Political Accountability and Democratic Institutions: An Experimental Assessment

José Gabriel Castillo*

John Hamman†

December 20, 2019

1 Introduction

Modern societies enforce collective action through delegation of sanctioning duties, which relies on the legitimacy of authority to promote socially desired outcomes.[1] And yet, whether for general economic outcomes or organizational behavior and performance, we know little about "the precise causal mechanisms through which the type of governance affects individual behavior" (Hargreaves Heap et al., 2015). In particular, do democratic institutions succeed by selecting the best delegates, or does participative democracy have an effect on accountability beyond selection? In this paper, we experimentally control for this selection effect to examine whether electoral political accountability functions as a source of institutional legitimacy to promote collective action.

Much research on democratic institutions focuses on direct democracy, in which participants vote to directly implement group-wide decisions. Walker et al. (2000) and Kroll et al. (2007) find that direct democracy increases contributions to public goods, though DeAngelo et al. (2018) show that majority coalitions may use direct democracy to exacerbate inequality. A similar strand

*jcastil@espol.edu.ec - Director, Center for Economic Investigation and Professor, Department of Humanities and Social Sciences, Escuela Superior Politécnica del Litoral (Guayaquil, Ecuador).

†jhamman@fsu.edu. Associate Professor, xs/fs research cluster and Hilton Center, Department of Economics. Florida State University (Tallahassee, Florida, USA).

[1]For a recent review of this literature, see Van Lange et al. (2014).

of research explores direct voting over sanctioning institutions and other consequential actions (Botelho et al., 2005; Ertan et al., 2009; Cinyabuguma et al., 2005; Tyran and Feld, 2006; Ambrus and Greiner, 2015). Sutter et al. (2010) compare the effectiveness of sanctions and rewards as determined via exogenous or endogenous selection, finding that both incentives have a larger effect when endogenously chosen by participants (see also Sefton et al., 2007).

A smaller literature examines democratic and other forms of delegation of authority. Democratic delegation of group contributions in collective action environments has been shown to effectively resolve the free-riding problem (Hamman et al., 2011; Bolle and Vogel, 2011; Fleiß and Palan, 2013). Other recent work shows that a centralized sanctioning authority brought in from outside the group can lead to more efficient outcomes than decentralized sanctioning between group members (Andreoni and Gee, 2012; Baldassarri and Grossman, 2011). One question that arises from these studies, which we address using laboratory experiments, is to what extent the method of appointing a central authority or whether the authority is chosen from within the group matters.

Contrary to the common belief about the causal influence of democratic institutions on collective action and economic outcomes, experimental evidence on delegated enforcement remains inconclusive. Baldassarri and Grossman (2011) and Grossman and Baldassarri (2012) use a lab-in-the-field public goods game with centralized punishment, finding that democratically appointed sanctioning agents brought in from outside the group increase public good contributions relative to randomly appointed external agents. However, Castillo et al. (2018) find no difference between elected and exogenously appointed sanctioning agents in two experimental environments with different levels of sanctions' efficacy, and Hargreaves Heap et al. (2015) similarly find no difference in group outcomes under democratic and dictatorial decision rules. Beyond known experimental nuances, the source of legitimacy in democratic institutional arrangements remains a challenge for causal inference.

In the political science literature, several mechanisms for legitimacy have been analyzed. These include the type of leadership (Grossman, 2014), quality of governance and public information (Adserà et al., 2003), and political competition (Zudenkova, 2018). Less is known about the in-

fluence of the mechanisms of political accountability in building a leader's legitimacy - in fact, Ferejohn (1999) acknowledges that informational advantages that leaders have can be used to reduce responsiveness without reducing legitimacy. Under representative institutions, political accountability rises as the primary mechanism to hold leaders responsible for their political agenda and assure their actions remain aligned with the public's best interests (Grossman and Baldassarri, 2012; A. Huber and Gordon, 2004). If citizens dislike the incumbent's performance, they may seek a replacement in the following election. In line with such retrospective voting, we hypothesize that political accountability builds legitimacy only through a democratic institutional scheme, that is, only if subjects are called to act through voting; hence, they are politically involved in the leader's selection process, as opposed to an automatic (exogenous) political selection by the end of the incumbent's term.

To study this relationship between democratic selection and political accountability, we conduct a laboratory experiment using a hybrid two-by-two design where subjects play a public goods game with and without punishment opportunities where the authority is selected from within the group. This captures the fact that in local or municipal government, agents are frequently selected as residents of the area. In one treatment dimension, we vary whether the central sanctioning authority is elected by the group or exogenously appointed. Here we make a novel contribution to the study of Hobbesian versus democratic institutions by designing this random process - described in greater detail in section 2 - to control for the selection effects of voting (i.e., signaling). The second dimension varies the frequency with which selections are made, either once for the duration of the session or every three periods. This allows us to study one commonly attributed benefit to democratic processes, in which re-election concern incentivizes the authority to act in the electorate's best interests (Ferejohn, 1986).

We find that democratic selection impacts political accountability only for an official's actions, but not for the behavior of their constituents. Specifically, when democratically chosen authorities must face repeated elections, they contribute as much as constituents to the public good. In all other treatments, we observe strong free riding by the sanctioning authority. We see no such differences

in contributions by other group members across treatments. In line with Castillo et al. (2018) and Hargreaves Heap et al. (2015), our results suggest muted effects of democratic systems once we control for the quality of the appointed leader.

In the following section we provide the details for our experimental design and specify our analytical approach. We report our results in section 3 and offer concluding remarks in section 4.

2 Experimental Design

The design extends the framework of Fehr and Gächter (2000), using the centralized tax/punishment environment proposed by Castillo et al. (2018) with a hybrid within-between-subjects design.

