Discussion Paper

VOTING AND CONTRIBUTING WHEN THE GROUP IS WATCHING

Emeric Henry and Charles Louis-Sidois

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2018-11

Voting and contributing when the group is watching

Emeric Henry and Charles Louis-Sidois*

September 12, 2018

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, from both a positive and normative perspective, the interaction between the way members vote on rules and their subsequent contribution decisions. We show how endogenous norms can emerge. We study in particular the role played by the visibility of individual actions, votes or contributions. While making votes visible always increases welfare in our setting, making contributions public 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

In May 2009, the elected members of the French Parliament (Assemblée Nationale) voted for a law imposing sanctions for those among them not attending weekly meetings of committees. Similarly, 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. In this paper we provide a

*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.

positive and normative analysis of group organization and describe how members vote on rules and subsequently contribute 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, are typically observed by the rest of the group. For other types of public goods, such as group projects, the individual contributions may be harder to observe. In terms of votes, the organization can have in place open or secret voting. 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. Image concerns have been shown empirically to be an important driver of contributions to common tasks.[1]

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.

In the benchmark model studied in Section 3, we consider the case where neither votes nor contributions to the public good are 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 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.

We then turn to environments where either individual votes or contributions are visible. The visibility of actions affects the interaction between voting behavior and

[1] 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 (2015) for evidence from the field

contribution decisions. When contributions are visible, a sanction, by increasing overall contributions, decreases the honor derived from being seen contributing. Thus voting affects the calculus of reputation. In turn, when votes are public, concern for image will affect voting behavior. Uncovering the subtle interactions inherent in these environments is a key focus of the paper.

We first consider in Section 4.1, groups where contributions are public and votes are secret. In the public good contribution phase, the same three categories (always, never and swing participants) emerge. The composition of these groups is however affected since 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 participant always voted in favor of sanctions, we show that it can now be a dominant strategy for them to vote against. Indeed, even though the sanction will never apply to them, they lose in reputation as contributing is no longer so rare 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.

The second key message is that there can exist multiple equilibria in the voting stage, that can be interpreted as corresponding to different norms of voting. Some equilibria have a high voting cutoff, i.e groups tending to vote against sanctions under the self realized expectation that the gain in additional public good is low. Other equilibria correspond to norms with a low cutoff where group members are more likely to vote for sanctions as they expect higher benefits. Technically, the multiplicity is linked to the information aggregated when voters consider the case where their vote is pivotal.

In Section 4.2, we turn our attention to groups where contributions are secret but votes are public. In this environment, the equilibrium of the contribution stage is identical to the benchmark case since in both environments contributions are hidden. However, the reputation of individuals is based on their votes and thus affects the voting stage since voters care not only about the event where their vote is pivotal, but also about the other realizations. In this setting there is an additional source of multiplicity of norms due to pivotality considerations. There could exist a norm of general support for sanctions, self sustained because the chance of being pivotal is small if everyone votes in the same way. Conversely there could be norms where individuals have more incentives to vote against sustained by the fact they are more likely to be pivotal.

The visibility of actions affects comparative statics. For instance, while in the benchmark case a higher sanction tends to be more likely to be accepted as it increases the expected contribution of others, it is no longer necessarily the case when contributions are visible since a higher sanction reduces the honor obtained by those who contribute. While in the benchmark case the size of the group plays no role, in the case of public votes, when groups are very large any level of sanction will be approved since the probability of being pivotal converges to zero.[2]

In Section 5, we turn to the welfare analysis of the benchmark case and to the welfare comparison of the different visibility setups. 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? We show that making votes public is unambiguously welfare increasing. Indeed, when individual votes are visible, group members are more inclined to vote in favor of the sanction and the planner can choose a sanction closer to the first best. On the contrary, making contributions visible can decrease welfare in certain circumstances. Making votes visible may push never 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.

Related literature

In this paper, we study a group of players voting on rules that apply to a second stage of the game involving the same group. In that sense we are connected to the literature on endogenous constitutions, that establishes conditions to guarantee stability of rules (seminal paper by Barbera and Jackson 2004, followed for instance by Acemoglu et al. 2012). One defining feature of our model, that differentiates it

[2]This is consistent with a large body of anecdotal evidence suggesting that the shift from secret to public evidence increases the consensus in voting (for instance Elster 2015 on the EU council of ministers).

from that literature, is that members are privately informed about their propensity to contribute. In a related paper, Godefroy and Perez-Richet (2013) consider a sequence of two elections in a committee with privately informed voters, the first to select the issue to be submitted to a vote and a second to vote on approval of this issue versus status quo.

A key element in the voting stage is information aggregation, and our paper is thus closely related to the literature on strategic voting, where the initial motivation was to revisit the Condorcet jury theorem when including strategic concerns (Austen-Smith and Banks, 1996, Feddersen and Pesendorfer, 1996, 1997, 1998 and Levine and Palfrey 2007 for empirical 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 literature.[3]

In Section 4, we consider environments where actions are visible which in our setting where group members are image concerned implies subtle interaction between the voting and contribution stage. 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 5, and this voting stage sets the norm. Acemoglu et al. (2012) also examine the interaction between laws and norms in settings where laws are endogenously enforced by the community. Levine and Mattozzi (2017) consider the endogenous setting of norms by party leaders to encourage turnout (see also Ali and Lin 2013).

The setting we consider in Section 4 where members care about the image they project also connects us to the literature on aggregation of information in committees when individuals have career concerns. The key distinction between our environment and the type of setups the career concern literature (Ottaviani and Sørensen, 2001, Visser and Swank, 2007) is that our model can be seen as a first model where agents

[3] One exception is Callander (2008) who examines a model where privately informed voter want to elect the best candidate but also want to vote for the winner. The source of multiplicity is very different: if the rest of the group is more likely to vote against their signal, it also makes sense to do so as there is a desire to be on the winning side.

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.

In Section 4 we analyze different environments that vary in terms of what actions (votes and contributions) are visible and we compare these environments in Section 5. Regarding votes there is a literature examining whether votes in committees should be made public or kept secret when members have career concerns. Levy (2007) shows that when votes are secret, committees are more likely to vote for the choice favored by the prior (see also Gersbach and Hahn 2008 who show that secrecy can be beneficial). Mattozzi and Nakaguma (2017) show that in a setting where individuals differ both in competence and bias, transparency might induce incompetent members to vote rather than abstain, thus decreasing overall welfare. Our setting differs along several dimensions, in particular we examine the effect of secrecy both of votes and contributions.

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.

A sizeable experimental literature 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 setting. 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 $2N + 1$ 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.[4]

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 $a = \{a_1, a_2, ..., a_{2N+1}\}$, the utility of player i is given by:

$$U_i = (v_i - c)a_i - s(1 - a_i) + e\frac{\sum_{j \neq i} a_j}{2N} + \mu E[v_i|y_i]. \tag{1}$$

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.[5] The v_i are i.i.d. drawn from the continuously differentiable density $f(v)$ with support $[v_{min}, v_{max}]$ and privately observed.

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 $s(1 - a_i)$.[6] In addition, individuals benefit from the contributions of other group members, i.e there is an externality gain $e\frac{\sum_{j \neq i} a_j}{2N}$.

Finally, agents are image concerned and want to be perceived as intrinsically motivated, which is captured by the component $\mu E[v_i|y_i]$. 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 3, we consider groups where no individual action, neither the vote nor the contribution, is observable. A key contribution of the paper is then to examine different observability setups: the case where the vote is secret but contributions are observable, i.e $y_i = a_i$ in Section 4.1 and the case where individual votes are revealed but the individual contributions are kept

[4]Note that the case of a bonus for contributing would lead to very similar results.

[5]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.

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

secret, i.e $y_i = b_i$ in Section 4.2. The case where both are public is discussed in the Supplementary Appendix.

