Voting and Contributing when the Group Is Watching

Emeric Henry, Charles Louis-Sidois

Voting and contributing when the group is watching

Emeric Henry and Charles Louis-Sidois*

June 11, 2019

Abstract

Members of groups and organizations often have to decide on rules that regulate their contributions to common tasks. They typically differ in their propensity to contribute and often care about the image they project: in particular, they want to be perceived by other group members as being high contributors. In such environments we study the interaction between how members vote on rules and their subsequent contribution decisions. We show that making contributions visible affects the calculus of reputation and the voting decisions, and can be welfare decreasing as it makes some rules more likely to be rejected.

JEL Classification: D71, D72, H41, D23

Keywords: image concern, voting, public good

1 Introduction

Most members of groups and organizations (firms, NGOs, academic departments...) choose the rules that govern their interactions, in particular those regulating tasks with group externalities, such as attending meetings, writing reports or participating in team work. Similarly, many countries vote in referenda on sanctions for those who impose negative externalities on others. In this paper we study how individuals vote on rules that regulate their contributions (and those of other group members) to a public good.

A key feature that varies across organizations is whether actions of group members are visible. Individual public good contributions, such as meeting attendance,

*Henry: Sciences Po Paris, 28 rue des Saints-Pères, 75007 Paris, France. Louis-Sidois: University of Mannheim, Collaborative Research Center 884 “Political Economy of Reforms”, B6, 30-32, 68131 Mannheim, Germany.

are typically observed by the rest of the group. For other types of public goods, like group projects, the individual contributions may be harder to observe. The visibility of actions is particularly relevant in our setting where we assume that individuals are image concerned, i.e. care about how other group members perceive them.[1] We show how the visibility of actions affects the interaction between voting behavior and contribution decisions. When contributions are visible, a sanction, by increasing overall contributions, decreases the honor derived from being seen contributing. Thus, the outcome of the vote affects the calculus of reputation.

To study the organization of such groups, we analyze a model that features two stages involving the same group of players. In the second stage, players simultaneously choose whether to contribute or not to a public good. Each individual costly contribution provides a positive externality to the rest of the group. Group members are heterogenous in their propensity to contribute, what we call their type. In the first stage, the same players vote on a given sanction, to be imposed in the second stage on non contributors.

Section 3 considers a simple example (special case of our general model), featuring a group of two players and two possible types. We highlight three main lessons. First, high contributors can vote against a sanction when contributions are visible, to avoid losing their good reputation. Second, visibility of contributions, even though it encourages contributions in the second stage, can make sanctions more difficult to adopt because of reputational payoffs. Third, as a consequence, visibility of actions may have a detrimental effect on welfare.

To determine under which conditions these conclusions hold in the general model, we first set the stage in Section 4 and study the benchmark case where contributions to the public good are not observable by other group members. In the public good contribution stage, for a given level of sanction, three categories of members emerge. Those with a high type, called always-participants, contribute regardless of whether the sanction was voted or not. Those with a low type, called never-participants, never contribute. Intermediate types, called swing-participants, contribute if and only if the sanction was approved.

In the voting stage, members are inclined to vote in favor to benefit from the

[1]Image concerns have been shown empirically to be an important driver of contributions to common tasks. For instance Ariely et al. (2009) show that efforts made to contribute to a good cause are much higher when individuals are observed by others. See also DellaVigna et al. (2012) and Andreoni and Petrie (2004), Rege and Telle (2004), Samek and Sheremeta (2014), Henry and Sonntag (2019) for evidence from the field.

increased contributions of other group members, but swing and never participants trade this gain off against the cost of paying the sanction or of contributing. We show that the equilibrium is of the cutoff form where members vote for the sanction if and only if their type is above a cutoff value.

The main focus of the paper is to see how visibility of actions impacts voting, contributions to the public good and ultimately welfare. We thus turn in Section 5 to environments where contributions are public. In the public good contribution phase, the same three categories (always, never and swing participants) emerge. The composition of these groups is however affected. The contribution cutoffs are lower than in the benchmark model: because of image concerns, group members are more inclined to participate.

In the voting phase, the behavior is very different. While in the benchmark, always participants voted in favor of sanctions, we show that it can be a dominant strategy for them to vote against when contributions are public. Indeed, even though the sanction will never apply to them, they lose in reputation since contributing is no longer such a rare event when a sanction is in place that it signals a high intrinsic value. If the sensitivity to reputation is sufficiently high, these members vote against the sanction. However, when the externality gain is big enough, they vote in favor and we show that the equilibrium of the voting game is still characterized as in the benchmark case by a cutoff.

In Section 6, we turn we analyze how the visibility of actions affects overall welfare. We consider a social planner who chooses the sanction before the start of the game without observing individual types. If the sanction was not submitted to a vote, this would be a classical problem of regulation of an externality and the planner would choose a sanction equal to the externality e. However, since we consider environments where sanctions are approved by a vote, the planner chooses a sanction higher than e to increase the probability of acceptance, at the cost of potentially making some members inefficiently contribute. How does visibility affect welfare? Making contributions visible may push participants to vote against the sanction since they lose in reputation if it is accepted. By affecting the calculus of reputation, visibility of contributions can be welfare reducing.

Finally, we consider in Section 7 a variation of the model where members vote on a bonus for contributors instead of a sanction for free riders. We show that the intuitions are similar and that our main results still hold. In particular, although all members are in favor of bonuses in the all secret environment, high contributors might vote against when contributions are public in order to preserve their reputation. This also implies that making contributions public can decrease welfare.

Related literature

Our paper considers environments where actions are visible which in our setting where group members are image concerned implies subtle interaction between the voting and contribution stages.[2] The voting stage can be seen as shaping the social norm that governs the second stage. In that sense we are closely connected to Bénabou and Tirole (2011) who examine a public good problem, very similar to the second stage of our model, and show how the calculus of honor and stigma can be derived. Their key focus is on how an informed principal can optimally set incentives. The key distinction is that in our setting the sanction is submitted to a vote, even if optimally chosen by the planner as in Section 6, and this voting stage sets the norm.[3] In a closely related paper Ali and Bénabou (2018), also building on the framework of Bénabou and Tirole (2011), examine whether a social planner, not perfectly informed about societal values, should use social image (praising or shaming) to spur contributions to a public good. Visibility in their context creates a tradeoff between a positive effect on contributions due to image concerns, but a signal jamming effect since image concerns prevent individuals from expressing their true motivations. While we abstract from the learning effect, our focus is on the interaction between visibility of contributions and voting incentives.[4]

The fact group members in our model care about the image they project connects us to the literature on aggregation of information in committees when individuals have career concerns. The key distinction between our environment and the career concern literature (Ottaviani and Sørensen, 2001, Visser and Swank, 2007) is that in our model agents take initial actions in order to enhance the reputational value of future actions. For instance in Midjord et al. (2017), privately informed agents vote on an approval decision and get a negative reputation payoff (of fixed value) if the outcome is to approve and the state was in fact bad.[5]

[2]There is also a related literature on whether votes should be made public or kept secret, see Levy (2007), Gersbach and Hahn (2008) and Mattozzi and Nakaguma (2017).

[3]See also Acemoglu et al. (2012) who examine the interaction between laws and norms in settings where laws are endogenously enforced by the community and Levine and Mattozzi (2017) who consider the endogenous setting of norms by party leaders to encourage turnout or Ali and Lin 2013

[4]See also Godefroy and Perez-Richet (2013).

[5]Our paper is also connected to the literature on endogenous constitutions (seminal paper by Barbera and Jackson 2004, followed for instance by Acemoglu et al. 2012). One defining feature of

Finally, there exists a sizeable experimental literature that studies the difference between exogenously and endogenously set sanctions on future behavior. Part of the literature (Galbiati and Vertova, 2008 and Galbiati et al., 2013) examines the case where the designer who decides on the sanction is informed, contrary to our model. Tyran and Feld (2006) consider an experimental setting closer to our model and show that if the group votes for the sanction (rather than have a sanction exogenously imposed), it is followed by higher contributions.

2 Model

We consider a two stage game involving a group of $n+1$ players. In the first stage, a rule (or law in certain contexts) is submitted to a vote. The rule specifies a sanction $s > 0$ (given to the group) that will be imposed in case of free riding in the public good stage that follows.[6]

In the first stage, all players cast their vote simultaneously. The voting decision of individual i is denoted $b_i \in \{0, 1\}$ (where b stands for ballot). If strictly more than k players vote in favor, the sanction is adopted. The outcome of the vote is publicly revealed and the players then simultaneously decide, in a second stage of the game, whether to contribute or not to the public good. Individual i's contribution is denoted a_i, where $a_i \in \{0, 1\}$.

For a given approved sanction s and a given vector of contributions to the public good $a = \{a_1, a_2, ..., a_{n+1}\}$, the utility of player i is given by:

$$U_i = (v_i - c)a_i - s(1 - a_i) + e\frac{\sum_{j \neq i} a_j}{n} + \mu\mathbb{E}[v_i|y_i]. \tag{1}$$

Individual i gets an intrinsic benefit of contributing to the public good, denoted v_i, which characterizes the type of the individual. This intrinsic motivation (as in Bénabou and Tirole, 2011) can in particular be linked to the player's level of altruism, since contributions benefit other group members.[7] The v_i are i.i.d. drawn from the continuously differentiable density $f(v)$ with support $[v_{min}, v_{max}]$ and privately observed.

our model, that differentiates it from that literature, is that members are privately informed about their propensity to contribute.