2.1 General framework

Table 1 summarizes the sample and treatments. In each session, participants face two stages: first, a standard "linear" voluntary contribution mechanism (VCM, henceforth) that runs for 10 identical rounds, and a centralized punishment institution that runs for 10 rounds in the single selection treatments and 12 rounds in the multiple selection treatments. Instructions for the second stage are distributed only after the first stage finishes, to avoid strategic decisions. We have treatments that vary on two between-subjects dimensions for stage 2: the power delegation mechanism and the political accountability institution.

Participants in the first stage receive an endowment of $w = 20$ experimental units (EU) in each decision round. They can contribute c to a "public account" which constitutes a pool with all group members' contributions, yielding revenue defined by a multiplier (m); in our experiment contributions increase by a multiplier of two ($m = 2$) and are divided equally among n group participants ($n = 5$). This implies a marginal per capita return of 0.4 ($MPCR(\alpha) = m/n$).[2] Each subject faces the trade off between keeping the endowment and free ride on contributions from his partners or contribute to the public account; that is, he faces three alternatives: $c = 0$, which

[2]The MPCR satisfies a basic condition: $0 < \alpha < 1 < n\alpha$; hence, it is socially efficient to contribute all the endowment to the public good if $n\alpha > 1$.

represents the dominant strategy (i.e. Nash equilibrium); $c = w$, which constitutes the socially optimal decision (i.e. Pareto solution), and $0 < c < w$.

The individual's i payoff function in period t can be summarized in the following:

$$\pi_{it} = 20 - c_{it} + 0.4C_t \tag{1}$$

where $C_t = \sum_{j=1}^{n} c_{jt}$, the sum of all members' contributions to the group account.

In the second stage, the centralized tax/punishment institution, each period has two parts. During the first part, subjects face the standard VCM from stage 1. In the second part, one subject, which we call "the manager," administers the management account funded by a tax of two EU ($\tau = 2$), automatically collected from each group member; that is, there are 10 points available in each round. The manager decides whether to punishment his fellow group members, and; if so, how many points to allocate and to whom they will be directed. There is no institutional inefficiency and unused points from the management account return to each group member equally. To allow for better punishment efficacy, punishment points "assigned" are transformed through a convex punishment cost function to punishment points "reduced" as follows:[3]

Points ASSIGNED (p)	0	1	2	3	4	5	6	7	8	9	10
Points REDUCED (p^*)	0	1	2	4	6	9	12	16	20	25	30

Note that, in extreme cases, negative earnings in a round are possible. To reduce the impact of negative payoffs, subjects are allowed to lose either the payoff result or the number of punishment points assigned (not reduced), whichever is lower in absolute terms. Yet, we do not observe any such instances.

The individual payoff function for the second stage is:

[3] Although Nikiforakis and Normann (2008) show its relevance, under the same decision environment as in this document, with no political accountability, Castillo et al. (2018) show that punishment efficacy is inconsequential.

$$\pi_{it} = \begin{cases} \underbrace{20 - c_{it} + 0.4C_t}_{\text{VCM}} - \overbrace{2 - p^*_{it} + \frac{1}{5}\left(10 - p_{jt}\right)}^{\text{tax/punishment mechanism}} & \text{,if } \pi_{it} \geq 0 \\ \min\left\{|p_{it}|, |\pi_{it}|\right\} & \text{,if } \pi_{it} < 0 \end{cases} \tag{2}$$

The general framework is one of *ex-post* full information. All group members, including the manager, see their actions and payoffs following each round, including any reduction in earnings resulting from punishment points. Participants also receive feedback about others' contributions and profits, anonymously in each period. They also observe the total punishment points used in the round, but not to whom they were targeted.

Subjects in this environment may desire to become managers due primarily to the fact that they can decide over punishment points assigned to others but cannot self-inflict punishment. Hence, they not only avoid the probability of reduction of each round's gains, but avoid any risk of bankruptcy; as a result, they face stronger incentives to free ride. This allows us to observe how they trade off these benefits with long-term incentives to maintain accountability.[4]

2.2 Treatments and Procedures

The first treatment dimension corresponds to the *centralized power delegation mechanism*. Here we analyze whether the manager selection mechanism affects the behavior of group members. In the exogenous power delegation mechanism, namely the *Leviathan*, one group member is selected as a manager by the experimenter. The selection process is calibrated in order to isolate the potential effect of the signaling of the manager's quality on participant's behavior. Based on calibrations from previous data (Castillo et al., 2018), the probability of choosing the highest contributor in

[4]This tension is similar to that used in Cooper et al. (2019), which studies leadership in situations involving incentive conflict.

stage 1 is 75%.[5] Subjects are informed of the probability (along with a pie chart of the selection probability as visual aid) and observe a complete contribution history and the average in stage 1 for their group. The second mechanism corresponds to endogenous power delegation, or *Democracy*. By plurality vote, subjects select one group member, after observing their contributions during the first stage (the VCM).[6] Votes are cast simultaneously and anonymously, with ties randomly broken by the software.

The second dimension is the *political accountability institution*, in which we compare *Single* and *Multiple* selections. Managers in *Single* are selected by one of the described mechanisms and, once chosen, they remain in-office permanently. Our *Multiple* framework allows for manager selection every three rounds; to even the decision rounds we extend the periods to 12. Feedback for these treatments is based on the contribution performance of every player on the previous three rounds, again anonymously, and their average contribution during the first stage, except for the first round of stage 2 where selection information is based on the 10 rounds of the VCM in the first stage (see Appendix for more experimental details).

Sessions were conducted in the Laboratory for Experimental and Behavioral Economics (L.E.E.) at ESPOL-Polytechnic University, in Guayaquil-Ecuador, between January and September 2018. We used O-Tree (Chen et al., 2016) as the computer interface, and the recruitment process was performed through ORSEE (Greiner, 2015).