We make the behavioral assumption that image $E[v_i|y_i]$ 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. Levy (2007) considers such inferences in a model with career concerns. The problem is further complicated in our setting as the players can update both based on aggregate results of the vote but also aggregate contributions, with intricate interactions between the two.[7] We discuss at different points of the paper how our results would be affected if we allowed group members to base the reputation on more sophisticated inferences.

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.

3 Voting on sanctions

We start by studying the behavior in groups voting on sanctions to spur public good contributions in the case where neither individual votes nor individual contributions are 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 hh (denoting the case where both votes and 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

[7]The case where both votes and contributions are public, discussed in supplementary appendix A1, provides some intuitions for those additional interactions.

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:

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

$$v_s^{hh} = c - s. \tag{2}$$

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.

We now turn to the first stage of the game. If the sanction s is implemented, players use in the contribution phase a strategy with cutoff v_s^{hh}, as derived above. On the contrary if the sanction is rejected, the players use in equilibrium a strategy with cutoff denoted v_0^{hh} (defined by equation (2) for $s = 0$) with $v_s^{hh} < v_0^{hh}$.

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 $v_i < v_s^{hh}$.
- *Swing-participants* who contribute if and only if the sanction is voted: members with $v_s^{hh} \leq v_i \leq v_0^{hh}$.
- *Always-participants* who always contribute regardless of the outcome of the vote: members with $v_i > v_0^{hh}$.

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 that obtained without:

$$G = e\frac{\mathbb{E}\left[\sum_{j\neq i} a_j | s > 0\right]}{2N} - e\frac{\mathbb{E}\left[\sum_{j\neq i} a_j | s = 0\right]}{2N}.$$

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 $D(v_i) \equiv R(v_i) + G$.[8] 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^{hh} such that a type v_i votes in favor if and only if $v_i \geq V^{hh}$.

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[9] (the x axis is v_i for R and V for G). The equilibrium cutoff V^{hh} is such that $-R(V^{hh}) + G(V^{hh}) = 0$ and thus corresponds to the intersection of $-R(v_i)$ and $G(V)$.

As in the literature on information aggregation in voting (Austen-Smith and Banks, 1996, Feddersen and Pesendorfer, 1996), voters consider only the case where

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

[9] Parameters used to plot all graphs used in the paper are presented at the end of the Appendix.

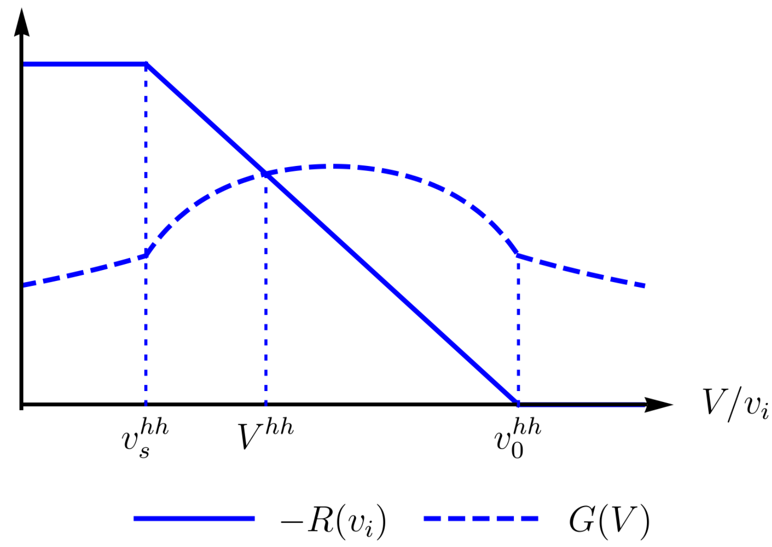

Figure 1: Voting cutoff in all secret environment

their vote is pivotal. In equilibrium a pivotal voter can infer additional information about the distribution of types. Under majority rule, a player is pivotal when there are exactly N yes-voters and N no-voters among the $2N$ 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 $\frac{F(V)-F(v_s^{hh})}{F(V)}$. 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 $\frac{F(v_0^{hh})-F(V)}{1-F(V)}$. This probability is decreasing in V. Overall, the expected externality gain is thus given by the following expression[10]:

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

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

As a function of V, G is first increasing and then decreasing. When V is close to v_s^{hh}, the positive effect on the probability that a no voter is a swing participant dominates. As V moves closer to v_0^{hh}, 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^{hh}, v_0^{hh}).

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^{hh}.

Furthermore,

1. *the probability of approval is increasing in* e*,*
2. *the probability of approval is increasing in* s *if the equilibrium is interior,*
3. *there exists* $\tilde{e}^{hh}$ *such that the voting cutoff is decreasing in* K *if and only if* $e \geq \tilde{e}^{hh}$.

All equilibria are of the cutoff form, i.e. such that individual i votes in favor if and only if $v_i \geq V^{hh}$. 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 unicity 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

---

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

the equilibrium is interior, i.e. the cutoff is within the swing voters, an increase in s increases G and does not affect R.[11]

The supermajority requirement K affects the calculation of G. Consider an equilibrium such that $V^{hh} < v_s^{hh}$, in other words the cutoff is in the group of never-participants. The pivotal voter, to calculate the expected externality, needs to build an expectation on the number of swing participants. Only the yes-voters can be swing participants in such an equilibrium. As K increases, the number of yes voters is higher when pivotal and thus the expected size of the externality is larger, which makes the pivotal voter more inclined to vote in favor (i.e V^{hh} decreases). In this case a more stringent voting rule, namely an increase in the supermajority rule, makes people more inclined to vote in favor. These cases occur when V^{hh} is initially small, or equivalently e is high.[12] In such instances, the overall effect on the probability of acceptance of a sanction is ambiguous since each individual voter is more likely to vote in favor, but more yes votes are required to pass the sanction. However we show there are instances where increasing the supermajority rule can increase the probability of approval. We present such a case in Supplementary Appendix A2.

To conclude this section, we highlight that our behavioral assumption that image depends only on individually observed actions y_i implies in this benchmark case that image concerns do not play a role. If we had assumed that aggregate results (vote outcome or total level of contribution) were used to update the image of individuals, this would affect the results in two ways. First it would imply that players' beliefs on others' types, and thus on G, depend on their own actions, which is not the case in our main model. To see that, consider a group of 3 players $\{i, j, k\}$ where the sanction is adopted if at least two members vote in favor. Suppose that i and j vote in favor of the sanction while k votes against and that players only observe whether the sanction was adopted. In this example, player k knows that i and j have voted for the sanction. However, i and j do not know the identity and number of other yes voters. Second, the inference should be based on the combination of two aggregate indicators, the result of the vote and the level of total contributions, implying intricate updating of beliefs. To avoid diverting the attention of the reader

[11] 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.

[12] On the contrary, if the pivotal voter is a never participant, he only expects no voters to be swing participants. In this case, as K increases, the number of no-voter decreases in the pivotal case and so does the expected externality. The pivotal voter is then less inclined to vote in favor (i.e V^{hh} increases).

from the main point of the paper, we assume that the behavioral component of image depends only on individually observed actions, as explained in Section 2.

4 Visible actions

A defining feature that varies across groups and organizations, is whether group members can observe individual actions of others. In many instances it is natural to think that individual contributions are observed by the rest of the group. This is the case for participation in meetings, as in the motivating example of the French Parliament.[13] However, for other types of tasks, such as team work, individual contributions are often hard to identify. Similarly for votes, some organizations use open voting while others keep votes secret. We therefore examine environments that differ in the visibility of actions. We study in Section 4.1 organizations where votes are secret and contributions public, an environment we call *public contribution* before considering in Section 4.2 the *public vote* case. The case where both are public is considered in the Supplementary Appendix.