[6]The case of a bonus for contributing is discussed in Section 7.

[7]It could also represent the efficiency of the individual in providing the public good. The only important feature is that a higher value of v_i is viewed positively by the rest of the group.

The utility function presented in (1) also includes a cost of contribution c common to the whole population. If a sanction is in place, there will be an additional cost for those not contributing $s(1-a_i)$.[8] In addition, individuals benefit from the contributions of other group members, i.e there is an externality gain $e\frac{\sum_{j\neq i} a_j}{n}$.

Finally, agents are image concerned and want to be perceived as intrinsically motivated, which is captured by the component $\mu\mathbb{E}[v_i|y_i]$. Individual actions y_i reveal information on the underlying value of v_i, the intrinsic motivation of each agent. In the benchmark model of Section 4, we consider groups where no individual action, neither the vote nor the contribution, is observable. In section 5 we study the case where the vote is secret but contributions are observable, i.e $y_i = a_i$.

We make the behavioral assumption that image $\mathbb{E}[v_i|y_i]$ is based only on observed individual actions y_i and not on inferences based on aggregate outcomes. For instance we assume that if individual votes are not observable, the inferences on v_i that could be drawn from the overall result of the vote are not used to update the image. Similarly if individual contributions are kept secret, inferences based on the aggregate level of contributions are not used. Instead, Levy (2007) considers a career concerns model where an outside principal uses the aggregate result of the vote to infer the type of players. The problem is further complicated in our setting since the players can update both based on aggregate results of the vote but also aggregate contributions, with intricate interactions between the two. We show in the context of the simple example that our results are not affected by removing this behavioral assumption and also highlight the technical complications it induces.

To sum up, the timing of the game is the following:

1. Types v_i are i.i.d drawn and privately observed.

2. Players vote on the rule with no abstention. The outcome of the vote is publicly revealed.

3. Players then simultaneously decide on their contribution decision.

We focus on symmetric Perfect Bayesian Nash equilibria where players with the same type choose the same strategy.

[8]From the point of view of group members, the sanction is a pure loss, in particular is not redistributed to the group.

3 A simple example

To provide the main intuitions of our results, we start by studying a special case of our model. Consider a group made up of two members who can be either of type H with value $v_H > 0$ from contributing or type L with $v_L = 0$ (i.e the density $f(v)$ is degenerate). The prior probability of type H is p. A sanction $s > c$ is submitted to a vote and is adopted if at least one of the group members votes in favor. As in the rest of the analysis we compare the situation where contributions are secret to the case with public contributions. We consider two restrictions on parameters

1. $v_H > c$, i.e type H has private incentives to contribute when contributions are hidden.

2. $c > e$, i.e the externality generated for the other player does not compensate for the cost of contributing.

Second stage: contribution decision

Since $v_H > c$, there is an equilibrium where type H contributes, regardless of whether sanctions were voted, and independently of the visibility of contributions. Type L contributes if the sanction was voted, since the sanction is higher than the cost of contributing ($s > c$). However, in the case where the sanction was rejected, the decision of the low type could depend on the visibility of actions. If contributions are secret, type L does not contribute. On the contrary, if contributions are visible, image concerns can potentially induce the low type to contribute. However, to keep the exposition simple, we rule out this possibility by imposing:

$$-c + \mu v_H < 0 \tag{2}$$

i.e. even if the player obtains the reputation payoff of the high type, this is not sufficient to cover the cost of contributing.

In summary, given our restrictions on parameters, type H always contributes and type L contributes if and only if the sanction is passed.

First stage: voting decision

We now examine the voting stage. Consider first the case of secret contributions. In this case, it is a dominant strategy for type H to vote in favor of the sanction, since it induces the low type to contribute. Given the restriction $c > e$, it is a dominant strategy for the low type L to vote against, since it forces the player to incur the cost c of contributing and in return get at most a gain of e from inducing the other player to contribute if she is of type L.

We now turn to the situation where contributions are visible. We show that there exists an equilibrium where both types vote against the sanction. In such an equilibrium, type H knows her vote is pivotal since all players vote against the sanction. If she votes in favor, the sanction is accepted, and her payoff is:

$$v_H - c + \mu\bar{v} + e \tag{3}$$

where $\bar{v} = pv_H$ is the expected value of v. By voting in favor, she guarantees herself e since the other player contributes, but obtains the average reputation.[9]

If she votes against, the sanction is turned down, and she obtains

$$v_H - c + \mu v_H + pe \tag{4}$$

i.e. her reputation from contributing is higher since only high types contribute in the absence of sanctions, but she does not benefit from the externality if the other player is type L.

Thus, under the condition below, there is an equilibrium where types H vote against the sanction:

$$\begin{aligned} \mu v_H + pe &> \mu\bar{v} + e \\ \Leftrightarrow \mu v_H &> e \end{aligned}$$

In summary, we identified conditions such that types H can vote against the sanction when contributions are public to preserve their good image in the contribution phase, while they vote in favor when contributions are secret.

[9] In this example we clearly see that our behavioral assumption that image depends only on individually observed actions y_i plays a key role. In Supplementary Appendix A1, we relax this assumption in two directions in the context of the simple example. First, we assume that players use the vote tally to infer the vote of the other player (but does not observe the aggregate contributions). We show that this is equivalent to public voting as players can then perfectly infer the type of the other group member. Second, we study a setup where players want to signal their type to an outside observer who bases her beliefs on aggregate outcomes. In both cases, we can identify an equilibrium where our results still go through, but we face a larger multiplicity of equilibria. Moreover, in larger groups, inferences based on aggregate behaviors would be even more complicated. Our behavioral assumption allows us to focus on our key message on the impact of the visibility of contributions.

General lessons

Under the condition $e < \mu v_H < c < v_H$, there is an equilibrium where the sanction is always rejected when contributions are visible[10], while it is accepted with hidden contributions if there is at least one type H in the group. In this case with only two players, welfare is in fact higher under public contributions, since forcing the low type to contribute is not socially beneficial (cost c is greater than benefit e for the other player), but we provide a version with 3 players (Appendix A2) where the opposite is true.

This simple example highlights three main results of our analysis. First, high contributors can vote against a sanction when contributions are visible, to avoid losing their good reputation. Second, visibility of contributions, even though they encourage contributions in the second stage, can make sanctions more difficult to adopt in the first stage because of reputational payoffs. Third, as a consequence, visibility of actions may have a detrimental effect on welfare, as shown in the version with 3 group members. We now show how these ideas generalize for a general distribution of types and groups of arbitrary size.

4 Voting on sanctions

We start by studying the behavior in groups voting on sanctions to spur public good contributions in the case where individual contributions are not visible to other group members. This also corresponds to a benchmark model without image concerns ($\mu = 0$). In terms of notation, we use the superscript h (denoting the case where contributions are hidden) to describe all relevant equilibrium parameters.

We solve the game backwards and start with the second stage. For a given sanction s (where $s = 0$ corresponds to the case where voters turned down the sanction), contributing yields intrinsic benefits and costs. Not contributing on the contrary exposes individual i to the sanction. The equilibrium of the voting game is characterized as follows:

[10]In fact, we can show that with additional restrictions, all players voting against the sanction is the only equilibrium when contributions are public. First, we must have $v_H > c + \mu(pv_H)$. Our condition $v_H > c$ is not sufficient to rule out the following equilibrium of the contribution subgame: no player contribute and if a player deviates and contributes, he is perceived as a low type. Moreover, we must have $e < \mu v_H(1-p)$. When $\mu v_H(1-p) \leq e \leq \mu v_H$, there is also an equilibrium where v_H vote in favor while v_L vote against if contributions are visible. In such a strategy profile, v_H get $v_H - c + e + \mu\bar{v}$ while if they deviate and vote against, they get $v_H - c + p(e + \mu\bar{v}) + (1-p)\mu v_H$. The deviation is profitable if $e < \mu v_H(1-p)$.

Lemma 1 *The unique symmetric Perfect Bayesian Nash equilibrium of the public good stage is such that player i contributes if and only if $v_i \geq v_s^h$ where the cutoff is defined by*

$$v_s^h = c - s. \tag{5}$$

The cutoff is increasing in c, as a more costly contribution reduces the incentives to participate and decreasing in s, as a higher sanction raises the material cost of free-riding.

If the sanction s is implemented, players use in the contribution phase a strategy with cutoff v_s^h, as derived above. On the contrary if the sanction is rejected, the players use in equilibrium a strategy with cutoff denoted v_0^h (defined by equation (5) for $s = 0$) with $v_s^h < v_0^h$.

Given the equilibrium behavior in the public good stage, players can be grouped in three categories:

- *Never-participants* who do not contribute regardless of the outcome of the vote: members with $v_i < v_s^h$.
- *Swing-participants* who contribute if and only if the sanction is voted: members with $v_s^h \leq v_i \leq v_0^h$.
- *Always-participants* who always contribute regardless of the outcome of the vote: members with $v_i > v_0^h$.

These three categories of individuals have different motivations in voting, but they all benefit from the increased contribution of other group members, i.e from the expected externality gain G, defined as the difference between the expected externality obtained with a sanction and the expected externality obtained without:

$$G = e\frac{\mathbb{E}\left[\sum_{j\neq i} a_j | s > 0\right]}{n} - e\frac{\mathbb{E}\left[\sum_{j\neq i} a_j | s = 0\right]}{n}.$$

In equilibrium, the expected externality G is the same for all group members, regardless of their type (i.e. the same for a never, always or swing participant). Indeed, types are i.i.d drawn and therefore the expectation about other players actions $a_j, j \neq i$ are independent of i's type. We describe later in this section the exact calculation of G in equilibrium.