[5]This feature, we argue, is crucial in the design. In the endogenous treatment individuals act upon the contribution information provided by voting, hence this decision reflects the preference over the perceived quality of the manager, while in the exogenous selection they are only informed on the appointment result. A fully random assignment in *Leviathan* would bias towards free riding behavior since subjects might perceive a higher risk of a bad quality manager in office. As we discuss further in the conclusion, this feature may detour from certain non-democratic institutions outside the lab, in which a lack of transparency over the selection rule may impact legitimacy of an appointed leader.

[6]We confirm this information is relevant for the decision making process by asking the participants a set of open questions at the end of the experiment. The majority of subjects focus on the number of points contributed to the public good as the reference for the selection process and the manager's quality. Some other interesting expectations over high contributors are: subjects expect high contributors to manage better the public account and to think more on others' wellbeing; also, they are attributed some personal traits such as intelligence and generosity, giving the sense of deservedness of the appointment.

3 Empirical Analysis

Table 1 summarizes the sample distribution in each cell. A total of 435 subjects were recruited; 200 subjects for the single selection treatment and 235 subjects for the multiple selections treatment. Subjects were undergraduate students that had not participated in an experiment before.[7] Sessions lasted for around 90 minutes and subjects received a show-up fee of USD 2.00, for an average total earnings of around USD 13.00.[8]

Table 1: Experimental design and (preliminary) sample description

Treatments	Single *n=subjects (groups)*	Multiple *n=subjects (groups)*
Leviathan	100 (20)	115 (23)
Democracy	100 (20)	120 (24)
Total sample	**200 (40)**	**235 (47)**

3.1 Empirical approach

Participants in our experiment are students from a relatively diverse background. Forty five percent are women with mean age of 21. Thirty five percent are economics majors, with the rest distributed among careers in engineering and STEM. We also collected some individual information on socioeconomic background and preferences.[9] Table 2 provide a quick description of the main information.

To provide a complete empirical analysis, we employ both nonparametric tests and formal econometric methods.

Since each treatment is randomly administered by session, we analyze mean differences between treatments and stages, directly through the Mann-Whitney U-test (Wilcoxon-Mann-Whitney)

[7] Data for replication is available from Castillo and Hamman (2019)

[8] The minimum basic salary in Ecuador (USD 394) implies an hourly wage of USD 2.46. The average experimental payment therefore represents 2/3 of the basic daily salary.

[9] Individual characteristics will become irrelevant in the models due to individual fixed effects; hence, we do not extend on their exposition.

Table 2: Data Summary

	N	Mean/Proportion	SD	Min	Max
Woman	435	0.45	0.50	0	1
Age	435	21.60	2.28	17	32
Income	435	2.52	1.16	1	5
Risk Aversion	435	6.03	1.38	2	10
Economics and social sci.	435	0.35	0.48	0	1
Communication	435	0.03	0.18	0	1
Natural sci. and mathematics	435	0.10	0.30	0	1
Life sciences	435	0.04	0.20	0	1
Earth sciences	435	0.07	0.25	0	1
Electrical engineering	435	0.21	0.41	0	1
Maritime and science of the sea	435	0.05	0.21	0	1
Mechanics and production sci.	435	0.15	0.36	0	1

Notes: Income levels (5): 1, i < $ 364; 2, $ 365 < i < $ 600; 3, $ 601 < i < $ 1000; 4, $ 1001 < i < $ 1600; and, 5, i > $ 1600. Risk aversion is the self-reported measure for a 10 points Likert scale, included in the questionnaire.

at group level. This is the main approach for our results on the differences between the two political accountability institutions.

We extend the analysis econometrically to control for possible confounds within each treatment of the power delegation mechanism. We include a fully saturated specification with several longitudinal controls and fixed effects at various levels. This is a Difference-in-Difference approach; to analyze the contribution determinants we estimate an equation as follows:

$$\begin{aligned} C_{igt} =& \alpha_1 Democracy(D)_i + \alpha_2 Punishment(P)_t + \alpha_3 (D*P)_{it} \\ &+ X'_{ig}\Lambda + Z'_i\Gamma + \phi_g + \tau_t + \varepsilon_{igt} \end{aligned}$$

where C_{igt} is the contribution level of subject i, in group g, in period t. α_2 represents the average effect of the democratic power delegation; α_3 is the average effect of a centralized punishment institution. The coefficient of interest for the average treatment effect (ATE) of the endogenous power distribution under centralized punishment institution is $\hat{\alpha_1}$. Λ is a vector controls for individual behavior within each group; Γ is the vector of individual controls (individual fixed effects

in the most flexible case); ϕ_g are group fixed effects; τ_t are dynamic time trends within each stage; and ε_{igt} the $i.i.d.$ idiosyncratic error term.

3.2 Results

3.2.1 Do democratic elections incentivize collective action?

The first thing to note in our analysis is that an endogenous (i.e., democratic) power delegation does not trigger intrinsic motivation to improve contribution behavior in a centralized management environment, in line with other recent work (Hargreaves Heap et al., 2015; Castillo et al., 2018).

Figure 1 shows the average contribution dynamics of our experiment. Note first that the centralized punishment institution yields results similar to the literature on decentralized punishment (Fehr and Gächter, 2000; Ledyard, 1995; Putterman et al., 2011; Nikiforakis and Normann, 2008). Once imposed, the mechanism promotes higher and more stable levels of cooperation than without punishment opportunities. Second, we observe no significant differences between power delegation mechanisms, whether or not the political accountability institution is imposed.

Table 3 summarizes the main results. Each cell of Panel A shows the average difference between the centralized punishment institution and the VCM. We answer the first question by comparing vertically. The *Democracy-Leviathan* row shows the average difference between the power delegation mechanisms. We observe that differences between *Democracy* and *Leviathan* are not statistically significant, regardless of the frequency of selections. In Table 4 we extend the analysis econometrically to show that results are consistent under alternative specifications. As in the seminal paper of Fehr and Gächter (2000), a significant effect comes from the punishment institution imposed, but there are no differences of the power delegation mechanism whether in *Single* or *Multiple*. Also, the higher the past contributions of others within a group, the higher the observed contribution, which again aligns with prior findings of conditional cooperation as an emergent social phenomenon.

