Obedience to Rules with Mild Sanctions: The Roles of Peer Punishment and Voting

Chen, Josie I

Brown University

15 April 2014

Obedience to Rules with Mild Sanctions: The Roles of Peer Punishment and Voting

Josie I Chen[1]

Abstract. Governments sometimes promote rules backed by sanctions too weak to make obedience privately optimal. Factors that may help make such rules effective include the presence of informal sanctions by peers, and implementation through voting. I study the impact of non-deterrent formal sanctions on voluntary contributions to a public good in a laboratory experiment. The effect is studied both in the presence and absence of informal sanctions, under fully exogenous implementation and after both implemented and randomly overridden voting. I find that informal sanctions strengthen the effect of formal ones in most conditions. However, voted implementation has no clear effect on non-deterrent formal sanction in my data, which suggests a reason for caution when studying exogenous implementation by a random vote override procedure.

Keywords: experiment, voluntary contribution, public goods, formal sanctions, informal sanctions, voting

[1] Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912 (email: I_Chen@brown.edu). TEL: +1-401-935-2689. I am grateful to Louis Putterman for his invaluable guidance and discussion. I am also grateful to Jean-Robert Tyran, Pedro Dal Bó, Roberto Serrano and the seminar audience at Brown University for helpful suggestions. I thank Kenju Kamei and Tai-sen He for sharing their z-tree files with me. In addition, I thank Owen Powell, Pablo Torija, Joe Rieff and Anna Albert for their assistance in the experiment. This project was supported by a grant from the Vienna Center for Experimental Economics (VCEE).

1. Introduction

Possibly the most common tool used to induce people to follow laws and fulfill legal requirements, such as paying taxes or not throwing trash on the street, is threatening non-compliers with pecuniary sanctions such as fines or penalties. The sanctions imposed are rarely sufficient on their own to cause compliance to be in the private material interest of most individuals, given low probability of detection and consequent difficulties in enforcement. Two likely reasons why many people comply with such mildly enforced laws, nonetheless, are (a) the fear that others who agree with the rationale behind the laws are willing to impose informal sanctions, including forms of social disapproval, on those who violate them, and (b) that selection of the laws and penalties by vote builds support for them and/or imbues them with moral authority. The impact of voting and of the fear of informal sanctions may also interact, if majority vote outcomes raise the perceived likelihood that others will impose informal sanctions when rules are violated.

How people respond to formal sanctions (FS) such as centrally administered fines and penalties, and to informal sanctions administered horizontally by peers (IS), have been studied under controlled conditions in numerous laboratory decision-making experiments. The effects of combining IS and FS, and the effects of voting on the impact of non-deterrent formal sanctions, have only recently begun to be studied, however. Tyran and Feld (2006) and Kamei (2013) found that non-deterrent formal sanctions (NFS) significantly raised contributions to a public good when chosen by vote but not when imposed exogenously by the experimenter, supporting (b) above. Kube and Traxler (2011) found the combination of IS with NFS to be more effective than IS-only in encouraging contributions to a public good. But they did not study the performance of NFS-only, so the proposition that IS aids the efficacy of NFS ((a) above) remains

untested, to my knowledge. The influence of voting on the effectiveness of combined NFS+IS has also not been studied. In the current study, I begin to fill these gaps.

Determining whether implementation of a policy or institution by voting influences its effectiveness is complicated by the need to control for the self-selection of those using it in the voted condition. A methodological innovation by Dal Bó, Foster and Putterman (2010, hereafter DFP), also adopted by Kamei (2013), has subjects vote on an institution but a computer then randomly decide whether the vote will be binding. If the random outcome sets the vote aside, the computer randomly assigns or does not assign the institution. The method thus yields observations of groups with identical shares of votes for the institution, and of individual members of such groups who voted identically, under both the endogenously and the exogenously assigned institution. Upon finding that the institution (in the DFP case, a payoff modification changing a prisoners' dilemma to a coordination game) leads to significantly more cooperation when adopted by vote, even after controlling for selection, DFP investigated whether that outcome might be due to a signaling effect of knowing how the group's majority voted. The authors implemented a condition in which groups whose vote was overridden had the majority vote reported to them, and found that cooperation was greater when the vote counted than when vote did not count, the computer randomly implemented the institution, and subjects had the same information about the group vote. Thus, most of the effect of endogenous implementation by the majority appeared not to be a signaling effect. In Kamei's study, subjects played a voluntary contribution game and voted on whether to adopt NFS. Results on the impact of voting, after controlling for selection, and findings on whether that result is attributable to signaling, were qualitatively similar to those of DFP.

To study the effect of voting and whether that effect (if present) is attributable to signaling, in the case of NFS and its combination with IS, I follow essentially the same procedure as DFP. Given my interest in comparing the combined impact of NFS+IS to that of NFS-only, this means observing the three conditions of play—simple voluntary contribution mechanism (VCM, also called the linear public goods game), NFS, and NFS+IS—under three choice situations: voted choice, exogenous overriding of vote with vote feedback, and exogenous overriding of vote without vote feedback. For completeness, I also study an IS-only condition under the three situations. In addition, I add a fourth situation not studied by DFP: I observe the effects of each condition when assignment is entirely exogenous and there is neither any voting nor any mention of a vote. I do this not only because it makes exogenous implementation more straightforward, but also because it affords an opportunity to check whether the steps---of deciding how to vote, voting, and learning that the vote has been overridden---have their own effects on the performance of the institutions being studied.

My results confirm (a) above: the availability of IS strengthens the impact of NFS. This is so not only in the fully exogenous condition, but also when chosen by vote and when assigned following a vote override, whether with or without vote feedback. The effectiveness of IS in this respect is easily understood in material terms: my data show that when added to the formal sanction, the cost of contributing to one's private account due to expected informal sanctions makes contributing to the public account the more profitable alternative at least until one matches the group average contribution. Somewhat unlike Tyran and Feld, however, I find that NFS-only has a significant short-term effect even when assigned without vote. The effect of NFS quickly decays without the support of IS, however, regardless of whether NFS is chosen by vote.

My findings regarding the effects of voting also raise a note of caution about the method employed by DFP: I find that contributions are higher under NFS with a binding vote than with vote override, an apparent democracy effect in the framework of DFP. However, contributions are higher in groups assigned NFS without voting than in groups assigned NFS when the vote does not count and there is no vote feedback. This last finding suggests that vote overrides may engender negative emotions, the effects of which could be misinterpreted as positive effects of democracy.

2. Background and Literature

In modern societies, penalties for non-compliance usually reinforce laws. But, because the punishments required when the apprehension of violators is uncertain would be viewed as too harsh by prevailing standards, the penalties actually imposed are often too small to change the privately optimal behavior on their own. That mildly enforced laws nevertheless often meet with substantial levels of compliance may be explained by a number of factors, including the possibly complementary effects of informal sanctions and the normative and informational effects of voting.

A simple model of additive linear utilities captures the gist of the issue. Consider a social norm compliance with which requires that an individual forgo a private benefit x. To help induce individuals to comply with the norm, the state imposes a penalty of expected value y on those failing to comply, where y can be understood as the product of the penalty if caught and the probability of detection. I call it a formal sanction because it is imposed by the state or some other central body, and I call it a non-deterrent formal sanction (NFS) because $y < x$ in the cases that concern me. If opportunities exist to impose informal sanctions (IS), those failing to comply

with the norm may anticipate an average punishment, z, by peers. Some also may internalize the norm, feeling it important to comply with it, either due to its inherent moral worth, the perceived support of others for it, or a sense of commitment to the decision from having voted for it. These individuals incur a cost w if they violate the norm.

An individual can be assumed to comply with the norm if $(y + z + w) > x$. The latter condition may occur even when $y = 0$, if $(w + z) > x$, or when $w = 0$, if $(y + z) > x$. The value of w may be individual-specific and may be conditional on whether the requirement is determined by majority vote, whether the majority is known to favor the project, and on whether others are perceived to support the requirement, as indicated by their own compliance levels.

Until recently, there was little contact between the literature on formal and that on informal sanctions. Formal sanctions were studied mainly in relation to tax compliance, with experimental studies of the topic focusing on the possibly differential effects of variation in the size of penalties versus the probability of detection (Torgler, 2002; Anderson and Stafford, 2003).[2] The topic of informal sanctions attracted the attention of those interested in voluntary collective action, with publications in psychology (Yamagishi, 1986) and political science journals (Ostrom, Walker and Gardner, 1992) preceding those in economics by a decade or more.

Fehr and Gächter (2000) pioneered a specialized literature on voluntary contributions under the threat of informal sanctions [see Gächter and Herrmann (2008) and Chaudhuri (2011) for reviews]. Initially, literature focused on both the conventionally unpredicted willingness of

[2] Anderson and Stafford (2003) study the effect of formal sanctions in a VCM game, varying both probability of detection and size of fine across treatments. Although the combinations studied include expected sanction levels that should deter free riding, other combinations fall in the non-deterrent range. Within that range, they find that contributions increase as the expected sanction rises and that subjects are more responsive to the size of the sanction than to the probability.

individuals to punish low contributors to a public good and on the power of anticipated punishment to replace the decaying contribution trends of earlier voluntary contribution experiments (Ledyard, 1997) with sustained or even rising contributions. Later, however, some contributions focused on the mixed effects of punishment on efficiency, on the presence of misdirected punishment, on the implications of opportunities to counter-punish, and on whether groups would freely choose to subject themselves to informal sanctions. Recognizing the ubiquity of misdirected sanctions (Cinyabuguma et al., 2006; Herrmann et al., 2008), Ertan et al. (2009) studied subjects' inclination to to permit informal sanctions to be imposed on free riders only.

