$\rho_{1,n} = V(1 - F(\rho_{1,n}))^{n-1}.$$

A unique solution  $\rho_{1,n}$  exists due to Proposition 1. That  $\rho_{1,n}$  decreases in  $n$  follows from the fact that  $(1 - F(\rho))^{n-1}$  is decreasing in  $n$  and weakly decreasing

in  $\rho$ . To see that  $\rho_{1,n} \rightarrow 0$ , suppose otherwise, i.e., there exists  $\kappa > 0$  and a subsequence  $n_j$  such that  $\rho_{1,n_j} \geq \kappa$  for all  $j$ . Then

$$\rho_{1,n_j} = V(1 - F(\rho_{1,n_j}))^{n_j-1} \leq V(1 - F(\kappa))^{n_j-1}$$

and the fact that the right hand side converges to 0 in  $j$  gives a contradiction. Moreover, we have

$$Q_{1,n} = 1 - (1 - F(\rho_{1,n}))^n = 1 - (1 - F(\rho_{1,n})) \frac{\rho_{1,n}}{V} \rightarrow 1.$$

□

*Proof of Proposition 13.* For  $s(y) \equiv 1$  and  $k > 1$  the equilibrium condition (3) becomes

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

Substituting  $\bar{n} = n-1$ ,  $\bar{k} = k-1$  and  $\bar{V} = V/\bar{k}$ , this becomes

$$\bar{k}\bar{V} b(\bar{n}, \bar{k}, F(\rho_{k,n})) = \rho_{k,n} \sum_{j=\bar{k}}^{\bar{n}} b(\bar{n}, j, F(\rho_{k,n})).$$

Since this condition is identical to condition (5) satisfied by  $\theta_{\bar{k}, \bar{n}}(\bar{V})$ , the result follows. □

*Proof of Proposition 14.* Define  $\bar{V} = V/(k-1)$ . We know from Proposition 7 that  $\theta_{k-1, n-1}(\bar{V}) \rightarrow 0$  and  $(n-1)F(\theta_{k-1, n-1}(\bar{V})) \rightarrow \infty$ . This immediately yields  $\rho_{k,n}(V) = \theta_{k-1, n-1}(\bar{V}) \rightarrow 0$  and  $nF(\rho_{k,n}(V)) \geq (n-1)F(\theta_{k-1, n-1}(\bar{V})) \rightarrow \infty$ . This proves (i) and (ii). In general, we have  $Q_{k,n}(V) \neq P_{k-1, n-1}(\bar{V})$  and thus the behavior of  $Q$  cannot simply be concluded from that of  $P$ . However, arguing exactly as in the proof of Proposition 8 we can conclude (iii) from (i) and (ii). □

## References

- [1] Admati, Anat R., and Motty Perry. Joint projects without commitment. *Review of Economic Studies*, 58, 259-276, 1991.

- [2] Austen-Smith, David, and Jeremy Banks. Information Aggregation, Rationality and the Condorcet Jury Theorem. *American Political Science Review*, 90, 34-45, 1996.
- [3] Bandura, Albert. Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193-209, 1999.
- [4] 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.
- [5] Bartling, Björn, and Urs Fischbacher. Shifting the Blame: On Delegation and Responsibility. *Review of Economic Studies* 79, 67-87, 2012.
- [6] Battigalli, Pierpaolo, Gary Charness, and Martin Dufwenberg. Deception: The Role of Guilt, *Journal of Economic Behavior and Organization*, 93, 227-232, 2013.
- [7] Battigalli, Pierpaolo, and Martin Dufwenberg. Guilt in Games, *American Economic Review*, 97, 170-76, 2007.
- [8] Baumeister, Roy F., Arlene M. Stillwell, and Todd F. Heatherton. Guilt: An Interpersonal Approach. *Psychological Bulletin*, 115, 243-267, 1994.
- [9] Bénabou, Roland, and Jean Tirole. Identity, Morals and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, 126, 805-855, 2011.
- [10] Bowles, Samuel. Endogenous Preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature* 36, 75-111, 1998.
- [11] Browning, Christopher R. *The Origins of the Final Solution*. William Heinemann, London, 2004.
- [12] Clauset, Aaron, Cosma R. Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. *SIAM Review* 51, 661-703, 2009.
- [13] 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.

- [14] Dana, Jason, Roberto A. Weber, and Jason X. Kuang. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33, 67-80, 2007.
- [15] Dufwenberg, Martin, and Amrish Patel. Reciprocity networks and the participation problem. *Games and Economic Behavior*, forthcoming, 2015.
- [16] Falk, Armin, and Nora Szech. Morals and Markets. *Science*, 340 (6133), 707-711, 2013a.
- [17] Falk, Armin, and Nora Szech. Organizations, Diffused Pivotality and Immoral Outcomes. Working Paper, 2013b.
- [18] Feddersen, Timothy, and Wolfgang Pesendorfer. Voting Behavior and Information Aggregation in Elections with Private Information. *Econometrica*, 65, 1029-1058, 1997.
- [19] Fischer, Peter, Joachim I. 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.
- [20] Geanakoplos, John, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1), 60-79, 1989.
- [21] Goldhagen, Daniel, Christopher Browning, and Leon Wieseltier. The 'Willing Executioners'/'Ordinary Men' Debate. In: *Selections from the Symposium*, United States Holocaust Research Institute, Washington D.C., 1996.
- [22] Grossman, Dave. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Little Brown, Toronto, 1996
- [23] 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.



- [24] Hamman, John R., George Loewenstein, and Roberto A. Weber. Self-interest through delegation: An additional rationale for the principal-agent relationship. *American Economic Review*, 100, 1826-46, 2010.
- [25] Harrington, Joseph E. A simple game-theoretic explanation for the relationship between group size and helping. *Journal of Mathematical Psychology*, 45, 389-392, 2001.
- [26] Huck, Steffen, and Kai A. Konrad. Moral cost, commitment and committee size. *Journal of Institutional and Theoretical Economics*, 161(4), 575-588, 2005.
- [27] Johnson, Norman L., Samuel Kotz, and Adrienne W. Kemp. *Univariate discrete distributions*. 2nd edition, Wiley, New York, 1992.
- [28] Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. *Continuous univariate distributions, Vol.2*. 2nd edition, Wiley, New York, 1994.
- [29] Latané, Bibb, and Steve Nida. Ten years of research on group size and helping. *Psychological Bulletin*, 89, 308-324, 1981.
- [30] Li, Hao, and Wing Suen. Decision-Making in Committees. *Canadian Journal of Economics*, 42, 359-392, 2009.
- [31] Lindbeck, Assar, Sten Nyberg, and Jörgen W. Weibull. Social norms and economic incentives in the welfare state. *Quarterly Journal of Economics*, 114, 1-35, 1999.
- [32] Nitzan, Shmuel, and Richard E. Romano. Private provision of a discrete public good with uncertain cost. *Journal of Public Economics*, 42, 357-370, 1990.
- [33] Osborne, Martin J. *An introduction to game theory*. Oxford University Press, Oxford, 2004.
- [34] Palfrey, Thomas R., and Howard Rosenthal. Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics*, 24, 171-193, 1984.
- [35] Sobel, Joel. *Markets and Other-Regarding Preferences*. Working Paper, 2010.

- [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.