Observed results for the first dimension of the analysis can not be explained by differences in punishment behavior. Figure 2 shows various punishment measures for both treatments, which

Table 3: Average Performance Comparison

Panel A: Contributions (points)

Punishment-VCM

Treatments	Single	Multiple	Single-Multiple
Leviathan	2.863 (1.035)	2.599 (0.880)	0.264 (0.772)
	$[p = 0.029]$	$[p = 0.006]$	$[p = 0.961]$
Democracy	3.046 (0.967)	1.931 (1.013)	1.114 (0.870)
	$[p = 0.005]$	$[p = 0.046]$	$[p = 0.195]$
Democracy-Leviathan	0.183 (0.892)	−0.667 (0.760)	
	$[p = 0.850]$	$[p = 0.395]$	

Panel B: Punishment (points)

Democracy-Leviathan

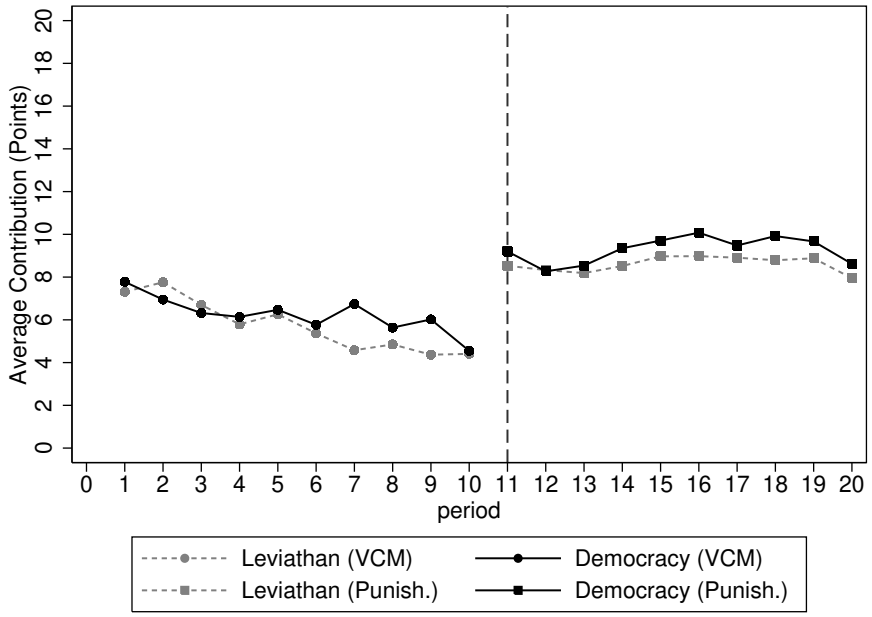
Experiments	Single	Multiple	Single-Multiple
	−0.076 (0.147)	0.027 (0.125)	−0.060 (0.208)
	$[p = 0.560]$	$[p = 0.831]$	$[p = 0.664]$

Panel C: Profits (points)

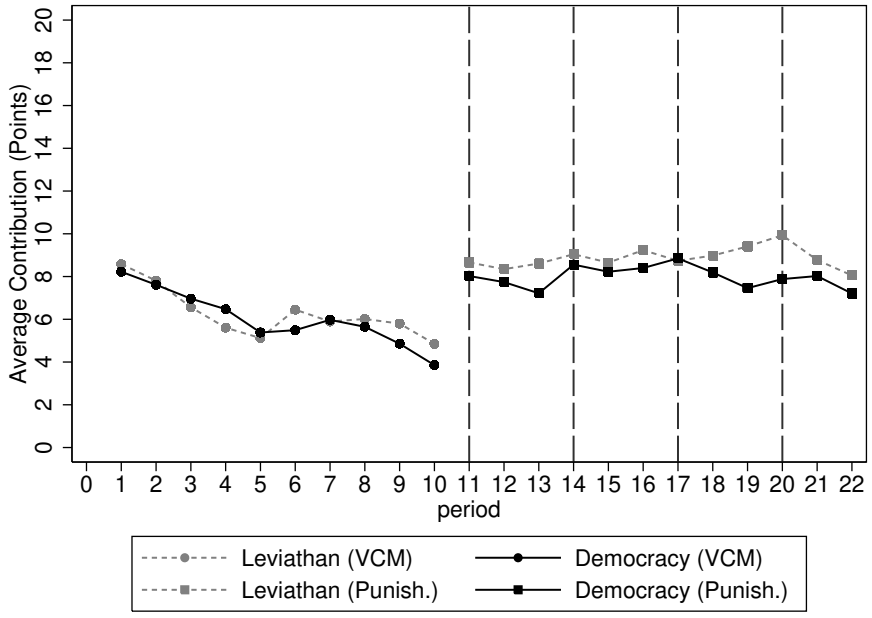
Punishment-VCM

Treatments	Single	Multiple	Single-Multiple
Leviathan	0.576 (0.983)	0.295 (0.846)	0.426 (0.417)
	$[p = 0.956]$	$[p = 0.684]$	$[p = 0.733]$
Democracy	0.833 (0.873)	−0.474 (1.043)	1.307 (0.853)
	$[p = 0.304]$	$[p = 0.327]$	$[p = 0.157]$
Democracy-Leviathan	0.256 (0.916)	−0.770 (0.787)	
	$[p = 0.903]$	$[p = 0.371]$	

Notes: Panels A and C report within subjects differences between stage 2 (punishment) and stage 1 (VCM). Panel B reports differences between treatments.
Group-clustered standard errors in parentheses.
p-values are reported in brackets for a Mann-Whitney U tests. Two-sided t-tests report similar results.