Parallel to Ertan et al. (2009), Putterman et al. (2011) conducted a study of voting on the targeting and level of formal sanctions. Most recently, Kamei et al. (forthcoming) and Markussen et al. (2014) investigated subjects' preferences between formal and informal sanction regimes, with both studies finding the choice to hinge on the fixed cost of using a formal sanction regime. Kube and Traxler (2011) and Andreoni and Gee (2011) study the coexistence of FS and IS, on which I comment further below.

In studies that the use of formal sanction is put to a vote of between an option of only FS and an option of only IS, a potential problem is that insofar as informal sanctions take forms such as social disapproval, they may be considered a default phenomenon which is costly to suppress and perhaps impossible to eliminate completely. If so, the relevant choice in many settings may not be between informal and formal sanctions, but rather between the default condition of informal sanctions and the additional imposition of costly formal sanctions co-existing with IS.[3]

[3] To be sure, often the formal authorities undertake to suppress many forms of peer punishment when a formal sanction regime is put in place. "Civilization," after all, may be perceived as being about the replacement of

Moreover, its co-existence with IS may be key to what makes NFS effective, as Kube and Traxler (2011) point out. Further study of NFS in the presence of IS is accordingly called for.

As I raise in my discussion above, there is a rather general possibility of a direct subjective payoff *w* from adhering to a norm or law, but such a payoff is investigated specifically in my paper only insofar as it is related to the process and outcome of voting. An often assumed advantage of democracy is that a greater level of compliance with laws may be achieved with less expenditure on enforcement and punishment because democratically-determined law carries greater legitimacy, especially among those who voted for it. In addition to such direct effects of democracy, vote outcomes may affect the compliance of conditional cooperators, that is, individuals whose willingness to comply with a norm or law is an increasing function of the number of peers whom they expect to comply (Fischbacher and Gächter 2010). To such individuals, the fact that a majority favored a law may suggest that many will comply with it, increasing their own willingness to comply. This factor implies that a vote can, in principal, affect compliance even if it is merely advisory or is for some reason overridden and the rule is imposed by a different mechanism.

Studying the impact of democracy on compliance with laws, or on the effects of policies or institutions more generally,generally is difficult. This is due to the fact that comparing a policy, mechanism or law's effect on groups experiencing it due to exogenous imposition to that in groups experiencing it thanks to their members' votes runs the risk of confusing the effect of voting. It potentially conflates the effect of voting with the fact that the groups experiencing the policy (etc.) by vote are made up of individuals who may differ in some unmeasured respect

---

punishment by individuals with a formal and hopefully impartial system of justice. Since the degree to which IS is suppressed may depend on how much the authorities spend on achieving its suppression, realistic models could involve inexpensive FS co-exiting with unconstrained IS, more expensive FS entailing suppression of much but not all IS, and so on.

from those experiencing it by imposition. Tyran and Feld (2006) attempt to address this issue by comparing observable characteristics of their exogenously and endogenously exposed subject groups. The alternative approach of Dal Bó et al. (2010) is adopted in Kamei (2013) and in the present paper but, as discussed above, I also compare schemes chosen by vote to ones imposed exogenously without voting, partly to check whether voting influences behaviors even when the vote is not counted. The details of my procedure are discussed in section 3.

3. Design

I study the potential interaction of NFS and IS both with and without voting using the familiar social dilemma design of the finitely repeated linear voluntary contribution mechanism. I prefer a repeated to a one-shot game because learning may be important and because the difference between the trends in contributions is one of the most distinctive differences of voluntary contribution behavior without and with IS. I prefer finite repetition for its simple predictions under classical assumptions of self-interest, rationality, and common knowledge of these. I use a partner design to statistically isolate the subjects in each group from others within their multi-group experimental sessions. Each group has 5 subjects so that tied votes cannot occur and because I anticipated richer interactions with slightly larger subject groups. I have each group first play six periods of a standard VCM with neither formal nor informal sanctions so that in voting treatments, votes on whether to use NFS will be informed by some understanding of voluntary contribution dynamics. All treatments are accordingly identical until after period 6, when subjects receive new instructions referring to the institutions that might be available in their treatment for the second and final set of six periods. I use a familiar

endowment, 20, but an MPCR of 0.3, which seems adequate considering the group size.[4] The payoff of subject *i* in a period of Phase 1 is given by

$$\pi_i = 20 - g_i + 0.3\sum_{i=1}^{5} g_i \qquad (1)$$

where 20 is the endowment, g_i is the amount subject put in the group account, 0.3 is the MPCR and 5 is the group size,.

Beginning in period 7, subjects play the remaining six periods under one of four possible conditions: simple VCM, IS, NFS, and IS + NFS. My design takes the presence or absence of IS as a strictly exogenous treatment variable, in part to simplify what are already a large number of conditions and situations to be studied, and in part because I wanted to take seriously the argument that IS may be a default condition, not an institution subject to choice. In half of my treatments, the possibility of informal sanctions is never mentioned in the instructions, there are no opportunities to engage in IS, and the only change possible when transitioning from Phase 1 (which includes periods 1 – 6) to Phase 2 (which includes periods 7 – 12) is replacement of VCM condition by NFS. In the other half of the treatments, there are definitely opportunities to give informal sanctions in Phase 2, and the transition from Phase 1 to Phase 2 involves either replacement of VCM by IS or replacement of VCM by IS + NFS. The instructions read to and by subjects after Phase 1 explain the nature of NFS, in treatments without IS, or the nature of both NFS and IS, in treatments with IS. When IS is available, it costs a subject 1 point to reduce the earnings of the targeted individual by 2 points. I chose a relatively low punishment effectiveness (see Nikiforakis and Normann, 2008) so that IS would be less likely to render NFS strictly redundant. I added the constraint that punishment received cannot drive the recipient's

[4] An MPCR of 0.4 and a group size of 4 has been common beginning with Fehr and Gächter (2000).

earnings for the period below zero, so as to reduce the possibility that subjects would need to pay the experimenter. However, I require subjects to pay for punishment they choose to give even if it drives their earnings below zero, due to the importance from a theoretical standpoint of having punishment be costly to give.[5] When subjects play under IS-only, the payoff of a subject *i* is given by

$$\pi_i^{IS} = \max \{0, 20 - g_i + 0.3\sum_{i=1}^{5} g_{i\_} - 2\Sigma_j s_{ji} \} - \Sigma_j s_{ij} \qquad (2)$$

where 2 is the Punishment effectiveness (cost to receiver), s_{ji} is the expenditure for a subject $j \neq i$ to punish subject *i*, and s_{ij} is the expenditure for subject *i* to punish subject *j* $(j \neq i)$.

In both IS and no-IS treatments subjects are uncertain whether their group will interact in Phase 2 with or without NFS, when the instructions that follow Phase 1 end. That question is determined in one of three ways, yielding the three-way division of treatments shown by the rows in Table 1. The top row shows treatments in which assignment of NFS for Phase 2 is determined randomly and there is neither voting nor mention of the possibility of voting.[6] The remaining two rows show treatments in which the question is determined following a vote, although whether or not the vote counts (i.e., determines what institution the group plays under) is itself randomly determined by the computer, as in DFP. (In DFP, subjects knew about the

[5] In the event, periods with IS saw only 8 out of a total of 480 period-by-subject observations in which first stage earnings minus punishment received was negative before invoking the zero minimum, and 20 observations in which earnings for a period after subtracting off costs of giving punishment were negative. Periods with IS+NFS saw only 6 out of 690 period-by-subject observations in which first stage earnings minus punishment and fines received were negative before invoking the zero minimum, and 13 observations in which earnings for a period after subtracting off costs of giving punishment were negative. The instructions factually informed subjects that negative earnings in any period would be covered by the positive earnings of other periods.

[6] To assure as near as possible to equal numbers of groups playing with and without NFS, I programmed the computer to randomly assign groups to a fixed number of predetermined statuses, rather than to conduct an independent random draw of status for each group. For example, in a session having 4 groups, it was determined in advance that two groups would end up using NFS and two groups would not use it. Each group thus had an equal chance of playing under each condition, ex ante. Deviations from equal splits occurred only because low show-up rates reduced the number of groups in some sessions.

possibility of an override in advance.) The treatments of the middle and bottom rows differ only with respect to whether group members learn what the majority voted for if the computer overrides the vote. In middle row treatments, subjects receive no information about how their group's majority voted if the vote is overridden, while in the bottom row treatments, they were told whether their group's majority had voted for or against the use of NFS. Subjects in both the feedback and the no feedback treatments were told about the possibility of feedback so that, before the outcome of the decision on overriding or not overriding the vote was announced, their situations were identical. Thus, groups from both no feedback and feedback treatments whose votes determine assignment or not of NFS are pooled in my analysis.[7] The three rows in Table 1 thus correspond to four ways of assigning or not assigning NFS for Phase 2: (1) fully exogenously, with no voting; (2) by vote (vote counts and is not overridden in either the no feedback or the feedback treatment); (3) by vote override without feedback; and (4) by vote override with feedback as to the majority's vote.