The never participants do not change their contribution decision even if the sanction is in place: a sanction implies for them a financial cost s. For these group members, the difference in expected utility comparing the situation with a sanction to the one without, that we denote $D(v_i)$, is given by $D(v_i) = -s + G$. For the always participants, the difference in expected utility is simply $D(v_i) = G > 0$; voting for the regulation is a dominant strategy for the always participants since they don't pay for the sanction but benefit from the increased contribution of other group members. Finally, for the swing participants $D(v_i) = v_i - c + G$.

For all types, the difference in utilities $D(v_i)$ expressed in the above conditions can be written as $D(v_i) \equiv R(v_i) + G$.[11] In Figure 1, we plot the function $-R(v_i)$, a decreasing function which suggests that symmetric Perfect Bayesian equilibria should be of the cutoff form, i.e equilibria characterized by a cutoff V^h such that a type v_i votes in favor if and only if $v_i \geq V^h$.

However, since all never participants, regardless of their particular type v_i, have the same voting incentive, in an equilibrium where never participants are indifferent between voting in favor or against, the identity of those voting for would not be uniquely pinned down. To limit the multiplicity of equilibria, we thus impose for the rest of the paper the following restriction:

Restriction A (tie breaking): *If in equilibrium two types $v_i > v_i^{'}$ are indifferent between voting in favor or against the sanction, then if type $v_i^{'}$ votes in favor, so does type v_i.*

We now describe the construction of G in equilibrium. As explained above, for a given voting cutoff V, G takes a unique value, identical for all groups. However, G is not necessarily monotonic in the voting cutoff V. We illustrate this in Figure 1 where in addition to $-R(v_i)$ we also plot the function $G(V)$ for the case where f is uniform[12] (the x axis is v_i for R and V for G). The equilibrium cutoff V^h is such that $-R(V^h) + G(V^h) = 0$ and thus corresponds to the intersection of $-R(v_i)$ and $G(V)$.

As in the literature on information aggregation in voting, voters consider only the case where their vote is pivotal.[13] Let's suppose that there is an odd number of

[11]with $R(v_i) = -s$ for never participants, $v_i - c$ for swing and 0 for always participants.

[12]Parameters used to plot all graphs used in the paper are presented at the end of the proof section.

[13]See for instance Austen-Smith and Banks, 1996, Feddersen and Pesendorfer, 1996, 1997, 1998 for the theory or Levine and Palfrey 2007, Battaglini et al. 2008 and Battaglini et al. 2010 for empirical or experimental evidence. To the best of our knowledge, in most of the papers in this literature, the benefits of the law submitted to a vote are exogenously given (but not publicly observed). In our public good setting, the benefit of the sanction is endogenously determined by how voters react to it. This leads to a multiplicity of equilibria not present in the rest of the

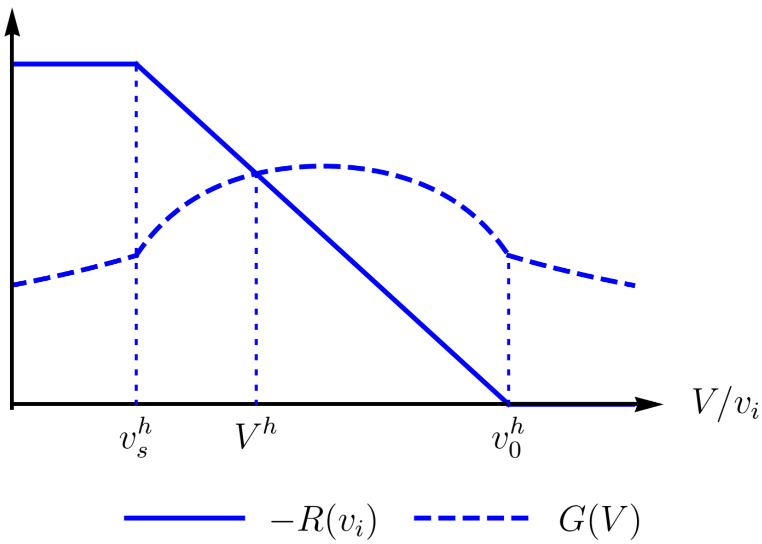

Figure 1: Voting cutoff in all secret environment

group members, i.e n is even. Under majority rule, a player is pivotal when there are exactly $n/2$ yes-voters and $n/2$ no-voters among the n other players. However, to determine the expected externality gain G (difference in externality with and without sanction), each voter only needs to determine the expected number of swing participants. Indeed they are the only types who change behavior based on whether the sanction is approved or not and they thus determine the added value of having a sanction in place.

Consider the case where the equilibrium cutoff is in the swing participant group, what we describe as an *interior equilibrium*. No-voters can either be swing participants or never participants. Specifically, given a voting cutoff V, the probability that a no voter is a swing participant is given by $\frac{F(V)-F(v_s^h)}{F(V)}$. As V increases, it becomes more likely that a no voter is in fact a swing participant. On the other hand, the probability that a yes voter is a swing participant (and not an always participant), is given by $\frac{F(v_0^h)-F(V)}{1-F(V)}$. This probability is decreasing in V. Overall, the expected externality gain is thus given by the following expression[14]:

literature with the exception of Callander (2008).

[14]In the same spirit, if the cutoff is among the never participants, i.e $V \leq v_s^h$, $G(V) = \frac{1}{2}e\left[\frac{F(v_0^h)-F(v_s^h)}{1-F(V)}\right]$ while if it is in the always participants, $V > v_0^h$, we have $G(V) = \frac{1}{2}e\left[\frac{F(v_0^h)-F(v_s^h)}{F(V)}\right]$.

$$G(V) = \frac{1}{n} e \left[\frac{n}{2} \left(\frac{F(V) - F(v_s^h)}{F(V)} \right) + \frac{n}{2} \left(\frac{F(v_0^h) - F(V)}{1 - F(V)} \right) \right].$$

As a function of V, G is first increasing and then decreasing. When V is close to v_s^h, the positive effect on the probability that a no voter is a swing participant dominates. As V moves closer to v_0^h, this first consideration becomes weaker and the negative effect on the probability that a yes voter is a swing participant drives the decrease in $G(V)$.

Depending on the shape of G, there could be a potentially large multiplicity of equilibria. To limit this multiplicity, we impose the following restriction on f, which implies that the externality gain $G(V)$ is a concave function on the interval (v_s^h, v_0^h).

Restriction B (type distribution): $\frac{f}{1-F}(v)$ *is weakly increasing and* $\frac{f}{F}(v)$ *weakly decreasing.*

Under Restriction A and B, that we impose for the rest of the paper, we obtain the following result:

Proposition 1 *When all actions are secret, there exists a unique symmetric Perfect Bayesian equilibrium, characterized by a cutoff* V^h.

Furthermore,

1. *the probability of approval is increasing in* e,
2. *the probability of approval is increasing in* s *if the equilibrium is interior.*

All equilibria are of the cutoff form, i.e. such that individual i votes in favor if and only if $v_i \geq V^h$. Given that $-R(v_i)$ is a (weakly) decreasing function, there is at most one equilibrium in the region where $G(V)$ is increasing. Restriction B that implies concavity of G thus guarantees the uniqueness of the equilibrium.

Proposition 1 also presents comparative statics on the voting cutoff. The cutoff naturally decreases in the externality parameter e since an increase in e increases G and thus makes voters more likely to vote in favor of the sanction. Similarly if the equilibrium is interior, i.e. the cutoff is within the swing voters, an increase in s increases G and does not affect R.[15]

[15] This would not necessarily be the case if the cutoff was in the never participant group since in that case an increase in s would also directly make the regulation more costly for the individual at the cutoff.

5 Visible contributions and image concerns

We now consider the case where individual contributions are visible by the rest of the group. In the second stage of the *public contribution* environment, players now consider the impact of their action on their reputation. We denote by the superscript p (public contribution) the equilibrium parameters in this case. We use the notation $\triangle(v_s^p) = \mathbb{E}[v|v > v_s^p] - \mathbb{E}[v|v < v_s^p]$ (used in Bénabou and Tirole 2011) for the net reputational incentive of being perceived as having a v_i above the participation cutoff v_s^p. As in the benchmark case, the equilibrium involves a cutoff such that only high types contribute, but the precise cutoff value is affected by the visibility of contributions.

Lemma 2 *The unique symmetric Perfect Bayesian Nash equilibrium of the public good stage is such that player i contributes if and only if $v_i \geq v_s^p$ where the cutoff is defined by*

$$v_s^p = c - s - \mu\triangle(v_s^p). \tag{6}$$

As in Bénabou and Tirole (2011), we impose the condition $1 + \mu\triangle'(v) > 0$ so that the voting cutoff is decreasing in s. The cutoff also decreases with the visibility of contribution (or taste for reputation) μ, since more pressure worsens the stigma attached to free-riding and thus provides incentives to contribute. The equilibrium of the public good stage is, like in the benchmark case, also characterized by three participation groups. However, since the participation cutoffs v_s^p and v_0^p are shifted to the left, the composition of the groups is now altered. There are now more always participants and fewer never participants while the impact on the size of the swing participants group is ambiguous.