(a) Single



(b) Multiple

Figure 1: Average contribution dynamics

reveal no evidence of significant differences. Panel B of Table 3 and Table 5 support this conclusion. The only significant difference observed is on the extensive margin; that is, the probability of being punished slightly diminishes in *Single* under a democratic scheme; however, this result

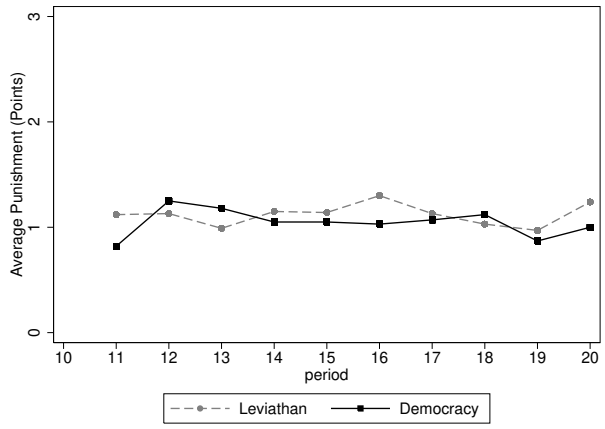
Table 4: Determinants of Contributions

	Single		Multiple	
Dep. Variable=Contributions (points)	FEgt	FEgtc	FEgt	FEgtc
Democracy vs. Leviathan (P*D)	0.1785	0.1636	-0.6672	-0.4069
	(0.9035)	(0.5920)	(0.7671)	(0.5224)
Punishment	2.8675***	2.1510***	3.9856***	3.1360***
	(0.6080)	(0.3940)	(0.4184)	(0.3144)
Other member's Av. Contribution (t-1)		0.3883***		0.3558***
		(0.0479)		(0.0406)
Punishment Received (t-1)		0.3658		-0.0816
		(0.4247)		(0.1109)
Punishment in the group (t-1)		-0.1574		0.1586
		(0.1316)		(0.3283)
Other controls	No	No	No	No
Individual FE	Yes	Yes	Yes	Yes
Group FE	Yes	Yes	Yes	Yes
Trend within stage	Yes	Yes	Yes	Yes
R.squared	0.4233	0.4712	0.4369	0.4736
Observations	4000	3800	5170	4935

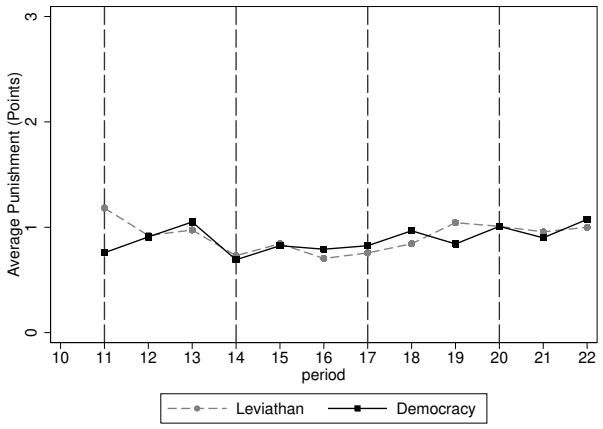
Notes: Dummy variable for the Democracy treatment (D) excluded since it is time invariant; hence it has a null coefficient for a Fixed Effect (FE) estimation.
Standard errors clustered at group level in parentheses.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

disappears in the presence of political accountability opportunities (i.e. *Multiple*. See Figure 2e and Figure 2f). There are neither differences in the intensive margin, punishment points used are not statistically different between power delegation mechanisms (see Figure 2a and Figure 2b); nor in the manager's use of punishment (see Figure 2c and Figure 2d). Deviations from the social norm (i.e. the group's average contribution) intervene in the probability of being punished in the expected way; negative deviations increase the probability and intensity of punishment under both *Single* and *Multiple*.

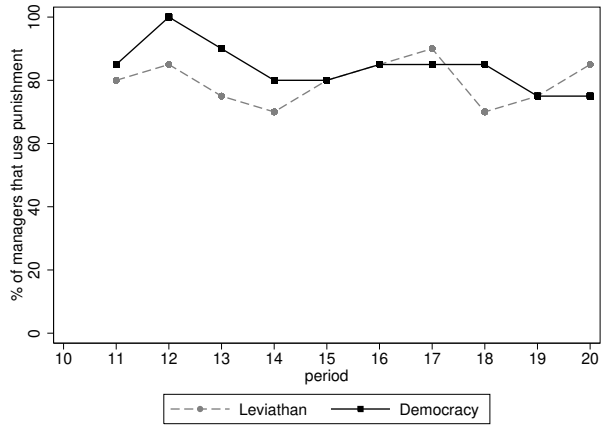
Putting things together, the welfare measure of our framework can be summarized in the subject's profit, that is, the net payoff received after punishment. Panel C of Table 3 shows these



