

The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda

Guy Grossman University of Pennsylvania
Delia Baldassarri Princeton University

Communities often rely on sanctioning to induce public goods contributions. Past studies focus on how external agencies or peer sanctioning induce cooperation. In this article, we focus instead on the role played by centralized authorities, internal to the community. Combining “lab-in-the-field” experiments with observational data on 1,541 Ugandan farmers from 50 communities, we demonstrate the positive effect of internal centralized sanctioning authorities on cooperative behavior. We also show that the size of this effect depends on the political process by which authority is granted: subjects electing leaders contribute more to public goods than subjects who were assigned leaders through a lottery. To test the ecological validity of our findings, we relate farmers’ behavior in the experiment to their level of cooperation in their community organization. We show that deference to authority in the controlled setting predicts cooperative behavior in the farmers’ natural environment, in which they face a similar social dilemma.

Contributions to the provision of local public goods are commonly framed as a problem of cooperation. Since public goods are nonrivalrous and nonexcludable, rational self-interested group members would rather free ride on others’ contributions than bear the costs of cooperation (Olson 1965). Sanctioning is widely considered a viable solution to this type of social dilemma (Fehr and Gächter 2002; Sigmund 2007). Most generally, scholars have identified two forms of sanctioning solutions to the public goods problem. In the first solution, *central authorities* that are *external* to the group are the locus of coordination and enforcement of cooperative efforts. Examples of these solutions include theories of the state (Scholz and Gray 1997). In the second solution, cooperation emerges from uncoordinated and *decentralized* punishment efforts that are *internal* to the group. Examples of these solutions include peer-sanctioning regimes (Ostrom 1990).

These two solutions, however, do not account for all relevant situations. In fact, even small-size groups and communities are characterized by some level of social differentiation from which *internal centralized authorities* emerge (King, Johnson, and Van Vugt 2009).¹ For example, merchants in medieval Europe created guilds to adjudicate disputes (Greif, Milgrom, and Weingast 1994), and villagers in rural Kenya created school committees to sanction parents who fail to contribute to local public goods (Miguel and Gugerty 2005). The first contribution of this article is to analyze the effectiveness of internal centralized-sanctioning institutions in fostering cooperation.

In the past two decades, formal and experimental research have focused almost exclusively on peer-sanctioning institutions. Within this framework, scholars have demonstrated that the threat of sanctioning induces greater cooperation by changing individuals’ payoff

Guy Grossman is Assistant Professor, Department of Political Science, 225 Stiteler Hall, 208 S. 37th Street, Philadelphia, PA 19104-6215 (ggros@sas.upenn.edu). Delia Baldassarri is Associate Professor, Department of Sociology, 147 Wallace Hall, Princeton University, Princeton, NJ 08544 (dbalda@princeton.edu).

We are grateful to Alex Barnard, Eliana Horn, Vivian Lu, and Sylvie Hoster and to our local team of enumerators for excellent research assistance. We thank Diego Gambetta, Benjamin Goodrich, Elizabeth Sperber, Kristin Michelitch, Laura Paler, Timothy Frye, Kimuli Kasara, Michael Hechter, Macartan Humphreys, David Laitin, David Stasavage, three anonymous reviewers, and participants in various seminars and conferences for helpful comments. D.B. gratefully acknowledges support from the NSF Grant SES(10S)-0924778 and the Princeton Institute for International and Regional Studies. G.G. gratefully acknowledges support from the NSF Doctoral Dissertation Improvement Grant SES-0921204 and Princeton University’s Office of Population Research. Data for replication can be found at <http://dvn.iq.harvard.edu/dvn/dv/guygrossman>.

¹Social differentiation denotes the tendency of groups and communities to develop hierarchies, in which social roles are defined as a set of rights and duties members are expected to fulfill (Eguíluz et al. 2005).

American Journal of Political Science, Vol. 56, No. 4, October 2012, Pp. 964–985

©2012, Midwest Political Science Association

DOI: 10.1111/j.1540-5907.2012.00596.x

functions (Fehr and Gächter 2002; Gintis et al. 2005).² If we consider, however, centralized sanctioning, institutional effectiveness may not rely exclusively on the threat of punishment, but also on the extent to which it is perceived as legitimate (Dickson, Gordon, and Huber 2009; Eckel, Fatas, and Wilson 2010). The legitimacy of an authority is defined as certain dispositions, beliefs, perceptions, and attitudes that increase the willingness to obey an authority. Those dispositions, which cannot be reduced to monetary incentives or to the threat of brute force, refer to the authority's right to govern (Rawls 1971) and the subject's obligation to obey it (Levi, Sacks, and Tyler 2009). According to Weber (1922), legitimacy can be derived from different sources, such as charismatic, traditional, or rational-legal authority. Focusing on the latter, the second contribution of this article is to demonstrate that the *political process* by which an authority originally acquires its sanctioning powers is consequential for cooperation.

To investigate these aspects, we combine "lab-in-the-field" behavioral experiments with observational data on 1,541 producers from 50 Ugandan farmer associations. We developed a novel adaptation of the public goods game (PGG), which is the conventional behavioral experiment used to study the conditions under which groups can overcome individual incentives to defect (Camerer 2003). The experimental setup allows us to attest to the positive impact of centralized-sanctioning institutions on cooperative behavior as well as to demonstrate that the size of this effect depends on the process by which these institutions are established. To assess the *ecological validity* of our findings, we then relate the behavior of subjects in the PGG to their behavior in their natural setting as members of farmer groups, in which they face a similar social dilemma.³

The article unfolds as follows. After discussing our theoretical framework, we describe the research design and provide information on the research site. Following a summary of our experimental manipulation, we present a first set of results. To study the impact of internal centralized-sanctioning authorities, we compare the cooperative behavior of subjects who do not face a threat of punishment with that of subjects in a context where a monitor is granted monopoly over sanctioning powers.

²The main focus of this strand of literature is in identifying conditions for overcoming the second-order collective action problem inherent in costly punishment. See Fowler (2005) and Boyd et al. (2003).

³A similar strategy has been used in a gift-exchange experiment (List 2006) and a donation experiment (Benz and Meier 2008). Whether individuals' prosocial behavior in experiments correlates with their behavior in natural settings is largely an open question.

We find that the introduction of a centralized-sanctioning authority has a strong positive impact on cooperation. To study whether the political process through which local authorities acquire their powers is consequential, we compare the cooperative behavior of subjects who elected their monitor with the behavior of subjects who were assigned a monitor selected at random. We find that the way in which centralized authorities obtain their sanctioning powers has an independent impact on cooperation: participants are more responsive to the authority of elected monitors.

Following a brief summary of our observational data, we present a second set of results. First, as in the controlled setting of the experiment, we find a strong association between the perceived legitimacy of the managers of the farmer cooperatives and the level of cooperation of group members. Second, we show that when the experimental conditions reproduce key features of the subjects' natural setting, cooperative behavior in the *experiment* can predict a level of cooperation in the farmer group. We conclude that the experimental setting captures institutional conditions and group dynamics that are relevant for determining levels of cooperation, at least in the context of Ugandan community organizations.

Theoretical Framework and Hypotheses

Experimental evidence shows that peer sanctioning is a successful strategy for increasing cooperation. In PGGs, subjects anonymously decide how to split an endowment between private and public accounts. What subjects put in the private account remains theirs, while what is contributed to the public account is doubled (or otherwise multiplied) and redistributed evenly among all group members regardless of their level of contribution. The most profitable outcome for the group occurs when all subjects contribute their entire endowment. Nonetheless, the most profitable strategy for the individual is to keep the entire endowment and benefit from what *everyone else* contributes to the public account. Designed to induce a social dilemma, PGGs capture how individuals balance self-interest and the well-being of the group.

In PGGs, participants initially contribute, on average, between 40 and 60% of their endowment. However, in repeated games, it is common to observe a drop in contributions in subsequent rounds, as conditional cooperators, who wish to avoid being exploited by free riders, gradually refrain from cooperation (Fischbacher, Gächter, and Fehr 2001). By contrast, when participants are allowed

to punish other subjects, overall levels of contribution increase, since conditional cooperators can discipline defectors (Fehr and Gächter 2002). Peer punishment provides a possibility of targeted interaction, thus fostering cooperation through mechanisms of direct and indirect reciprocity (Lubell and Scholz 2001).

Peer sanctioning, however, is only effective under very restrictive conditions (Sigmund 2007; Taylor 1982). It can only sustain cooperation in relatively small-size groups, where the cost of punishment is likely to be recuperated. In such groups, self-interested contributors may choose to punish defectors at a personal cost, as long as they have reasons to believe that punishment will increase the future contributions of the “targets.” This, in turn, depends on the frequency of interaction between members (Boyd, Gintis, and Bowles 2010). As the number of members increases and interactions become infrequent, bilateral punishment becomes unlikely to sustain cooperation because future gains from punishment cannot be internalized (Greif 1993).

The limited scope of peer sanctioning induces groups, organizations, and communities to delegate sanctioning powers to internal centralized authorities. These institutions are likely to be more efficient than peer punishment (Erikson and Parent 2007; Guth et al. 2007; O’Gorman, Henrich, and Van Vugt 2009), since they are better positioned to overcome coordination failures and free riding problems, although they might experience flaws in information, thus leading to enforcement errors (Dickson, Gordon, and Huber 2009). The centralization of sanctioning is the likely outcome of an *endogenous* process of social differentiation: virtually all social groups, even those characterized by low levels of complexity, experience elementary forms of division of labor that lead to the emergence of hierarchical structures and leadership roles (Baldassarri and Grossman 2011).⁴ To incorporate these intuitions into theories of public goods provision, we study how group members behave when a single individual is given a monopoly over sanctioning. Specifically, we test the following hypotheses:

H1: Internal centralized authorities that are given a monopoly over sanctioning decisions will punish defectors at a personal cost.

H2: Centralized sanctioning induces greater contribution toward public goods provision.

⁴Sanctioning, of course, is only one way in which leaders impact cooperative behavior. Other means include persuasion (Henrich and Gil-White 2001), coordination (Wilson and Rhodes 1997), provision of information (Dewan and Myatt 2008), and leading by example (Levati, Sutter, and van der Heijden 2007).

In the past two decades, social scientists working within a “new institutionalist” framework have focused on demonstrating how individuals’ expectations and behavior are shaped by *incentives* embedded within formal and informal rules (Levitsky and Murillo 2009). The various expansions of Douglas North’s basic approach—which include distinguishing between formal and informal, weak and strong, or exogenous versus endogenous institutions—continue to place incentives at the center of the theory’s microfoundation. We complement and enrich this framework by showing that the *process* by which institutions are put into place has a causal effect on individuals’ behavior. Namely, our second research question is whether and how the political process through which internal centralized authorities obtain their sanctioning powers is consequential for cooperation.⁵

Differently from peer-sanctioning systems in which the right to punish defectors comes hand-in-hand with group membership, in a centralized-sanctioning regime it is important to distinguish between the effect of sanctioning and the effect of the way in which sanctioning powers are granted. In this article, we focus on testing whether elections have a positive impact on subjects’ cooperative behavior. Notably, our focus on leader-selection methods addresses a major gap in the current literature on the impact of leaders on cooperation: the tendency to treat the emergence of leadership institutions as wholly exogenous (Ahlquist and Levi 2011).⁶

There are several complementary mechanisms through which the participation of group members in the selection of a sanctioning authority would induce greater cooperation. Our focus here is on testing whether the electoral procedure itself can increase public goods contribution via its impact on the perceived legitimacy of the centralized authority.

