American Economic Review 2015, 105(2): 747–783
http://dx.doi.org/10.1257/aer.20120700

Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia†

By Michael Kosfeld and Devesh Rustagi*

We conduct a social dilemma experiment in which real-world leaders can punish group members as a third party. Despite facing an identical environment, leaders are found to take remarkably different punishment approaches. The different leader types revealed experimentally explain the relative success of groups in managing their forest commons. Leaders who emphasize equality and efficiency see positive forest outcomes. Antisocial leaders, who punish indiscriminately, see relatively negative forest outcomes. Our results highlight the importance of leaders in collective action, and more generally the idiosyncratic but powerful roles that leaders may play, leading to substantial variation in group cooperation outcomes. (*JEL* C93, D03, O13, Q23)

Whoever imposes severe punishment becomes repulsive to the people; while he who awards mild punishment becomes contemptible. But whoever imposes punishment as deserved becomes respectable. For punishment when awarded with due consideration, makes the people devoted to righteousness and to works productive of wealth and enjoyment; while punishment, when ill-awarded under the influence of greed and anger or owing to ignorance, excites fury even among hermits and ascetics dwelling in forests, not to speak of householders.

— Chanakya, *Arthashastra*, Book I, Chapter IV

Economists have long speculated why persistent and substantial differences in cooperation outcomes exist among groups engaged in collective action. Examples run from common property management (Ostrom 1990) to teamwork in organizations (Gittel 2000) and collective neighborhood efforts in reducing violent

* Kosfeld: Faculty of Economics and Business Administration, Goethe University Frankfurt, Grüneburgplatz 1, D-60323 Frankfurt am Main, Germany (e-mail: kosfeld@econ.uni-frankfurt.de); Rustagi: Faculty of Economics and Business Administration, Goethe University Frankfurt, Grüneburgplatz 1, D-60323 Frankfurt am Main, Germany (e-mail: rustagi@econ.uni-frankfurt.de). We especially thank the communities that participated in our research and our research assistants, Alazar Ruffo, Muhammad, and Ayano. We are grateful to four anonymous referees as well as numerous seminar participants for helpful comments and suggestions. Kosfeld acknowledges support from the Alfons und Gertrud Kassel-Stiftung and Rustagi acknowledges support from ETH Zurich while conducting this research. The study was financially supported by the North-South Centre at ETH Zurich, the German Federal Ministry of Education and Research (BMBF) (BioTEAM 01LM0201), and the German Federal Ministry of Economic Cooperation and Development (BMZ) via the Centre for Development Research, Bonn. We also acknowledge kind support from the German Organization for International Cooperation (GIZ) and the Oromia State Forest Enterprises Agency. The authors declare that they have no relevant or material financial interests that relate to the research described in the paper.

† Go to http://dx.doi.org/10.1257/aer.20120700 to visit the article page for additional materials and author disclosure statement(s).

crime (Sampson, Raudenbush, and Earls 1997). In this paper, we investigate the role that group leaders play in achieving cooperation. In a natural field setting we study whether leaders with the authority to enforce cooperation vary in their intrinsic motivation to punish group members and whether the observed variation can explain why some groups are more successful than others in achieving and sustaining cooperation.

Leader-based punishment institutions are widely present in many organizations to bolster group cooperation. A key concern that arises in implementing these institutions is the motivation for a leader to enforce cooperation by punishing group members who free ride, especially given that punishment is typically costly for the leader. A prevailing view among economists is that leaders have pure self-regarding preferences and are motivated to punish only if they have explicit incentives to do so (Alchian and Demsetz 1972; Holmstrom 1982; Bianco and Bates 1990). However, in recent years, profound empirical evidence has shown that many individuals have also other-regarding preferences and that these affect cooperation outcomes both in the lab (Fehr and Gächter 2000; Herrmann, Thöni, and Gächter 2008; Fischbacher and Gächter 2010) and in the field (Bandiera, Barankay, and Rasul 2005a; Karlan 2005; Rustagi, Engel, and Kosfeld 2010). Important motives that have been documented to play a role are egalitarian motives (Fehr and Schmidt 1999; Dawes et al. 2007), efficiency concerns (Charness and Rabin 2002; Engelmann and Strobel 2004), and spite and envy (Fehr, Hoff, and Kshetramade 2008; Herrmann, Thöni, and Gächter 2008). This suggests that besides any explicit incentives that are shaped by the institutional environment, a leader's intrinsic, other-regarding motivation may also be important.