We now turn to the voting stage. Even if the vote is secret, image concerns are still relevant to determine the equilibrium strategies: whether a sanction is voted or not shapes social norms. For the always participants, a sanction decreases the honor they derive from doing the right thing since more types will contribute in equilibrium. For this group, the incentives to vote in favor of the regulation, $D(v_i)$ is given by:

$$D(v_i) = \mu \underbrace{\left(\mathbb{E}[v_i|v_i > v_s^p] - \mathbb{E}[v_i|v_i > v_0^p]\right)}_{reputation\ loss\ <0} + G^p,$$

where G^p is the equivalent of G in the case of public contributions. Note that the functional forms are identical in the two cases but, given that G takes as arguments

$v_s^\cdot$ and $v_0^\cdot$, the values are different.

As opposed to the benchmark case with unobservable actions (where $D(v_i) = G$), always participants may now have an incentive to vote against the regulation in order to preserve their image. When considering their voting decision, they tradeoff the externality gain that a sanction would bring against the decrease in reputation. If e is low enough, the second effect dominates:

Proposition 2 *For any sanction s, there exists a value $\overline{e}(s)$ such that if $e \leq \overline{e}(s)$, it is a weakly dominant strategy for the always-participants to vote against the sanction.*

Proposition 2 shows that group members who in any case contribute to the public good, have a motive to vote against a sanction that would force the others to participate as well. From a policy perspective this result is important: even if the conditions for Proposition 2 are not met, the fact that these individuals always suffer from a loss of reputation if the sanction is passed, means that they have fewer incentives to support regulation than what could be expected at first sight.

Turning to the other groups, for the never participants, the regulation increases the stigma attached to not contributing because fewer people free-ride when the sanction is implemented:

$$D(v_i) = \mu \underbrace{(\mathbb{E}[v_i|v_i < v_s^p] - \mathbb{E}[v_i|v_i < v_0^p])}_{reputation\ loss\ <0} - s + G^p.$$

Finally, for the swing participants, the sanction implies a reputation gain. When not in place, they pool with the never participants and when it is implemented they cannot be distinguished from the always participants:

$$D(v_i) = \mu \underbrace{(\mathbb{E}[v_i|v_i > v_s^p] - \mathbb{E}[v_i|v_i < v_0^p])}_{reputation\ gain\ >0} + v_i - c + G^p.$$

Which group has the most incentives to vote in favor of the sanction? The answer is not straightforward. Consider for instance the comparison between never and always participants. It could a priori be the case that the loss in reputation for the always-participants be greater than for the never participants. We however show that in equilibrium, even if that were the case, the difference in reputation cannot be greater than s and the equilibrium of the voting stage is still characterized by a cutoff:

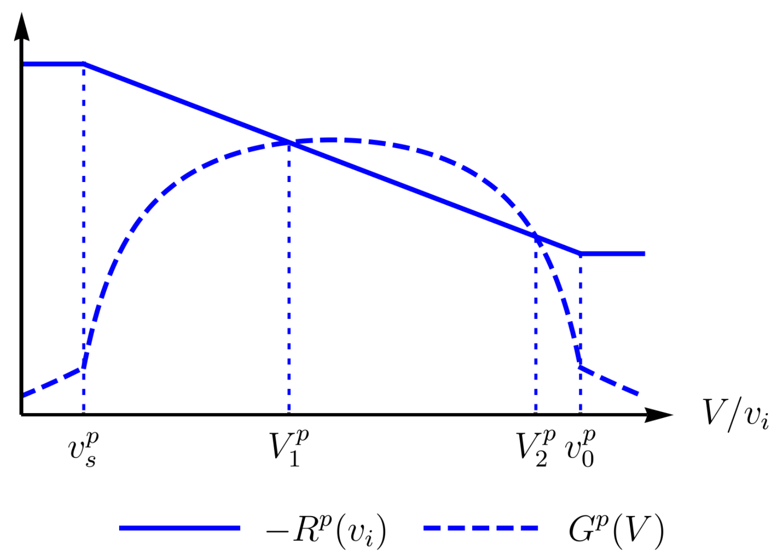

Figure 2: Voting cutoff in the public contribution environment

Proposition 3 *In the public contribution environment, all symmetric Perfect Bayesian equilibria are cutoff equilibria where players vote in favor if and only if $v_i \geq V^p$. Moreover there are at most two such equilibria and a unique stable interior equilibrium.*

Furthermore, there exists a benchmark $\bar{e}^p$ such that, if the stable equilibrium is interior, the voting cutoff V^p:

1. *is decreasing in e,*

2. *is decreasing in the level of sanction s if and only if $e \geq \bar{e}^p$.*

While restrictions A and B guaranteed a unique equilibrium in the benchmark case, we might now have a second equilibrium as illustrated in Figure 2, that represents a case with two equilibria with cutoffs V_1^p and V_2^p corresponding to the intersections of the functions $-R^p(v_i)$ and $G^p(V)$.[16] In equilibrium with cutoff V_1^p, the pivotal voter expects a large portion of yes voters (to the right of V_1^p) and of no voters (to the left of V_1^p) to be swing participants. The expected externality is thus

[16]The multiplicity of equilibrium was not possible in Section 4 because always participants had $-R(v_i) = 0$. Thus, the concavity of $G(V)$ was sufficient to guarantee the uniqueness of the equilibrium. Now we have $-R^p(v_i) = \mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_s^p] > 0$, which implies that a second equilibrium can exist.

large and justifies the low voting cutoff. On the contrary in the case of V_2^p, it is very unlikely that the yes voters are swing participants, the expected externality is thus lower, justifying the higher cutoff. These different equilibria can be understood as corresponding to different norms of voting. A norm of opposition to sanctions (high cutoff V_2^p) might prevail and would be based on a self realized expectation of low externality gain when a player is pivotal. There could also exist norms of voting more favorable to sanctions (lower cutoff V_1^p) based on an expectation of a high externality gain. Both these norms would be self sustained due to the mechanisms of information aggregation described above.

Proposition 3 also presents comparative statics on how voting cutoff and approval probabilities vary with s. The effect of s follows a different logic as in the case without reputation. An increase in s, like in that case, increases contributions in the second phase and thus increases the expected externality gain in the voting phase. Reputation however creates a countervailing effect as increasing s decreases the reputation gain enjoyed by a swing voter. This effect on reputation decreases incentives to vote in favor when the sanction is higher. Overall, the balance between these two effects is determined by the size of the externality e as expressed in Result 3.2: if e is small, increasing the sanction decreases the probability of acceptance.

6 Welfare analysis

After having described the equilibrium organization of groups voting on their own rules, we now turn to the welfare analysis of the different environments. We consider a planner who chooses, prior to the start of the game, the level of the sanction submitted to a vote but who does not observe the individual types of group members. We focus on sanctions that create no deadweight loss and require no enforcement costs to focus on the main tradeoffs. Finally we assume that the planner maximizes total welfare net of reputation concerns.[17] We start by a welfare analysis of the benchmark case where actions are not visible before studying whether visibility is welfare enhancing.

[17]This assumption is innocuous in the benchmark model where reputation does not matter. However, in the public contribution case, reputation is not zero sum because of our assumption that the aggregate result of the vote is not taken into account to update reputation. We did not want our welfare results to be driven by this assumption.

6.1 Benchmark model

To fix ideas, suppose the planner chooses the sanction without submitting it to a vote. This is a classic problem of regulation of an externality. Each individual contribution creates a positive externality of level e for the group. The first best requires that a group member i contributes if and only if $v_i + e \geq c$ and can thus be implemented in the decentralized equilibrium without voting using a sanction $s = e$.

When the planner needs to submit the sanction to a vote, the choice of the socially optimal sanction is affected in two ways. First, the level of the sanction affects the probability of approval. Second, conditional on acceptance, the expected composition of the group and thus the expected effect of sanctions on welfare depends on the level of s. In this case with no deadweight loss, submitting the sanction to a vote weakly decreases welfare.

Proposition 4 *In the* all secret *environment, the socially optimal sanction is always weakly higher than the optimal level without voting:* $s^h \geq e$. *Under unanimity rule, it is strictly higher* $s^h > e$.

In terms of welfare, setting a sanction different from e imposes an ex post cost in the contribution phase as it deviates from the socially optimal level without voting. However from an ex ante point of view, setting a sanction different from e is beneficial. To see that, first notice that if the sanction is set at the socially optimal level without voting, $s = e$, the voting cutoff is necessarily among the swing participants. Indeed, the expected externality gain G can never be greater than e, so the never participants, who would pay a sanction $s = e$, will necessarily vote against. The comparative static of Proposition 1 thus applies and the probability of acceptance of the sanction is increasing in s. A direct consequence is that the optimal sanction submitted to a vote is necessarily weakly greater than e.

In the case of unanimity rule, we show that the optimal sanction is in fact strictly greater than e. Given that unanimity is required, if the sanction is approved the group has to be such that all members have a type greater than V^h. Since $V^h > c - e$, there is no ex post cost from setting a higher sanction: those who would inefficiently contribute in the ex post phase because the sanction is set higher than e will vote against the sanction in the ex ante phase and thus can never be part of a group that approves. It follows that setting a sanction strictly higher than e is socially optimal. A similar logic applies as long as the majority required is sufficiently large.

6.2 The impact of visibility

We now derive the probability of acceptance of a sanction and ultimately welfare when contributions are observable, and compare the results with those in the benchmark case.