Why should we expect elections to deem leaders more legitimate? First, elections may have a ritualistic or symbolic value that confers on leaders greater authority. Closely related, in many groups and societies, *explicit consent* via elections gives people a sense of agency

⁵Similarly, Hibbing and Alford (2004) show that acceptance of binding decisions depends partially on the procedure through which decisions were made, and Dal Bo, Foster, and Putterman (2010) show that the effect of a policy on cooperation is greater when it is chosen democratically by the subjects rather than being exogenously imposed.

⁶Past studies allowed participants (Casari and Luini 2009) and external monitors (Dickson, Gordon, and Huber 2009) to endogenously select some features of the punishment institution. The monitors in those studies, however, were exogenously determined. Our study expands the small experimental literature on endogenous institutions by allowing participants to select the identity of their monitor.

and control over the selection process.⁷ Second, elections can contribute to the legitimacy of centralized authorities through a “peer effect” (Zelditch 2001). In short, election results signal to subjects the worthiness of the chosen authority, affecting the way subjects encode information from leaders, and the importance they give to their messages. Third, elections may increase people’s sense of *obligation* to follow the leader (Greif 2006). In addition, a core argument of a large political science and social psychology literature is that the fairness of the procedure through which authorities gain power and/or exercise power shapes the willingness of subjects to defer to their authority.⁸ This claim is widespread in legal (Gibson, Caldeira, and Spence 2005), political (Levi and Sacks 2009), and managerial settings (Hoffmann 2005). Our experimental design does not allow for a test of hypotheses concerning specific internal psychological dispositions, but it will nonetheless enable us to focus on the *behavioral manifestation of legitimacy*. Namely, we test whether subjects are more likely to defer to a leader and accept her punishment decisions if they are allowed to participate in her selection.

Two other notable mechanisms might account for the relationship between elections and cooperative behavior: a leadership-selection effect and accountability. First, direct elections may enable group members to select “better” leaders, namely, leaders whose *status or characteristics* make them more effective in triggering cooperation (Eckel, Fatas, and Wilson 2010). Second, periodic elections may induce elected leaders to sanction more stringently in order to be reelected, thereby increasing cooperative behavior (Huber and Gordon 2004). Since these mechanisms may confound the effect of legitimacy, we designed our experiments in order to minimize leader-selection effects and control for this aspect in the analysis. We also eliminate accountability effects.⁹ Building on the above framework, our experiment was designed to test the following hypotheses:

H3: Outcome. The process by which a centralized authority acquires its sanctioning power is consequential for

⁷The idea that consent is needed to legitimize authority is usually accredited to Locke (1990). That elections are the most appropriate way to elect local leaders was forcefully put forth by rural Ugandans in dozens of interviews we have conducted during our field work.

⁸The evaluation of the fairness of a political process may vary over time, space, and contextual conditions.

⁹Accountability effects are eliminated by having the monitor selected “once and for all” in subsequent rounds. Leaders’ selection effects were minimized by randomly sampling our experimental subjects from six different villages and from a pool of members who did not hold any leadership roles in their associations. We return to these points in the research design and analysis sections.

cooperation: contribution to public goods provision is higher when authorities are elected rather than selected at random.

H4: Mechanism. Elections increase cooperation through a “legitimacy effect”—individuals are more likely to commit to a leader’s authority if they participate in her selection.

Our third research question concerns the extent to which subjects’ behavior in the experimental setting resembles their behavior in the natural setting as farmer group members. Morton and Williams (2010, 264–65) refer to such congruence between the experimental setting and the target population’s natural environment as the *ecological validity* of an experiment. The approach we took to address this concern consists of two steps. First, we conducted individual-level surveys with the experiment’s subjects in which we collected information about their cooperative behavior as farmer association members. We then use those observational data to test whether there exists a positive relation between a respondent’s evaluation of his group leader’s legitimacy and the respondent’s contribution toward the public good in his farmer group. Secondly, we test whether group members’ level of cooperation in the PGGs can predict their cooperative behavior in their natural setting. We test the following hypotheses:

H5: A positive relation between the perceived legitimacy of leaders and levels of cooperation exists in the subjects’ natural setting.

H6: The more the experimental conditions reproduce key features of the natural setting, the more cooperative behavior in the experiment will predict cooperation in the farmer association.

Research Site, Sampling, and Experimental Design

Our research design entailed taking a behavioral experiment, typically performed in a laboratory environment, to rural Uganda and conducting our research with members of farmer associations who face collective action problems on a regular basis. Though the move from the lab to the “field” entails some loss of control by the experimenter, conducting the PGGs with members of producer organizations in one of the world’s least developed countries has several notable benefits. First, it extends the scope conditions of public goods experiments to new cultures and areas of the world. PGGs are assumed to capture individuals’ behavior in real-world social dilemmas, yet

the universal applicability of these experiments has been limited by the fact that almost all past studies have relied on students from western universities (Cardenas and Carpenter 2008).¹⁰

Second, subjects in our experiments interacted with subjects who share membership in the same farmer organization. It has been argued that *preexisting groups* are the ideal setting to test theories about the emergence of cooperative behavior (Nowak 2006; Rand et al. 2009). This is because laboratory experiments—which undoubtedly have contributed immensely to the understanding of human behavior—strip context away and are limited in their ability to replicate the mutual trust, past experience, shared norms, and group identity that are central for balancing tension between private and public interests (Baldassarri 2009; Burnham and Johnson 2005; de Rooij, Green, and Gerber 2009; Henrich et al. 2004). In addition, as mentioned, such research design allows us to relate the subjects' cooperative behavior in the experiment to their behavior in the natural setting, in which they face a similar social dilemma.

Research Site

The farmer associations we study were created as part of one of Uganda's largest recent rural development interventions: the Agriculture Productivity Enhancement Project (APEP).¹¹ APEP's goal was to support the integration of smallholder producers into commercial farming. Between 2004 and 2009, it helped organize over 60,000 farmers into about 2,500 village-level groups (known as producer organizations, or POs), which were further organized into more than 200 farmer associations. Serving, on average, 200 members from 10 neighboring POs, the farmer associations (known as depot committees, or DCs) were designed to exploit economies of scale and to bargain for better prices based on quality and volume.

Studying the APEP groups presents many advantages. First, the project's scope and size allow us to conduct a large-scale quantitative study within the boundaries of a single nation, thus securing the homogeneity of the political and legal environments, as well as many project-related factors. Moreover, the process of group formation occurred under the lead of a few project field trainers.

¹⁰Dictator and ultimatum games are widely used in field settings (Henrich et al. 2004). By contrast, PGs in field settings are rare (cf. Fearon, Humphreys, and Weinstein 2009 and Habyarimana et al. 2007).

¹¹APEP was funded by USAID and implemented by Chemonics, a Washington, DC, consultancy.

As a consequence, APEP groups have similar governance structures and leadership positions whose roles and functions are comparable across sites. Each farmer association has an executive committee, comprised of a manager, chairperson, secretary, and treasurer. Operationally, the manager is the leader of the association. His most important responsibilities include organizing crop collection, searching for buyers, and negotiating output and input prices. Additionally, managers are involved in coordinating activities, facilitating information diffusion, and overseeing the sanctioning members who do not follow the association's rules and bylaws.

Farmer associations provide members with several services (e.g., training, input procurement), the most important of which is securing higher output prices through collective marketing.¹² Though highly valuable, collective marketing is subject to a social dilemma.¹³ Because of the high costs of transportation and market information in many developing countries, dispersed smallholder farmers are restricted to selling their crops through local middlemen, who likely exploit asymmetries in information and bargaining power, offering unorganized farmers below-market prices. Organized farmers, by contrast, can obtain higher prices by increasing their bargaining power and by reducing buyers' transaction costs (Staatz 1987).

Once a farmer group is in place, however, middlemen tend to raise prices to remain competitive. Since middlemen, unlike most farmer groups, collect crops at the farmers' gate and pay cash on delivery, members have a private interest in selling to middlemen. The private gain of selling to middlemen ("defecting"), however, is conditional on a sufficient number of *other members* selling their crops via the farmer group ("cooperating"). This is because the price offered by middlemen depends on the price that the farmer group secures ("yardstick effect"), which itself crucially depends on volume. If too many members defect, collective marketing collapses. Some groups manage to overcome this tension between private and group interests, while many others fail.¹⁴

¹²The vast literature on the potential of farmer organizations as engines of growth generally suggests that farmer cooperatives in developing countries can play an important role in poverty alleviation (Narayan-Parker 2002).

¹³APEP was designed to help farmers overcome the collective action problem of creating an organization. It did not, however, eliminate other collective action problems such as the one inherent in collective marketing.

¹⁴Farmer groups that manage to overcome this social dilemma can have a strong, positive, and significant impact on their members' welfare (Grossman and Hanlon 2011).

Sampling and Data Collection

We used a stratified, random, multistage cluster design to select our sample.¹⁵ The use of random samples is not common in behavioral experiments, since their goal is usually to test general causal statements and not to determine the probability that a certain event will occur in a particular population (Berkowitz and Donnerstein 1982). Drawing a representative sample from each sampled farmer association, nonetheless, had two benefits. First, knowing they are interacting with “average” co-members, the experimental subjects were better able to form consistent beliefs about the behaviors of the individuals with whom they were playing (see Habyarimana et al. 2007). Second, it allowed us to make inferences from the behavior of our sample to the groups from which our subjects were drawn (see Levitt and List 2007).

Within each of the 50 sampled associations, different types of data were collected. We interviewed the four DC executives to gather information at the cooperative level. Data on the DCs’ economic activities were also assembled from the associations’ books and records. In each association, we sampled six producer organizations (POs), for a total of 287.¹⁶ An interview with the leaders of the sampled POs allowed us to collect additional data at that level. We also collected individual-level data. From each sampled PO, we further sampled, on average, six members, for a total of 36 members per association. Sampled members were surveyed in person by trained interviewers in the respondents’ language, for a total of 1,781 surveys.¹⁷

Experimental Design

To test Hypotheses 1 through 4, we designed a novel adaptation of the PGG. In each round of play, subjects received an endowment of 10 coins of 100 USH—10 monetary units (MUs)—which is the equivalent of about half a daily wage in rural Uganda. Subjects then had to decide, anonymously, how to split this endowment between a private and a public account. What subjects put in the private account remained theirs, while what was contributed to

the public account was doubled and redistributed evenly among all group members.