So far, the effect of a leader's intrinsic motivation to enforce cooperation on group outcomes has largely remained unexplored. One reason is that a reliable measure for the motivation of a leader to punish is hard to obtain. Often, there exist multiple motives to punish which operate at the same time. Further, observed punishment and group outcomes are typically simultaneously determined which renders the analysis of their relationship complicated.[1] These considerations suggest that, at best, an experimental approach is warranted in which the determination of variables can be controlled and different motives can be carefully separated. Conducting behavioral experiments with actual leaders, however, is difficult due to limited accessibility, reputational concerns between leaders and group members, and the need for high stakes.[2]

In this study, we resolve these problems by combining behavioral experiments with real-world leaders in a natural field setting with performance data from the same groups these leaders are in charge of. The field setting, in which our study takes place, is a large forest commons management program in Ethiopia launched in 2000 by the Ethiopian government in collaboration with the Deutsche Gesellschaft

[1] This is akin to the problem that accrues in estimating the effect of police on crime rate, given that cities with higher crime rates also deploy a higher number of policemen.

[2] With the exception of Fehr and List (2004), we are not aware of any study that uses behavioral experiments to measure preferences of actual leaders. Fehr and List (2004) measure trust and trustworthiness of CEOs of small-scale firms in Costa Rica but do not relate these outcomes to firm performance. Evidence from experiments with students (in which randomly selected participants become so-called "leaders"), on the other hand, is limited as there is no such thing as a randomly selected leader outside the laboratory and leaders typically differ from non-leaders across several dimensions (Fehr and List 2004; Lazear 2012).

für Internationale Zusammenarbeit (GIZ). In this program, which at the time of our study comprised 56 different forest user groups, group members confront collective action dilemmas that are similar to the problem of cooperation and moral hazard in teams. Groups rely on leaders to monitor and enforce cooperation, but leaders are not materially incentivized to perform these activities. The field setting thus represents a unique environment to analyze the effect of leaders' intrinsic punishment motivation on group cooperation.

We elicit a leader's motivation to punish by means of an anonymous, one-shot third-party punishment game (TPPG) in which the leader observes the cooperation decisions of his group members in a public goods game and then has the option to punish members as a third party at a personal cost.[3] We employ the strategy method for the leader to draw robust conclusions on the leader's punishment motivation. Further, we take additional measures to assuage reputational concerns between leaders and group members by precluding that group members can learn the behavior of their leader in the TPPG.

Despite facing an identical experimental environment, we find that leaders differ widely in their propensity and motivation to punish in the TPPG. Based on key motives outlined in the literature, a detailed analysis reveals that leaders can be characterized as either non-punishing (L_{NP}), equality driven (L_{EQ}), equality and efficiency driven, (L_{EQEF}), or antisocial (L_{AS}). While L_{NP} confirm the standard economic prediction of zero punishment under one-shot, anonymous conditions, L_{EQ} exclusively punish deviations from the equality norm, and L_{EQEF} punish deviations from both equality and efficiency norms.[4] Most notably, we identify antisocial leaders (L_{AS}), who punish even cooperators, i.e., players who contribute their full endowment to the public good. We can rule out that the observed variation in motivation is driven by differences in ethnicity, religion, or institutional environment, as these are held constant across groups. We also exclude lack of understanding as an explanatory factor, because we took several measures to ensure that leaders understand all details of the experiment. Supplementary results on the rating of individual leaders from a household survey with 495 members of forest user groups corroborate our leader characterization. In the survey, antisocial leaders are significantly more likely to be rated as "bad" compared to other leader types.