4.1 Secret votes and public contributions

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 hp (hidden vote, public contribution) the equilibrium parameters in this case. We use the notation $\triangle(v_s^{hp}) = E[v|v > v_s^{hp}] - E[v|v < v_s^{hp}]$ (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^{hp}. 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 $v_i \geq v_s^{hp}$ where the cutoff is defined by*

$$v_s^{hp} = c - s - \mu\triangle(v_s^{hp}). \tag{3}$$

[13] Le Bihan and Monnery (2018) analyse the reform of attendance in the French Parliament mentioned in the introduction. They show that both monetary sanctions (introduced in 2009) and later changes that made attendance more visible both had a sizeable impact on attendance.

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 equilibrium of the public good stage is thus, like in the benchmark case, also characterized by three participation groups. However, since the participation cutoffs v_s^{hp} and v_0^{hp} 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:

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

where G^{hp} 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 $\bar{e}(s)$ such that if $e \leq \bar{e}(s)$, 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

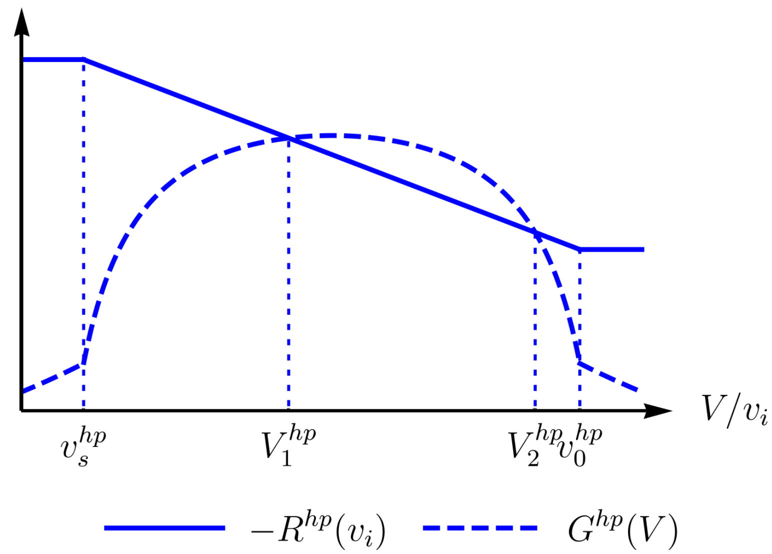

Figure 2: Voting cutoff in the public contribution environment

sanction is implemented:

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

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:

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

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:

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

Furthermore, there exists benchmarks $\bar{e}^{hp}$ and $\tilde{e}^{hp}$ such that, if the stable equilibrium is interior, the voting cutoff V^{hp}:

1. *is decreasing in e,*
2. *is decreasing in the level of sanction s if and only if $e \geq \bar{e}^{hp}*$,*
3. *is decreasing in K if and only if $e \geq \tilde{e}^{hp}$.*

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^{hp} and V_2^{hp} corresponding to the intersections of the functions $-R^{hp}(v_i)$ and $G^{hp}(V)$.[14] In equilibrium with cutoff V_1^{hp}, the pivotal voter expects a large portion of yes voters (to the right of V_1^{hp}) and of no voters (to the left of V_1^{hp}) to be swing participants. The expected externality is thus large and justifies the low voting cutoff. On the contrary in the case of V_2^{hp}, 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^{hp}) 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^{hp}) 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.

This outcome with multiple equilibria is a feature of our model that is, to the best of our knowledge, not present in the literature on aggregation of information in voting. Consider the classic case where voters get information on an underlying state of the world and the expected payoff is increasing in this state. The type of a voter is the signal he obtains, and like in our model the equilibrium takes the form of a cutoff strategy. In this case, the expected payoff is increasing in the cutoff: a

[14] The multiplicity of equilibrium was not possible in Section 3 because always participants had $-R(v_i) = 0$. Thus, the concavity of $G(V)$ was sufficient to guarantee the unicity of the equilibrium. Now we have $-R^{hp}(v_i) = E[v_i|v_i > v_0^{hp}] - E[v_i|v_i > v_s^{hp}] > 0$, which implies that a second equilibrium can exist.

higher cutoff means the information obtained when pivotal indicates a higher state of the world. In our model there can be non monotonicities as suggested above.

Proposition 3 also presents comparative statics on how voting cutoff and approval probabilities vary with s and K. The intuition behind Result 3.3 is the same as in the benchmark case, though the cutoff value $\tilde{e}^{hp}$ is different. The effect of s follows a different logic. As in the case without reputation, an increase in s 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 of the sanction 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.

4.2 Public votes and secret contributions

We now consider the *public vote* environment where the individual votes are public but the contribution decisions are not observed by group members. This could be the case if the individual contributions to the public good are hard to identify, which is common for team work. Of course, to impose the sanctions, the designer/manager has to observe individual actions. We thus consider cases where he commits not to disclose individual actions (or is unable to credibly do so), as in the case of bonuses in firms.

In this environment, the equilibrium of the contribution stage is identical to the benchmark case described in Lemma 1 since contributions are not observable. The three categories in equilibrium, always participants, swing participants and never participants are unaffected.

However, the reputation of individuals is based on their votes and thus affects the voting stage. All voters care in the same way about reputation, so would vote the same way if they knew they were not pivotal. However, in the pivotal case, those with lower v_i have higher incentives to vote against the sanction.

We denote Piv the event of being pivotal when voting and $\triangle^* \equiv E[v_i|b_i = 1] - E[v_i|b_i = 0]$ the reputation derived in equilibrium from voting in favor of the sanction rather than against. Both these measures are determined in equilibrium and do not depend on the individual types of players.

In equilibrium, the net benefit of the never participants to vote in favor of the

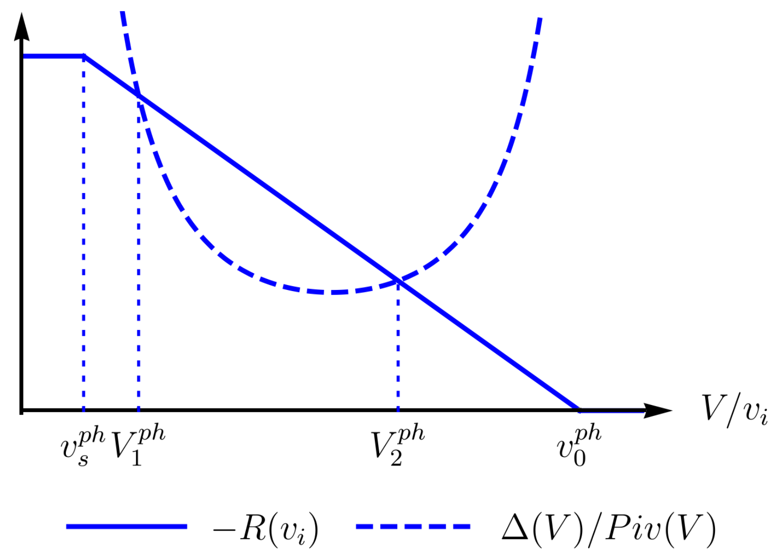

Figure 3: Voting cutoff in the public vote environment

sanction is given by

$$\mu\triangle^* + P\left[Piv\right]\left(-s + G\right),$$

for swing participants by

$$\mu\triangle^* + P\left[Piv\right]\left(v_i - c + G\right),$$

and for always participants by

$$\mu\triangle^* + P\left[Piv\right]\left(G\right),$$

where G is the expected externality gain from the sanction, conditional on the event of being pivotal. Given that the contribution cutoffs v_0^{ph} and v_s^{ph} are identical to the benchmark parameters v_0^{hh} and v_s^{hh}, $G(V)$ is defined as in Section 3. All symmetric Perfect Bayesian equilibria are also of the cutoff form.