Our experiment consists of three different variants of the PGG: baseline, random monitor, and elected monitor. Subjects assigned to the baseline participated in six rounds of a PGG *without* sanctioning. In the two monitoring treatments, we introduced a centralized-sanctioning institution. Differently from peer-sanctioning settings, in which subjects may punish each other, we gave sanctioning powers to a single authority. Namely, after two preliminary rounds, one of the subjects was assigned the role of a monitor. Monitors received the same endowment as the other subjects, but could not contribute to the PGG, nor receive part of the public account. Instead, monitors were able to spend 1 MU to take away 3 MUs from subjects whose contribution level they disapproved.¹⁸ Monitors’ payoff did not depend on the group’s level of cooperation, but only on their sanctioning decisions. A monitor’s payoff in round t is, therefore, 10 (MUs) minus the number of subjects sanctioned in that round. Subjects’ payoff is calculated as $\pi_{it} = (10 - x_{it}) + \frac{2 \sum x_{it}}{n} - P_{it} \times 3$, where $x_{it} \in \{0, 1, \dots, 10\}$ is the contribution to the public account, and $P \in \{0, 1\}$ indicates whether player i was sanctioned at round t .¹⁹

The two sanctioning treatments differed only in the procedure for selecting the monitor. In the random-monitor treatment (T_r), the monitor was selected through a *lottery*. Comparing the baseline with the random-monitor condition allows for an assessment of the causal effect of instituting a centralized-sanctioning regime. In the elected-monitor treatment (T_e), participants *elected* their monitor using a secret ballot.²⁰ A comparison of the random and elected-monitor treatments

¹⁸This study follows the convention in PGGs, according to which the threat of punishment is credible and substantial: sanctioned subjects cannot refrain from paying their fine.

¹⁹To ensure the credibility of the leadership-selection process, as well as the practical execution of the PGG in a field setting, the experiment took place in a single room, where participants would decide their contributions behind screens. In this setting, the physical appearance of the monitor, as well as that of the other participants, was known. We designed the game and tested it extensively to make sure that the lack of complete anonymity would not induce any kind of hostility or retaliation among the participants, especially toward the monitor. We also conceived several features in the design of the PGG to make sure that participants would not interpret the action of the monitor as directly oriented at damaging their own welfare. In particular, the monitor did not sanction individuals directly, but simply levels of contribution. Moreover, the actual monetary remuneration occurred at the end of the day, after participants had played several other games, and the payoff of this game was a small share of their total gains.

²⁰The voting procedure guaranteed anonymity: each player wrote on a piece of paper the ID number of the player she would like to serve as a monitor. Subjects could see each other but were not

¹⁵A detailed description of the sampling scheme, including a map showing the farmer groups’ location, can be found online in the Supporting Information (SI) appendix.

¹⁶When a farmer association had fewer than seven POs, we selected all its village-level groups.

¹⁷Only 1,541 of the 1,781 sampled members participated in the PGGs. This gap is due to the fact that the experiments were conducted, in each DC, in a single day in a central location, while to reduce attrition, interviewers returned to sampled villages several times to locate members who did not show up on the data-collection day.

TABLE 1 Experimental Design

Game Variant	Centralized Sanctioning	Elections	Sessions	Avg. Subjects per Session	Rounds per Session
Baseline (BL)			50	10	6
Random Monitor (T_r)	X		48	10	2 prelim + 4
Elected Monitor (T_e)	X	X	49	10	2 prelim + 4

allows an estimate of the independent effect of the process by which the monitor has obtained her sanctioning powers. In each farmer association, 30 members, on average, took part in the experiment. Each of those participants was randomly assigned to only one of the three game variants, though all three variants were played in each of the 50 farmer groups.²¹ Table 1 summarizes the experimental design.²²

Stages of the PGG

- **Stage 1: Contribution to the PG.** Players decide *anonymously* how to divide 10 MUs between a private and a public account. To ensure anonymity, players make their allocation decisions behind three-sided cardboard screens.
- **Stage 2: Contributions become common knowledge.** Research assistants (RAs) display publicly *all* the contributions to the public account (in USH), from the lowest to the highest. Players are unable, however, to match between contributions and players' identity.
- **Stage 3: Payoffs.** RAs calculate *publicly* the mean contribution to the public account and the size of the social return. Using this information, RAs display the payoff (private + social returns) for each contribution level displayed on the public board.

allowed to talk and were not given any information about other subjects. Subjects were sampled from six different villages and, therefore, did not know, on average, more than one or two other participants (covillagers). They knew, however, that all session subjects shared membership in their farmer cooperative. Importantly, none of the subjects held a leadership position in the cooperative.

²¹The number of subjects per session ranged from 8 to 12. This is higher than in most PGGs (2 to 4), and was required in order to protect our subjects' anonymity. In two occasions the number of sampled members was too small to conduct all three variants due to a funeral that took place on the data-collection day.

²²For additional information on the experimental design, including scripts, we encourage readers to consult the online supplementary materials.

Variants

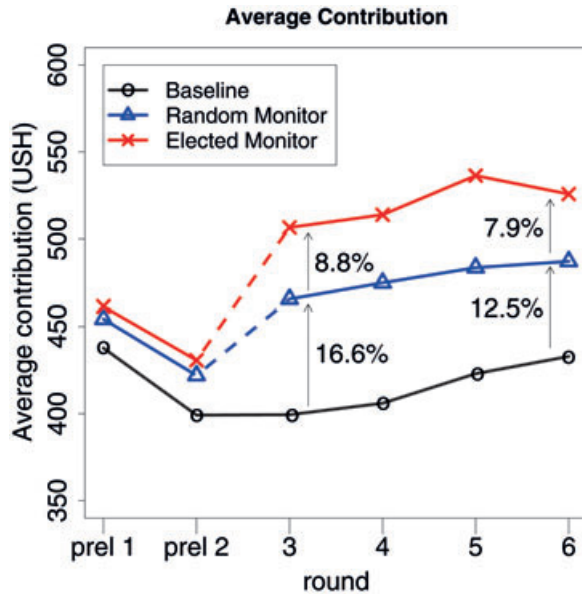
- In BL , stages 1–3 are repeated for six rounds.
- In T_r and T_e , stages 1–3 are repeated for two rounds, then, at the end of preliminary round 2, monitors are selected/elected. In the subsequent four rounds (rounds 3 to 6), subjects repeat stages 1–5.

- **Stage 4: Punishment decisions.** Monitors, standing in front of the participants, point to the contribution(s) to which they want to assign “reduction points.” Monitors *do not identify the individual players* who are sanctioned, but only the level of contribution. Similarly, players know what levels of contribution are punished, but cannot match sanctions to players' identity.
- **Stage 5: Payoffs recalculated.** Following the monitors' sanctioning decision, RAs reduce the payoffs of sanctioned contributions by 3 MUs (300 USH). Play repeats for four rounds under a sanctioning regime.

Experimental Findings

A descriptive summary of our two major experimental findings is offered in Figure 1, in which we report the trend in the average contribution to the public good for each of the three variants. Consistent with previous findings, in the first two preliminary rounds, subjects contributed between 40 and 45% of their endowment, and there were no differences between variants in the preliminary rounds, as one would expect, since all subjects participated in a PGG without sanctioning. In contrast, significant differences in contributions between BL and both T_r and T_e are observed as soon as the threat of punishment is introduced, even *before observing monitors' behavior*. In round 3, subjects in T_r contributed to the public account 16.6% ($P = 0.000$) and in T_e 24.4% more ($P = 0.000$) than subjects in BL , suggesting that participants acted under the expectation that monitors would punish defectors. Similar differences are observed in subsequent rounds, confirming

FIGURE 1 Average Contribution to the Public Good by Treatment



Note: For rounds 3 and 6, the graph reports the percentage increase in contributions comparing random with baseline condition, and elected with random. Reproduced from Baldassarri and Grossman (2011).

the effectiveness of centralized sanctioning. Secondly, subjects in T_e contributed to the public account, on average, 9% more than subjects in T_r ($P = 0.002$), thus providing evidence that elections have a positive impact on contribution levels, above and beyond the mere threat of punishment.²³

We test the statistical significance of our results estimating various specifications of the following three-level random-intercept model.

$$y_{ijt} = \alpha + \delta T + \gamma t + \zeta_j^{(3)} + \zeta_{ij}^{(2)} + \varepsilon_{ijt} \quad (1)$$

where y_{ijt} is the public goods contribution of subject i from the farmer cooperative j at round t ; $T \in \{0, 1\}$ is a treatment assignment indicator, such that δ is the average treatment effect of interest; $T \in \{3, \dots, 6\}$ indicates the round of play; $\zeta_j^{(3)}$ is the random intercept for farmer cooperative j and $\zeta_{ij}^{(2)}$ is the random intercept for individual i nested within farmer cooperative j ; and, finally, ε_{ijt} is the residual error term. We make the following assumptions regarding the random intercepts and the residual

error term:

$$\begin{aligned} \zeta_j^{(3)} | T, t &\sim N(0, \psi^{(3)}), \zeta_{ij}^{(2)} | T, t, \zeta_j^{(3)} \sim N(0, \psi^{(2)}), \\ \varepsilon_{ijt} | T, t, \zeta_j^{(3)}, \zeta_{ij}^{(2)} &\sim N(0, \theta). \end{aligned}$$

The random intercepts and residual error are also assumed to be mutually independent; however, given the panel nature of the data, we further assume that the residual errors have an autoregressive structure of order $p = 2$. To increase precision we also estimate the following model:

$$y_{ijt} = \alpha + \delta T + \gamma t + \beta X_{ij} + \zeta_j^{(3)} + \zeta_{ij}^{(2)} + \varepsilon_{ijt} \quad (2)$$

where X_{ij} is an $n \times K$ matrix of individual covariates and β is a vector of coefficients. Table A1 in the appendix reports the main results of the experiment. First, corroborating Hypothesis 2, in the presence of a centralized-sanctioning authority, subjects significantly increased their contribution to the public good. The average treatment effect (ATE) for T_r compared to BL is .5 MU, and the ATE for T_e compared to BL is almost 1 MU.²⁴

Second, confirming Hypothesis 3, the political process through which monitors obtain their sanctioning power is consequential. The ATE for T_e compared to T_r is .4 MU.²⁵ Having provided strong evidence that elections positively influence contribution levels, we now turn to explore the mechanisms that might account for such an effect.

According to Hypothesis 4, we expect greater contributions in the elected monitor to be due to a “legitimacy effect”: subjects should express greater deference to monitors whose legitimacy has been certified through elections. There are two instances in our research design where people are met with the opportunity to change their behavior in compliance with monitors’ authority: (1) when they enter a sanctioning environment in round 3, in *anticipation* of possible sanctioning and (2) when they are sanctioned in any subsequent round. In both cases, participants’ adaptive behavior is a function of the threat of punishment, which changes the structure

²⁴This finding is equivalent to other lab-in-the-field PGGs, where peer sanctioning increased cooperation relative to the baseline (see Barr 2001 and Carpenter 2004). The modest decline in baseline is consistent with findings from PGG studies in nonwestern settings. Whereas cooperation declines significantly with college-aged participants in the United States, cooperation rates remain higher and are sustained longer with African and Asian subjects (Cardenas and Carpenter 2008).