6.2.1 Simple example

To fix ideas, we present in Appendix A2, a version of the simple example of Section 3, but with 3 members in the group. As in Section 3, there are two types of members $v \in \{v_L, v_H\}, Pr(v = v_H) = p$ and $v_L = 0$. Moreover, $v_H > c$ so that high types always contribute. We also impose $-c + \mu v_h < 0$. We suppose that the externality is such that $e/2 < c < e$.

In the example, the first best is for all players to contribute since $c < e$ and can be attained with a sanction $s > e$. Such a sanction may be approved when contributions are hidden. However, in the case where contributions are public, even types H vote against the sanction as they lose too much in terms of reputation.

6.2.2 General case

We now examine how these results extend to the general case. For the same expected externality, a voter in the never participant group is more inclined to vote against the sanction if contributions are visible than if they are hidden because this voter loses in terms of reputation. This can be seen in Figure 3: $-R^p$ is above $-R$ for low v_i. The same is true for always participants. As expressed in Proposition 2, they are more inclined to vote against the sanction when contributions are visible since they derive less honor from contributing when the sanction is in place: $-R^p$ is above $-R$ for high v_i. In the intermediate zone, the ordering is reversed because swing participants benefit from a gain in reputation when contributions are public.

Visibility of contributions also affects the expected externality gain. The comparison of G^p and G depends on the expected number of swing participants among others conditional on a given voter being pivotal. To clarify the tradeoffs we focus in the next proposition on the case where f is uniformly distributed, which guarantees that the size of the swing participant group is the same with and without visibility of contributions: $v_0^h - v_s^h = v_0^p - v_s^p = s$. This implies, as shown in Figure 3 that G^p and G are equal when $v < v_s^p$ and when $v > v_0^h$, and that in the intermediate zone G^p is above for low values of v_i and below for high values. For instance when V is

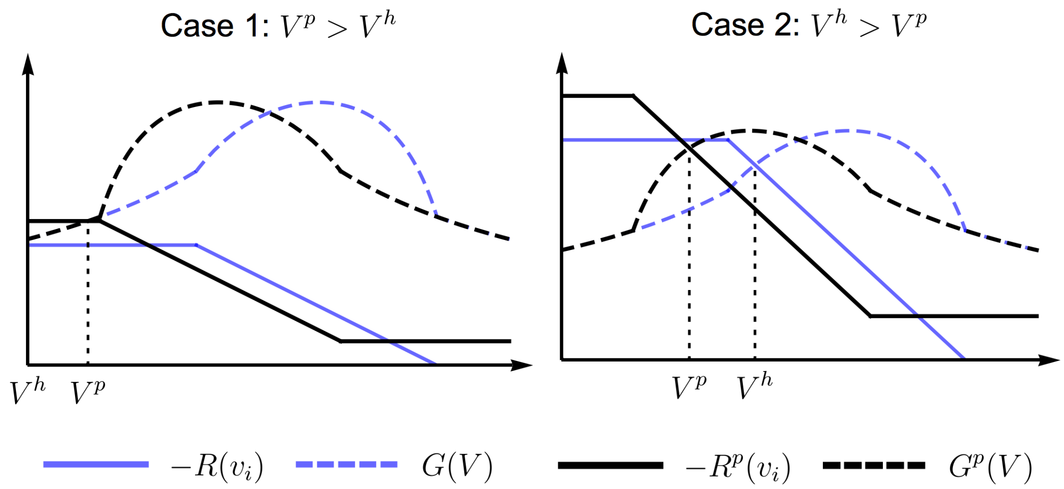

Figure 3: Comparing the cases *all secret* and *public contribution*

just above v_s^p and below v_s^h, in the public contribution case, some of the no voters can be swing participants and because of information aggregation, this increases the expected externality gain in this environment compared to the case of secrecy. This leads us to our formal result comparing the two environments.

Proposition 5 *Comparing the public contribution and all secret environments, if f follows a uniform distribution:*

1. *There exists e_l, e_m and e_h such that:*

   (a) *If $e > e_m$, the voting cutoff is lower under all secret ($V^h \leq V^p$) and strictly lower if $e \in (e_m, e_h)$*

   (b) *If $e \in (e_l, e_m)$, the voting cutoff is strictly higher under all secret: $V^h > V^p$*

2. *There exists $\tilde{e}_m$ and $\tilde{e}_h$ such that, if $e \in (\tilde{e}_m, \tilde{e}_h)$, welfare is strictly higher under the all secret environment.*

Proposition 5.1 shows that the comparison of the equilibrium voting cutoffs depends on the externality e. In particular, there are situations where, for a given sanction s, a proposal is more likely to be rejected under public contributions. For e very large, G and G^p are always above $-R$ and $-R^p$, which implies that all members vote in favor of the proposal ($V^h = V^p = v_{min}$). Decreasing e, we reach situations where all voters under secrecy still accept the proposal, but members with the lowest v_i reject it when contributions are public because they lose too much in reputation, as illustrated in the left panel of Figure 3. In this range the proposal is more likely to be rejected when contributions are visible ($V^h < V^p$). Finally, when e is lower, the voting cutoff moves to the swing participant group under visibility of contributions (case represented in the right panel of Figure 3). In this case, two forces decrease the voting cutoff when contributions are public. First, swing participants actually benefit in terms of reputation ($-R^p$ moves below $-R$). Second, information aggregation makes voters more confident that the externality from adopting the sanction is large (G^p above G). Overall, voters are more inclined to adopt the sanction when contributions are visible in this range ($V^h > V^p$).[18]

We assume in Proposition 5 that the distribution of types is uniform. For a general distribution, we have $G(0) = \frac{1}{2}e\left[F(v_0^h) - F(v_s^h)\right]$, i.e calculating the expected externality gain when the sanction is certain to pass is equivalent to determining the probability that a random voter is a swing participant. Similarly, $G^p(0) = \frac{1}{2}e\,[F(v_0^p) - F(v_s^p)]$. There is thus no systematic ordering of G and G^p. Nevertheless, more general conditions on f would guarantee that Proposition 5 holds. For instance a sufficient condition for Proposition 5.1.(a) to hold, is that $G^p(0) > G(0)$.

Proposition 5.1 shows that in certain circumstances, sanctions are more likely to be accepted when contributions are secret rather than public. This leads us, in Proposition 5.2, to identify a range for the externality parameter e such that welfare is higher when contributions are secret. Note that we identify here only a sufficient condition. Specifically we consider a case where $V^h = v_{min}$ when the socially optimal sanction $s = e$ is submitted to a vote. The first best without voting is thus always achieved when contributions are secret. On the contrary, in the *public contribution* environment, when the socially optimal sanction is submitted to a vote, in this range of parameters, the sanction could be rejected.[19] Thus we have identified conditions where rendering contributions visible is welfare decreasing.

[18] As e is further decreased, depending on parameters, there could be several other inversions of the ranking between voting cutoffs in the two environments.

[19] Note that the socially optimal sanction is different in this case: $s = e - \mu\Delta(v_s^p)$.

7 Voting on bonuses

In many groups, contributions to the public good are not constrained by sanctions but rather encouraged by bonuses. We consider this possibility in this section and show that our main results carry through.

We consider an alternative model where players vote on a bonus b for contributors. The utility function becomes:[20]

$$U_i = (v_i - c + b)a_i + e\frac{\sum_{j\neq i} a_j}{n} + \mu\mathbb{E}[e_i|y_i].$$

As in the main model, we compare the results when contributions are hidden versus public.

All hidden

The participation cutoffs v_0^h and v_b^h are similar to the cutoffs with a sanction of the same magnitude:

$$v_b^h = c - b$$

The participation groups are thus defined as in the main model. When contributions are secret, the difference in utility between a vote in favor and a vote against is:

$$D(v_i) = \begin{cases} G & \text{if } v_i < v_b^h \\ G + v_i - c + b & \text{if } v_b^h \leq v_i \leq v_0^h \\ G + b & \text{if } v_i > v_0^h \end{cases}$$

We can see that $D(v_i) > 0$ for all v_i. A bonus impacts positively the utility of contributors but does not decrease the payoff of free-riders. Moreover, it increases the expected payoff of all agents through the externality gain. As a result, a bonus is always accepted when contributions are hidden.

Public contributions

When contributions are public, the participation cutoffs v_0^p and v_b^p satisfy:

[20] Notice that we do not model how bonuses are funded.

$$v_b^p = c - b - \mu \triangle(v^p)$$

where $\triangle(v^p) = \mathbb{E}[v_i|v_i > v^p] - \mathbb{E}[v_i|v_i < v^p]$.

In the voting stage, we have:

$$D(v_i) = \begin{cases} G - \mu(\mathbb{E}[v_i|v_i < v_0^p] - \mathbb{E}[v_i|v_i < v_b^p]) & \text{if } v_i < v_b^p \\ G + v_i - c + b + \mu(\mathbb{E}[v_i|v_i > v_b^p] - \mathbb{E}[v_i|v_i < v_0^p]) & \text{if } v_b^p \leq v_i \leq v_0^p \\ G + b - \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_b^p]) & \text{if } v_i > v_0^p \end{cases}$$

As in the case of sanctions, always and never participants now suffer from a loss in reputation when the bonus is implemented. If contributions are visible, some players can vote against bonuses. The next proposition naturally extends Proposition 2:

Proposition 6 *If $b < \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_s^p])$, there exists $\bar{e}$ such that, it is a weakly dominant strategy for the always-participants to vote against the bonus if $e < \bar{e}$.*

With bonuses, any agent necessarily votes in favor if b is large enough. If $b < \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_b^p])$, always participants vote against the sanction if they do not expect any externality gain ($G = 0$). In such a case, they turn down the bonus when e is small enough in order to preserve their reputation.