There is however an additional source of multiplicity of equilibria compared to the previous sections. This multiplicity is present even when we impose $e = 0$ to abstract from the externality effect, as illustrated in Figure 3. For a given N, the probability of being pivotal is non monotonic in V. For V lower than the median, the probability of being pivotal is increasing in V while it is decreasing otherwise. There could thus be two potential equilibria: one where V^{ph} is low and one where it is close to the median of the distribution of types. In Figure 3, when V_1^{ph} is played, players have high incentives to vote in favor because the probability of being pivotal is low and the reputation effect dominates. This is a self sustaining norm of general support for sanctions, self sustained because the chance of being pivotal is small if everyone votes in the same way. In the second equilibrium, players have more incentives to vote against and the high cutoff V_2^{ph} is coherent with a higher probability of being pivotal.

Proposition 4 *In the public vote environment, all symmetric Perfect Bayesian equilibria are cutoff equilibria where players vote in favor if and only if $v_i \geq V^{ph}$.*

Furthermore in all interior stable equilibria,

1. *The probability of acceptance is increasing in s and e.*
2. *If $e = 0$, there exists $\overline{V}^{ph}$ such that V^{ph} is increasing in K if and only if $V^{ph} \leq \overline{V}^{ph}$.*

When $N \to +\infty$, in all sequential equilibria, any sanction $s > 0$ is approved with probability converging to one. If $\Delta(V)$ is decreasing in V, then an increase in the size of the organization N decreases the probability of acceptance.

As in the benchmark model, Result 1 of Proposition 4 shows that s and e increase the expected externality G and thus the probability of acceptance. Result 2 on the effect of K follows a logic specific to the case of public votes that relies on pivotality considerations[15]. When V is small, a member taken at random is likely to vote in favor of the sanction and the probability of being pivotal is therefore small. An increase in the supermajority requirement will in this case increase the probability of being pivotal and therefore make voters less likely to vote for the sanction. The opposite logic applies for large V.

In Proposition 4, we consider comparative static results with respect to N. Note that in the environments considered in the previous sections, the size of the group

[15] The restriction $e = 0$ allows us to focus on the effect of K on the pivotal probability, which was not present in the previous sections. When $e > 0$, K also impacts the calculus of $G(V)$. As explained in Propositions 1 and 3, increasing K increases the expected externality gain for low cutoffs, which balances the effect described in Result 4.2.

N played no role. Here, N affects behavior through its impact on the probability of being pivotal. The intuition of the second point when N becomes very large is quite clear. Low types vote against the sanction when the probability of being pivotal is high enough that it compensates for the loss of reputation. When N becomes large, the probability of being pivotal goes to zero and the proportion of people ready to vote against shrinks. This result generalizes Feddersen et al. (2009) who consider the case where one alternative is exogenously given as the ethical outcome. The last part of Proposition 4 however shows that this is only a limit result, and qualifies the finding of Feddersen et al. (2009). Increasing N can actually decrease the probability of acceptance of the sanction in the case where $\triangle(V)$ is a decreasing function at V^{ph}.

5 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 can choose, 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 consider 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.[16] We start by a welfare analysis of the benchmark case where actions are not visible before studying whether visibility is welfare enhancing.

5.1 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 $s = e$.

[16] This assumption is innocuous in the benchmark model where reputation is constant and in the public vote case where reputation is a zero sum game. 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.

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 sanction on welfare depends on the level of s. In this case with no deadweight loss of sanctions, submitting the sanction to a vote weakly decreases welfare.

Proposition 5 *In the* all secret *environment, the socially optimal sanction is always weakly higher than the optimal level without voting: $s^{hh} \geq e$. Under unanimity rule, it is strictly higher $s^{hh} > e$.*

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^{hh}. Since $V^{hh} > c - e$, 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.

5.2 The impact of visibility

We now compare the probability of acceptance of a sanction and ultimately welfare depending on which actions are observable.

5.2.1 Welfare comparison: public votes

We first compare the benchmark case with the public vote environment.

Proposition 6 *Comparing the public vote and all secret environments:*

1. *Any equilibrium voting cutoff in the public vote environment is smaller than the voting cutoff in the all secret environment.*

2. *Social welfare is larger in the public vote environment than in all secret.*

Result 1 of Proposition 6 shows that for any given sanction, the probability of acceptance is higher when votes are public. The intuition is the following. In the case where the voter is pivotal, the voting incentives are identical in the two environments since group members behave in the same way in the public good stage. However, when votes are public, voting in favor gives a reputation payoff that is absent in the *all secret* environment. Group members are thus always more inclined to vote in favor.

Result 2 is a direct consequence or Result 1: in the benchmark, the planner trades off the probability of acceptance against the risk of inducing ex post inefficient contributions. Since the probability of approval of any sanction is larger when the vote is public, the social planner can propose a smaller sanction and increase welfare with respect to the benchmark.[17] This lower sanction does not affect the acceptance of the sanction but decreases the likelihood of ex post inefficient contributions.

5.2.2 Welfare comparison: public contributions

As opposed to making votes visible, the effect on the voting cutoff of making contributions public is not unambiguous since visibility of contributions affects the calculus of reputation. 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 4: $-R^{hp}$ 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: $-R^{hp}$ is above $-R$ for high v_i. In the intermediate zone,

[17] This strategy increases welfare with respect to the benchmark but it is not necessarily the optimal sanction in the *public vote* environment.

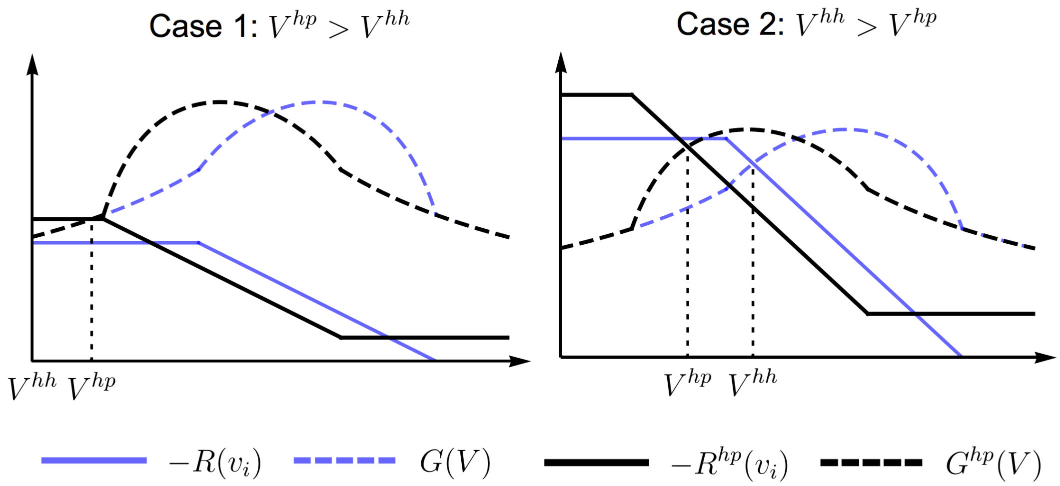

Figure 4: Comparing the cases *all secret* and *public contribution*

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^{hp} 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: $v_0^{hh} - v_s^{hh} = v_0^{hp} - v_s^{hp} = s$. This implies, as shown in Figure 4 that G^{hp} and G are equal when $v < v_s^{hp}$ and when $v > v_0^{hh}$, and that in the intermediate zone G^{hp} is above for low values of v_i and below for high values. For instance when V is just above v_s^{hp} and below v_s^{hh}, 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 7 *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:*

   - *If $e > e_m$, the voting cutoff is lower under all secret ($V^{hh} \leq V^{hp}$) 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: $V^{hh} > V^{hp}$*

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

The first result of Proposition 7 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^{hp} are always above $-R$ and $-R^{hp}$, which implies that all members vote in favor of the proposal ($V^{hh} = V^{hp} = v_{min}$). 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 4. In this range the proposal is more likely to be rejected when contributions are visible ($V^{hh} < V^{hp}$). 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 4). In this case, two forces decrease the voting cutoff when contributions are public. First, swing participants actually benefit in terms of reputation ($-R^{hp}$ moves below $-R$). Second, information aggregation makes voters more confident that the externality from adopting the sanction is large (G^{hp} above G). Overall, voters are more inclined to adopt the sanction when contributions are visible in this range ($V^{hh} > V^{hp}$).[18]

We assume in Proposition 7 that the distribution of types is uniform. For a general distribution, we have $G(0) = \frac{1}{2}e\left[F(v_0^{hh}) - F(v_s^{hh})\right]$, 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, $G^{hp}(0) = \frac{1}{2}e\left[F(v_0^{hp}) - F(v_s^{hp})\right]$. There is thus no systematic ordering of G and G^{hp}. Nevertheless, more general conditions on f would guarantee that Proposition 7 holds. For instance a sufficient condition for Result 1, first bullet point, is that $G^{hp}(0) > G(0)$.