We then investigate the association between leader type and group performance in commons management by comparing outcomes across groups with different leader types. Following the literature, we hypothesize the association to be positive for L_{EQEF}, positive but potentially weaker for L_{EQ}, and negative for L_{AS}. Our results confirm this. We find that with respect to non-punishing leaders (L_{NP}) as a benchmark, equality and efficiency driven leaders (L_{EQEF}) have a significantly positive association with the forest management outcome, but the association with antisocial leaders (L_{AS}) is significantly negative. We find no significant association for leaders who focus on equality only (L_{EQ}). Considering data from a recent second round of forest assessment, we find that the association of L_{EQ} leaders with group outcomes becomes positive and significant. On average, L_{EQEF} leaders have nearly 29 more

[3] Previous papers that have used third-party punishment games include Fehr and Fischbacher (2004), Bernhard, Fischbacher, and Fehr (2006), Henrich et al. (2006), and Carpenter and Matthews (2012).

[4] We do not observe leaders who punish only deviations from the efficiency norm but not from the equality norm.

young trees per hectare whereas groups with a L_{AS} leader have about 20 less young trees per hectare. The effects are robust and quantitatively large relative to the mean outcome of 67 young trees per hectare across all groups. Adding leader types to a model of common property management controlling for other factors leads to an increase in adjusted R^2 by 11 percentage points.

Interpretation of the above results as causal could be problematic if groups with better outcomes selected better leaders due to reverse causality, or if a leader's type reflects some underlying conditions in groups leading to an omitted variables bias. Although, with cross-sectional data like ours, it is difficult to get entirely rid of these problems, we further combine unique features of the field setting with a rich econometric specification to mitigate these concerns. In our setting, besides the leader election predating the outcome assessment, the ethnolinguistic, religious, and occupational similarity across groups together with the traditional age-based leader election process ensure that groups headed by different leader types as well as leader types themselves are comparable in many observables characteristics. In addition, our econometric specification includes a rich set of group and leader controls including experimental measures of members' other-regarding preferences and village fixed effects that absorb unobserved heterogeneity across villages the groups are part of.

Our paper relates to several strands in the literature. First, the results complement recent works in economics that highlight the importance of leadership in the performance of groups. These works focus either on the general importance of leaders (Jones and Olken 2005) or specific leader characteristics including management style (Bertrand and Schoar 2003), gender (Chattopadhyay and Duflo 2004), overconfidence (Malmendier and Tate 2005), management practices (Bloom and van Reneen 2007), and superstar status (Malmendier and Tate 2009). Our study can be interpreted as lending support to a leader's intrinsic and other-regarding motivation for norm enforcement as an additional dimension. Second, our results fill an important gap in the literature on the role of sanctioning in commons management (Ostrom 1990; Tang 1992; Baland and Platteau 1996). Leadership is understood as an important factor in this domain (Platteau and Seki 2007), yet few empirical works have actually tested its importance. Given the increasing reliance of many development programs on local participation in meeting success (Björkman and Svensson 2009), our findings provide an additional dimension in explaining why such programs may succeed or fail. Third, previous studies have documented negative effects of antisocial punishment on cooperation in laboratory experiments. By linking actual leader punishment behavior observed in an experiment with outcomes on forest commons management, our study provides important external validity to these results (Levitt and List 2007; Falk and Heckman 2009; Camerer forthcoming).

The rest of the paper is organized as follows. Section I provides hypotheses for why a leader's motivation to punish should matter for group performance. Section II offers a brief introduction to the institutional setting and describes the data on group performance and group and individual characteristics. Section III describes the behavioral experiment we used to measure a leader's motivation to punish. Section IV identifies motives behind leader punishment and uses these to characterize leaders into different types. Section V enlists the empirical methodology and quantifies the association between leader types and group performance, followed by robustness checks. Finally, Section VI summarizes and offers concluding remarks.

I. Why Should a Leader's Propensity to Punish Matter?

Many groups engaged in collective action incorporate leader punishment with the objective to deter individuals from free riding on the cooperation of others and to bolster overall group performance. However, because punishment is typically costly to the punisher, a key concern is the motivation of a leader to enforce norms of cooperation. Based on the assumption of pure self-regarding preferences, leaders enforce cooperation at a cost to themselves only if institutional conditions are such that doing so yields them material or reputational gains. As per this view, all leaders are homogeneous in their propensity to punish and therefore individual leaders are unlikely to matter for group performance unless explicit incentives for punishment differ.