Because the VCM permits various degrees of cooperation, I let formal sanctions received under NFS vary proportionately with points allocated to a subject's private account. The subject earns 0.3 points per point allocated to the group account versus 1 point per point she assigns to her private account, but she loses 0.4 points for each point so assigned under NFS. With $1 - 0.4 = 0.6 > 0.3$, the formal sanction in and of itself is clearly non-deterrent. Modification of payoffs (1) for NFS is straightforward, so no equation need be shown, conserving space. When both

[7] In other words, the middle and bottom row treatments differ ONLY if the vote does *not* count; they are otherwise identical and put subjects whose votes count (determine the Phase 2 institution) in an indistinguishable situation. To obtain as near as possible to equal numbers of groups in the vote count, vote override without feedback, and vote override with feedback situations, I pre-determined (to the extent show-up and thus number of groups permitted) that two-thirds of groups in each session would have a vote override outcome and one-third a vote count outcome, since after pooling of the thirds from feedback and no feedback treatments, the numbers of groups in each situation would be the same. Random assignment of groups to predetermined slots as opposed to independent random determination of each group's ex post outcome affects ex ante probabilities in the same qualitative fashion as discussed in the previous note.

NFS and IS are in place, the rule that earnings prior to one's cost of punishment cannot fall below zero holds, and the payoff function becomes

$$\pi_i^{NFS+IS} = max\ \{0,\ [(1-0.4)*(20 - g_i) + 0.3\sum_{i=1}^{5} g_i - 2\Sigma_j s_{ji}]\ \} - \Sigma_j s_{ij} \qquad (3)$$

where 0.4 is the fine rate.[8]

The four assignment methods listed above interact with the division between IS and no IS treatments to partition my data into sixteen categories of Phase 2 play. In Phase 2, that is, subjects can be playing under one of four conditions (VCM, IS, NFS, IS+NFS) reached in one of four ways (fully exogenously, by effective vote, by vote override with feedback, by vote override without feedback). When analyzing behaviors at the group level, it is also important in those cases in which feedback is given to distinguish "feedback for" (FgF) and "feedback against" (FgA)—i.e., groups learning that their majority vote, although overridden, was for NFS versus those receiving the alternative (always accurate) information. Since the number of cases to be observed in each category is endogenous to voting behavior, I defer further discussion to the Result part.

Andreoni and Gee (2011) also study the interaction of formal and informal sanctions. Their design resembles mine in that they observe a full 2x2 set of combinations with and without informal sanctions and with and without formal sanctions. They let the use or not of formal sanctions be determined endogenously by the subjects, as I do in some treatments. However, my own design differs not only in detail, but also in the fundamental respect that the authors deal

[8] It can be argued that greater realism would be achieved if a fixed amount were collected from subjects in periods played under NFS, representing a cost of having a formal sanction system in place (Markussen et al., 2012; Kamei et al., forthcoming). I omitted this element from the present design so as not to further reduce the likelihood of observing voted NFS. Also, the association of NFS with fixed costs may rarely in practice deter decisions to use it, since the state exists whether or not NFS is used to back one additional law. It is rare for discrete adjustments of enforcement capacity, and therefore of taxation, to be tied to the passage of any one law.

only with a deterrent formal sanction, so the idea that IS might be complementary to FS makes little sense in their setting (deterrent FS should in principle be self-sufficient). In line with this, their interpretive stance is that the crowding out of IS by FS is self-evidently desirable, much the opposite of the idea that non-deterrent FS and IS are likely to be complements, as I conjecture.[9]

Table 1. Treatments, groups, and subjects

	No IS (Phase 2 never includes IS)		**IS** (Phase 2 always includes IS)	
No voting; Phase 2 conditions assigned by computer.	15 groups 75 subjects	Group outcomes 8 NFS[c] 7 No NFS[a]	15 groups 75 subjects	Group outcomes 8 NFS[d] 7 No NFS[b]
Voting; if vote is overridden, Phase 2 condition is assigned by computer and subjects receive **no feedback** regarding the vote.	12 groups 60 subjects	Group outcomes 1 VC NFS[c] 3 VC No NFS[a] 6 VO NFS[c] 2 VO No NFS[a]	12 groups 60 subjects	Group outcomes 0 VC NFS[d] 4 VC No NFS[b] 6 VO NFS[d] 2 VO No NFS[b]
Voting; if vote is overridden, Phase 2 condition is assigned by computer and subjects receive **feedback** about how majority voted.	12 groups 60 subjects	Group outcomes 4 VC NFS[c] 0 VC No NFS[a] 6 VO NFS[c] 2 VO No NFS[a]	12 groups 60 subjects	Group outcomes 3 VC NFS[d] 1 VC No NFS[b] 6 VO NFS[d] 2 VO No NFS[b]

Note: VC = vote counts, VO = vote overridden; a = VCM (no IS, no NFS); b = IS only; c = NFS only; d = NFS+IS.

[9] Other detailed differences between Andreoni and Gee and this study include (a) the fact that their deterrent FS punishes only the lowest contributor, not all who contribute to their private accounts, and (b) that the way in which they endogenize adoption of FS is not by voting but by making implementation of FS a threshold public good, the provision of which depends on achieving a certain level of voluntary contributions.

4. Predictions

Predictions under classical assumptions of rationality and self-interest as well as common knowledge of these assumptions are straightforward but worth reviewing briefly. Rational selfish subjects who assume others to be of the same type would contribute nothing to the group account in the VCM since $0.3 < 1$. They would spend nothing on punishing under IS since it would be known that none would punish in the last period and thus punishing low contributors to induce higher contributions by threat cannot be credible. So, contributions would again be zero under IS. Contributions are again predicted to be zero under NFS, since $0.3 < (1 - 0.4) = 0.6$. Since one earns $20x0.6 = 12$ with NFS rather than 20 per period with VCM, and since one cannot rule out casting the decisive vote given the absence of communication, universal voting against NFS is the weakly dominant strategy for each individual and is predicted. Giving informal sanctions when NFS is in place is ruled out by the same logic as in the IS-only case. Hence, there would also be zero contributions under IS+NFS. Accordingly, it is also weakly dominant strategy to vote no in the IS treatments.

From dozens of past VCM experiments, one can safely predict that most actual contributions will initially be positive, averaging around half of the endowment, and that they will decline with repetition if not boosted by mechanisms such as IS, voting, etc.[10] When IS is available, I expect to see substantial numbers of subjects taking on the expense of punishing others, mainly but not only lower contributors. Accordingly I expect to see a slower decay of contributions. Far fewer observations of the effects of NFS are available in existing the literature. Based on Tyran and Feld (2006) and Kamei (2013), I might expect that NFS will

[10] While it is a potentially interesting exercise to predict behaviors from an explicit model, for example that of Fehr and Schmidt (1999), I avoid doing so because it is unclear *a priori* which of the numerous social preference models proposed in recent years will best organize my data and because adequately characterizing my results on a descriptive level leaves little space for such discussion in a paper of conventional length.

mildly, but probably not significantly, boost contributions when implemented exogenously. However, when NFS is chosen in a majority vote, I expect it will significantly boost contributions. One reason is that subjects may get utility w from acting in accord with what they take to be the spirit of the group vote and $(y + w) > x$, in the notation of Section 2. Another reason is that the vote may send a signal of willingness to cooperate if others do (Fischbacher, Gächter and Fehr, 2001).

Based on Kube and Traxler (2011), I can expect that contributions are higher under NFS+IS than under IS-only, and given that informal sanctions aimed at low contributors are so common in other experiments, it seems reasonable to expect that contributions will also be higher under NFS+IS than under NFS-only. How contributions under NFS-only will compare with contributions under IS-only is more difficult to predict. Taking both Tyran and Feld (2006)'s result and the result of the much larger number of IS experiments into account, there is reason to suppose that IS-only may be more effective at raising contributions than exogenously imposed NFS. But, my low punishment effectiveness gives reason for caution here (Nikiforakis and Normann, 2008). I cannot predict how contributions will compare between voted NFS and IS-only, which is never implemented by vote in my design.

Contrary to the standard predictions in this section's first paragraph, substantial numbers of subjects may vote for NFS.[11] Whether more or fewer subjects will vote for NFS when IS is also present is somewhat unclear *a priori*. If I modify standard theory by only accepting as a stylized fact that there is substantial punishing of low contributors, then I might expect more to

[11] Again, either expectations of a democracy effect *per se* (as captured by w above) or belief in the possibility of signaling conditional cooperation or both, may lead to such votes. Conditional cooperators are subjects willing to contribute most or all of their endowment to the public good, provided that they believe that others will do so. Voting for NFS can be a signal of both this willingness and the belief, since one can only lose money if NFS is implemented without cooperation.

vote for NFS when IS is also present, since the combination of NFS with IS can be expected to be more effective than NFS-only. Indeed, the selfishly rational choice may be tipped towards contributing if the average sanction per point *not* contributed in my setting exceeds 0.3, since then $(y + z) > x$, in the notation of Section 2. It is possible, however, that some subjects anticipate that IS-only may suffice to generate cooperation. In addition, those so inclined may hope to save the potential losses from NFS by voting against NFS when IS is present. There are no past observations to guide predictions regarding the question of the impact of IS on voting for NFS.