Result 1 shows that in certain circumstances, sanctions are more likely to be accepted when contributions are secret rather than public. This leads us, in Result 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 condi-

[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.

tion. Specifically we consider a case where $V^{hh} = v_{min}$ 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, whereas making votes public always increases welfare, we identify conditions where rendering contributions visible is welfare decreasing.

6 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 votes public increases welfare while making contributions public may have a detrimental effect.

We considered separately the case where contributions and votes are made public. In the Supplementary Appendix, we discuss some results in the case where both are visible. Image is then based on both observed choices and the equilibrium of the voting stage is no longer necessarily characterized by a cutoff. In particular there exists an equilibrium where low types vote against the sanction and do not participate in the second phase, intermediate types vote for and do not participate and high types vote against and participate. In this equilibrium, intermediate types are ready to take the risk of voting in favor of the sanction and potentially losing if pivotal, to benefit from the increased reputation. If the sanction ends up passing, they however prefer not contributing. The fact that members might substitute one way of gaining a positive image against the other suggests it is not necessarily welfare increasing to make both decisions visible. Analyzing more generally multiple signaling channels to influence image could be the object of interesting future work.

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

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.

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.

Elster, J. (2015). *Secrecy and publicity in votes and debates*. Cambridge University Press.

Feddersen, T., S. Gailmard, and A. Sandroni (2009). Moral bias in large elections: theory and experimental evidence. *American Political Science Review 103*(02), 175–192.

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 (2015). Measuring image concerns. *CEPR Discussion Paper No. DP10831*.

Le Bihan, M. and B. Monnery (2018). Can public and private sanctions discipline politicians? evidence from the french parliament. *working paper*.

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 $V \in (v_s^{hh}, v_0^{hh})$.

In this region, as established in the main text:

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

The second derivative is thus given by:

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

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

$$f'(V)F(V) - 2(f(V))^2 \leq 0, \tag{4}$$

and

$$-f'(V)(1-F(V)) - 2(f(V))^2 \leq 0. \tag{5}$$

The first restriction of Condition B ($\frac{f}{1-F}(v)$ weakly increasing) implies condition (5) and the second restriction ($\frac{f}{F}(v)$ weakly decreasing) implies condition (4). 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 $[0, v_s^{hh}]$ ($-R(v) = s$), decreasing on $[v_s^{hh}, v_0^{hh}]$ ($-R(v) = v_i - c$) and equal to 0 if $v > v_0^{hh}$.
- $G(V)$ is increasing on $[0, v_s^{hh}]$ ($G(V) = \frac{1}{2}e\left[\frac{F(v_0^{hh})-F(v_s^{hh})}{1-F(V)}\right]$ as given in the main text), concave on $[v_s^{hh}, v_0^{hh}]$ (according to Step 1) and decreasing if $V > v_0^{hh}$ ($G(V) = \frac{1}{2}e\left[\frac{F(v_0^{hh})-F(v_s^{hh})}{F(V)}\right]$ 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, $V^{hh} = 0$.

ii $G(.)$ and $-R(.)$ intersect for $V^{hh} < v_s^{hh}$. In such a case, the concavity of $G(.)$ guarantees that a second crossing cannot exist. Indeed, the second crossing could only be in $[v_s^{hh}, v_0^{hh}]$ 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 $V > v_0^{hh}$. 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 $[v_s^{hh}, v_0^{hh}]$. 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^{hh} (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^{hh} (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^{hh}, leaves v_0^{hh} unaffected and thus increases $G(V)$ for all V. Moreover, $R(v)$ does not depend on s for $v \in [v_s^{hh}, v_0^{hh}]$, which is the case by definition if the equilibrium is interior. The voting cutoff V^{hh} is thus decreasing in s, which implies that the probability of approval is increasing.

3. For any majority requirement K, an interior equilibrium satisfies:

$$V^{hh} + e\left[\frac{K}{2N} \times \frac{F(v_0^{hh}) - F(V^{hh})}{1 - F(V^{hh})} + \frac{2N-K}{2N} \times \frac{F(V^{hh}) - F(v_s^{hh})}{F(V^{hh})}\right] - c = 0. \tag{6}$$

Considering K as a continuous variable, we can apply the implicit function theorem:

$$\frac{\partial V^{hh}}{\partial K} = \frac{\frac{e}{2N}\left[\frac{F(V^{hh})-F(v_s^{hh})}{F(V^{hh})} - \frac{F(v_0^{hh})-F(V^{hh})}{1-F(V^{hh})}\right]}{1 + \frac{e}{2N} f(V^{hh})\left[(2N-K)\frac{F(v_s^{hh})}{(F(V^{hh}))^2} + K\frac{F(v_0^{hh})-1}{(1-F(V^{hh}))^2}\right]}.$$

Given that in the unique equilibrium, G is increasing, the denominator is positive. Thus $\frac{\partial V^{hh}}{\partial K}$ is of the same sign as $\frac{F(V^{hh})-F(v_s^{hh})}{F(V^{hh})} - \frac{F(v_0^{hh})-F(V^{hh})}{1-F(V^{hh})}$, which is an increasing function of V^{hh}, negative at v_s^{hh} and positive at v_0^{hh}. By the intermediate value theorem, there exists a unique value $\widehat{V}^{hh}$ defined by

$$\frac{F(\widehat{V}^{hh}) - F(v_s^{hh})}{F(\widehat{V}^{hh})} = \frac{F(v_0^{hh}) - F(\widehat{V}^{hh})}{1 - F(\widehat{V}^{hh})},$$

such that if $V^{hh} \leq \widehat{V}^{hh}$, V^{hh} is decreasing in K (notice that the voting cutoff is also decreasing in K when $V^{hh} < v_s^{hh}$) and increasing in K if $V^{hh} > \widehat{V}^{hh}$.

Since V^{hh} is decreasing in e according to Result 1, we can express the result as a function of e: there exists $\tilde{e}^{hh}$ such that for $e > \tilde{e}^{hh}$, $V^{hh} < \widehat{V}^{hh}$, and the equilibrium voting cutoff is decreasing in K.

Proposition 2

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

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

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

Proposition 3

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

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

Thus there can be at most two equilibria, defined as the intersections of G^{hp} and $-R^{hp}$. Only one of these equilibria is stable, i.e. is such that the intersection occurs on a portion where G^{hp} is increasing. The different cases are illustrated in Figure 5 for $f \sim U[0, 1]$.