This has similar consequences on welfare as in the case of subsidies. We consider the problem of a social planner interested in maximizing the sum of utilities net of reputation and bonuses. When contributions are secret, we have seen that all players vote for the bonus. This implies that the social planner can submit a bonus $b = e$ to the vote and achieve the first best. On the contrary, with public contributions, the optimal bonus might be rejected with some probability, thus reducing welfare.

8 Conclusion

In this paper we examined from a positive and normative point of view the organization of groups voting on their own rules. We have shown that there is a close interaction between voting and contribution choices and that the visibility of actions affects this interaction. When the social planner sets the sanction to be submitted to a vote, making contributions public may have a detrimental effect on welfare.

In some environments, not only contributions but also individual votes might be publicly observed. This would affect the equilibrium as the players would no longer only care about the case where their vote is pivotal, since votes would have an effect on reputation regardless of their impact on the outcome. This would tend to encourage group members to vote in favor of sanctions and thus facilitate approval.

References

Acemoglu, D., G. Egorov, and K. Sonin (2012). Dynamics and stability of constitutions, coalitions and clubs. *The American Economic Review 102*(4), 1446–1476.

Ali, S. N. and R. Bénabou (2018). Image versus information: changing societal norms and optimal policy. *working paper*.

Ali, S. N. and C. Lin (2013). Why people vote: ethical motives and social incentives. *The American Economic Review 105*(2), 1446–1476.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics 88*(7), 1605–1623.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *The American Economic Review 99*(1), 544–555.

Austen-Smith, D. and J. S. Banks (1996). Information aggregation, rationality, and the condorcet jury theorem. *American Political Science Review 90*(01), 34–45.

Barbera, S. and M. Jackson (2004). Choosing how to choose: self-stable majority rules and constitutions. *The Quarterly Journal of Economics 119*(3), 1011–1048.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2008). Information aggregation and strategic abstention in large laboratory elections. *The American Economic Review 98*(2), 194–200.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2010). The swing voter's curse in the laboratory. *The Review of Economic Studies 77*(1), 61–89.

Bénabou, R. and J. Tirole (2011). Laws and norms. Technical report, National Bureau of Economic Research.

Callander, S. (2008). Majority rule when voters like to win. *Games and Economic Behavior 64*(2), 393–420.

DellaVigna, S., J. A. List, and U. Malmendier (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics 127*(1), 1–56.

Feddersen, T. and W. Pesendorfer (1997). Voting behavior and information aggregation in elections with private information. *Econometrica: Journal of the Econometric Society*, 1029–1058.

Feddersen, T. and W. Pesendorfer (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review 92*(1), 23–35.

Feddersen, T. J. and W. Pesendorfer (1996). The swing voter's curse. *The American Economic Review 86*(0), 408–424.

Galbiati, R., K. H. Schlag, and J. J. Van Der Weele (2013). Sanctions that signal: An experiment. *Journal of Economic Behavior & Organization 94*, 34–51.

Galbiati, R. and P. Vertova (2008). Obligations and cooperative behaviour in public good games. *Games and Economic Behavior 64*(1), 146–170.

Gersbach, H. and V. Hahn (2008). Should the individual voting records of central bankers be published? *Social Choice and Welfare 30*, 655–683.

Godefroy, R. and E. Perez-Richet (2013). Choosing choices: Agenda selection with uncertain issues. *Econometrica 81*(1), 221–253.

Henry, E. and J. Sonntag (2019). Measuring image concerns. *Journal of Economic Behavior and Organization 160*, 19–39.

Levine, D. K. and A. Mattozzi (2017). Voter turnout with peer punishment. *WP EUI*.

Levine, D. K. and T. Palfrey (2007). The paradox of voter participation? a laboratory study. *The American Political Science Review 101*(1), 143–158.

Levy, G. (2007). Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review 97*(1), 150–168.

Mattozzi, A. and M. Nakaguma (2017). Public versus secret voting in committees. *WP EUI*.

Midjord, R., T. R. Barraquer, and J. Valasek (2017). Voting in large committees with disesteem payoffs: A 'state of the art'model. *Games and Economic Behavior*.

Ottaviani, M. and P. Sørensen (2001). Information aggregation in debate: who should speak first? *Journal of Public Economics 81*(3), 393–421.

Rege, M. and K. Telle (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics 88*(7), 1625–1644.

Samek, A. S. and R. M. Sheremeta (2014). Recognizing contributors: an experiment on public goods. *Experimental Economics 17*(4), 673–690.

Tyran, J.-R. and L. P. Feld (2006). Achieving compliance when legal sanctions are non-deterrent. *The Scandinavian Journal of Economics 108*(1), 135–156.

Visser, B. and O. H. Swank (2007). On committees of experts. *The Quarterly Journal of Economics 122*(1), 337–372.

APPENDIX

Proposition 1

Step 1: Under Restriction B, $G(V)$ is concave for $V \in (v_s^h, v_0^h)$.

In this region, as established in the main text:

$$G(V) = \frac{1}{2}e\left[\frac{F(V)-F(v_s^h)}{F(V)} + \frac{F(v_0^h)-F(V)}{1-F(V)}\right].$$

The second derivative is thus given by:

$$\begin{aligned} G^{''}(V) &= \frac{1}{2}e[\frac{F(v_s^h)F(V)(f'(V)F(V)-2(f(V))^2)}{(F(V))^4} \\ &+\frac{(1-F(v_0^h))(1-F(V))(-f'(V)(1-F(V))-2(f(V))^2)}{(1-F(V))^4}]. \end{aligned}$$

Thus, the two following conditions are sufficient to establish $G^{''}(V) < 0$:

$$f'(V)F(V) - 2(f(V))^2 \leq 0, \tag{7}$$

and

$$-f'(V)(1-F(V)) - 2(f(V))^2 \leq 0. \tag{8}$$

The first restriction of Condition B ($\frac{f}{1-F}(v)$ weakly increasing) implies condition (8) and the second restriction ($\frac{f}{F}(v)$ weakly decreasing) implies condition (7). This establishes the first step.

Step 2: there exists a unique symmetric Perfect Bayesian equilibrium.

As represented in Figure 1:

- $-R(v)$ is constant on the interval $[0, v_s^h]$ ($-R(v) = s$), decreasing on $[v_s^h, v_0^h]$ ($-R(v) = v_i - c$) and equal to 0 if $v > v_0^h$.
- $G(V)$ is increasing on $[0, v_s^h]$ ($G(V) = \frac{1}{2}e\left[\frac{F(v_0^h)-F(v_s^h)}{1-F(V)}\right]$ as given in the main text), concave on $[v_s^h, v_0^h]$ (according to Step 1) and decreasing if $V > v_0^h$ ($G(V) = \frac{1}{2}e\left[\frac{F(v_0^h)-F(v_s^h)}{F(V)}\right]$ as derived in the main text).

Given the shape of the functions $-R(v)$ and $G(V)$, and the fact the equilibrium is defined by the intersection of G and $-R$, we have 3 possible cases:

i If $G(0) > -R(0)$, the functions never cross and all players vote in favor, $V^h = 0$.

ii $G(.)$ and $-R(.)$ intersect for $V^h < v_s^h$. In such a case, the concavity of $G(.)$ guarantees that a second crossing cannot exist. Indeed, the second crossing could only be in $[v_s^h, v_0^h]$ in the region where $G(.)$ is decreasing. However, if a second crossing exists, we must also have a third crossing since $G(.)$ lies above $R(.)$ for $V > v_0^h$. But the third crossing cannot exist given that $G(.)$ is concave by step 1 and $R(.)$ is linearly decreasing. Thus the equilibrium needs to be unique.

iii $G(.)$ and $-R(.)$ intersect for $[v_s^h, v_0^h]$. By the above argument, there must be an odd number of crossings and multiple equilibria violate the concavity of $G(.)$, following the same reasoning as in case 2. The equilibrium also needs to be unique.

In all cases, the equilibrium is unique and is defined by a cutoff V^h (given Restriction A). We now prove the comparative static results:

1. An increase in e shifts $G(V)$ upwards and does not affect $R(v)$. The voting cutoff V^h (defined as the intersection of G and $-R$) is therefore decreasing in e, which implies that the probability of approval is increasing.

2. An increase in s decreases v_s^h, leaves v_0^h unaffected and thus increases $G(V)$ for all V. Moreover, $R(v)$ does not depend on s for $v \in [v_s^h, v_0^h]$, which is the case by definition if the equilibrium is interior. The voting cutoff V^h is thus decreasing in s, which implies that the probability of approval is increasing.

Proposition 2

For any voting cutoff, we must have $G^p(V) \leq e$ ($G^p = e$ if everyone is a swing participant). Thus, for the always participants, the net benefit of voting for the sanction is given by:

$$D(v_i) \leq \mu(\mathbb{E}[v_i|v_i > v_s^p] - \mathbb{E}[v_i|v_i > v_0^p]) + e.$$

Define $\overline{e}(s) \equiv -\mu(\mathbb{E}[v_i|v_i > v_s^p] - \mathbb{E}[v_i|v_i > v_0^p])$. Voting for the sanction is thus a weakly dominated strategy if $e \leq \overline{e}(s)$.