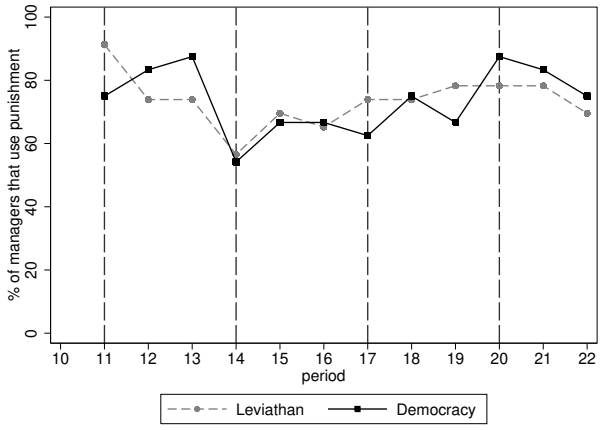
(a) Single: punishment assigned



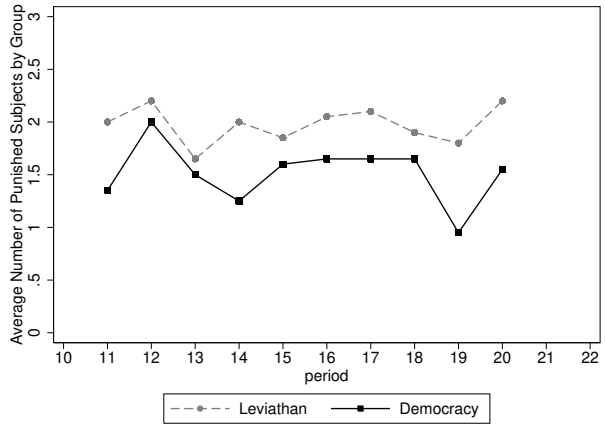
(b) Multiple: punishment assigned



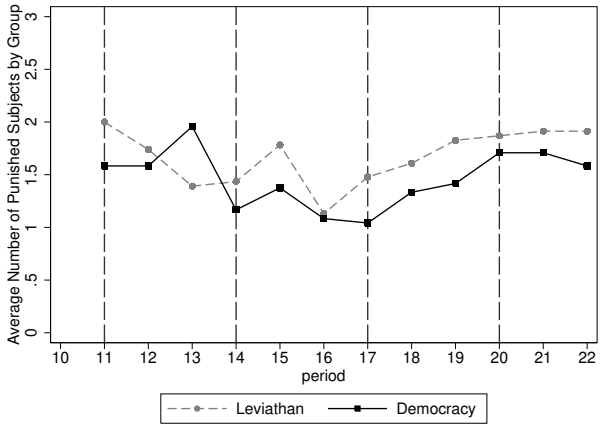
(c) Single: manager's use of punishment



(d) Multiple: manager's use of punishment



(e) Single: subjects punished



(f) Multiple: subjects punished

Figure 2: Punishment Behavior

results. Conclusions remain.

3.2.2 Does political accountability of sanctioning authority affect behavior?

This question can be sliced into two different aspects of the framework's incentives: contribution behavior of the group, and contribution behavior of the managers.

To answer the first part, we return to the main results in Table 3, only this time we concentrate on the comparison across columns. In the previous section we show that the democratic election has no effect on the general contribution behavior regardless of the frequency of selection; in other words, political accountability does not add any differential incentive to the democratic process to promote collective action. Results in the third column in Panel A show whether there are behavioral differences across treatments on the second dimension of the analysis, that is, the institution of political accountability. Reinforcing the previous conclusions, observed differences between *Single* and *Multiple* are not statistically significant, regardless of the power delegation mechanism in place.

Political accountability adds little to the contribution dynamics. Every three periods, contributions tend to increase slightly on the electoral period; this is more clearly seen in *Democracy* (Figure 1b). Given the design conditions, subjects in both treatments can enhance their selection probabilities by increasing their contributions, which signal either their peers, in *Democracy*, or the experimentalist's selection rule, in *Leviathan*. Again, observed differences are not statistically significant, in particular once netting out the first stage behavior (i.e. VCM). In other words, subjects in both treatments resolve equivalently their social dilemma between contributing –raising the probabilities of being in office–, and free riding. Once signaling opportunities are adequately controlled, we argue, *Democracy* does not offer improvements in institutional legitimacy, and incentives towards collective action do not play a differential role.

The second part of the question sheds some light on the relationship of the political accountability institution and the manager's selection mechanism. Figure 3 decomposes the contribution dynamics of manager's and non-manager's by each dimension of the experimental design. As

Table 5: Punishment Decision

	Single		Multiple	
	(1)	(2)	(3)	(4)
Dependent variable:	Punished=1	Punishment Points	Punished=1	Punishment Points
Democracy (D)	-0.0843**	-0.0190	-0.0361	0.1173
	(0.0380)	(0.2791)	(0.0254)	(0.1962)
OMC negative deviation	0.0425***	0.3795***	0.0300***	0.3474***
	(0.0033)	(0.0315)	(0.0031)	(0.0303)
OMC positive deviation	-0.0294***	0.0448	-0.0229***	0.0079
	(0.0042)	(0.0542)	(0.0036)	(0.0245)
R.squared (overall)		0.444		0.3450
Observations	1999	698	2820	883

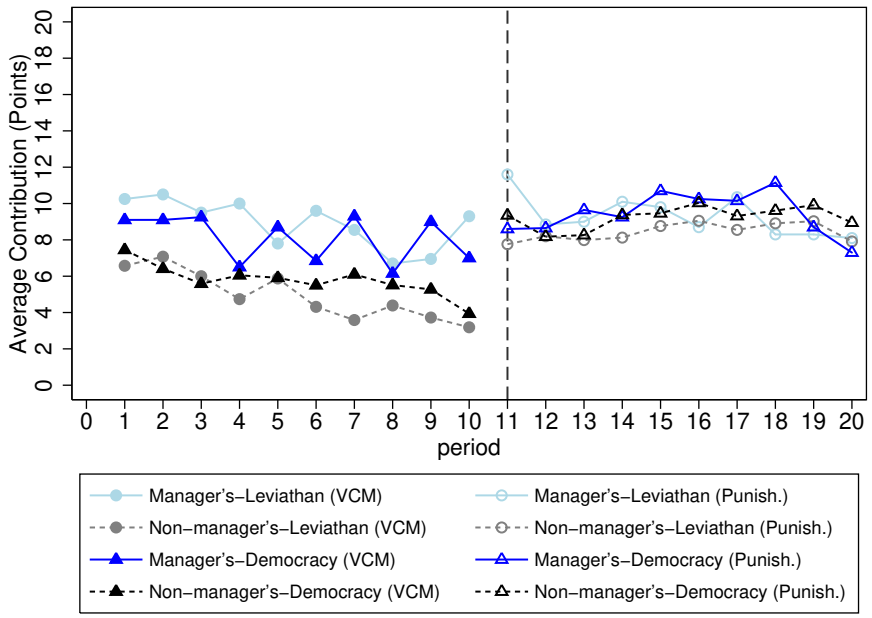
Notes: Coefficients in models 1 and 3 report the marginal effects (at means) of the probability of being punished for a Panel Data Probit model to capture the within individual correlation.
OMC=Other members' Average Contribution.
Standard errors clustered at group level in parentheses for models 2 and 4. Robust standard errors for models 1 and 3.