We now prove the comparative static results:

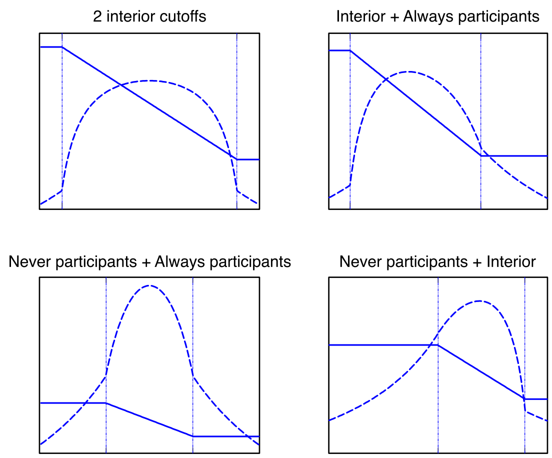

Figure 5: Four possible outcomes

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

2. We consider the case where the equilibrium is interior, i.e. V^{hp} is in the swing participants group. In that case, we have:

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

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

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

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

   which can be reexpressed:

$$\frac{e}{F(V^{hp})} < 2\frac{\mu}{(1 - F(v_s^{hp}))}(E[v_i|v_i > v_s^{hp}] - v_s^{hp}). \tag{7}$$

The right hand side of expression (7) is positive and does not depend on e. The left hand side is (strictly) increasing in e since, according to Result 1, V^{hp} 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^{hp} converges to 0). By the intermediate value theorem, there exists a unique value $\bar{e}^{hp}$ such that equation (7) holds if and only if $e > \bar{e}^{hp}$. In such a case V^{hp} is increasing in s.

3. In this case the interior voting cutoff is defined by:

$$V^{hp} + e\left[\frac{K}{2N} \times \frac{F(v_0^{hp})-F(V^{hp})}{1-F(V^{hp})} + \frac{2N-K}{2N} \times \frac{F(V^{hp})-F(v_s^{hp})}{F(V^{hp})}\right]$$
$$-c + \mu(E[v_i|v_i > v_s^{hp}] - E[v_i|v_i < v_0^{hp}]) = 0.$$

Compared to the expression (6), in the proof of Proposition 1, there is an additional term $E[v_i|v_i > v_s^{hp}] - E[v_i|v_i < v_0^{hp}]$. However, this term does not depend on K, and the expression of the derivative, and the following arguments, are thus the same as in the proof of Proposition 1.

Proposition 4

The equilibrium condition can be expressed as:

$$\frac{\Delta(V^{ph})}{Piv(V^{ph})} + G(V^{ph}) = -R(V^{ph}) \qquad (8)$$

$G(V)$ and $-R(v)$ are identical to the baseline model. Compared to the equilibrium condition in the benchmark case, the left hand side of (8) includes the additional term $\frac{\Delta}{Piv}(V)$, which is positive and goes to infinity for $V \to v_{min}$ and $V \to v_{max}$. Moreover, in any stable equilibrium, the left hand side must cross $-R(v)$ from below.

We now derive the comparative static results for interior and stable equilibria

1. Since neither $\Delta(V^{ph})$ nor $Piv(V^{ph})$ depend on s or e, the proof is identical to the proof of Results 1 and 2 of Proposition 1.

2. For $e = 0$, the equilibrium condition is $\frac{\Delta(V^{ph})}{Piv(V^{ph})} = -R(V^{ph})$

The pivotal probability is given by:

$$Piv(V) = \binom{K-1}{2N}(1-F(V))^{K-1}(F(V))^{2N+1-K}.$$

Taking the derivative:

$$\begin{aligned}\frac{\partial Piv(V)}{\partial K} =& \binom{K-1}{2N}(1-F(V))^{K-1}(F(V))^{2N+1-K} \\ & \times [\log(1-F(V)) - \log(F(V)) + \psi^{(0)}(K) - \psi^{(0)}(K-2N)]\end{aligned}$$

Where $\psi^{(0)}$ is the polygamma function of order 0 defined as $\psi^{(0)} = \frac{\Gamma'(z)}{\Gamma(z)}$, $\Gamma(z)$ being the gamma function. $\frac{\partial Piv(V)}{\partial K}$ thus has the sign of $\log(1-F(V)) - \log(F(V)) + \psi^{(0)}(K) - \psi^{(0)}(K-2N)$ which is decreasing in V. Thus there exists $\hat{V}^{ph}$ such that $Piv(V)$ is decreasing in K if and only if $V > \hat{V}^{ph}$. Since neither $\Delta(.)$ nor $R(.)$ depend on K:

- For $V \leq \hat{V}^{ph}$, an increase in K shifts downwards $\frac{\Delta(V^{ph})}{Piv(V^{ph})}$ and for all stable equilibria the intersection with $-R$ occurs at a higher cutoff.
- For $V > \hat{V}^{ph}$, an increase in K shifts upwards $\frac{\Delta(V^{ph})}{Piv(V^{ph})}$ and for all stable equilibria the intersection with $-R$ occurs at a lower cutoff.

To conclude the proof, we establish the comparative statics with respect to N.

- We first show that the equilibrium where all types vote against the sanction is not a sequential equilibrium. Indeed for all totally mixed strategies, it has to be the case that $E[v|b_i = 1] - E[v|b_i = 0] > 0$, because of the cutoff property. As a result, in all sequential equilibria, a vote in favor enhances reputation and always participants vote in favor of the sanction to benefit from the externality gain.

  We now show that when $N \to +\infty$ no player can vote against the sanction.

  In equilibrium, the net benefit of the never participants to vote in favor of the sanction are given by:

  $$\mu\left(E[v|b_i = 1] - E[v|b_i = 0]\right) + Piv\left(-s + G\right).$$

  $-s + G$ is bounded and $Piv(V)$ converges to zero when $N \to +\infty$ for all $V \in [v_{min}, v_{max}]$. As $E[v|b_i = 1] - E[v|b_i = 0] > 0$, never participants (and thus all players due to the cutoff property) vote for the sanction when N is large enough.

- For all intervals, the indifference condition characterizing the equilibrium can be rewritten:

$$\mu\triangle^* = Piv\Lambda,$$

  where Λ can take different values depending on which interval the equilibrium cutoff belongs to, but does not depend on N.

  An increase in N, for a given V decreases Piv. Thus, if $\triangle^*$ is decreasing in V for $V = V^{ph}$, we see that an increase in N leads to an increase in V^{ph} (i.e decrease in probability of acceptance).

Proposition 5

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 $v_i \geq V^{hh} > c - e$.

Proposition 6

1. Equilibrium is given by

$$H(V) = -R(V)$$

   where:

   - in the benchmark model (all secret environment) $H(V) = G(V)$
   - in the public vote environment $H(V) = \frac{\Delta(V)}{Piv(V)} + G(V)$

Given that the voting cutoffs in the contribution phase are identical, the function G is the same. Thus given that $\frac{\Delta}{Piv} > 0$, all intersections between H and $-R(V)$ happen for lower cutoffs in the public vote environment compared to the unique intersection for the all secret environment.

2. Consider the socially optimal sanction in the all secret environment s^{hh}. If the same sanction is submitted to the vote in the *public vote* environment, according to Result 1, the probability of approval is strictly larger. Considering lower values of s, one of the following must hold:

   i There exists $\bar{s} : e \leq \bar{s} < s^{hh}$ such that $V^{ph}(\bar{s}) = V^{hh}(s^{hh})$.

   ii No such $\bar{s}$ exists and $V^{ph}(e) < V^{hh}(s^{hh})$

   When i. is true, the sanction s^{hh} is accepted in the *all hidden* environment by the groups that would accept $\bar{s}$ with *public vote*. But since $e \leq \bar{s} < s^{hh}$, less ex post inefficient contributions are induced in the second case and welfare is larger when votes are public.

   If ii. holds, suppose that the planner proposes $s = e$ with *public vote*. The sanction is accepted by the groups which would accept s^{hh} under *all hidden* (and also in some other groups). Moreover, the sanction induces no ex post inefficient contributions. Welfare is thus also larger in *public vote*.