This leaves the question of the impact of vote overrides and feedback. The simplest method to look for differences between voted and exogenous NFS, or NFS+IS, is to compare behaviors under each condition in the fully exogenous treatments without voting against those in the voting treatments in which the vote determines the condition. Although I cannot tell for sure how each subject in the first treatment would have voted and cannot rule out selection effects, I may partially avoid this problem by using initial contributions and debriefing information, such as gender, for a sense similarity or difference of population characteristics. The firmest prediction, assuming that the conclusions of Kenju (2012) are applicable, is that contributions will be higher under voted NFS than under NFS with vote override in groups with identical numbers of yes votes, regardless of whether there is feedback about the vote. A likely corollary is that contributions will be higher the more yes votes there are in the group, and that contributions by yes voters will tend to be higher than those by no voters. I see no reason to rule out, a priori, that there is a positive effect of knowing that the group voted for NFS. Thus, NFS may be more effective when imposed by group override with feedback of a favorable majority

than without feedback. This can be the case even if there is also a pure democracy effect of the kind found by DFP (2010) and Kamei (2013).

Finally, it is worth remembering that in my design whatever condition subjects play under in Phase 2 follows six periods of play in a VCM and a period of instructions with or without voting. Even when NFS is not implemented, contributions can be expected to be higher in the first period after the break than in the last period before it due to the familiar restart effect first reported by Andreoni (1988). The presence of restart effects are likely to make it important to look at contribution trends in later periods of Phase 2 to properly distinguish the effects of both different conditions and different methods for determining them.

5. Results

A total of 390 subjects, all of whom were undergraduate or masters degree students at various universities in Vienna, participated in sessions lasting an average of 90 minutes at the Vienna Center for Experimental Economics at the University of Vienna. The subjects earned an average of €13.69 each, with a minimum of €7.00 and a maximum of €18.80.[12] The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) in English.[13]

In this paper, I begin by comparing behaviors under the VCM, IS, NFS, and NFS+IS conditions, initially emphasizing the fully exogenous situations without voting. I then analyze the votes, consider how voting affected the performance of conditions NFS and NFS+IS, and

[12] 48 % of the subjects were female, about 36% were majoring in business, management or economics, 32% in social sciences and humanities, and the remainder in natural sciences, mathematics, engineering and other fields. Roughly two thirds were enrolled at the University of Vienna, the rest at other universities in Vienna, mainly Vienna University of Technology, University of Natural Resources and Applied Life Sciences, Vienna and Vienna University of Economics and Business. Subjects were recruited using the ORSEE system (Griener, 2004). As this was the first economics experiment conducted at the university in some time, almost all subjects can be assumed to have been inexperienced.

[13] Subjects were informed that they would need strong English language skills to participate.

discuss the evidence for pure democracy and signaling effects as well as the possibility of a "disappointment effect" due to the overriding of the vote.

5.1 Conditions compared

Figure 1 compares average contributions in each period by condition, with panel (a) showing the data of the fully exogenous treatments, (b) the data of groups whose condition was determined by their majority vote, (c) the data of groups whose condition was determined by vote override (here, the data of groups receiving no feedback and those receiving feedback of each possible majority outcome are combined to save space), and (d) all of the data, combined. Recall that during periods 1 – 6, all groups are in the identical VCM condition without detailed knowledge of how the second phase will be played, so differences in Phase 1 are neither treatment nor condition specific. I show the Phase 1 curves to reassure the reader that they adhere to the usual properties of (a) beginning in the neighborhood of half of the endowment and (b) trending downwards with repetition. While the figure makes clear that there are some non-trivial, albeit random, differences as to how Phase 1 was played by groups randomly assigned to different treatments, I reserve discussion of controlling for unintended selection of subjects with differing disposition for later. Now, I only consider Phase 2 comparisons without reference to differences in prior experience.

The focus of my paper is on the effects of IS and of voting choice on NFS. With respect to the first issue, the figures appear to tell a fairly consistent story. In the aggregate (panel d) and under the exogenous and endogenous situations taken individually (panels a and b), there is considerable support for the expectation that the combination of NFS with IS leads to higher and more sustained contributions to the public good than either IS-only or NFS-only, although the

differences with respect to NFS-only are less consistent.[14] In the vote override situations (panel c), NFS, IS and NFS+IS have a similar effect of raising contributions, overall, although the trends over time differ as remarked below.[15]

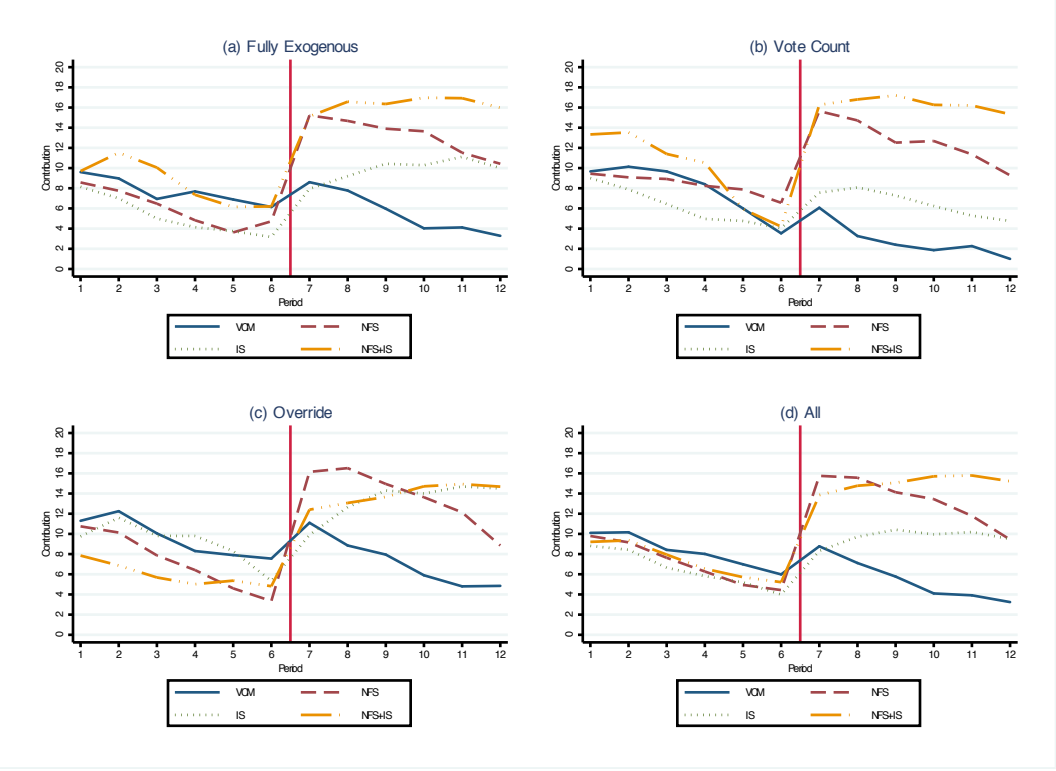

Figure 1. Average contribution by period and situation

[14] Mann-Whitney tests pooling group level data from all situations find the difference in contributions between NFS+IS and IS to be significant at the 1% level both in Period 7 and in Phase 2 as a whole. Similar test results are obtained for groups in the fully exogenous treatment only and for groups in the vote counts situation of the endogenous treatments only, although the results for groups in override situations are mostly insignificant. In comparisons of NFS+IS to NFS-only, the latter often shows higher contributions in the first period of Phase 2, Period 7, with the difference significant at the 5% level or better for the pooled data and for the groups of the fully exogenous treatment only and those in the vote counts situation of the endogenous treatments only. Thereafter, contributions under NFS+IS tend to catch up with and overtake those under NFS-only as the former exhibit an upward and the latter a downward trend. While overall Phase 2 contributions are higher under NFS+IS than under NFS-only for all groups pooled, the difference falls short of significance at the 10% level in a two-tailed test (p = .129, as opposed to p = .065 if a one-tailed test is performed). See Table A.3.

[15] In Mann-Whitney tests using groups in all override situations without regard to whether there was feedback and what the group vote outcome was, contributions for Phase 2 as a whole are significantly higher under IS than under VCM, with $p \approx .08$, under NFS than under VCM, with $p \approx .02$, and under NFS+IS than under VCM, with $p \approx .01$. Out of all conditions, NFS is the only condition that has contributions significantly higher than those under VCM in Period 7, with $p \approx .01$.

For the other conditions, my anticipation based on previous experimental results that contributions would be higher under IS than in the ordinary VCM condition also appear to be supported, although the difference at group level is significant only when all data are pooled.[16] As in Tyran and Feld (2006), voted NFS appears to raise contributions relative to VCM, but in contrast to their results, this is true also for exogenously imposed NFS, both in the absence of voting and following a vote override.[17] Subjects continuing in the VCM condition in Phase 2 show the usual Period 7 uptick of contributions (conventionally called a "restart effect") before continuing their downward slide.

A noticeable feature of the contribution graphs for Phase 2 is that whereas subjects in the VCM and NFS treatments tend to display some decay of contribution with repetition, as is typical in VCM experiments generally, those in the IS and NFS+IS conditions tend to show at least some initial upward movement, with either no or less overall decay. I estimated regressions and tested hypotheses that linear time trend coefficients for VCM (the omitted category), NFS, IS, and NFS+IS significantly differ from zero and from each other in the fully exogenous, endogenous, override without feedback, and override with feedback situations, and in the pooled data of all four situations. The tests (shown in Appendix Table A.1) confirm that there are significant downward trends in the VCM and NFS conditions, with the exception of VCM condition in override with feedback situations, which lacks a significant trend. They also show

[16] A Mann-Whitney test using data of all groups under VCM and all groups under IS finds average Phase 2 contribution to be higher with IS with $p = 0.03$, but there is no difference for Period 7 alone and there are also no significant differences of contribution in VCM versus IS groups if the data of single treatments (fully exogenous) or situations (e.g., vote counts) are taken alone. See Table A.3.