In contrast, profound empirical evidence shows that individuals regularly deviate from the pure self-interest hypothesis, in particular in the context of cooperation, revealing important other-regarding motives such as egalitarian concerns (Fehr and Schmidt 1999; Dawes et al. 2007), efficiency motives (Charness and Rabin 2002; Engelmann and Strobel 2004), but also envy and spite (Fehr, Hoff, and Kshetramade 2008; Herrmann, Thöni, and Gächter 2008). This opens a new channel for group leaders to play a role as the heterogeneity in leaders' propensity to punish and enforce cooperation may be due to differences in leaders' underlying other-regarding motivation.

If leaders differ in their punishment propensity, research suggests that this will lead to different group cooperation outcomes. Several studies have documented that punishment—both monetary and nonmonetary—that is targeted at non-cooperators significantly increases cooperation levels (Yamagishi 1986; Ostrom, Walker, and Gardner 1992; Fehr and Gächter 2000; Masclet et al. 2003; Gürerk, Irlenbusch, and Rockenbach 2006). In these findings the incentives for conditional cooperators (i.e., group members who cooperate voluntarily given that others cooperate as well) has been shown to play an important role (Fischbacher and Gächter 2010; Gächter and Thöni 2005, 2011). The punishment of non-cooperators directly provides an incentive to cooperate, and also indirectly, via upholding members' belief about others' cooperation, motivates conditional cooperators to cooperate.[5]

Punishment that is motivated by either equality or efficiency concerns is targeted at non-cooperators. Hence, both motives are expected to be positively associated with group cooperation outcomes. However, there exists an important difference between these two motives. While efficiency motives induce punishment of no cooperation *independent* of other group members' behavior, egalitarian motives induce punishment *only if there are other group members who cooperate more*. In case nobody cooperates, equality norms demand no punishment, while efficiency norms do. Thus, while both equality- and efficiency-motivated punishment induce full cooperation (assuming that punishment is strong enough), only efficiency-motivated punishment is able to eliminate the classic free-riding as well as any other symmetric low-cooperation equilibrium. Strategically, equality-motivated punishment leads to multiple Nash equilibria whereas efficiency-motivated punishment creates a unique Nash equilibrium. This suggests that efficiency concerns may play

[5] See also Bowles and Polonía-Reyes (2012).

a superior role in leader punishment and that equality motives alone may be less effective in increasing group cooperation.

Lab experiments have also identified detrimental effects of punishment on group cooperation. In a seminal paper, Herrmann, Thöni, and Gächter (2008) show that so-called "antisocial punishment," i.e., the punishment of cooperators, is prevalent across different cultures and strongly negatively related to observed levels of cooperation. In a similar vein, Falk, Fehr, and Fischbacher (2005); Cinyabuguma, Page, and Putterman (2006); Nikiforakis (2008); Fehr, Hoff, and Kshetramade (2008); and Gächter and Herrmann (2009, 2011) document negative effects of antisocial punishment on group cooperation. While the ultimate motivations behind antisocial punishment are still up to debate, evidence suggests that retaliation (Nikiforakis 2008), reduction of relative payoffs (Falk, Fehr, and Fischbacher 2005), and deviation from descriptive norms (Irwin and Horne 2013) play an important role. Notably, antisocial punishment is not a repeated-game phenomenon but is observed in one-shot experiments, too (Gächter and Herrmann 2009, 2011).

Together, the evidence suggests that differences in leaders' other-regarding motivation to invest personal resources for the punishment of norm violators are associated with differences in group cooperation outcomes. Leaders who punish free riders at a personal cost due to equality and efficiency motives are likely to achieve better cooperation outcomes than leaders who are not willing to exert costs on punishment. Further, leaders who punish inequality only may be less effective than leaders who punish also equal but inefficient outcomes. Finally, leaders who punish even cooperators are expected to have negative effects on cooperation outcomes. In the following, we provide an analysis of these hypotheses in the context of a large natural field setting.

II. Institutional Setting and Data

A. *The Participatory Forest Management Program*

Ethiopia saw rapid deforestation in the 1990s, causing the government to seek alternatives to centralized forest management. One such mechanism is Participatory Forest Management (PFM), in which local groups are given legal property rights to use and manage their forests. Currently, the government together with many international agencies is implementing PFM as a major alternative approach to conserve its last remaining forests (Kubsa and Tadesse 2002).