²⁵This result is consistent with Fearon, Humphreys, and Weinstein (2009), who find an increase in contribution in a PGG in Liberian communities that were exposed to the political and social participatory components of Community Driven Reconstruction (CDR) programs.

²³A brief summary of these experimental results has appeared in Baldassarri and Grossman (2011).

TABLE 2 DIF in DIF: Anticipation of Monitors' Behavior

	(A)	(B)	(C)	(D)
Round 3	36.49*	38.83*	34.77*	36.34*
	(12.03)	(12.48)	(12.69)	(12.90)
Elected monitor	-5.14	0.82	23.35	26.01
	(30.32)	(28.76)	(26.84)	(26.43)
Round 3 × Elected monitor	46.78*	44.36*	46.99*	46.61*
	(17.68)	(18.04)	(18.71)	(19.22)
Individual controls		X		X
Monitor profile			X	X
Intercept	428.49*	461.22*	338.59*	368.23*
	(17.49)	(50.41)	(125.33)	(123.88)
Observations	1951	1862	1858	1782

Standard errors clustered at the farmer association level in parentheses.* $p > 0.05$.

Note: Difference-in-difference estimation of the change in contribution from preliminary round 2 to round 3—the first round under a sanctioning regime, but *before* the behavior of the monitor has been revealed. Round 3 is binary: it equals zero to indicate round 2 and equals one to indicate round 3. Elected monitor is also binary: it equals zero to indicate random monitor condition and equals one to indicate elected monitor condition. The coefficient on the interaction (Round 3 × Elected monitor) is the difference-in-difference estimator. The t-statistic on the DIF regression coefficient is the t-test for equality of the differences.

of monetary incentives in the same way in both T_r and T_e , and a function of legitimacy, which we expect to be greater in T_e . In measurement terms, the effect of legitimacy is the change in behavior that cannot be reduced to the threat of punishment. If legitimacy has no effect, we would simply see similar patterns of adaptive behavior in both T_r and T_e .

First, we consider subjects' anticipation by looking at changes in contributions from the second preliminary round to round 3, before monitors' decisions over sanctioning take place. While in the preliminary rounds average contributions to the public account in T_r and T_e are similar, in round 3 subjects in T_e contributed significantly more (+8.8%) than subjects in T_r . Difference-in-difference regression models confirm the significance of these results (Table 2). Whether controlling for individual covariates (Model B), monitors' profiles (Model C), or both (Model D), the *change* in contribution in round 3 under an elected monitor is between two to three times higher than the change in behavior under a random monitor.

Second, we consider players' reactions to punishment as further evidence of the greater deference to elected monitors. Parameter estimates come from a three-level random-intercept model as in equation (2), only here the

dependent variable is the *change* in player i 's contribution from $t-1$ to t as a function of whether player i has been sanctioned at $t-1$, the type of monitor, and the interaction between those variables. As shown in the Appendix, Table A2, having been punished at round $t-1$ increases subjects' contribution at round t by 20–24 USH under a random monitor but *two to three times higher* under an elected monitor.

Having found evidence consistent with a “legitimacy effect,” we rule out the possibility of a confounding impact due to leadership selection. Namely, we test whether higher levels of cooperation are due to the characteristics of the elected monitors. We find, on one hand, that subjects elected monitors with socially dominant profiles—elected monitors were more likely to be male, wealthier, more educated, and more likely to have been born locally, compared to the pool of eligible monitors (online SI, Figure 1). On the other hand, knowing a monitor's profile—his or her gender, education, age, wealth, place of birth, and religiosity—does not improve our capacity to predict subjects' contributions. When regressing subjects' contribution on monitors' profile in T_r , the characteristics of monitors, whether tested separately or jointly, do not have a significant effect on contributions (online SI, Table 2). In addition, the sociodemographic profile of monitors does not affect subjects' change in contributions from preliminary round 2 to round 3 (SI, Table 2), nor their reaction to sanctioning (Appendix, Table A2).

These findings, cumulatively, weaken the possibility of a leadership-selection effect.²⁶ It is possible, however, that elected monitors have certain attributes that induce cooperation, which are unobserved to the research team but visible to the experimental subjects. Though the possibility of unobserved heterogeneity cannot be ruled out, it does not seem to play a decisive role in this experiment. For one, elected monitors are not more *public-spirited* than random monitors, at least as this is reflected in their contributions to the public good in the preliminary rounds.²⁷ Second, as we demonstrate below, we do not find much evidence suggesting that elected monitors were enforcing cooperation more ardently than random monitors. Third, elected monitors' *religiosity* is not higher than random monitors, at least as this is reflected in

²⁶These results do not question the role that leaders' qualities play in solving collective action problems (Grossman and Hanlon 2011). Rather, they confirm that leaders' selection effects are not likely to play a role in the context of this experiment.

²⁷In preliminary round 1, random monitors contributed to the public account, on average, 5.61 MUs, whereas elected monitors contributed 5.02. In preliminary round 2, random monitors contributed 4.98 MUs and elected monitors 3.54. Contributions were made anonymously, before monitors were selected.

church attendance.²⁸ Finally, recall that our subjects were drawn from six neighboring villages, none held leadership positions in the farmer association, and they were not allowed to talk throughout the entire course of the experiment. These design features further reduce the possibility that participants had private knowledge of how well other subjects would perform as monitors. In the next section, we analyze monitors' sanctioning behavior and test whether the different criteria for monitor selection may have affected monitors' sanctioning strategies, and thus, indirectly, cooperation.

Monitors' Sanctioning Behavior

In our experimental setting, monitors did not gain any monetary benefit from higher levels of contribution to the public account, and had to sacrifice part of their endowment to sanction members. Nonetheless, and in accordance with Hypothesis 1, both elected and random monitors sanctioned "defectors": from about an average of 2.5 subjects sanctioned in round 3 to 1.5 subjects in round 6. We therefore conclude that, at least in a situation in which reputation is at stake (the identity of the monitors was known to all participants), internal centralized authorities will punish defectors at a personal cost.

Next, we consider whether monitors in T_r and T_e follow different sanctioning strategies. Though elected and random monitors sanction, on average, the same number of subjects per round, they vary in the maximum contribution for which subjects are punished. In round 3, monitors in T_r and T_e sanctioned similar levels of contribution. In subsequent rounds, the average maximum contribution sanctioned by random monitors gradually declined, while it increased for elected monitors. By round 6, subjects in T_r who contributed more than 25% of their endowment were not punished, while subjects in T_e were sanctioned for contributing up to 37% of their endowment ($P = 0.022$).

Comparing the behavior of monitors in T_r and T_e is, however, complicated by the fact that the distributions of contributions faced by elected monitors are different from those faced by random monitors. In fact, the PGG was designed to study subjects' levels of cooperation under different monitoring treatments, whereas the distribution of contributions was generated endogenously and was not experimentally controlled. PGGs studies that analyze punishment behavior have generally dealt with this problem by ignoring it. Yet, since the contribution

levels that monitors face are *not* under the control of the experimenter, they should be treated as observational data. Accordingly, we used the Kullback–Leibler (K–L) divergence measure (Kullback and Leibler 1951) to match the distribution of contributions that an elected monitor faced with the closest distribution of contributions that a random monitor faced.²⁹ Using the matched pairs, we were then able to assess the extent to which the behavior of elected and random monitors differs.³⁰

To assess whether elected monitors enforce stronger norms of cooperation, we rely on two measures: (1) the number of subjects and (2) the maximum contribution sanctioned by the monitors. Plots in the top row of Figure 2 report, for each round t , the number of subjects punished by elected monitors (y-axis), as a function of the number of subjects punished by random monitors (x-axis), for all matched pairs. Dots that are above (below) the 45° line are matched pairs in which the elected monitor sanctioned more (less) subjects than the matched random monitor. Similarly, plots in the bottom row report the maximum contribution sanctioned in round t for matched pairs of monitors. In both cases, the number of dots above the 45° line is roughly similar to the number of dots below the line, suggesting that there is no difference between random and elected monitors with respect to the number of players and maximum contribution sanctioned. This finding is confirmed by Wilcoxon matched-pairs signed-rank tests (p-values reported in Figure 2).

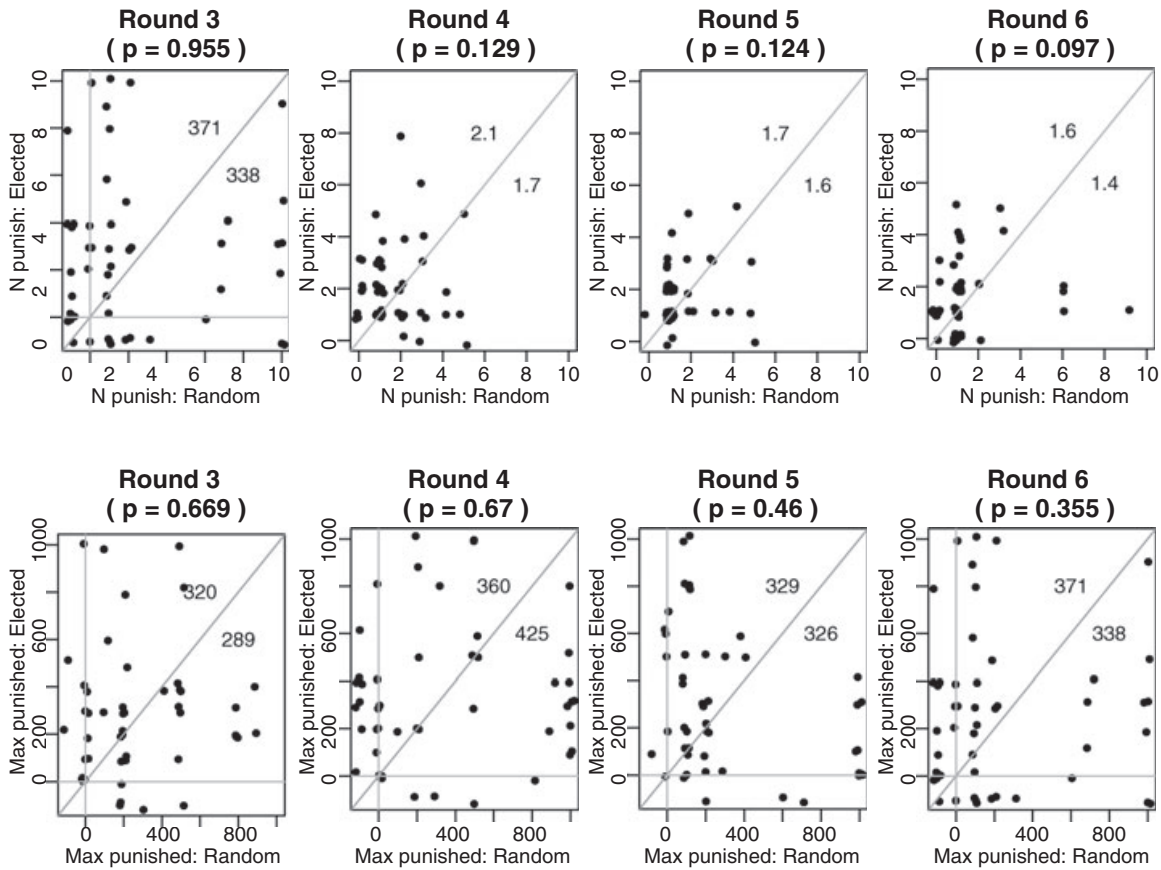
In conclusion, when facing similar distributions, elected and random monitors adopt similar sanctioning strategies. We find no evidence that elected monitors enforced norms of cooperation more forcefully. In addition, we find that for both monitors, the frequency of punishment is not related to a group's average contribution (online SI, Figure 2). Taken together, these findings are consistent with a model in which monitors do not punish according to some predefined acceptable level of contribution. Instead, both elected and random monitors consider subjects' contribution relative to the contribution of others and follow a heuristic strategy of sanctioning a few contributors at the bottom of the distribution. Our analysis of monitors' sanctioning strategy strengthens our confidence that leader-selection effects and unobserved heterogeneity are not confounding the effect of legitimacy. Finally, considering that subjects' contributions are not