[17] For groups choosing between NFS and VCM by vote, contributions are significantly different both in Period 7 and in Phase 2 as a whole with $p = 0.03$. For those to whom NFS or VCM are imposed exogenously without voting, both differences are significant with $p = 0.02$. For groups with a vote override and no feedback, only the Phase 2 difference is significant, with $p = 0.05$; for those with a vote override and feedback that the majority favored NFS, both differences are significant with $p = 0.10$. For all observations in vote override situations, contributions are higher with NFS than VCM with $p \approx 0.01$. And for all groups pooled, NFS contributions exceed those in VCM in both the period and the phase with $p < 0.001$. See Table A.3.

that the trends in IS and NFS+IS conditions generally differ significantly from those in the NFS-only condition, with the exception of the trend in IS condition in endogenous situations being significantly different from the trend in NFS+IS condition in endogenous situations. The trend under IS-only is not significantly different from zero (a flat trend) except in the override with feedback situations, where contributions are significantly increasing. The tests find flat trends for contributions in the NFS+IS condition in fully exogenous, override without feedback and endogenous situations, but a significant increasing trend in both override with feedback and in the pooled data under this combined condition.

An obvious explanation would be targeting of informal sanctions at low contributors, spurring them to contribute more. Appendix Figure A.1 confirms that substantial amounts of informal sanctions were indeed given in both IS and NFS+IS conditions, and that roughly three quarters (in IS) and two thirds (in NFS+IS) of all informal sanctions given were directed at group members who contributed less than the group average during the period in question. Overall, a smaller amount was spent on sanctioning in the NFS+IS than in the IS treatment, with the difference most pronounced in the middle periods of the phase—a difference that might be explained by perceptions of complementarity between or even redundancy of IS in the presence of NFS.[18] I estimate regressions which follow a specification first used by Fehr and Gächter (2000). Results, shown in Appendix Table A.2, indicate that for below-average contributors in all conditions permitting informal sanctions, punishment received was increasing in the difference between own and other group members' average contribution, significant at the 1% level.

[18] The fact that a higher proportion of informal sanctions were perversely "misdirected" at high contributors in the NFS+IS condition than under IS-only might be explained by pro-social subjects tending to believe NFS sufficient in the NFS+IS condition, leaving more of the sanctions in that condition to be the work of the minority of subjects inclined to resist pressures to cooperate.

Since the formal sanction, y, is 0.4 and the amount gained by allocating a point to one's private rather than public account, x, is $1 - 0.3 = 0.7$, the condition under which contributing another point is profitable under NFS+IS, $(w + y + z) > x$, is met provided that (a) the regression coefficient, z, exceeds 0.3 and (b) the subjective cost of norm-violation, w, is not negative. The table shows significant coefficients exceeding 0.3 for all pooled observations and for those separated by treatment and situation.[19] Under IS-only, it is better to put a point in the group account if $(w + z) > x$, which, if $w = 0$, requires that the coefficient exceeds 0.7, and this is the case according to all of the regressions except that for observations in the override without feedback situation. The fact that the coefficients on the positive deviation term are either insignificant or positive means that there is no incentive to contribute more than the group's average.[20] But with low contributors raising their contributions to avoid punishment, that average is an upwardly moving target, which helps to explain the upward trends in contributions.

5.2 Voting

240 subjects voted on whether their group should operate under NFS in Phase 2, understanding that a decision randomly taken by the computer would determine whether their group's majority vote decides their Phase 2 condition, that the computer will randomly determine the condition if the vote does not count, and that in the latter event how the majority voted may or may not be made known to group members.[21] Exactly half were in the IS treatment and were

[19] Note that the dependent variable is the loss to the subject receiving the sanction and thus already doubles the number of sanction points given by the punisher.

[20] A positive significant coefficient on positive deviation could be a sign of perverse punishment: the further the contribution is above the group average, the more likely is one to be punished. Similar results are found for some treatments in Önes and Putterman (2007).

[21] I provided subjects no specific information about the probability with which each outcome would occur for various reasons, including the fact that I needed to vary probabilities across some sessions and that overrides had to be about twice as common as votes counting in a given session (since results in vote count outcomes are pooled from both 'no feedback' and 'vote feedback' treatment sessions), In the event, no subject asked what the probabilities were, suggesting that the matter did not raise concerns for them.

thus choosing between NFS+IS and IS-only; the other half were in the treatment with neither opportunity for nor mention of IS and were choosing between NFS and VCM. 43.3% of those in the IS treatment voted for NFS versus 68.5% of those in the no-IS treatment, a difference significant at the $p < .001$ level in a chi-square test ($\chi^2(1) = 12.253$). Possibly considerable numbers of subjects expected IS to boost contributions sufficiently without the help of NFS. Despite the fact that those voting for NFS were a minority overall in the treatments with IS, the majority favored NFS+IS in 8 of the 24 groups in IS treatments. There were majorities favoring NFS in 19 of the 24 groups in no-IS treatments. Subjects in 5 groups ended up playing under endogenously chosen NFS (i.e., their vote counted) and those in 3 groups ended up playing under endogenously chosen NFS+IS.

In Table 2, I report probit regressions to explain individuals' votes, with errors clustered by group. Pooling all 240 voting observations, I find in column (1) a large and highly significant negative effect of being in treatments with IS, a smaller highly significant positive effect of own contribution in Period 1, and generally negative but insignificant coefficients on other group members' average contribution in Phase 1. Separate estimations for only individuals in treatments without IS and for only those in treatments with IS, shown in columns (3) and (5) respectively, confirm the significant positive effect of own initial contribution. The result on the effect of IS accords with the aggregate outcomes mentioned above. The result on own initial contribution suggests that subjects who came into the experiment with more cooperative inclinations and perhaps with more optimistic beliefs about others' willingness to cooperate were significantly more likely to vote for NFS, perhaps because optimism about the ability to encourage cooperation, preference for cooperation, or both are associated with contributing more in the initial VCM condition. Columns (2), (4) and (6) show versions of the same regressions in

Table 2: probit regression of vote for NFS

Dependent Variable: votefor						
	All Voting Treatments		Treatments without IS		Treatments with IS	
	(1)	(2)	(3)	(4)	(5)	(6)
IS	-0.565*** (0.18)	-0.546*** (0.20)				
Own Period 1 contribution	0.0533*** (0.01)	0.0577*** (0.01)	0.0508*** (0.01)	0.0474*** (0.01)	0.0559*** (0.02)	0.0804*** (0.02)
Others' Phase 1 contribution	-0.0258 (0.03)	-0.0126 (0.03)	-0.0266 (0.04)	-0.0276 (0.04)	-0.0248 (0.05)	0.0141 (0.06)
Constant	0.0857 (0.25)	0.287 (0.52)	0.116 (0.29)	0.424 (0.71)	-0.511 (0.33)	-0.86 (0.81)
Demographic Controls	no	yes	no	yes	no	yes
Observations	240	220	120	120	120	100

Notes: All results are from probit regressions and clustered by group. Numbers in parenthesis are standard errors. The dependent variable is Votefor, which is an indicator variable for whether subject voted for NFS. The independent variables are as follows: IS (=1 for subjects in a treatment with IS; =0 otherwise), Own period 1 contribution (individual's Period 1 contribution) and Others' phase 1 contribution (average Phase 1 contribution of others in the individual's group)
Besides these independent variables, the female dummy (=1 if female; 0 otherwise), Year of Study (1 = freshman to 5=master's students), Political View (1= very conservative to 7= very liberal) and Major in Econ or Business dummy are included in columns (2), (4) and (6). I omitted the estimated coefficients of these variables to conserve space, since these are not related to our hypothesis.
*, **, and *** indicate significance at the .10 level, at the 0.05 level and at the .01 level, respectively.

which are included additional controls for four personal characteristics on which data were obtained in the end-of-session survey: gender, an economics or business major dummy, year of study, and self-reported political orientation. Their addition has no important effects on the other coefficients.

5.3 Is NFS more effective when chosen by voting? The No IS case.

I observe individual and group play with only NFS under five distinct situations, namely (1) fully exogenous, (2) endogenous (subjects voted and the vote counts), (3) after a vote override when subjects have no feedback about the vote, (4) after a vote override with feedback that the majority voted *for* NFS, and (5) after a vote override with feedback that the majority

voted *against* NFS. Further controlled comparisons require that I control also for each individual's own vote (which could be seen as sub-partitioning conditions (2) – (5) into eight individual-level categories). Controls for prior experience and perhaps for personal characteristics are desirable to more fully rule out selection effects. Although non-parametric tests have the advantage of not requiring distributional assumptions, I also use regression analysis so as to allow for such controls.

In analyzing the impact of having adoption of NFS be determined by vote, I focus on three steps: [1] check for a "democracy effect" or "endogeneity premium" in the sense of DFP (2010) by comparing contributions in situation (2) to those in situation (3) above, which parallels DFP's principal method; [2] check for an informational effect of voting by comparing contributions in situations (4) and (3), paralleling the additional exercise in DFP,[22] and [3] as another way of checking for a democracy effect, where those assigned the scheme exogenously have no experience of voting and of the vote not counting, compare contributions in situation (2) to those in (1). Check [3] differs from checks [1] and [2] because, in the absence of votes, controls for selection can only be based on information about other individual characteristics including the propensity towards cooperation exhibited in Phase 1, and characteristics such as gender.