We conducted our study with forest users involved in the largest PFM program in Ethiopia, which was launched in June 2000 to conserve degraded forests in Adaba and Dodola districts (*Woredas*) of the Bale Zone, Oromia State. These forests were found to be shrinking at a high rate of 3 percent per annum, posing a major threat to people's livelihood as well as wildlife in the Bale Mountains (Kubsa and Tadesse 2002). Under the program, a forest area was subdivided into forest blocks with an average size of 360 hectares each. Negotiations were held with forest users to reach a consensus on block boundaries. Based on the carrying capacity of 12 hectares per household, each forest block was assigned to a group not exceeding 30 members. These groups were given complete jurisdiction to manage their forests as a common property resource. The membership in a group was not voluntary but was assigned mainly on the basis of a household's customary right of forest use. By 2005, there were 56 forest user groups: 17 groups in two villages of the Adaba district and 39 groups in three villages of the Dodola district.[6]

Among the program benefits, group members are allowed to harvest fodder, timber, and non-timber forest products for both self-consumption and sale, as well as maintain existing farm plots and homesteads inside the forest. In return, members are required to maintain the existing forest cover, restrict further agricultural and settlement expansion inside the forest, and prevent nonmembers from accessing the forest.[7] The performance of each group in managing its forest is planned to be assessed every fifth year by the local forest administration with support from group members. The assessment is based on the number of young trees in a group-managed forest (see details in Section IIB) and is chiefly used to calculate the rent a group has to pay annually to the local forest administration. This is expected to provide incentives for groups to perform better. The rent is progressive and groups with fewer young trees per hectare pay a higher rent. For instance, there is a rent of 2 Ethiopian Birr for each hectare with at least 40 young trees and 8 Birr (roughly the same as the daily wage[8] in the study area) for less than 20 young trees per hectare. To ensure higher group performance and hence lower rents, each group is allowed to implement local rules regarding forest use, such as the amount of timber a member is allowed to harvest for self-consumption and sale.

While managing their forest as a common property resource, members confront cooperation dilemmas. Each member is better off when every member in the group cooperates by adhering to internal rules on forest management; however, because violating the rules leads to higher payoffs, for instance, from the sale of extra timber, individual members have little incentive to cooperate. Even though groups are penalized for poor performance by paying a higher rent, because the rent is levied at the group level it does not target individual free riding. As in any organization, groups thus have to rely on monitoring and punishment of free riders to overcome these dilemmas.

Forest management decisions in a group are made by the general assembly, which comprises all recognized members. The assembly elects democratically an executive committee comprising a leader, a vice-leader, a secretary, a cashier, and a member usually for a period of five years, but our data suggests that this is often longer. The power and function of the executive committee as well as its individual members is specifically listed in the model bylaw (Kidan and Tadesse 2004) and there is no inter-group variation in this aspect or the organizational structure. For instance, the executive committee is responsible for formulating forest management plans including forest patrols, developing measures to protect the forest from nonmembers, and dealing with rule violation. The executive committee at the time of the study in all the groups was elected at the onset of the program, before the outcomes were assessed.

[6] A forest user group is a small settlement within a village, such that a village comprises many forest user groups plus other groups that are not part of the PFM program. The villages include Ejersa and Bubisa in Adaba, and Barisa, Denaba, and Bura-Adele in Dodola.

[7] Nonmembers are individuals who do not have legal access to use the forest and live in separate settlements located within the village but far from the groups involved in forest commons management.

[8] The daily wage in the study area at the time of the study was 7 Birr for urban areas.

Among the committee members, leaders and vice-leaders occupy the highest position and are chiefly responsible for implementing key decisions, such as supervising forest patrols and punishing rule violators.[9] Punishment of rule violators varies from a warning and fine to even exclusion from the program. According to the model bylaw, the vice-leader is required not only to assist the leader in all his activities but also to take over all the duties of the leader in case of his absence or illness. Thus, like the leader, the vice-leader is responsible for supervision and punishment. Because these activities are costly and leaders and vice-leaders are not paid to perform these activities, we expect their willingness to punish norm violators to play an important role in the detection as well as the punishment of free riders.[10]

B. *Data Description*

The data used in this paper come from four different sources: (i) a database maintained by the program office on group performance and group-level characteristics; (ii) two community surveys covering group- and individual-level characteristics not included in the program database and participation in forest management activities; (iii) a household survey covering individual socioeconomic characteristics, participation in forest management activities, and feedback on the group leader; and (iv) behavioral experiments to measure a leader's propensity to punish norm violators, as well as members' propensity to cooperate conditional on the behavior of other members.[11] This section describes data on group performance, group level characteristics, and individual socioeconomic characteristics. The experiments are described in Section III.