Proposition 3

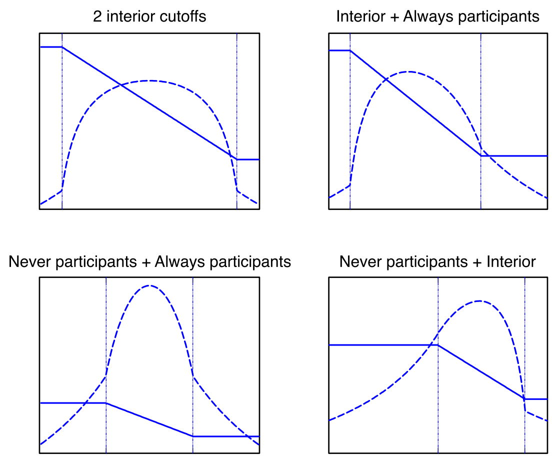

Figure 4: Four possible outcomes

The shape of the function G^p is the same as G in the benchmark case. In particular, as established in Step 1 of the proof of Proposition 1, $G^p(V)$ is concave on the interval (v_s^p, v_0^p), increasing on $[0, v_s^h]$ and decreasing if $V > v_0^h$.

On the contrary, the function R is modified. $-R^p(v)$ is constant on the interval $[0, v_s^h]$: $-R^p(v) = s + \mu(\mathbb{E}[v_i|v_i < v_0^p] - \mathbb{E}[v_i|v_i < v_s^p])$, decreasing on $[v_s^h, v_0^h]$: $-R^p(v) = v_i - c - \mu(\mathbb{E}[v_i|v_i > v_s^p] - \mathbb{E}[v_i|v_i < v_0^p])$ and equal to a constant, $-R^p(v) = \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_s^p])$, different from 0 as opposed to the benchmark case, if $v > v_0^h$.

Thus there can be at most two equilibria, defined as the intersections of G^p and $-R^p$. Only one of these equilibria is stable, i.e. is such that the intersection occurs on a portion where G^p is increasing. The different cases are illustrated in Figure 4 for $f \sim U[0, 1]$.

We now prove the comparative static results:

1. An increase in e shifts $G^p(V)$ upwards and does not affect $R^p(v)$. The stable voting cutoff V^h is therefore decreasing in e (as $G^p(V)$ needs to be increasing at $V = V^h$), which implies that the probability of approval is increasing.

2. We consider the case where the equilibrium is interior, i.e. V^p is in the swing

participants group. In that case, we have:

$$\frac{\partial V^p}{\partial s} = \frac{\frac{\partial v_s^p}{\partial s}\left[\frac{e}{2}\frac{f(v_s^p)}{F(V^p)} - \mu\frac{f(v_s^p)}{1-F(v_s^p)}(\mathbb{E}[v_i|v_i > v_s^p] - v_s^p)\right]}{1 + \frac{e}{2}f(V^p)\left[\frac{F(v_s^p)}{(F(V^p))^2} + \frac{F(v_0^p)-1}{(1-F(V^p))^2}\right]}.$$

In stable equilibria, $G^p(V)$ is increasing in V. This guarantees that $\frac{F(v_s^p)}{(F(V^p))^2} + \frac{F(v_0^p)-1}{(1-F(V^p))^2}$ is positive.

Thus, since $\frac{\partial v_s^p}{\partial s} < 0$ we have that V^p is increasing in s if and only if:

$$\frac{e}{2}\frac{f(v_s^p)}{F(V^p)} - \mu\frac{f(v_s^p)}{1 - F(v_s^p)}(\mathbb{E}[v_i|v_i > v_s^p] - v_s^p) < 0,$$

which can be reexpressed:

$$\frac{e}{F(V^p)} < 2\frac{\mu}{(1 - F(v_s^p))}(\mathbb{E}[v_i|v_i > v_s^p] - v_s^p). \tag{9}$$

The right hand side of expression (9) is positive and does not depend on e. The left hand side is (strictly) increasing in e since, according to Result 1, V^p is decreasing in e. Moreover, the left hand side converges to 0 when e converges to 0 and to infinity when e becomes large (V^p converges to 0). By the intermediate value theorem, there exists a unique value $\bar{e}^p$ such that equation (9) holds if and only if $e > \bar{e}^p$. In such a case V^p is increasing in s.

Proposition 4

According to Proposition 1, submitting $s < e$ to the vote decreases the probability of acceptance. Moreover, in the contribution phase, regardless of the composition of the group, having $s = e$ leads to higher welfare. For the planner, choosing $s < e$ always leads to lower welfare than choosing exactly $s = e$.

We now show that for unanimity rule, it is optimal to choose a sanction strictly greater than e. According to Proposition 1, it strictly increases the probability of acceptance. Furthermore, the ex post cost represented by the fact that players with v_i below $c - e$ would be forced to contribute due to the high sanction, is zero since the proposal is accepted only if all group members have $v_i \geq V^h > c - e$.

Proposition 5

1. G and G^p are increasing in e while $-R$ and $-R^p$ are independent of e. We can therefore define e_h as the value of e such that $G^p(0) = -R^p(0)$, and $\tilde{e}_h$ as the value of e such that $G(0) = -R(0)$. Furthermore, we have as described in the main text $-R^p(0) > -R(0)$ and $G(0) = G^p(0)$, so that $e_h > \tilde{e}_h$. Given the definition given above, we have

   - For $e \geq e_h$ $V^h = V^p = v_{min}$
   - For $e \in (\tilde{e}_h, e_h)$, $V^h = v_{min} < V^p$

   Decreasing e further, we can reverse the inequality and get $V^h < V^p$. Consider the value of $\bar{V}$ such that $-R^h(\bar{V})$ and $-R(\bar{V})$ intersect. Since G^p is increasing in e, we can find a value of $\tilde{e}_l$ such that $V^p = \bar{V}$, in other words, $G(V^p) = -R(V^p) = -R^p(V^p)$. For this value we have $G^p(V^p) > G(V^p)$, so that the intersection of G and $-R$ is such that $V^h < V^p$. Thus there exists an intermediate value $e_m \in (\tilde{e}_l, \tilde{e}_h)$ used in the statement of the proposition, such that

   - If $e > e_m$, the voting cutoff is lower under all secret ($V^h \leq V^p$) and strictly lower if $e \in (e_m, e_h)$
   - If $e \in (e_l, e_m)$, the voting cutoff is strictly higher under all secret: $V^h > V^p$

2. *Note: To simplify the notation, we consider $f \sim U[0,1]$ for the proof of this Result. The extension to other uniform distributions is straightforward.*

   Consider the *all secret* environment. As explained in the main text, the first best would be achieved for $s^h = e$. Suppose that this sanction is submitted to a vote. We have:
$$G(0) = \frac{e}{2}[v_0^h - v_s^h] = \frac{e}{2}s = \frac{e^2}{2}$$
   and
$$-R(0) = e.$$

   If $e > 2$, we have $G(0) > -R(0)$ which implies $V^h = 0$ and the first best is always implemented.

   Now consider the *public contributions* setup. If a planner were to set the sanction without voting, he would still make players contribute if and only if $v_i > c - e$ (reputation is a zero-sum game). However, he must also take into account the impact of reputation on contributions. The first best would be achieved for $s^p = e - \frac{\mu}{2}$. If this sanction is submitted to a vote, we now have:

$$G^p(0) = \frac{e}{2}[v_0^h - v_s^h] = \frac{e}{2}[e - \frac{\mu}{2}]$$

and the $-R(v_i)$ function now includes a reputation term:

$$\begin{aligned}-R(0) =& e - \frac{\mu}{2} + \mu(\mathbb{E}[v_i|v_i < v_0^p] - \mathbb{E}[v_i|v_i < v_s^p]) \\ =& e + \frac{\mu}{2}(e - 1 - \frac{\mu}{2})\end{aligned}$$

This implies that $V^p = 0$ is an equilibrium if:

$$(-2 + e - \mu)(2e - \mu) \geq 0$$

Which holds if $e \geq \mu + 2$.

As a result, if $2 < e < \mu + 2$, the planner can always implement the first best in the all secret environment ($s^h = e$ is always accepted) while in the *public contributions* setup $V^p > 0$ if $s^p = e - \frac{\mu}{2}$ is submitted to a vote and thus the sanction that would lead to the first best is rejected with some probability.

Proposition 6

In the case of bonuses, the net benefit for the always participants of voting for the sanction satisfies:

$$D(v_i) \leq b - \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_b^p]) + e,$$

where e is the maximum externality gain that a member can expect. Define $\overline{e}(b) \equiv \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_b^p]) - b$.

Under the condition:

$$b < \mu(\mathbb{E}[v_i|v_i > v_0^p] - \mathbb{E}[v_i|v_i > v_b^p]),$$

$\overline{e}(b) > 0$ and voting for the sanction is thus a weakly dominated strategy if $e \leq \overline{e}(b)$.

Parameters used in the different figures in the paper

In all figures we considered the case where $f \sim U[0,1]$ and simple majority with an odd number of group members ($k = n/2$). The other parameters are given as follows:

Figure	e	s	c	μ
1	0.7	0.6	0.8	0
2	2.7	0.8	3	4.2
3-1	2.1	0.5	0.85	0.4
3-2	1	0.5	0.85	0.4

Supplementary Appendix A1: Simple example with observable aggregate outcomes

We revisit the simple example and relax the behavioral assumption that image depends only on individually observed actions y_i. We first consider a case where players use the vote tally to infer the type of the other group member. Moreover, we also study an environment where both vote tally and aggregate contributions are used.