Table 3 shows the number of groups observed in each of situations (1) – (5) above, average contribution in the groups in the first period of Phase 2, average contribution in Phase 2 as a whole, and average contribution for the same groups in Phase 1, which may help to reveal selection problems. It is worth bearing in mind that overall, NFS is associated with startlingly high Phase 2 contributions given the theoretical prediction of no contributions, and that

[22] See Section IV.B in DFP.

Table 3. Average contribution in groups that play under NFS without IS in Phase 2.

	Exogenous (no vote) (1)	Endogenous (vote counts) (2)	Vote over-ride, no feedback (3)		Vote over-ride, feed-back FOR (4)	Vote over-ride, feed-back AGAINST (5)
			FOR	AGAINST		
Period 7 only	15.23	15.64	15.47		16.83	n.a.
			15.32	16.2		
Phase 2 as a whole	13.23	12.70	13.26		14.15	n.a.
			12.93	14.93		
Same groups in Phase 1 as a whole	6.00	8.35	7.68		6.69	n.a.
			7.57	8.23		
Number of groups	8	5	6		6	0
			5	1		

Note: I show results for situation (3)—vote override, no feedback—both as a whole, in the top number of each cell of this column, and separately for groups in which the majority voted for NFS, to the left, and groups in which the majority voted against NFS, to the right.

contributions are also significantly higher under NFS than those in the baseline VCM condition in the aggregate and in every situation allowing comparison. However, the effect of NFS shows little sign in my data of being greater when implemented by vote. If anything, NFS seems to perform best when selected by vote override but with feedback of a favorable vote (situation 4). The prospect of a democracy effect appears further diminished when considering Phase 1 behavior, since groups that use NFS by vote in Phase 2 seem if anything to have been more cooperative than others from the outset. As for other effects, there is a faint indication of an information effect (comparing situations (4) and (3)), but the difference is not significant, and since the same information is present in situation (2), concluding that a positive information effect is present would imply that the pure effect of democracy (over and above that of

information) must itself be negative. Nor is there any indication that disappointment with an overridden vote depresses contributions in situations (4) or (3) relative to that in the fully exogenous condition (1). Non-parametric tests for differences by situation using group level observations find no statistically significant differences.[23]

My regression analysis of data from the treatments that permit controlling for own vote uses the 120 individual-level period 7 contributions or period 7 – 12 average contributions of situations (2) – (5), in alternate columns adding also controls for own Period 1 and others' Phase 1 contributions. Treatments without IS, discussed in the present sub-section, are shown in Panel A of Table 4. Interestingly, own Period 1 contribution is a significant positive predictor of Period 7 and average Phase 2 contribution, whereas average Phase 1 contribution of others in one's group returns positive but insignificant coefficients.

I next focus on the more complete specifications and on the coefficients pertaining to cases in which NFS is implemented and to voters who favored NFS. Consider first the difference in contribution of Yes voters in the vote count vs. override without feedback situations. According to Wald tests, in fact, in no two situations in which NFS is implemented without IS are coefficients significantly different for Yes voters whose situations differ with respect to the vote counting, feedback being given, or the content of the feedback.

To test for differences with the fully exogenous treatment, I perform similar exercises in which I include the 75 fully exogenous treatment observations. In this case, however, I have to drop controls for own vote, which I do by combining the situation dummy variables that were

[23] See the Appendix, Table A.4.

Table 4 : The EFFECT of DEMOCRACY (INDIVIDUAL LEVEL DATA)

(A) no-IS treatments

	(1)	(2)	(3)	(4)
EndoNFSn	17.75***	15.19***	14.19***	11.06***
	(2.23)	(2.63)	(1.53)	(2.22)
EndoNoNFSn	6.222***	3.668	2.704**	-0.31
	(2.10)	(2.47)	(1.21)	(2.18)
vExoNFS_nFn	11.89***	9.102***	10.78***	7.814***
	(2.10)	(2.37)	(2.80)	(2.74)
vExoNoNFS_nFn	7.667***	4.44	7.167***	3.073
	(2.57)	(3.23)	(1.85)	(2.82)
vExoNFS_FgFn	14.33***	12.17***	14.64***	12.24***
	(2.57)	(2.67)	(3.05)	(3.37)
vExoNoNFS_FgFn	1.667	0.472	2	0.0425
	(3.64)	(3.67)	(1.85)	(1.39)
EndoNFSy	14.65***	11.35***	12.00***	8.454***
	(1.53)	(2.13)	(1.94)	(2.79)
EndoNoNFSy	5.833**	1.867	2.972**	-0.952
	(2.57)	(2.91)	(1.29)	(1.72)
vExoNFS_nFy	17***	12.94***	14.33***	10.27***
	(1.38)	(2.10)	(1.27)	(1.82)
vExoNoNFS_nFy	17.50***	11.96***	8.208***	2.829
	(3.15)	(3.67)	(1.94)	(2.73)
vExoNFS_FgFy	17.46***	14.03***	14.03***	10.59***
	(1.29)	(1.86)	(1.34)	(2.11)
vExoNoNFS_FgFy	14.43***	10.43***	9***	4.978*
	(2.38)	(2.80)	(2.60)	(2.43)
Own Period1 contribution	-	0.301***	-	0.207**
		(0.08)		(0.08)
Others' Phase 1 contribution	-	0.0262	-	0.175
		(0.17)		(0.16)
clustered by group	no	no	yes	yes
Observations	120	120	120	120
R-squared	0.857	0.873	0.851	0.866
Tests of differences of contribution rate by situation for yes-voters				
EndoNFSy= vExoNFS_nFy	0.2552	0.4181	0.3256	0.4014
EndoNFSy= vExoNFS_FgFy	0.1624	0.1636	0.3978	0.3955
EndoNFSy= vExoNFS_FgAy	n.a.	n.a.	n.a.	n.a.
vExoNFS_nFy= vExoNFS_FgFy	0.8082	0.5494	0.8731	0.8507
vExoNFS_nFy= vExoNFS_FgAy	n.a.	n.a.	n.a.	n.a.
vExoNFS_FgFy= vExoNFS_FgAy	n.a.	n.a.	n.a.	n.a.

Note: All results are from OLS regressions. Numbers in parenthesis are standard errors. N and y denote the individual vote of the subject (against or for NFS). The p-values correspond to Wald tests based on the regression results. I omitted the estimated coefficients of some control variables (vExoNFS_FgAn, vExoNoNFS_FgAn, vExoNFS_FgAy and vExoNoNFS_FgAy) because no subjects were in these situation in the event. Regressions in column 2 and 4 control for the individual's contribution in period 1 and their group members' average contribution in Phase 1. *, **, and *** indicate significance at the .10 level, at the 0.05 level and at the .01 level, respectively.

separated by individual voter type in the previous analysis. To further control for selection effects, in alternate specifications of Appendix Table A.5 I add controls for the same individual characteristics as were included in some specifications in Table 2.[24] A third variable, year of study, is added as a check, but only in alternative estimates because its inclusion causes the loss of additional observations.[25]

As in Table 4(A), own period 1 contribution remains a significant positive contributor of Period 7 and Phase 2 average contribution. With fully exogenous treatment observations added and without controls for voting, other group members' average Phase 1 contribution also shows an effect on own average contribution in Phase 2, significant at the 10% level. The Wald tests, however, find no statistically significant differences in contributions under NFS (without IS) between subjects in the fully exogenous treatment, those in endogenous treatments whose votes were counted, or subjects in exogenous treatments whose votes were overridden, whether without feedback, with feedback of a favorable majority, or with feedback of an unfavorable majority.

Summarizing, both for Period 7 and for Phase 2 as a whole, and both as indicated by non-parametric tests and by regressions controlling for vote (for endogenous treatments) and by regressions not controlling for vote (for exogenous and endogenous treatments combined), I find no sign that the process that leads to adoption of NFS affects the level of contributions under NFS when IS is not present. There is no sign of an endogeneity premium, nor is there an

[24] Versions of the Table 4 regressions were also estimated with added controls for personal characteristics. Since their inclusion has no significant effects on the other coefficients, I have not reported them. Controls for individual characteristics are more critical to the analysis in Table A.5 given the inability to control for selection effects by means of individual vote in this case.

[25] Due to a mishap in the handling of data files, 2 subjects' observations in the IS treatments are recorded without gender, an additional 2 observations are missing major and an additional 16 observations are missing year of study and political view. I report estimates for each subsample with and without the controls, so that the effect of adding controls can be distinguished from that of dropping observations.

indication the informational feedback increased contributions, nor is the result of exogenous implementation different among subjects who voted but had their vote overridden and subjects who never voted.[26]

5.4 Is NFS+IS more effective when NFS is chosen by voting?

I can look for observations of NFS+IS under the same five situations as were listed in section 5.3 for NFS-only. Table 5 reports the information paralleling that in Table 3 for the NFS+IS groups by situation.