²⁹For two probability distributions P and Q of a discrete random variable, the K–L divergence is defined as $D_{KL} = (P||Q) = \sum P(i) \log \frac{P(i)}{Q(i)}$.

³⁰Within blocks defined by rounds, we used a nearest neighbor with replacement matching algorithm. Consult the online SI for additional information.

²⁸In Uganda, religiosity is considered an important attribute for those seeking public office.

FIGURE 2 Comparison of the Sanctioning Behavior of Matched Pairs of Monitors



Note: Plots in the first row show the number of players sanctioned in round t by elected monitors as a function of the number of players sanctioned by the random monitors, for all matched pairs. Plots in the second row show maximum contribution sanctioned in round t by elected monitors, as a function of the maximum contribution sanctioned by the random monitors, for all matched pairs. Note that negative values on the x- and y-axis refer to sessions in which the monitor did not sanction any contribution. Each graph reports p-values from Wilcoxon matched-pairs signed-rank tests. Dots are jittered.

affected by the number of other subjects or the maximum amount sanctioned (Appendix, Table A2), we conclude that monitors' major impact on cooperation has more to do with the anticipation and response to the sanctioning that they elicit than with their actual sanctioning strategy.

Discussion of the Experimental Part

Given that both types of monitors use a similar sanctioning strategy, and given that subjects are not swayed by elected monitors' dominant social profile, why would rational subjects contribute more toward the public goods under an elected monitor compared to subjects under a random monitor? One plausible explanation is that subjects simply miscalculated; i.e., they had false expectations regarding monitors' relative intensity of punish-

ment strategy. However, if this explanation were true, we would have seen contributions under random and elected monitors converge over time once subjects updated their beliefs following monitors' revealed behavior. Moreover, "miscalculation" does not explain why subjects who were punished by elected monitors increase their contributions to the public account in the subsequent round more than twice as much as subjects punished by random monitors. Instead, we argue that legitimacy is a mediator in the relationship between elections and subjects' cooperative behavior. If true, how then does legitimacy operate?

We have noted that the concept of legitimacy captures internal psychological dispositions, beliefs, and attitudes that increase the willingness to obey an authority (Levi, Sacks, and Tyler 2009). Since we do not have direct measures of these dispositions, but only observe their behavioral manifestations in the form of

compliance (referred to by Levi et al. as “behavioral legitimacy”), we can only sketch here some preliminary propositions. One plausible explanation builds on recent work by Hopfensitz and Reuben (2009) on response to punishment in PGGs. According to Hopfensitz and Reuben, for sanctions to be effective they must trigger the negative prosocial emotions of guilt or shame.

If legitimacy denotes people’s *sense of moral obligation* to follow the authority, and if moral standards denote “internalized rules of behavior from which deviations are psychologically costly” (Greif and Tadelis 2010), it follows that deviations from group rules may carry a *psychological cost* that is a function of the legitimacy of the authority. In other words, to be punished by a leader selected by the group is more shameful and/or triggers a stronger sense of guilt. Assuming that punishment by a legitimate authority carries a higher psychological cost fully accounts for the difference in subjects’ anticipation and response to sanctioning that we observed between T_r and T_e .³¹ Further research should explore more deeply the emotions and dispositions triggered by authorities that are endowed with sanctioning powers via different selection methods or via different sources of legitimacy, considering, for example, more traditional forms of authority.

In recent years, researchers have used PGGs to demonstrate that allowing subjects to punish their peers increases the contribution to public goods production. Our lab-in-the-field experiment builds on that intuition, but it expands the literature in two directions. First, consistent with Hypothesis 1, both elected and random monitors punished defectors. Second, consistent with Hypothesis 2, the experiment demonstrates that internal centralized-sanctioning authorities punish sufficiently to have a net positive impact on cooperation. Third, the experiment demonstrates that beyond the threat of punishment, cooperation is also conditional on the political process by which leaders acquire their authority (Hypothesis 3). Subjects were more responsive to a leader when given the possibility to participate in her selection.

Observational Data

The experimental findings point to the relevance of the political *process or procedure* through which leadership is

³¹At no point do we claim that the legitimacy of elected leaders depends necessarily on the fairness of the selection process. Lottery is indeed a fair selection method, as any experimentalist would attest. Recall that ancient Athens filled seats on its legislative council by drawing lots from among its citizens. See Manin (1997) for an excellent discussion.

selected as the basis for the legitimacy of internal centralized authorities. We further demonstrated that this sort of procedural legitimacy likely mediates the causal relation between leader-selection processes and members’ public goods contribution.³² In this section, we analyze observational data gathered specifically to attest to the ecological validity of the experimental findings: i.e., the extent to which the experiment captures features that are relevant to our subjects’ cooperative behavior in their *natural setting*. Our analysis proceeds in two steps. First, we test whether there exists a positive association between the perceived legitimacy of the managers of the farmer cooperatives and group members’ level of cooperation. Secondly, we test whether behavior in the controlled environment of the experiment can predict cooperative behavior in the farmer association. If so, this should increase our confidence that the experimental conditions reproduce key features of the subjects’ natural setting.

In our attempt at relating experimental findings to the ordinary life of farmer cooperatives, we focus our empirical analysis on a set of measures of procedural legitimacy. In general, measuring legitimacy is a difficult task. Following previous scholarship, we rely on a variety of survey questions that capture different aspects that might affect the respondent’s perception of the legitimacy of the group leader. We consider a total of eight proxy measures of legitimacy. Two are attitudinal and rather abstract: they are members’ assessment of (a) whether the DC manager is monitored and (b) whether the DC leadership is accountable. Two measures relate to the respondent’s level of information, namely (c) whether the member is able to name the DC manager or chairperson and (d) whether the member knows the method for electing the DC manager. Three measures are behavioral, namely (e) how often a member receives receipts when selling through the association;³³ (f) whether the respondent attended the last general assembly meeting; and (g) whether the respondent voted in the last farmer group elections. Finally, using principal component analysis, we also combined the above responses into (h) an index of “perceived legitimacy.”

Our key outcome variable is members’ level of participation in public goods production. Collective marketing is the central activity of farmer associations, and, as

³²See Rothstein (2009) for a fruitful discussion of the distinction between procedural and performance-based legitimacy.

³³In countries such as Uganda, credit-constrained farmer associations are usually unable to pay members cash-on-delivery. Instead, members receive payments only about three to five weeks after they had delivered their crops to their DC. Members, therefore, have a clear interest in receiving a receipt, specifying the date and volume delivered, since it can reduce the likelihood of being exploited.

explained above, it is a major collective action problem that farmer groups must overcome. We therefore measure cooperation by looking at the marketing decisions of group members. A high level of cooperation exists when members sell a large fraction of their crops via their farmer group.³⁴ We use two self-reported measures of members' marketing decisions to measure cooperation: (1) a binary indicator of whether a member sold his coffee via the association, at least once, in the past season; and (2) the proportion of a member's total seasonal coffee yield that was sold via the farmer group in the past season. The results of our analyses are similar using either measure of cooperation. Here, we report results using the continuous measure of cooperation, while results from the binary measure can be found in the online SI.

According to Hypothesis 4, we expect farmer associations in which the manager's perceived legitimacy is higher also to have higher levels of collective marketing. We test this hypothesis by running a set of multilevel random-intercept models, in which a member's marketing decision is modeled as a function of one of the eight measures of procedural legitimacy listed before. We use a regression equation of the form:

$$y_{ij} = \beta_0 + \beta_1 L_{ij} + X_{ij}\Gamma^{(1)} + F_j\Gamma^{(2)} + \zeta_j^{(2)} + \varepsilon_{ij} \quad (3)$$

where the dependent variable y_{ij} is the proportion of a member's total seasonal yield that was sold via the farmer group in the past season. Our main variable of interest is L_{ij} , which measures individual i 's perception of the manager's legitimacy; X_{ij} is a vector of individual-level controls, such as sex, age, education, log seasonal yield, years since joining the farmer groups, the rating of the local middleman's honesty, richness of associational life, and church attendance. F_j is a vector of group-level controls: the association's age, its membership size, manager's effort level, mean seasonal yield, and mean distance to the nearest trading center; $\zeta_j^{(2)}$ is the random intercept for farmer cooperative j ; and ε_{ij} is the residual error term.³⁵ To make the interpretations of coefficient easier, we follow Gelman's (2008) advice to center binary independent variables and rescale all other variables by subtracting their mean and dividing by two standard deviations. Regression results are reported in Table A3 (appendix).

³⁴This measure is better than alternative public goods measures such as price. First, it directly measures members' cooperative behavior. Second, price may be affected by a host of factors outside the control of the association.

³⁵We also tested model specifications that nested farmer cooperatives in strata and that allowed the legitimacy variables to vary across groups. A likelihood ratio test rejected those specifications in favor of the more parsimonious two-level random-intercept model.

Notably, a positive association between the leader's perceived legitimacy and members' participation in collective marketing is found in all eight proxy measures of legitimacy. For example, controlling for individual and for group-level variables, participation in group activities—e.g., attending the last general assembly meeting and voting in the last farmer group elections—is associated with more than a 10% increase in the share of a member's seasonal yield sold via her farmer association. In sum, across a wide range of legitimacy proxy measures, the change in the estimated probability of cooperation is positive, substantial, and significant by conventional standards.