A1.a Observable vote tally

Suppose that players observe the number of votes in favor and use this information to form expectations about the type of the other player. Given that there are only two players, member i can perfectly infer the vote of member $-i$ based on her own vote and on the aggregate outcome (i.e. it is as if votes were observable). Moreover, we suppose that player i cares about the belief of player $-i$ about her type.

All hidden

Let's consider the following strategy profile:

- types H vote in favor and always contribute,
- types L vote against and only contribute if the sanction is accepted,
- Players who vote in favor have reputation v_H, players who vote against v_L.

In the contribution stage, reputation does not matter since actions are not observable. Given that $c < v_H$, type H contributes, and given that $v_L = 0$, type L does not.

In the voting stage, votes reveal the type of players. If type L deviates and votes in favour of the sanction, she gets $e - c + \mu v_H$ while staying on the equilibrium path gives her $p(e - c)$. Under the condition:

$$\mu v_H < (1 - p)(c - e),$$

types L do not deviate. Types H do not want to deviate either because a vote against the sanction would decrease both reputation and externality gain. Thus, under the above condition, this strategy profile is an equilibrium.

Public contributions

In this case, reputation is based on the observation of both votes and contributions, that we denote (b_{-i}, a_{-i}). Moreover, beliefs need to be specified for each regulatory environment, i.e with and without the sanction. For instance $\mathbb{E}[v_{-i}|(1,1), s]$ refers to the observation that the other player voted for the sanction and contributed in a situation where the sanction was accepted.

Let's focus on symmetric beliefs and consider the following strategy profile:

- types H vote against and always contribute,
- types L vote against and only contribute if the sanction is accepted,
- beliefs on the equilibrium path are: $\mathbb{E}[v_{-i}|(0,1), s=0] = v_H$, $\mathbb{E}[v_{-i}|(0,0), s = 0] = 0$,
- beliefs off the equilibrium path are: $\mathbb{E}[v_{-i}|(0,1), s] = v_H$, $\mathbb{E}[v_{-i}|(0,0), s] = 0$, $\mathbb{E}[v_{-i}|(1,0), s] = 0$ and $\mathbb{E}[v_{-i}|(1,1), s] = 0$.

Let's consider a player of type L. On the equilibrium path, she plays (0,0) and gets payoff ep. The possible deviations would yield the following payoffs:

$$
\begin{aligned}
(1,0) &\rightarrow e - s \\
(0,1) &\rightarrow ep - c + \mu v_H \\
(1,1) &\rightarrow e - c
\end{aligned}
$$

We can see that there is no profitable deviation (recall that we impose $c > e$ and $c > \mu v_H$ in the main text).

We turn to types H. Such players choose (0,1) and get $v_H - c + ep + \mu v_H$. Instead, the deviation payoffs are:

$$
\begin{aligned}
(0,0) &\rightarrow ep \\
(1,0) &\rightarrow e - s \\
(1,1) &\rightarrow v_H - c + e
\end{aligned}
$$

(1,1) is thus the most profitable deviation and those players would stay on the equilibrium path if:

$$e(1-p) < \mu v_H.$$

As a result, if $e(1-p) < \mu v_H < (1-p)(c-e)$, there exists an equilibrium that has the same properties as in the main text where the sanction is less likely to be adopted when contributions are visible.

A1.b Observable vote tally + aggregate contributions

Suppose now that all aggregate behaviors (vote tally + aggregate contributions) are used to form expectations. Assume that members care about the beliefs of an outsider who observes the aggregate outcome as well as individual contribution decisions when the second stage is public.[21]

Second stage: contribution decisions

The assumptions $v_H > c$ and $\mu v_H < c$ imply that there is an equilibrium where, regardless of whether individual contributions are visible or not, type H always contribute while L only contribute if the sanction is accepted.

First stage: voting

Let's consider the all hidden environment. In this setup, the outsider still observes aggregate behaviors (number of votes in favor and number of contributors). As in the main text, suppose that type L votes against while type H votes for. In such a strategy profile, type H gets:

$$\begin{aligned} &p[e + v_H - c + \mu v_H] + (1-p)[e + v_H - c + \mu \frac{v_H}{2}] \\ =&e + v_H - c + \mu v_H \frac{1+p}{2}. \end{aligned}$$

If the other player is a type H (probability p), there are two votes in favor and the outsider infers perfectly the type of players. However, if the other player is a type L, there is exactly one vote in favor, the sanction passes and both players

[21]If players care about the belief of the other player, the problem is equivalent to the previous case with public contributions and the visibility of the second stage plays no role: in a group of two members, if each member could observe the aggregate actions, it would be equivalent to observing the other player's actions.

contribute. The outsider only knows that there is one player of each type and the reputation of both players is $\frac{v_H}{2}$.

Now suppose that type H deviates and votes against. His expected payoff is:

$$\begin{aligned}&p[e + v_H - c + \mu\frac{v_H}{2}] + (1-p)[v_H - c + \mu\frac{v_H}{2}]\\ =&pe + v_H - c + \mu\frac{v_H}{2}.\end{aligned}$$

If the other is type H, the sanction is still approved and both players contribute. Given that there is only one vote in favor, the observer believes that there is one player of each type. When the other player is of type L, the sanction is rejected. Type H still contributes. The observer assigns a reputation $\frac{V_H}{2}$ to each player[22]. Comparing the two expressions, we see that type H does not want to deviate.

In this strategy profile, type L gets a payoff:

$$p(e - c + \mu\frac{v_H}{2}) + (1-p) \times 0.$$

If he deviates and votes in favor, he necessarily has to contribute but he benefits from a better reputation:

$$p(e - c + \mu v_H) + (1-p)(e - c + \mu\frac{v_H}{2}).$$

If $\mu v_H < 2(c-e)(1-p)$, type L prefers to vote against and we have an equilibrium.

We turn to the public contributions environment and consider a strategy profile where all players vote against the sanction. Suppose that the observer believes that if a player deviates and votes in favor, her type is H. However, if the outsider cannot identify the deviator, she believes that each player has deviated with a probability 1/2. For type H, deviating and voting in favor thus gives:

$$e + v_H - c + \mu\frac{v_H + \bar{v}}{2}$$

where $\bar{v} = pv_H$ is the mean of v. The outsider infers that the deviator is a type H, but since both players contribute in the second stage, the deviator cannot be identified. The reputation is therefore $\frac{v_H+\bar{v}}{2}$.

If type H votes against, she gets:

[22]We arbitrarily set out-of-equilibrium beliefs such that a player who votes against and contributes is perceived to be a high type.

$$pe + v_H - c + \mu v_H.$$

Voting against implies to give up the externality gain with probability p but allows the player to perfectly reveal her type. Type H does not want to deviate if:

$$2e < \mu v_H.$$

If this condition holds, low types do not want to deviate either and we have an equilibrium.

To conclude, if $2e < \mu v_H < 2(1 - p)(c - e)$, we have identified an equilibrium where the sanction is less likely to be adopted when players contribute publicly.

Supplementary Appendix A2: Simple example with 3 group members

Consider a group with 3 members and simple majority. We still have $v \in \{v_L, v_H\}, Pr(v = v_H) = p$ and $v_L = 0$. We also impose $v_H > c$ and $-c + \mu v_h < 0$. However, we depart from the introductory example and consider larger externality gains: $e/2 < c < e$.

Suppose that a sanction $s > e$ is proposed to the group. We focus on trembling hand perfect equilibria, which rules out strategy profiles where players play weakly dominated strategies.

Contribution decisions

In the all hidden environment as well as in the public contributions setup, we focus on the equilibrium of the contribution subgame where v_H always contribute while v_L only contribute if the sanction is implemented.

Voting stage

Let's first consider the all hidden environment. For v_H, voting for the sanction is a weakly dominant strategy: it does not impact their contribution decision and provides an externality gain if there are some v_L in the group. For v_L, voting for the sanction is weakly dominated as $c > e$. In the only trembling hand perfect equilibrium with hidden contributions, v_H vote for and v_L vote against.

Now suppose that contributions are public. v_L still vote against the sanction in this setup. There is always an equilibrium where v_H also vote against. Players are never pivotal on the equilibrium path. If v_H happens to be accidentally pivotal, voting for could increase externality gain but decreases reputation. Thus this equilibrium is trembling hand perfect.

Instead, suppose v_H vote in favor. A player is pivotal if there is exactly one v_L among the other players. If a player of type v_H is pivotal, voting in favor gives:

$$v_H - c + \mu \bar{v} + e,$$

while if he votes against he gets:

$$v_H - c + \mu v_H + \frac{e}{2}.$$

As a result, if

$$\frac{e}{2} < \mu v_H(1-p),$$

this is not an equilibrium and v_H prefers to vote against. In such a case, all players vote against the sanction in the only trembling hand perfect equilibrium of the game.

Welfare

In this example, the sanction is never accepted when contributions are visible. v_L therefore never contributes, which is not optimal for the group as $c < e$. Instead, the sanction is sometimes implemented when contributions are hidden. When there is exactly one v_L in the group, this member is forced to contribute because the other two players vote in favor of the sanction. We can therefore conclude that making contributions visible decreases welfare.