Proposition 7

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

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

   Decreasing e further, we can reverse the inequality and get $V^{hh} < V^{hp}$. Consider the value of $\bar{V}$ such that $-R^{hh}(\bar{V})$ and $-R(\bar{V})$ intersect. Since G^{hp} is increasing in e, we can find a value of $\tilde{e}_l$ such that $V^{hp} = \bar{V}$, in other words, $G(V^{hp}) = -R(V^{hp}) = -R^{hp}(V^{hp})$. For this value we have $G^{hp}(V^{hp}) > G(V^{hp})$, so that the intersection of G and $-R$ is such that $V^{hh} < V^{hp}$. 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 ($V^{hh} \leq V^{hp}$) 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: $V^{hh} > V^{hp}$

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 $s^{hh} = e$. Suppose that this sanction is submitted to a vote. We have:

$$G(0) = \frac{e}{2}[v_0^{hh} - v_s^{hh}] = \frac{e}{2}s = \frac{e^2}{2}$$

   and

$$-R(0) = e.$$

   If $e > 2$, we have $G(0) > -R(0)$ which implies $V^{hh} = 0$ 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 $s^{hp} = e - \frac{\mu}{2}$. If this sanction is submitted to a vote, we now have:

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

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

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

This implies that $V^{hp} = 0$ 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 ($s^{hh} = e$ is always accepted) while in the *public contributions* setup $V^{hp} > 0$ if $s^{hp} = e - \frac{\mu}{2}$ is submitted to a vote and thus the sanction that would lead to the first best is rejected with some probability.

Parameters used in the different figures in the paper

In all figures we considered the case where $f \sim U[0, 1]$. The other parameters are given as follows:

Figure	e	s	c	μ	K	$2N+1$
1	0.7	0.6	0.8	0	N	$2N+1$
2	2.7	0.8	3	4.2	N	$2N+1$
3	0	0.8	0.9	0.1	2	3
4-1	2.1	0.5	0.85	0.4	N	$2N+1$
4-2	1	0.5	0.85	0.4	N	$2N+1$

Supplementary Appendix A1: Observing votes and contributions

We consider the case where both contributions and votes are observable. To clarify the forces at play, we focus on the case where the sanction s submitted to a vote is large, $s > v_{min} - c$, so that if the sanction is voted, all types participate. Consider the case where the sanction was not voted in the first stage of the game. The behavior of the players will depend on the way they voted in the first phase. Conditional on a vote, the players will choose a cutoff strategy. We denote the cutoff $v_0^{pp}(1)$ for the players who voted in favor of the sanction in the first phase and $v_0^{pp}(0)$ for those who voted against.

Using the notation $E_0(b, a) = E\left[v_i | (b_i = b) \cap (a_i = a) \cap (s = 0)\right]$ (for instance $E_0(1, 0)$ is the expected value of v given that the player voted for the sanction, the sanction was not passed and he did not participate), the cutoff is defined by

$$v_0^{pp}(i) = c - \mu \left[E_0[i, 1] - E_0[i, 0]\right]$$

There is no clear ordering between $v_0^{pp}(0)$ and $v_0^{pp}(1)$. On the one hand, those who already sent a bad signal by voting against the sanction, might have little to lose by not participating. On the other hand, those who already voted for the sanction, can afford to send a bad signal of not participating. The ranking will depend on inferences made in equilibrium.

Proposition 8 below presents properties of equilibria such that certain types vote in favor of the sanction and some against. The full set of equilibria is described in the proof. The first key property is that the voting strategy is not necessarily of the cutoff form. In particular there is an equilibrium where low types vote against and do not participate, intermediate types vote for and do not participate and high types vote against and participate. In this equilibrium, the low types do not want to vote in favor because contributing is too costly and they consider the case where their vote can be pivotal. The intermediate types are ready to take the risk of voting in favor and potentially losing if pivotal, to benefit from the increased reputation. They however do not want to deviate to action (0,1) since contributing is still too costly.

Proposition 8 *In the voting phase:*

1. *There exist equilibria where* $v_0^{pp}(0) < v_0^{pp}(1)$ *and others where the ordering is reversed*

2. *The equilibrium voting strategy is not necessarily a cutoff strategy. In particular there exists an equilibrium with two cutoffs $\underline{v}$ and $\overline{v}$ such that:*

   - *for $v < \underline{v}$, players vote against the sanction and do not participate, i.e choose (0,0)*
   - *for $\underline{v} \leq v \leq \overline{v}$, players vote for the sanction and do not participate (1,0)*
   - *for $v > \overline{v}$, players vote against the sanction and participate (0,1)*

Proof: As indicated in the main text, in the public good stage, players will use cutoff strategies conditional on their voting behavior in the first stage. There are therefore two relevant cutoff: $v_0^{pp}(0)$ and $v_0^{pp}(1)$.

Conditional on a particular equilibrium, denote P_p the probability that the sanction is adopted independently of individual i's vote, P_r the probability that the sanction is rejected independently of individual i's vote, and P_{piv} the probability that the individual is pivotal. All these probabilities are independent of the player's actual type.

We also use the notation $E_0(b, a) = E\left[v|(b_i = b) \cap (a_i = a) \cap (s = 0)\right]$. For instance $E_0(1, 0)$ is the expected value of v given that the player voted for the sanction, the sanction was not passed and he did not participate

Consider group members with $v_i < min(v_0^{pp}(0), v_0^{pp}(1))$. If he votes in favor, his expected benefit is:

$$P_p(\mu E_s[1,1] + v_i - c + G_s) + P_r(\mu E_0[1,0] + G_0) + P_{piv}(\mu E_s[1,1] + v_i - c + G_s),$$

where G_s (resp. G_0) denotes the expected externality payoff when the sanction is passed (resp. not passed).

If the group member votes against, his expected benefit is:

$$P_p(\mu E_s[0,1] + v_i - c + eG_s) + P_r(\mu E_0[0,0] + eG_0) + P_{piv}(\mu E_0[0,0] + eG_0).$$

The net benefit of voting in favor for that individual is thus

$$D(v_i) = P_p\mu(E_s[1,1] - E_s[0,1]) + P_r\mu(E_0[1,0] - E_0[0,0]) + P_{piv}(\mu(E_s[1,1] - E_0[0,0]) + v_i - c + G^{pp})$$

$D(v_i)$ is increasing in v_i on that interval.

Similarly, if $v_i > max(v_0^{pp}(0), v_0^{pp}(1))$, the net benefit of voting in favor is given by

$$D(v_i) = P_p\mu(E_s[1,1] - E_s[0,1]) + P_r\mu(E_0[1,1] - E_0[0,1]) + P_{piv}(\mu(E_s[1,1] - E_0[0,1]) + G^{pp}), \quad (9)$$

which is independent of v_i.

We now consider the intermediate regions.

Suppose first $v_0^{pp}(0) < v_0^{pp}(1)$ and consider the case $v_0^{pp}(0) < v_i < v_0^{pp}(1)$. Such a group member participates when the sanction did not pass, if and only if he voted against the sanction.

If he votes in favor, his expected benefit is:

$$P_p(\mu E_s[1,1] + v_i - c + G_s) + P_r(\mu E_0[1,0] + G_0) + P_{piv}(\mu E_s[1,1] + v_i - c + G_s)$$

If he votes against, his expected benefit is:

$$P_p(\mu E_s[0,1] + v_i - c + G_s) + P_r(\mu E_0[0,1] + v_i - c + G_0) + P_{piv}(\mu E_0[0,1] + v_i - c + G_0)$$

The net benefit of voting in favor for that individual is thus

$$\begin{aligned} D(v_i) &= P_p\mu(E_s[1,1] - E_s[0,1]) + P_r(\mu(E_0[1,0] - E_0[0,1]) - (v_i - c)) \\ &+ P_{piv}(\mu(E_s[1,1] - E_0[0,1]) + G^{pp}). \end{aligned}$$

$D(v_i)$ is then decreasing in v_i on that interval.