At first glance, the data in Table 5 seem potentially consistent with a positive effect of the vote counting, especially when contributions in the fully exogenous condition (1) are not considered. Comparing vote count with vote override no feedback situations, there is an impression of a democracy effect in both Period 7 (average contribution 11.20 when vote is overridden without feedback versus 16.27 when vote counts) and Phase 2 as a whole (12.34 in vote override without feedback versus 16.34 when vote counts). The Period 7 difference is significant at the 5% level in group-level Mann-Whitney test, but the Phase 2 difference is not. Also, comparing the override with feedback of a majority for NFS, in column (4), it seems that much of this apparent democracy effect may be attributable to the informational impact of knowing how the group voted—i.e., average contribution when NFS is imposed by override but with feedback of a favorable vote lies between that in the vote count and that in the override no

[26] I also tested the differences between each situation by organizing my data slightly differently from the approach shown in Tables 4 and A.5. First, in regressions resembling those of Table 4, I checked for an information effect by comparing a dummy variable for subjects in the situation of a vote override with feedback that the group voted for NFS to another dummy variable for subjects only in those groups in the vote override without feedback situation that had majorities vote for NFS (although subjects were not so informed). As when all groups in the override without feedback situation are pooled under the same situation dummy, I found no significant difference. Second, in regressions paralleling those of Table A.5, I checked for a difference between subjects in the fully exogenous treatment and the pooled set of subjects in situations of vote override, including those in the override without feedback situation and those in both override with feedback situations. Here, too, no significant differences were found.

Table 5. Average contribution in groups that play under NFS+IS in Phase 2.

	Exogenous (no vote) (1)	Endogenous (vote counts) (2)	Vote over-ride, no feedback (3)		Vote override, feedback FOR (4)	Vote override, feedback AGAINST (5)
			For	Against		
Period 7 only	15.13	16.27	11.20		14.60	12.60
			13.60	10.72		
Phase 2 as a whole	16.32	16.34	12.34		15.77	15.20
			14.23	11.96		
Same groups in Phase 1 as a whole	8.50	9.82	5.32		6.70	6.40
			8.10	4.76		
Number of groups	8	3	6		3	3
			1	5		

feedback situations but closer to the former. But neither the information effect nor the residual pure endogeneity premium, which should together account for the difference between the vote count and override without feedback conditions, is individually significant in either Period 7 or Phase 2, according to Mann-Whitney tests.

Further complicating the picture is the fact that the indication of an endogeneity or democracy effect imparted by comparing situations (2) and (3)—a method paralleling that of DFP—is not supported when comparing groups using NFS+IS by vote (2) to those assigned it fully exogenously, without voting (1). There is no difference in contributions in these two situations, according to Mann-Whitney tests. Moreover, contributions in the override without feedback situations are noticeably lower than those in the fully exogenous situation, and these differences are significant according to Mann-Whitney tests—at the 1% level, for Period 7, and at the 5% level, for Phase 2 as a whole. Although in no way definitive before controls for

possible selection effects are applied, the initial impression is that the appearance of an endogeneity effect when comparing situations (2) and (3) might in fact be attributable to a negative influence on contributions from having voted and having the vote be overridden with no feedback as to its outcome. [27]

But these inferences from comparing contributions without additional controls are thrown into doubt by the evident differences in Phase 1 contributions shown in the third row of Table 5. Most noticeable is the low average Phase 1 contribution (5.32) of groups ending up in the vote override without feedback situation and the higher corresponding averages for groups in the fully exogenous condition (8.50) and those in the endogenous treatments whose votes were randomly chosen to count (9.82). These raise the possibility that the good Phase 2 performance of NFS+IS when the vote counts versus when the vote is overridden without feedback (columns (2) versus (3)) might be mainly attributable to chance differences in the inclinations to cooperate of the subjects randomly assigned to the different situations.

As before, I attempt to advance the analysis by controlling for individual votes and Phase 1 experience, estimating multivariate regressions at the individual level. I have 120 individual level observations for Period 7 and the same number of average individual contribution observations for Phase 2 as a whole on which I can estimate regressions that control for own vote in the treatments with IS. Again, Period 1 contribution is a significant positive predictor of contribution under NFS+IS in these regressions. For these treatments, other group members'

[27] The group level Mann-Whitney test results reported in Table A.4 show no significant differences in contributions under NFS+IS by situation except those between the fully exogenous or vote counts situations, on the one hand, and all override without feedback groups and especially those override without feedback groups in which the majority voted against NFS, on the other. Since differences with those override without feedback groups in which the majority voted for NFS are not significant, the differences with override without feedback groups in general are clearly driven by the negative majority groups that played in the override without feedback situation.

average contribution in Phase 1 also shows a significant effect on own average Phase 2 contribution.

Unlike the corresponding regressions for the treatments without IS, there is one pair of statistically significant Wald test results for differences in situation dummy variables (see Table 4(B)). Both for Period 7 and for Phase 2 as a whole, contributions are significantly higher ($p < 0.10$ and $p < 0.01$, respectively) among yes voters in the vote override situation with feedback that the group voted for NFS than for those in the vote override situation without feedback. A separate test, not shown, finds a significant difference from subjects in the override situation with feedback that group voted for and those in override with no feedback when only groups having a majority for NFS are included, although this holds for Phase 2 as a whole only ($p < 0.001$). Thus, the data support the presence of an information effect: despite the vote override, yes voters whose groups are told that their majority voted for NFS contribute more under NFS+IS than yes voters in the override with no feedback situations.

Regressions that drop controls for individual voting and add observations from the exogenous treatment with IS are shown in part (b) of Table A.5. These results provide further support for the presence of an information effect, as well as supporting the indications that the override without feedback situation reduced contributions compared to the fully exogenous treatment, perhaps due to some kind of disappointment or unhappiness with having voted but receiving no feedback about the overridden vote outcome. In the regression variants controlling only for gender and major subject (economics or business), among individual characteristics, the difference between contributions in the feedback that group voted for and in the no feedback condition is significant at the 10% level both for Period 7 and for Phase 2 as a whole. The difference is not significant, however, if only those groups in the no feedback situation that had

Table 4: The EFFECT of DEMOCRACY (INDIVIDUAL LEVEL DATA)

(B) IS treatments

	(5)	(6)	(7)	(8)
EndoNFSn	11.67***	9.766**	15.33***	9.249**
	(3.43)	(3.94)	(2.38)	(3.58)
EndoNoNFSn	6.438***	4.973**	5.833***	1.843
	(1.49)	(2.00)	(1.83)	(1.69)
vExoNFS_nFn	10.77***	9.326***	12.20***	8.644***
	(1.27)	(1.74)	(1.21)	(1.36)
vExoNoNFS_nFn	11***	9.345***	11.47***	5.504**
	(2.66)	(3.35)	(2.02)	(2.58)
vExoNFS_FgFn	8***	8.080***	11.30***	7.480**
	(2.66)	(3.01)	(2.82)	(3.34)
vExoNFS_FgAn	11.45***	9.435***	14.48***	9.600***
	(1.79)	(2.41)	(1.00)	(1.60)
vExoNoNFS_FgAn	8***	6.699**	14.03***	8.754***
	(2.42)	(3.03)	(0.36)	(1.75)
EndoNFSy	17.42***	14.23***	16.60***	9.905***
	(1.71)	(2.82)	(1.77)	(2.76)
EndoNoNFSy	9.556***	6.526***	7.778***	3.096
	(1.98)	(2.44)	(2.47)	(2.10)
vExoNFS_nFy	12.38***	10.09***	12.71***	8.980***
	(2.10)	(2.35)	(1.21)	(1.53)
vExoNoNFS_nFy	12.40***	10.05***	12.93***	7.239*
	(2.66)	(3.22)	(3.20)	(3.59)
vExoNFS_FgFy	17.90***	15.30***	18.00***	13.44***
	(1.88)	(2.35)	(0.38)	(1.44)
vExoNFS_FgAy	15.75***	14.62***	17.17***	14.82***
	(2.97)	(2.94)	(1.53)	(0.83)
vExoNoNFS_FgAy	8.250***	5.56	15.17***	9.460***
	(2.97)	(3.44)	(1.79)	(2.24)
Own Period1 contribution	0.302***	0.302***	0.173***	0.173***
	(0.08)	(0.08)	(0.05)	(0.05)
Others' Phase 1 contribution	-0.111	-0.111	0.434**	0.434**
	(0.21)	(0.21)	(0.17)	(0.17)
clustered by group	no	no	yes	yes
Observations	120	120	120	120
R-squared	0.822	0.842	0.907	0.92
Tests of differences of contribution rate by situation for yes-voters				
EndoNFSy= vExoNFS_nFy	0.0656*	0.1414	0.0830*	0.7116
EndoNFSy= vExoNFS_FgFy	0.8496	0.6753	0.4469	0.1338
EndoNFSy= vExoNFS_FgAy	0.6279	0.9131	0.8100	0.0515*
vExoNFS_nFy= vExoNFS_FgFy	0.0524*	0.0560*	0.0004***	0.0022***
vExoNFS_nFy= vExoNFS_FgAy	0.3555	0.1966	0.0319**	0.0001***
vExoNFS_FgFy= vExoNFS_FgAy	0.5419	0.8438	0.6024	0.1627

Note: All results are from OLS regressions. Numbers in parenthesis are standard errors. N and y denote the individual vote of the subject (against or for NFS). The p-values correspond to Wald tests based on the regression results. I omitted the estimated coefficients of some control variables (vExoNoNFS_FgFn and vExoNoNFS_FgFy) because no subjects were in these situations in the event. Regressions in column 6 and 8 control for the individual's contribution in period 1 and their group members' average contribution in Phase 1. *, **, and *** indicate significance at the .10 level, at the 0.05 level and at the .01 level, respectively.

in fact voted for NFS are included.[28] The variants that add a control for years of study, losing additional observations, are also significant but only in the case of Phase 2 as a whole. As for the "disappointment effect" represented by a difference between contributions in the fully exogenous treatment and those in the override without feedback situation of the endogenous treatments, the regressions support its significance at the 5% level in the Period 7 regression with no controls but not that with three personal characteristic controls, and at the 10% level in the Phase 2 regressions with no controls but significance falls just short of 10% ($p \approx 10.2\%$) with all four controls.