Comparing Behavior in the Experiment and Real Life

In the last part of the analysis, we turn to compare the subjects' cooperative behavior in the controlled experiment with their behavior in their natural environment, as farmer group members. To our knowledge, this is the first study to make such comparison in the context of PGGs.³⁶ In particular, we expect that cooperative behavior in the elected-monitor treatment, which we interpret as capturing individuals' deference to legitimate authority, will predict behavior in the farmers' natural setting, where we have shown a positive relation between procedural legitimacy and cooperation. While this would not be a direct proof of the causal impact of a leader's legitimacy on cooperation in the natural setting, a positive correlation can be considered as an indirect validation of such a causal statement. In other words, what works under the "petri dish" of a controlled experiment might be at work in the natural setting as well.

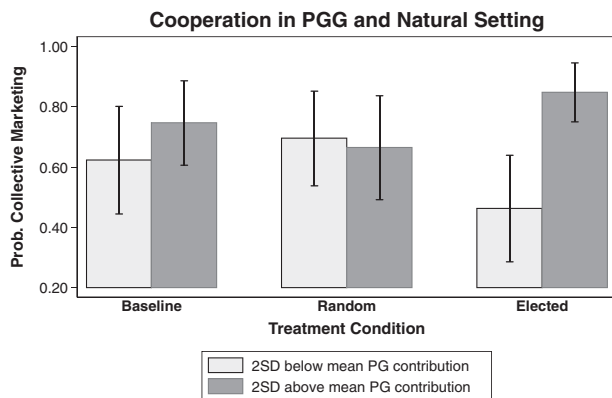
To test whether cooperative behavior in the controlled setting predicts behavior in the natural setting, we run the following multilevel random-intercept logistic regression for each of the three treatment conditions:

$$Pr(Y_{ij} = 1) = \beta_0 + \beta_1 \bar{C}_{ij} + X_{ij}\Gamma^{(1)} + \zeta_j^{(2)} + \varepsilon_{ij} \quad (4)$$

where Y_{ij} is an indicator of whether respondent i from group j contributed to the group public goods by selling her crops via the association, at least once, in the past season. Our key independent variable is \bar{C}_{ij} , the respondent's mean contribution to the public account in the PGG; X_{ij} is a vector of individual-level controls: contribution in the preliminary round, sex, age, education, total seasonal yield, years since joining the farmer group,

³⁶Laury and Taylor (2008) compare behavior in a PGG to subjects' willingness to donate from their experiment's earnings to a local NGO. Since such donation is not part of the subjects' natural activity, their results should be treated with some care.

FIGURE 3 Relation between Cooperative Behavior in the PGG and as Farmer Cooperatives Members



Note: Caps represent 95% confidence intervals.

richness of associational life, church attendance, honesty of the local middleman, and whether the respondent was born locally. $\zeta_j^{(2)}$ is the random intercept for farmer cooperative j , and \mathcal{E}_{ij} is the residual error term.³⁷ Here, too, binary independent variables were centered, and all other variables were rescaled by subtracting their mean and dividing by two standard deviations.

Results are presented in Figure 3 and in the online SI, Table 7.³⁸ For each treatment, we graph the predicted probability of selling via the farmer group ($Y_{ij} = 1$) for a farmer whose mean contribution in the PGG was two standard deviations below the grand mean (“defector”) and for a farmer whose mean contribution in the PGG was two standard deviations above the grand mean (“cooperator”), holding control variables at their mean. We find that in the *elected-monitor* condition, cooperative behavior in the PGG is positively and significantly related to behavior in the farmer group: whereas the predicted probability of a defector in the PGG to sell her crop via the farmer group is 46%, it is 85% for a PGG cooperator.

We interpret these findings to reflect the fact that the elected-monitor treatment approximates the process of legitimation of centralized authorities that occurs in farmer groups. Farmers who elected their monitor were able to draw on their past experiences and group norms.

³⁷The proportion of cooperators in the natural setting was equally balanced across the three treatments. See Table 11 in the online Supporting Information.

³⁸The results we present are robust to other specifications: e.g., running the model with and without controls; using other specifications of cooperation in the natural setting, such as share of total yield sold via DC, etc.

As a consequence, their game behavior reflects real-life patterns of behavior in a manner that is not visible in the other treatment conditions.

That people bring their experience and social norms into a laboratory environment has been demonstrated in past studies. For example, Henrich et al. (2004) report a large variation in cooperative behavior across communities, differences that they attribute to interactional patterns of everyday life and the social norms operating in those communities. Similarly, Goette, Huffman, and Meier (2006) show that individuals’ willingness to cooperate and enforce norms, in an experiment, is a function of their “natural” group affiliations. It is important to recall that such values, norms, and experiences are brought into the lab by the subjects and are not experimentally induced. As such, they constitute key information that people use in their decision making (Levitt and List 2007).

Our findings suggest that the legitimacy of internal centralized authorities is an important factor in determining the effectiveness of farmer groups, and likely of other similar community organizations. These findings underscore the merit of behavioral experiments, insofar as results suggest that the PGGs used in the current study were able to identify causal mechanisms that are likely to be at work in a meaningful real-world setting. The correlation between the behavior in the controlled setting (when exposed to the legitimacy stimulus/cue) and in the natural environment (where the stimulus is experienced and reported by the subjects) suggests that there is some continuity between the two contexts. This, indirectly, provides support to the idea that the mechanism we tested in the lab might be at work in the real world.

Conclusion

This study makes contributions to several distinct literatures. In experimental and formal research, peer punishment has been largely considered as the only alternative to the coercive power of an external agency (Ostrom, Walker, and Gardner 1992). However, complex societies are not sustainable on the basis of peer punishment alone. Using a novel modification of the PGG, this article incorporates the idea of internal centralized authorities into theories of public goods provision. In doing so, it contributes to the study of the role sanctioning plays in inducing cooperation in social dilemmas.

Centralized-sanctioning systems, however, cannot rely exclusively on coercive punishment. Indeed, institutions that are perceived as legitimate only rarely turn to brute force to enforce group norms. To minimize the use of brute force, which is costly, centralized authorities need

to be recognized as legitimate by the ruled (Ahlquist and Levi 2011). We conceive of legitimacy as the dispositions and attitudes that increase subjects' willingness to obey the authority of a centralized power, and we measure its manifestation as the extent to which subjects comply with its directives. The fact that sanctioned subjects who elected their monitor increased their contribution in subsequent rounds by more than double, compared to sanctioned subjects who faced an arbitrary authority, is indicative of the role legitimacy plays in sustaining cooperation. Demonstrating how legitimacy mediates the relationship between political processes and the effectiveness of leaders is, therefore, our second contribution.

Our article also contributes to the "new-institutionalism" literature, which places incentives at the center of the theory's microfoundation. We complement and enrich this framework by showing that the process by which institutions are put into place has an independent effect on an individual's behavior, in ways that *cannot be easily reduced to monetary incentives*. We have shown that beyond the threat of punishment, the political process through which centralized authorities acquire their sanctioning powers is consequential. When regulatory and sanctioning powers are granted through elections, leaders' directives are more likely to be followed, and their sanctioning decisions are more likely to be effective in inducing cooperation. These findings are consistent with recent evidence of a positive impact that democratic elections have on public goods provision at the village level in China (Martinez-Bravo et al. 2010). As such, our findings have important implications for our understanding of the development of local forms of democratic rule. Our study, therefore, calls for a closer examination of the independent role of leaders and their selection rules.³⁹

Our article also contributes to the study of legitimacy, which is a central concept in normative theory and applied research (Hechter 2009). Until the late 1980s, hardly any attention was paid to empirical investigation of legitimacy (McEwen and Maiman 1986). In recent years, however, scholars have increasingly sought to document the political and social effects of legitimacy. For example, legitimacy is thought to play a central role in compliance with health regulations during an epidemic (Lieberman 2007), citizens' support of war efforts (Levi 1997), and democratic transitions (Linz and Stepan 1996). Similarly, Gibson, Caldeira, and Spence (2005), who focus on courts, and Murphy (2005), who focuses on tax agencies, find that political institutions can gain acceptance for unpopular decisions when legitimate. Past

research focused almost exclusively on whether legitimacy encourages deference to laws and norms. This study expands this literature by examining the role legitimacy plays in securing cooperation in social dilemmas.

Finally, our article makes an important contribution to the study of political participation. We began by demonstrating experimentally something quite intuitive—that elections increase the value of a local public good. But as we began ruling out options commonly associated with why elections are deemed beneficial, we were left with an important finding. Elections increased the value of local public goods even after we eliminate incumbents' reelection considerations, and even when we minimize the information voters have on potential candidates, reducing their ability to select more able and more responsive leaders. We found evidence suggesting that something fundamental causes us to be more prosocial when we participate in key political process such as elections. That elections affect not only the behavior of incumbents but also the behavior of constituents who had participated in the electoral process is among the key findings of our study.

Turning our attention to the role of internal centralized authorities opens up a new set of questions for future research. We have found that in the presence of a centralized-sanctioning authority, groups can reach higher levels of cooperation and that monitors, at least in a situation in which their reputation is at stake, are willing to bear the cost of punishing in order to increase cooperation.⁴⁰ These results are qualitatively similar to those obtained using peer-sanctioning institutions, with the possible advantage that a centralized system of monitoring will be more efficient than a decentralized one. Future studies should investigate the relative efficiency of decentralized (i.e., peer) versus centralized-sanctioning regimes. In addition, more work is needed in order to understand the motivation behind the observed behavior of both regular subjects and monitors and to explore the relative effectiveness of different legitimation processes (e.g., democratic legitimacy versus more traditional forms of authority).

The study offers three core findings: (1) in an experimental setting, political process had a causal impact on individuals' cooperative behavior: participants contributed more to a public good when their monitor obtained her sanctioning powers through elections; (2) the relation between legitimate authority and cooperation exists also in the participants' natural

³⁹For recent examples, see Beath, Christia, and Enikolopov (2012); Grossman (2012); and Olken (2010).

⁴⁰In the experiment, monitors did not gain from higher levels of cooperation and had a monetary disincentive to sanction. Since their identity was known to the game participants, when analyzing their sanctioning decisions, we cannot decouple altruistic considerations from reputation considerations.

environment; and (3) greater cooperation in participants' natural environment translates into a greater response to the legitimacy cue: group members who participate in collective marketing contribute more, on average, in the elected-monitor condition than "defectors" who sell their coffee to local middlemen. Taken together, our findings suggest that the legitimacy of internal centralized authorities is an important factor in determining the success of farmer cooperatives, and likely of other similar organizations. Combining a lab-in-the-field experiment with observational data, we are able to isolate one of the elements that makes group members cooperate in real life.