Of the 56 groups established under the program, data on group performance was not available for 2 groups and a pilot study was conducted in 3 groups to test the experiments. These 5 groups were excluded from the final study, leaving a sample size of 51 groups: 17 from Adaba and 34 from Dodola.[12]

Group Performance.—Data on the performance of each group in managing its forest were obtained from program offices at Adaba and Dodola and were verified with data obtained from the head office in Addis Ababa. The first performance assessment, and the only one for all groups until January 2013, was carried out in 2005 and we use data from this assessment as the main outcome variable in our study.[13] The assessment involves counting the number of young trees per hectare in sampled plots and then preparing an average for the entire group-managed forest. The young trees included in the assessment are called "potential crop trees" (PCT)

[9] In addition to these activities, they are also responsible for organizing group meetings, to represent groups in district level meetings, signing the forest management agreement, preparing and presenting the forest management status report to the assembly, etc.

[10] In seven groups, vice-leaders participated in the experiment. Because of their comparable position in a group and the fact that none of our results are driven by the inclusion of vice-leaders (Section VC), we do not distinguish between leaders and vice-leaders in the remainder of the paper.

[11] The behavioral experiments were conducted together with the first community and household survey from March to May 2008. The second community survey was conducted from December 2012 to January 2013.

[12] Our sample size is in line with previous studies on this topic (e.g., Dayton-Johnson 2000; Bardhan 2000; Miguel and Gugerty 2005; Khwaja 2009).

[13] For a subsample of groups data from a second forest assessment is available. This data is analyzed in Section VD.

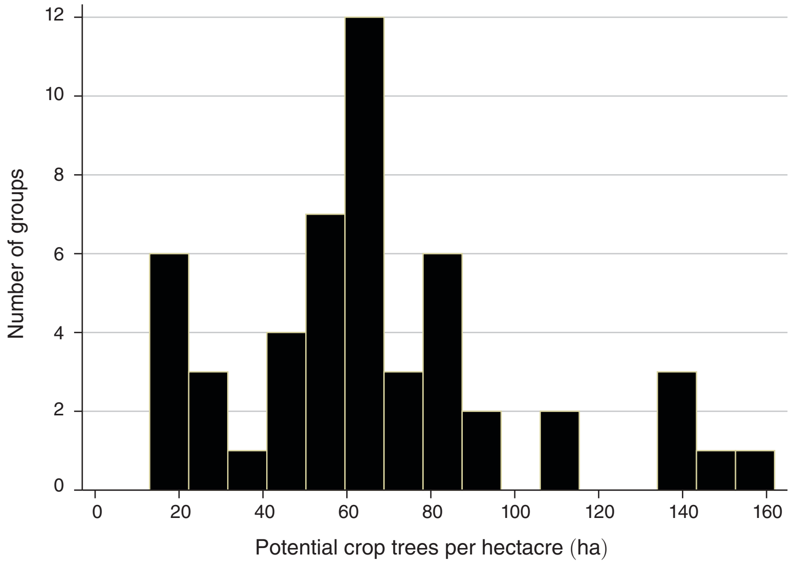

Figure 1. Distribution of Potential Crop Trees (PCT) per Hectare

and are at least two meters tall, have a diameter at breast height of $\leq$ 40 centimeters (cm) for group-I[14] and $\leq$ 25 cm for group-II,[15] with no damage to the main stem, and a distance of at least four meters separating them (Amente 2005).