Summing up, for treatments with IS there is indication of support for an endogeneity premium only for Period 7 and only for the group level non-parametric tests without controls for voting, experience, or personal characteristics. A finding of higher contributions in groups whose vote is overridden but that receive feedback of a majority for NFS than in those in the override without feedback situation, or a positive effect of information, is significantly supported in both regressions without control for own vote and those with such control. And a seeming "disappointment effect," whereby subjects who are assigned NFS+IS after voting but without feedback of their vote outcome contribute less than those assigned NFS+IS completely exogenously, without voting, is supported both by non-parametric tests and by multivariate regressions.

6. Concluding discussion

[28] While this finding slightly undermines the conclusion that a feedback effect is present, recall that the test with control for own vote did find a significant difference. The latter result, in a variant of Table 4, deserves more weight than the one reported in the present paragraph, which fails to control for own vote. The regression variants distinguishing no feedback groups according to their majority votes are not included in the main Appendix but are available on request in Appendix B.

I study the questions of whether availability of informal sanctioning (IS) opportunities adds to the efficacy of non-deterrent formal sanctions (NFS) and whether, if so, this effect is enhanced when those sanctions are chosen by vote. Along the way, I revisit two questions: whether adding NFS improves the effectiveness of IS, and whether being chosen by vote in and of itself enhances the effectiveness of NFS. I also take the opportunity to compare two methods for investigating the endogeneity premium question: the one introduced by Dal Bó, Foster and Putterman (2010) that permits controlling for selection effects with information on the individual's vote by generating observations of exogenously imposed institutions after voting and the overriding of votes; and the simpler method of comparing observations of subjects operating under the institution imposed without opportunity to vote to those of subjects in groups operating under it after its adoption by vote. The latter method does not permit controlling for individuals' votes, so I attempt to control for selection effects by using own initial behavior, experience, and personal characteristics in multivariate regression equations.

Taking all conditions together and in the fully exogenous treatments, my data support the hypothesis that adding IS strengthens the effect of NFS on contributions, at least when considering the six period sequence of Phase 2 interactions as a whole. This result did not find support in the vote count and vote override situations of my endogenous treatments taken by themselves, however. The reverse idea that adding NFS strengthens the effect of IS, supported in Kube and Traxler's exogenous one-shot experiment (2011), is also supported in my fully exogenous treatment, endogenous treatment with vote counting, and in my pooled data from all treatments and conditions.

Unlike Tyran and Feld (2006), my data support neither the idea that the effectiveness of NFS is increased by voting nor the idea that NFS is ineffective in the absence of voting. For

whatever reason, my subjects respond to NFS with significantly higher contributions even when it is imposed without a vote. I likewise do not find contributions under NFS+IS together to be significantly higher when NFS is adopted by vote than when it is imposed.

In my treatment with IS, I do find support for an effect of voting on effectiveness of NFS (that is, of the NFS+IS combination) when the comparison is to the vote override situation without feedback, paralleling the principal approach in DFP. However, that difference is only significant when I fail to control for individual vote and experience. Moreover, unlike DFP, I find evidence that most of the effect of voting in the treatment with IS, if there is one, is attributable to the signaling or informational effect of learning about the preferences of other group members from their votes; an effect is present, that is, even if the vote is declared void and the institution is imposed exogenously. And my results raise a potential warning about the method that DFP employ: I find contributions significantly lower in the override without feedback situation than in the fully exogenous treatment, suggesting that disappointment or some other negative affective impact of the vote override, rather than a positive effect of determining the scheme used by voting, may account for the contribution difference between subjects in the vote counts situation and those whose vote is overridden without feedback.

Although not all of these findings lend themselves to easy interpretation, many of my results are consistent with past findings and with intuition. I find the usual decay of contributions from a substantial initial level in a finitely repeated VCM, and the usual sustaining of contributions when informal sanctions are available, with the latter effect explained, as usual, by decisions to voluntarily impose costly sanctions mainly on low contributors. NFS+IS tends to perform better than NFS-only mainly because availability of IS allows contributions to be more sustained over time.

The original question motivating my study is: why do we observe in society so much use of sanctions too low to be deterrent in material terms, and why do such sanctions often appear to have effects on behavior? While I found support for the view that availability of informal sanctions helps to make non-deterrent formal ones more effective, the conjectures that NFS is effective only with the help of IS or only thanks to the empowerment, legitimacy, or signaling conveyed by voting, were not supported by my data. Possibly, my subjects understood themselves to be engaged in a situation in which the good of all is at odds with individual gains, and perhaps most were already characterized by a bias towards cooperation or a conditional willingness to contribute provided that others did. While raising the MPCR but leaving it below the return from private allocation should have no effect, according to standard theory, it is well known that higher MPCRs are associated with higher contributions to the public good in experiments (Zelmer, 2008). A higher MPCR lowers the private cost of contributing, though that cost remains positive, so it could tip the balance between the cost and benefit of contributing for subjects who assign small positive values to contributing or bestowing benefits on others.

The effect of a non-deterrent sanction, much like raising the MPCR, is also to reduce the opportunity cost of contributing. For most subjects, that reduction may have been sufficient to raise their subjectively optimal contributions substantially. Higher anticipated contributions by others also raise the preferred contributions of conditional cooperators through the indirect channel of conditional cooperation. Finally, introduction of a sanction for allocating funds to the private accounts could conceivably have had an "experimenter demand effect," reinforcing subjects' senses that contributing to the public good is a sign of virtue. But clearly more work is needed before we understand more fully why non-deterrent sanctions are often effective, and what roles reinforcing informal sanctions and public assent by voting play in the process.

References

Anderson, Lisa and Sarah Stafford, 2003, "Punishment in a Regulatory Setting: Experimental Evidence from the VCM," *Journal of Regulatory Economics* 24(1): 91-110.

Andreoni, James, 1988, "Why free ride?: Strategies and learning in public goods experiments." *Journal of Public Economics* 37(3): 291-304.

Andreoni, James and Laura Gee, 2011, "The Hired Gun Mechanism," *NBER Working Paper Series,* Working Paper 17032

Chaudhuri, Ananish, 2011, "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Experimental Economics* 14(1): 47 - 83.

Cinyabuguma, Matthias, Talbot Page, and Louis Putterman, 2005, "Cooperation under the threat of expulsion in a public goods experiment." *Journal of Public Economics* 89(8) : 1421-1435.

Dal Bó, Pedro, Andrew Foster and Louis Putterman, 2010, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review* 100(5): 2205-29.

Ertan, Arhan, Talbot Page, and Louis Putterman, 2009, "Who to punish? Individual decisions and majority rule in mitigating the free rider problem." *European Economic Review* 53(5): 495-511.

Fehr, Ernst, and Klaus M. Schmidt, 1999, "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics* 114(3): 817-868.

Fehr, Ernst and Simon Gächter, 2000, "Cooperation and Punishment in Public Goods Experiments," *American Economic* Review 90: 980 – 994.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr, 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters* 71(3): 397-404.

Fischbacher, Urs, 2007, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10(2), 171-178.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review*, 100(1): 541-56.

Gächter, Simon and Benedikt Herrmann, 2008, "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment," *Philosophical Transactions of the Royal Society B* doi:10.1098/rstb.2008.0275.

Greiner, Ben, 2004, "The Online Recruitment System ORSEE 2.0 - A Guide for the

Organization of Experiments in Economics," University of Cologne, Working Paper Series in Economics 10.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran, forthcoming, "State or nature? Formal vs. informal sanctioning in the voluntary provision of public goods." *conditionally accepted at Experimental Economics.*

Kube, Sebastian and Christian Traxler, 2011, "The Interaction of Legal and Social Norm Enforcement," *Journal of Public Economic Theory* 13(5): 639-660.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran, 2014, "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes." *The Review of Economic Studies* 81(1): 301-324.

Nikiforakis, Nikos and Hans-Theo Normann, 2008, "A Comparative Statics Analysis of Punishment in Public Goods Experiments," *Experimental Economics* 11(4): 358-369.

Önes, Umut and Louis Putterman, 2007, "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation," *Journal of Economic Behavior and Organization* 62: 495 – 521.

Ostrom, Elinor, James Walker and Roy Gardner, 1992, "Covenants with and without a Sword: Self-governance is Possible," *American Political Science Review* 86 (2): 404 – 17.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei, 2011, "Public goods and voting on formal sanction schemes." *Journal of Public Economics* 95(9): 1213-1222.

Torgler, Benno, 2002, "Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments," *Journal of Economic Surveys* 16 (5): 657-83.

Tyran, Jean-Robert, and Lars P. Feld, 2006, "Achieving Compliance when Legal Sanctions are Non-deterrent." *The Scandinavian Journal of Economics* 108(1): 135-156.

Yamagishi, Toshio, 1986, "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology* 51 (1): 110 – 16.

Zelmer, Jennifer. "Linear public goods experiments: A meta-analysis." *Experimental Economics* 6.3 (2003): 299-310.