* Significant at the 10 percent level.

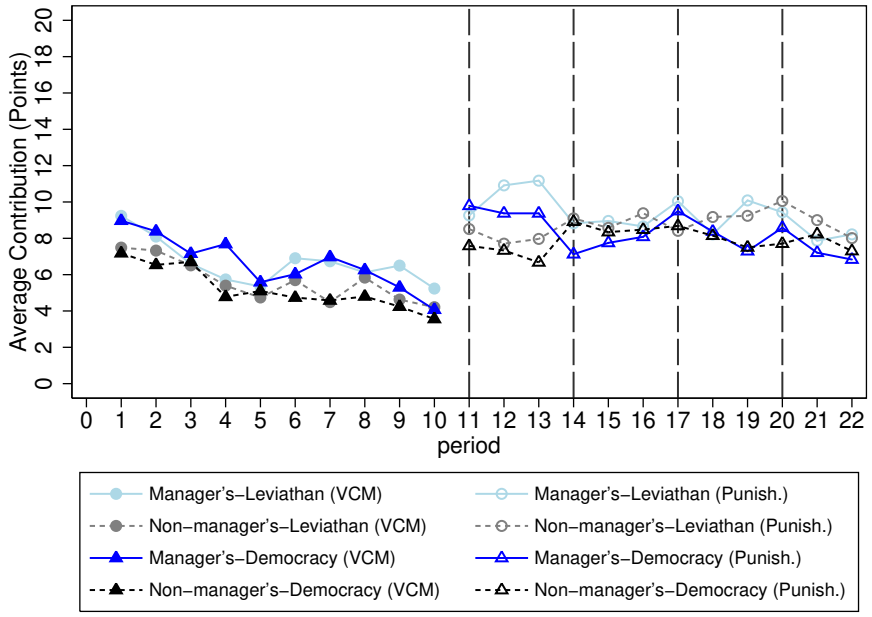
** Significant at the 5 percent level.

*** Significant at the 1 percent level.

expected, in the first stage (i.e. the VCM) subjects selected as manager's are usually the highest contributors in both power delegation schemes (see Figure 4); on the other hand, manager's contribution differences, observed in the first stage, do not translate into the second stage for a centralized punishment environment. The contribution dynamics of all treatments shrinks, regardless of the roles of the group members. Table 6 offers a formal test of the mean changes in contribution between stages within the two dimensions of the design, by roles subject's play within the group. The take-away from the table's results is that manager's, on average, exercise their edge in free riding opportunities and contribute significantly less in all treatments, except in the democratic power delegation under the political accountability institution; mean differences between managers and non-managers are not statistically significant in this treatment (last row). In other words, although political accountability does not have an effect on constituents, it does matter in terms of the manager's behavior, conditional on being in *Democracy*.



(a) Single



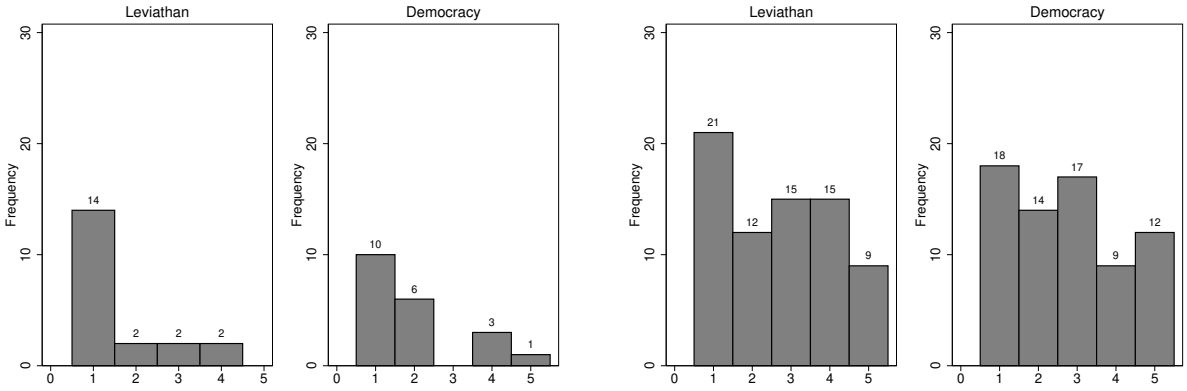
(b) Multiple

Figure 3: Contribution's dynamics by roles

Table 6: Leadership contribution analysis

	Punishment-VCM	Manager	Others	p-value Mann-Whitney U tests (H_0: equal means)
Single	Overall	0.870(0.370)	3.476(0.177)	0.000
	Leviathan	0.395(0.522)	3.482(0.231)	0.000
	Democracy	1.345(0.525)	3.471(0.269)	0.000
Multiple	Overall	2.006(0.315)	2.423(0.159)	0.061
	Leviathan	1.926(0.494)	2.881(0.231)	0.016
	Democracy	2.083(0.396)	1.985(0.216)	0.832

Notes: Standard errors clustered at group level, in parentheses. *Multiple* has 12 rounds in stage 2; hence, for equal comparison, we only take into account the difference in contribution until round 20.