This measure is a highly relevant indicator of group performance for two reasons. First, before the launch of the forest management program, members targeted young and medium-sized trees for logging as these were easier to cut, process, and transport. For trees used as non-timber forest product, members did not feel the need to wait until the trees were mature. Because of these selective practices, young and medium-sized trees were found to be missing from the forest resulting in a very low average annual forest increment of just 1 m^3/ha/year (Trainer 1996; Amente 2005). Second, Adaba-Dodola is a degraded forest and its recovery potential depends on the intensity of past extraction as well as the presence of young trees in the forest, especially given that silviculture is not practiced. Therefore, group performance based on the abundance of young trees with a potential to form a long-term final stand is an important indicator. In accordance, an inventory of the number of young trees was formally used as the criterion to evaluate the performance of a group. A tree height of two meters was chosen as a benchmark on the assumption that trees above this height are beyond the reach of animals and thus can escape browsing damage. Moreover, group-specific diameters were selected on the assumption that trees above this range will not have a significant impact on improving the forest quality. These features make the number of young trees a robust measure of group performance that covers forest quality as well as vitality (Amente 2005).

[14] Group-I species: *Hagenia abbyssinica, Podocarpus falcatus, Junipersu excelsa, Olea europea, Ekbergia capensis.*

[15] Group-II species: *Pittosporum viridiflorum, Hypericum lanceolatum, Rapania melanophloes, Erica arborea.*

Table 1—Summary Statistics on Group Characteristics

Group characteristics	Proxy variable	Mean	SD	Minimum	Maximum
Propensity to cooperate	CC share	0.34	0.21	0.00	0.87
Resource	Elevation	0.24	0.43	0.00	1.00
	Plantation	0.07	0.33	0.00	2.185
Group	Group size	26.45	4.55	16.00	30.00
Market	Market distance	2.34	0.88	0.75	4.50
Time	Time	0.49	0.50	0.00	1.00
Social heterogeneity	Female share	0.20	0.11	0.00	0.67
	Settlement	0.55	0.18	0.00	0.76
	Clan	0.39	0.23	0.00	0.78
Economic heterogeneity	Cattle	0.34	0.08	0.19	0.59
	Land	0.30	0.08	0.02	0.48

Notes: CC share is the proportion of group members with a propensity to cooperate conditional on the contribution of others in a public goods game (Rustagi, Engel, and Kosfeld 2010). Elevation is an indicator variable, which is 1 if the group-managed forest is located above 3,000 meters and 0 if it is below. Plantation is the ratio of historical plantation forest area to natural forest area. Group size is the number of members in a group. Female share is the proportion of group members who are females. Settlement and clan are indicators of social heterogeneity measured by a fractionalization index (cf. online Appendix A.II). Gini cattle and Gini land are indicators of economic heterogeneity and are measured using a Gini index. Market distance is the number of hours it takes a member to access the nearest market weighted by equine ownership. Time indicates the phase the program was launched, with 1 being phase 2000–2002 and 0 being phase 2003–2005. SD is the standard deviation. The number of observations is 51.

The data on group performance shows that groups vary widely in their success in managing their forest commons (Figure 1). The outcome varies from as low as 13 young trees per hectare to as high as 162 young trees per hectare (average $= 66.79$, standard deviation $= 34.61$). Because inventory studies conducted prior to the launch of the program found that young trees were almost entirely missing from the forest, the outcome indicates the success of a group in managing its commons.

Forest User Groups.—All 51 groups included in this study are inhabited by the Bale Oromo people, who are linguistically, occupationally, and religiously very similar. For instance, all members speak the Bale dialect of *Afan Oromo*, practice small-scale herding, forest gathering and subsistence agriculture, and follow Islam. In addition, all groups are remote and difficult to access. The level of socioeconomic development is low, with no access to tap water, electricity, irrigational facilities, roads, public health facilities, and veterinary care. Geographically, all groups lie on the northern slope of the Bale Mountains. Due to this, weather conditions (such as precipitation and temperature) are similar across groups from the two districts. Table 1 provides summary statistics on a range of important group-level variables that are used as controls in our analysis. These variables are described in detail in the online Appendix A.II.