References

- Ahlquist, John, and Margaret Levi. 2011. "Leadership: What It Means, What It Does, and What We Want to Know about It." *Annual Review of Political Science* 14(1): 1–24.
- Baldassarri, Delia. 2009. "Collective Action." In *The Oxford Handbook of Analytical Sociology*, ed. Peter Hedström and Peter Bearman. Oxford: Oxford University Press, 391–418.
- Baldassarri, Delia, and Guy Grossman. 2011. "Centralized Sanctioning and Legitimate Authority Promote Cooperation in Humans." *Proceedings of the National Academy of Sciences* 108(27): 11023–27.
- Barr, Abigail. 2001. "Social Dilemmas and Shame-Based Sanctions: Experimental Results from Rural Zimbabwe." Center for the Study of African Economies Working Paper.
- Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2012. "Power to the People? Experimental Evidence on Direct Democracy from Afghanistan." Massachusetts Institute of Technology, Political Science Department, Working Paper No. 2012-6.
- Benz, Matthias, and Stephan Meier. 2008. "Do People Behave in Experiments as in the Field?—Evidence from Donations." *Experimental Economics* 11(3): 268–81.
- Berkowitz, Leonard, and Edward Donnerstein. 1982. "External Validity Is More than Skin Deep: Some Answers to Criticisms of Laboratory Experiments." *American Psychologist* 37(3): 245–57.
- Boyd, Robert, Herbert Gintis, and Samuel Bowles. 2010. "Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare." *Science* 328: 617–20.
- Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences of the United States of America* 100(March): 3531–35.
- Burnham, Terence C., and Dominic Johnson. 2005. "The Biological and Evolutionary Logic of Human Cooperation." *Analyse & Kritik* 27(2): 113–35.
- Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell Sage Foundation.
- Cardenas, Juan Camilo, and Jeffrey Carpenter. 2008. "Behavioural Development Economics: Lessons from Field Labs in the Developing World." *Journal of Development Studies* 44(3): 311–38.
- Carpenter, Jeffrey P. 2004. "Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research." *Journal of Economic Behavior & Organization* 55(1): 93–99.
- Casari, Marco, and Luigi Luni. 2009. "Cooperation under Alternative Punishment Institutions: An Experiment." *Journal of Economic Behavior & Organization* 71(2): 273–82.
- Dal Bo, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100(5): 2205–29.
- de Rooij, Eline A., Donald P. Green, and Alan S. Gerber. 2009. "Field Experiments on Political Behavior and Collective Action." *Annual Review of Political Science* 12(May): 389–95.
- Dewan, Torun, and David P. Myatt. 2008. "The Qualities of Leadership: Direction, Communication, and Obfuscation." *American Political Science Review* 102(3): 351–68.
- Dickson, Eric S., Sanford C. Gordon, and Gregory A. Huber. 2009. "Enforcement and Compliance in an Uncertain World: An Experimental Investigation." *Journal of Politics* 71(4): 1357–78.
- Eckel, Catherine C., Enrique Fatas, and Rick Wilson. 2010. "Cooperation and Status in Organizations." *Journal of Public Economic Theory* 12(4): 737–62.
- Eguíluz, Víctor M., Martín G. Zimmermann, Camilo J. Cela Conde, and Maxi San Miguel. 2005. "Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks." *American Journal of Sociology* 110(4): 977–1008.
- Erikson, Emily, and Joseph M. Parent. 2007. "Central Authority and Order." *Sociological Theory* 25(3): 245–67.
- Fearon, James D., Macartan Humphreys, and Jeremy M. Weinstein. 2009. "Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia." *American Economic Review* 99(2): 287–91.
- Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415(6868): 137–40.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters* 71: 397–404.
- Fowler, James H. 2005. "Human Cooperation: Second-Order Free-Riding Problem Solved?" *Nature* 437(7058): E8.
- Gelman, Andrew. 2008. "Scaling Regression Inputs by Dividing by Two Standard Deviations." *Statistics in Medicine* 27(15): 2865–73.
- Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2005. "Why Do People Accept Public Policies They Oppose? Testing Legitimacy Theory with a Survey-Based Experiment." *Political Research Quarterly* 58(June): 187–201.
- Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2005. *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.
- Goette, Lorenz, David Huffman, and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *American Economic Review* 96(May): 212–16.

- Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83(June): 525–48.
- Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.
- Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102(4): 745–76.
- Greif, Avner, and Steven Tadelis. 2010. "A Theory of Moral Persistence: Crypto-Morality and Political Legitimacy." *Journal of Comparative Economics* 38: 229–44.
- Grossman, Guy. 2011. "Causal Effects of Leader Selection Rules on Leader Responsiveness and Cooperation: Evidence from Ugandan Community Organizations." Typescript. Columbia University.
- Grossman, Guy, and W. Walker Hanlon. 2011. "Do Better Monitoring Institutions Increase Leadership Quality in Community Organizations? Evidence from Uganda." Typescript. Columbia University.
- Guth, Werner, M. Vittoria Levati, Matthias Sutter, and Eline van der Heijden. 2007. "Leading by Example with and without Exclusion Power in Voluntary Contribution Experiments." *Journal of Public Economics* 91(5–6): 1023–42.
- Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4): 709–25.
- Hechter, Michael. 2009. "Legitimacy in the Modern World." *American Behavioral Scientist* 53(3): 279–88.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Henrich, Joseph, and Francisco J. Gil-White. 2001. "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission." *Evolution and Human Behavior* 22(3): 165–96.
- Hibbing, John R., and John R. Alford. 2004. "Accepting Authoritative Decisions: Humans as Wary Cooperators." *American Journal of Political Science* 48(1): 62–76.
- Hoffmann, Elizabeth A. 2005. "Dispute Resolution in a Worker Cooperative: Formal Procedures and Procedural Justice." *Law & Society Review* 39(1): 51–82.
- Hopfensitz, Astrid, and Ernesto Reuben. 2009. "The Importance of Emotions for the Effectiveness of Social Punishment." *The Economic Journal* 119(540): 1534–59.
- Huber, Gregory A., and Sanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind When It Runs for Office?" *American Journal of Political Science* 48(2): 247–63.
- King, Andrew J., Dominic D. P. Johnson, and Mark Van Vugt. 2009. "The Origins and Evolution of Leadership." *Current Biology* 19(19): R911–R916.
- Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." *Annals of Mathematical Statistics* 22(3): 79–86.
- Laury, Susan K., and Laura O. Taylor. 2008. "Altruism Spillovers: Are Behaviors in Context-Free Experiments Predictive of Altruism toward a Naturally Occurring Public Good?" *Journal of Economic Behavior & Organization* 65(1): 9–29.
- Levati, M. Vittoria, Matthias Sutter, and Eline van der Heijden. 2007. "Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information." *Journal of Conflict Resolution* 51(5): 793–818.
- Levi, M., and A. Sacks. 2009. "Legitimizing Beliefs: Sources and Indicators." *Regulation & Governance* 3(4): 311–33.
- Levi, Margaret. 1997. *Consent, Dissent, and Patriotism*. Cambridge: Cambridge University Press.
- Levi, Margaret, Audrey Sacks, and Tom Tyler. 2009. "Conceptualizing Legitimacy, Measuring Legitimizing Beliefs." *American Behavioral Scientist* 53(3): 354–75.
- Levitsky, Steven, and M. Victoria Murillo. 2009. "Variation in Institutional Strength: Causes and Implications." *Annual Review of Political Science* 12: 115–33.
- Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21(2): 153–74.
- Lieberman, Evan S. 2007. "Ethnic Politics, Risk, and Policy-Making." *Comparative Political Studies* 40(12): 1407–32.
- Linz, Juan J., and Alfred C. Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore: Johns Hopkins University Press.
- List, John A. 2006. "The Behaviorist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy* 114(1): 1–37.
- Locke, John. 1990. *Second Treatise on Civil Government [1690]*. Indianapolis, IN: Hackett.
- Lubell, Mark, and John T. Scholz. 2001. "Cooperation, Reciprocity, and the Collective-Action Heuristic." *American Journal of Political Science* 45(1): 160–78.
- Manin, Bernard. 1997. *The Principles of Representative Government*. Cambridge: Cambridge University Press.
- Martinez-Bravo, Monica, Gerard Padro-i Miquel, Nancy Qian, and Yang Yao. 2010. "Accountability in an Authoritarian Regime: The Impact of Local Electoral Reforms in Rural China." Presented at the annual meeting of the Northeast Universities Development Consortium (NEUDC), Boston.
- McEwen, Craig A., and Richard J. Maiman. 1986. "In Search of Legitimacy: Toward an Empirical Analysis." *Law & Policy* 8(3): 257–73.
- Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89(11–12): 2325–68.
- Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge: Cambridge University Press.
- Murphy, Kristina. 2005. "Regulating More Effectively: The Relationship between Procedural Justice, Legitimacy, and Tax Non-compliance." *Journal of Law and Society* 32(4): 562–89.
- Narayan-Parker, Deepa. 2002. *Empowerment and Poverty Reduction: A Sourcebook*. Washington, DC: World Bank.
- Nowak, Martin A. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314: 1560–63.
- O'Gorman, Rick, Joseph Henrich, and Mark Van Vugt. 2009. "Constraining Free Riding in Public Goods Games:

- Designated Solitary Punishers Can Sustain Human Cooperation." *Proceedings of the Royal Society B: Biological Sciences* 276(1655): 323–29.
- Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104(2): 243–67.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- 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.
- Rand, David, Anna Dreber, Tore Ellingsen, Drew Fudenberg, and Martin Nowak. 2009. "Positive Interactions Promote Public Cooperation." *Science* 325(5945): 1272–75.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Rothstein, Bo. 2009. "Creating Political Legitimacy." *American Behavioral Scientist* 53(3): 311–30.
- Scholz, John T., and Wayne B. Gray. 1997. "Can Government Facilitate Cooperation? An Informational Model of OSHA Enforcement." *American Journal of Political Science* 41(3): 693–717.
- Sigmund, Karl. 2007. "Punish or Perish? Retaliation and Collaboration among Humans." *Trends in Ecology & Evolution* 22(11): 593–600.
- Staatz, John M. 1987. "Recent Developments in the Theory of Agricultural Cooperation." *Journal of Agricultural Cooperation* 2: 74–95.
- Taylor, Michael. 1982. *Community, Anarchy, and Liberty*. Cambridge, MA: Cambridge University Press.
- Weber, Max. [1978] 1922. *Economy and Society*. Berkeley: University of California Press.
- Wilson, Rick K., and Carl M. Rhodes. 1997. "Leadership and Credibility in N-Person Coordination Games." *Journal of Conflict Resolution* 41(6): 767–91.
- Zelditch, Morris. 2001. "Processes of Legitimation: Recent Developments and New Directions." *Social Psychology Quarterly* 64(March): 4–17.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

- Supporting analysis, tables, and figures
- Analysis of contributions
- Monitors' characteristics
- Analysis of monitors' behavior
- Descriptive analysis
- Matching analysis based on the distributions of contributions
- Cooperation in the PGG and in the natural setting
- Broader scope of the research
- Sampling strategy
- Data sources
- Implementation
- Missing data
- Randomization procedure
- Balance
- Interviewer manual
- Script of the Public Goods Game: Elected monitor condition