Suppose on the contrary that $v_0^{pp}(1) < v_0^{pp}(0)$ and consider types with $v_0^{pp}(1) < v_i < v_0^{pp}(0)$, then the net benefit of voting in favor for that individual is thus

$$\begin{aligned} D(v_i) &= P_p\mu(E_s[1,1] - E_s[0,1]) + P_r\mu(E_0[1,1] - E_0[0,0] + (v_i - c)) \\ &+ P_{piv}(\mu(E_s[1,1] - E_0[0,0]) + (v_i - c) + G^{pp}), \end{aligned}$$

which is increasing in v_i

Consider case A: $v_0^{pp}(1) < v_0^{pp}(0)$.

In this case as shown above, if we impose restriction A as before,the voting strategy is a cutoff strategy with cutoff V^{pp} since the net benefit function $D(v_i)$ is

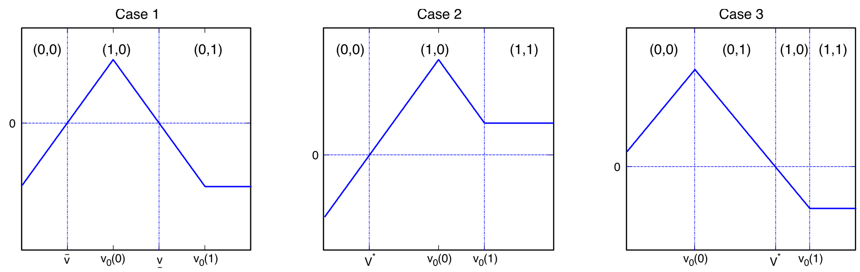

Figure 6: Comparing the cases *both secret* and *secret vote*

weakly increasing in v_i on all intervals and continuous. There are three situations

- $V^{pp} < v_0^{pp}(1)$ then there are three zones with respective outcomes (0,0), (1,0) and (1,1),
- $v_0^{pp}(1) < V^{pp} < v_0^{pp}(0)$ with two zones: (0,0) and (1,1),
- $V^{pp} > v_0^{pp}(0)$ with three zones (0,0), (0,1) and (1,1).

We now check whether these equilibria are compatible with the condition $v_0^{pp}(1) < v_0^{pp}(0)$. We have

$$v_0^{pp}(0) = c - \mu(E[0,1] - E[0,0]),$$
$$v_0^{pp}(1) = c - \mu(E[1,1] - E[1,0]).$$

So that $v_0^{pp}(1) < v_0^{pp}(0)$ is equivalent to:

$$E[1,1] - E[1,0] > E[0,1] - E[0,0]. \tag{10}$$

Only that values $E[1,1]$, $E[0,0]$ and $E[1,0]$ (in the second equilibrium) are pinned down in equilibrium, and thus $E[0,1]$ can be chosen low enough to guarantee that condition (10) is satisfied.

Consider case B: $v_0^{pp}(0) < v_0^{pp}(1)$. In this case as shown above, the voting strategy is no longer necessarily a cutoff strategy since the net benefit curve $D(v_i)$ first increases in v_i then decreases. There are then potentially three cases for an equilibrium with some types voting in favor and some against.

Consider the case $e = 0$, then there exists a value of μ such that $D(v_i)$ intersects the zero line twice and you thus have outcomes (0,0) for low values of v_i, outcomes

(1,0) for intermediate values and (0,1) for high values, as indicated in the result of the Proposition. This is represented in case 1 in Figure 6.

Suppose now that $D(v_i)$ intersects the zero line once for $v < v_0^{pp}(0)$ (case 2 in Figure 6). Agents will then choose (0,0) for low values, (1,0) for intermediate values and (1,1) for high values.

Finally, if $D(v_i)$ intersects the zero line once for $v_0^{pp}(0) < v < v_0^{pp}(1)$ (case 3 in Figure 6), there are four zones (0,0), (0,1), (1,0) and (1,1).

Note furthermore that the condition $v_0^{pp}(0) < v_0^{pp}(1)$ is equivalent to

$$E[1,1] - E[1,0] < E[0,1] - E[0,0]. \tag{11}$$

We can again find beliefs that will imply condition (11).

Supplementary Appendix A2: Approval of sanction may increase in K

We present an example where the sanction is more likely to be implemented when K increases. We consider that the distribution of types f is uniform on [0,1]. We use the following parameters: $c = 0.95$, $s = 0.4$, $e = 0.8$. In this group of 3 players we impose different supermajority requirements, namely $K = 2$ and $K = 3$ and we plot the results in Figure 7.

For $K = 2$, the equilibrium $V_{K=2}^{hh}$ is in the swing participants group. It solves:

$$\frac{e}{2}[\frac{V_{K=2}^{hh} - v_s^{hh}}{V_{K=2}^{hh}} + \frac{v_0^{hh} - V_{K=2}^{hh}}{1 - V_{K=2}^{hh}}] = c - V_{K=2}^{hh}$$

Numerical resolution reveals that $V_{K=2}^{hh} \approx 0.58$ and the sanction is accepted with probability:

$$\begin{aligned} P(\text{sanction is accepted}) &= P(3 \text{ yes votes}) + P(2 \text{ yes votes}) \\ &= (1 - V_{K=2}^{hh})^3 + V_{K=2}^{hh} \times (1 - V_{K=2}^{hh})^2 \times 3 \\ &\approx 0.38 \end{aligned}$$

Now consider the case $K = 3$. According to Figure 7, the cutoff is in the never participants group and is defined by:

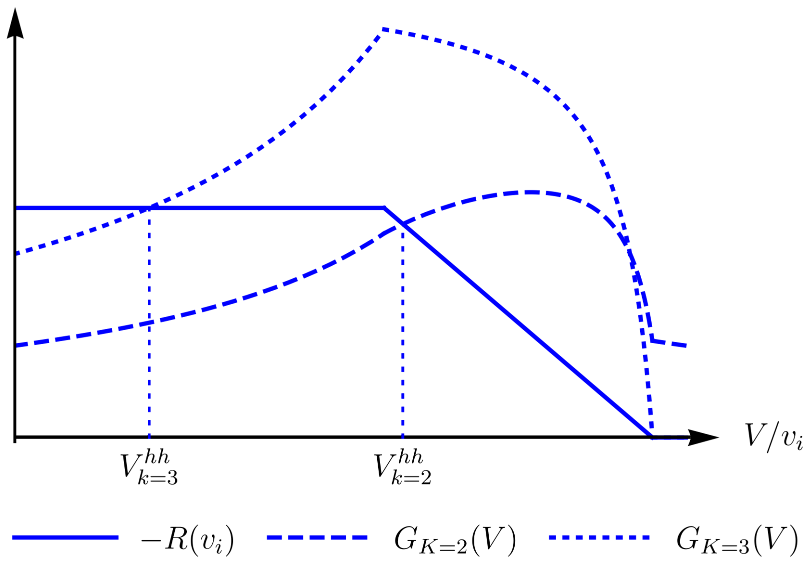

Figure 7: Changing the majority rule

$$e[\frac{v_0^{hh} - v_0^{hh}}{V_{K=3}^{hh}}] = s$$

Plugging the parameters, the voting cutoff that satisfies this equation is $V_{K=3}^{hh} = 0.2$. The probability of acceptance is thus:

$$\begin{aligned} P(\text{sanction is accepted}) &= P(3 \text{ yes votes}) + P(2 \text{ yes votes}) \\ &= (1 - V_{K=3}^{hh})^3 \\ &= 0.512 \end{aligned}$$

In this example, the sanction is more likely to be accepted when the required supermajority increases. This result is due to information aggregation: when $K = 3$, a lower cutoff can be sustained because players expect a higher number of swing participants (and thus a larger externality gain) when they condition on the pivotal event. The equilibrium voting cutoff is smaller when we require unanimity to pass the sanction, which makes the sanction more likely to be accepted even if an additional vote in favor is required.