(a) Single (b) Multiple

Figure 4: Manager's contribution rank within group

4 Conclusion

There is a rich empirical literature examining the benefits of democratic institutions in providing public goods, often through direct democracy. We contribute to this literature by examining the effects of democratic accountability on sanctioning authority. In contrast to democratically determined contributions, we find that democratically elected sanctioning authority has muted effects on group outcomes. When we control for the quality of the authority, we find no difference in group outcomes between democratic and exogenous mechanisms.

While it is tempting to conclude that the lack of selection differences drives these muted effects of democratic appointment, we cannot rule out that certain experimental design elements may also contribute to our findings. For example, our administrator may spend the punishment points in the group pool differently than if they were from a private account, as is done in prior studies. However, we do see punishment used in roughly similar amounts to authorities in Baldassarri and Grossman (2011). We simply see no difference in responses to punishment based on institution once selection is ruled out.

Because participants knew the exogenous selection rule, their beliefs about the quality of the chosen candidate may not differ between institutions. While selection criteria can be observed in democratic elections, many non-democratic institutions certainly lack transparency in selecting their leaders, which directly affects their legitimacy. This raises an important question for future study. Namely, would we continue to see similar behavior between institutions if we kept leader quality fixed, but did not make this transparent to voters?

Interestingly, we do find that democratically elected authorities facing repeated elections no longer free ride. Instead, their contributions are in line with those of other group members. In contrast, democratically elected authorities who do not face repeated election (i.e., in the absence of political accountability), as well as exogenously appointed authorities, contribute significantly less to the public good than their fellow group members.

We conjecture that strong beliefs over the advantages of democratic institutions in centralized power environments rely on features that either act jointly or are independent of the power dele-

gation mechanism. One important feature of modern governance is political accountability; when in place, it offers different incentives to the authorities, in particular, what we refer to as a *responsibility effect* reflected in higher contribution behavior. Important in our study results, this effect arises only under a democracy.

References

A. Huber, G. and Gordon, S. C. (2004). Accountability and coercion: Is justice blind when it runs for office? *American Journal of Political Science*, 48(2):247–263.

Adserà, A., Boix, C., and Payne, M. (2003). Are you being served? political accountability and quality of government. *The Journal of Law, Economics, and Organization*, 19(2):445–490.

Ambrus, A. and Greiner, B. (2015). Democratic punishment in public good games with perfect and imperfect observability. *Economic Research Initiatives at Duke (ERID) Working Paper*, (183).

Andreoni, J. and Gee, L. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, 96(11-12):1036–1046.

Baldassarri, D. and Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27):11023–11027.

Bolle, F. and Vogel, C. (2011). Power comes with responsibility—or does it? *Public Choice*, 148(3):459–470.

Botelho, A., Harrison, G. W., Pinto, L., and Rutström, E. E. (2005). Social norms and social choice. Technical report, Universidade do Minho. Núcleo de Investigação em Microeconomia Aplicada (NIMA).

Castillo, J. G. and Hamman, J. (2019). Replication Data for: Political Accountability and

Democratic Institutions: An Experimental Assessment. *Harvard Dataverse*. Availale at: https://doi.org/10.7910/DVN/RIEMMC.

Castillo, J. G., Xu, Z., Zhang, P., and Zhu, X. (2018). The effects of centralized power and institutional legitimacy on collective action. Technical report, ESPOL. Available at SSRN: https://ssrn.com/abstract=3127775.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree- an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9(Supplement C):88 – 97.

Cinyabuguma, M., Page, T., and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89(8):1421 – 1435.

Cooper, D. J., Hamman, J. R., and Weber, R. A. (2019). Fool me once: An experiment on credibility and leadership. *forthcoming in The Economic Journal*.

DeAngelo, G. J., Dubois, D., and Romaniuc, R. (2018). The perils of democracy. *Journal of Economic Behavior & Organization*.

Ertan, A., Page, T., and Putterman, L. (2009). Who to punish? individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, 53(5):495–511.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1):5–25.

Ferejohn, J. (1999). Accountability and authority: toward a theory of political accountability. *Democracy, accountability, and representation*, 131:133.

Fleiß, J. and Palan, S. (2013). Of coordinators and dictators: A public goods experiment. *Games*, 4(4):584–607.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Grossman, G. (2014). Do selection rules affect leader responsiveness? *Quarterly Journal of Political Science*, 9(1):1–44.

Grossman, G. and Baldassarri, D. (2012). The impact of elections on cooperation: Evidence from a lab-in-the-field experiment in uganda. *American Journal of Political Science*, 56(4):964–985.

Hamman, J. R., Weber, R. A., and Woon, J. (2011). An experimental investigation of electoral delegation and the provision of public goods. *American Journal of Political Science*, 55(4):738–752.

Hargreaves Heap, S. P., Tsutsui, K., and Zizzo, D. J. (2015). An experiment on democratic versus dictatorial collective decision making. Technical report, UEA, U Bath, U Queensland.

Kroll, S., Cherry, T. L., and Shogren, J. F. (2007). Voting, punishment, and public goods. *Economic Inquiry*, 45(3):557–570.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In Kagel, J. and Roth, A., editors, *Handbook of Experimental Economics*, Handbook of Experimental Economics. Princeton University Press.

Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.

Putterman, L., Tyran, J.-R., and Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics*, 95(9):1213–1222.

Sefton, M., Shupp, R., and Walker, J. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4):671–690.

Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, 77(4):1540–1566.

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

Van Lange, P. A., Rockenbach, B., and Yamagishi, T. (2014). *Reward and Punishment in Social Dilemmas*. Oxford University Press.

Walker, J. M., Gardner, R., Herr, A., and Ostrom, E. (2000). Collective choice in the commons: Experimental results on proposed allocation rules and votes. *The Economic Journal*, 110(460):212–234.

Zudenkova, G. (2018). Electoral accountability and interdistrict competition. *American Economic Journal: Microeconomics*, 10(3):143–76.