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix

Table A1 Public Goods Contribution in Round t by Treatment

	$T_r - BL$			$T_e - BL$			$T_e - T_r$		
	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
ATE	60.35*	46.66*	51.01*	106.82*	88.85*	91.60*	46.52*	38.95*	36.07*
	(15.16)	(11.72)	(11.99)	(15.58)	(12.69)	(12.96)	(14.56)	(11.41)	(11.76)
Round $t - 1$	9.66*	9.67*	8.26*	9.89*	9.90*	7.91*	7.58*	7.58*	6.39*
	(2.64)	(2.67)	(2.76)	(3.04)	(3.06)	(3.15)	(2.78)	(2.78)	(2.83)
N. subjects in session		2.70	3.98		-3.72	-2.04		0.21	-1.38
		(6.24)	(6.52)		(6.68)	(6.74)		(6.92)	(7.10)
Prelim contribution		0.59*	0.59*		0.51*	0.50*		0.50*	0.50*
		(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)
Male			-20.83			-17.36			-37.48*
			(13.61)			(14.45)			(13.13)
Age (units of 10)			9.01*			8.18			1.52
			(4.42)			(4.91)			(4.29)
Church attendance			-28.73*			-23.90*			-23.54*
			(11.18)			(11.94)			(11.10)
Education (Std.)			-4.35			-1.08			2.65
			(6.68)			(7.22)			(6.65)
Wealth (Std.)			-7.44			-11.31			-11.64
			(6.39)			(7.37)			(6.47)
Intercept	373.23*	109.93	197.43*	372.11*	209.71*	280.94*	445.68*	231.76*	349.53*
	(18.57)	(69.33)	(80.00)	(20.87)	(74.08)	(83.21)	(19.18)	(76.80)	(86.84)
$\sqrt{\Psi_{(3)}}$	68*	52*	56*	82*	56*	55*	74*	55*	56*
$\sqrt{\Psi_{(2)}}$	203*	127*	118*	163*	63*	0.03	176*	108*	102*
σ_e	190*	196*	199*	237*	243*	251*	195*	195*	198*
ϕ_1	0.27*	0.30*	0.32*	0.41*	0.43*	0.44*	0.33*	0.33*	0.35*
ϕ_2	0.12*	0.14*	0.15*	0.18*	0.19*	0.20*	0.11*	0.11*	0.13*
Observations	3964	3964	3724	3880	3880	3704	3708	3708	3540
Log Likelihood	-26934	-26674	-25043	-26591	-26384	-25164	-25099	-24860	-23712

Standard errors in parentheses. * $p > 0.05$.

Note: DV: Public goods contribution in round t . Table reports results from a series of three-level random intercept models, in which contributions in round t (level 1) are nested within individuals (level 2), who themselves are nested within farmer associations (level 3). $T_r - BL$ refers to the average treatment effects (ATE) of random monitor (T_r) compared to baseline (BL), $T_e - BL$ refers to the ATE of elected monitor (T_e) compared to baseline (BL), and $T_e - T_r$ refers to the ATE of random monitor (T_r) compared to elected (T_e). $\sqrt{\Psi_{(2)}}$ refers to between-subjects variability, $\sqrt{\Psi_{(3)}}$ refers to variability between farmer groups, and σ_e is the estimated standard deviation of the overall error term. Given the panel setup, the multilevel regression models further assume that the errors have an autoregressive structure of order 2 (captured by the parameters ϕ_1 and ϕ_2).

Table A2 Reaction to Monitor's Behavior: Response to Punishment

	(A)	(B)	(C)	(D)
Elected monitor condition	-5.06 (7.22)	-6.93 (7.34)	-8.41 (8.59)	-8.60 (8.84)
Subject sanctioned at $t - 1$	24.50* (12.15)	20.88 (12.29)	24.72 (12.63)	22.39 (12.77)
Sanctioned $t - 1 \times$ Elected monitor	41.21* (16.25)	40.70* (16.48)	41.98* (16.74)	40.61* (16.99)
Round $t - 1$	-2.50 (5.50)	-2.23 (5.55)	-2.67 (5.67)	-2.62 (5.70)
Prelim contribution	-0.02* (0.01)	-0.03* (0.01)	-0.02 (0.01)	-0.02* (0.01)
N. subjects in session j	-0.14 (3.39)	0.48 (3.48)	0.81 (3.76)	1.89 (3.86)
N. subjects sanctioned at $t - 1$	3.67 (2.88)	3.54 (2.98)	3.87 (3.08)	3.73 (3.18)
Max contribution sanctioned at $t - 1$	-0.02 (0.02)	-0.02 (0.02)	-0.03 (0.03)	-0.03 (0.03)
Min contribution sanctioned at $t - 1$	-0.04 (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.05 (0.03)
Individual controls		X		X
Monitor profile			X	X
Intercept	31.85 (45.88)	45.75 (50.23)	40.96 (63.43)	41.70 (66.47)
$\sqrt{\Psi_{(3)}}$	21*	22*	25*	26*
$\sqrt{\Psi_{(2)}}$	48*	54*	46*	51*
σ_e	217*	213*	218*	214*
ϕ_1	-0.67*	-0.74*	-0.66*	-0.71*
ϕ_2	-0.37*	-0.41*	-0.35*	-0.38*
Observations	2559	2443	2445	2339
Log Likelihood	-17260	-16425	-16485	-15721

Standard errors in parentheses. * $p > 0.05$.

Note: Table reports results from a series of three-level random intercept models, in which contributions in round t (level 1) are nested within individuals (level 2), themselves nested within farmer associations (level 3). The dependent variable, change in individual contribution from $t - 1$ to t , is modeled as a function of whether player i has been sanctioned at $t - 1$ and the type of monitor, controlling for monitor's sanctioning behavior at time $t - 1$ and player i 's preliminary contributions. In Model (A) we add individual controls, in Model (B) the monitor's sociodemographic characteristics, and in Model (D) we add both types of control. $\sqrt{\Psi_{(2)}}$ refers to between-subjects variability and $\sqrt{\Psi_{(3)}}$ refers to variability between farmer groups, and σ_e is the estimated standard deviation of the overall error term. All models assume that the errors have an autoregressive structure of order 2 (captured by the parameters ϕ_1 and ϕ_2).

Table A3 Procedural Legitimacy and Cooperation: Natural Setting

	Monitored (c)			Accountable (c)			Name (z)			Rule (c)			Receipts (c)			Attend (c)			Vote (c)			Index (z)		
	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3	F1	F2	F3	G1	G2	G3	H1	H2	
Legitimacy proxy	0.04 (0.02)	0.02 (0.01)	0.02* (0.01)	0.02* (0.01)	0.02 (0.01)	0.02* (0.01)	0.09* (0.03)	0.07* (0.03)	0.06* (0.03)	0.06* (0.03)	0.06* (0.03)	0.06* (0.03)	0.03* (0.01)	0.03* (0.01)	0.03* (0.01)	0.11* (0.02)	0.10* (0.02)	0.10* (0.02)	0.10* (0.02)	0.10* (0.02)	0.10* (0.02)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)
Male (c)	-0.06* (0.03)	-0.06* (0.03)	-0.07* (0.03)	-0.07* (0.03)	-0.07* (0.03)	-0.07* (0.03)	-0.06* (0.03)	-0.06* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.06* (0.03)	-0.06* (0.03)	-0.06* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.07* (0.03)	-0.07* (0.03)	-0.07* (0.03)
MM honesty (c)	0.23* (0.04)	0.22* (0.04)	0.24* (0.04)	0.24* (0.04)	0.23* (0.04)	0.23* (0.04)	0.26* (0.04)	0.25* (0.04)	0.26* (0.04)	0.26* (0.04)	0.25* (0.04)	0.25* (0.04)	0.22* (0.05)	0.21* (0.05)	0.21* (0.05)	0.25* (0.04)	0.24* (0.04)	0.24* (0.04)	0.25* (0.04)	0.24* (0.04)	0.24* (0.04)	0.23* (0.05)	0.23* (0.05)	0.22* (0.05)
Age (z)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Education (z)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Seasonal yield (log)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.00 (0.01)	-0.00 (0.01)	0.00 (0.01)
Years in group (z)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.01* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.01* (0.01)	0.02* (0.01)	0.02* (0.01)
Associational life (z)	0.01* (0.01)	0.02* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.02* (0.01)	0.02* (0.01)	0.01* (0.01)	0.02* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.02* (0.01)
Church attendance (z)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)
N. of members (z)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
DC seasonal yield (z)	-0.02 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)

(Continued)

Table A3 (Continued)

	Monitored (c)		Accountable (c)		Name (z)		Rule (c)		Receipts (c)		Attend (c)		Vote (c)		Index (z)	
	A1	A2	B1	B2	C1	C2	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2
Manager's effort (z)	0.06*	(0.01)	0.06*	(0.01)	0.05*	(0.01)	0.06*	(0.01)	0.05*	(0.01)	0.06*	(0.01)	0.06*	(0.01)	0.05*	(0.01)
Age of DC (z)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	-0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)
Distance to TC (z)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.02	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Intercept	0.48*	(0.03)	0.49*	(0.03)	0.48*	(0.03)	0.48*	(0.03)	0.51*	(0.03)	0.48*	(0.03)	0.48*	(0.03)	0.51*	(0.03)
$\sqrt{\psi_{(2)}}$	0.17*	(0.02)	0.17*	(0.02)	0.11*	(0.02)	0.16*	(0.02)	0.16*	(0.02)	0.16*	(0.02)	0.17*	(0.02)	0.16*	(0.02)
σ_e	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)	0.41*	(0.02)
Observations	1267		1237		1306		1306		1152		1301		1271		1102	
Log Likelihood	-735		-715		-753		-756		-671		-747		-727		-642	

Standard errors in parentheses. * $p > 0.05$.

Note: Table describes a set of random intercept regressions in which the dependent variable is the share of a member's total seasonal coffee yield that was sold via the farmer group in the past season. In all models, group members (level 1) are nested within their farmer association (level 2). The key independent variable is one of eight proxy measures of the DC manager's perceived legitimacy. "Monitored" and "Accountable" refer to whether the respondent believes the DC manager to be monitored and accountable, respectively; "Name" refers to whether the respondent was able to name the DC manager or chairman correctly; "Rule" refers to whether the respondent knows the method for selecting the DC manager; "Receipts" refers to whether the respondent receives receipts when participating in collective marketing; "Attend" refers to whether the respondent attended the last general assembly meeting; "Vote" refers to whether the respondent voted in the last group elections; and "Index" is the first principal component from a principal component analysis (PCA) that combines the above proxy variables into a single index. For each legitimacy proxy measure we fit a model that uses only individual-level controls (A1, B1, etc.) and a model that adds also DC-level controls. $\sqrt{\psi_{(2)}}$ refers to variability between farmer groups and σ_e is the estimated standard deviation of the overall error term. Binary variables were centered (c), while all other variables were rescaled by subtracting their mean and dividing by two standard deviations (z). DC = depot committee; MM = middleman; TC = trading center.