Table 2 reports demographic, social, and economic background on leaders and members. It shows that leaders are on average not only older, but also have higher education, family size, and wealth (as measured by number of cattle, horses, and land holding) than group members. In fact, the average age of a leader is in line with the once prominent Oromo age-based class system of leader selection called *Gada*, in which individuals in the age grade 40–48 years are considered as belonging to the leader grade (cf. Section VA below). As in any tribal society, in our setting too, leaders command a high social standing in their group, such that their recommendations

Table 2—Demographic, Social, and Economic Background of Leaders and Members

Variable	Obs.	Leaders (SD)	Obs.	Members (SD)	Difference (SE)
Age (years)	51	41.18	616	36.68	4.496**
		(7.72)		(13.81)	(1.196)
Education (years)	51	4.67	613	3.23	1.433***
		(2.75)		(3.10)	(0.448)
Family size	51	11.59	609	8.34	3.250***
		(4.36)		(4.13)	(0.604)
Number of cattle	48	14.73	616	11.17	3.557**
		(12.62)		(10.28)	(1.568)
Number of horses	48	3.02	617	2.55	0.467
		(1.77)		(2.07)	(0.307)
Land holding (hectares)	48	1.84	618	1.46	0.381**
		(1.02)		(1.11)	(0.165)
Conditional cooperation (dummy variable)	51	0.25	658	0.34	−0.089
		(0.44)		(0.48)	(0.069)
Conditional cooperation (coefficient)	51	0.40	658	0.50	−0.099
		(0.55)		(0.52)	(0.076)

Notes: Data on cattle, horses, and land holding was not available for three leaders. Difference is computed as the regression coefficient on leader dummy and reflects the average difference in variables between leaders and members. Results hold when vice-leaders are excluded.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

are sought not only in the realm of forest management but also in wider social and economic domains.[16]

III. Behavioral Experiments

The experimental data used in this paper come from two behavioral experiments: (i) a third-party punishment game (TPPG) to measure a leader's propensity to punish norm violators in a public goods dilemma, and (ii) a public goods game to measure an individual's propensity to cooperate conditionally on the cooperation of others. In this section, we describe the TPPG followed by a brief introduction to the public goods game. The experimental instructions for the TPPG and experimental procedures are described in online Appendix B; for experimental instructions on the public goods game, we refer the reader to Rustagi, Engel, and Kosfeld (2010).

A. *Third-Party Punishment Game*

A straightforward way to estimate the effect of a leader's punishment behavior on the forest management outcome would be to regress the latter variable on actual punishment behavior of leaders in the field, controlling for all the relevant observable group-level characteristics. There are two obvious problems with this

[16] As an example, in one instance, a leader was approached to decide on the punishment of a thief responsible for stealing group members' livestock. On another occasion, the saddle of a horse, which one of our field assistants was riding, was accidentally broken. The horse owner approached the leader of his group to settle the compensation that the field assistant should pay.

approach. First, in well-managed groups, punishment is less likely to be observed than in poorly managed groups, leading to the simultaneous determination of punishment and group performance. Second, in the field, leaders might punish because of multiple motives, making it difficult to separate different other-regarding motives from self-interest and reputational concerns. We circumvent these problems by conducting a behavioral experiment with group leaders that allows us to obtain an exogenous measure of leader punishment in which we can separate self- and other-regarding motives behind punishment. Emulating the field setting, we implemented a third-party punishment option in a one-shot public goods game to measure a leader's propensity to enforce cooperation. The game involved two decision stages:

Stage 1: Group members took part in an anonymous linear public goods game, in which they were randomly assigned to groups of two. Each player received an endowment of six bills of 1 Birr and had to decide simultaneously on how many bills to put in his pocket and how many to contribute to a public good called "project." Players could contribute any one of the following amounts: 0, 2, 4, and 6 Birr. Any amount in the project was multiplied by 1.5 and then distributed equally between the two players, regardless of their individual contribution. Players were aware that in the second stage of the game, the leader could take a decision that may reduce their income based on their contribution decision (the word "punishment" was never used). Formally, the payoff of player i, $i \in \{1, 2\}$, in the first stage of the game was equal to

$$\Pi_i^{first} = 6 - C_i + 0.75(C_1 + C_2), \tag{1}$$

where 6 is the player's endowment, C_i is the contribution of player i to the public good, 0.75 is the marginal per capita return, and $C_1 + C_2$ is the total contribution to the public good. Because the marginal per capita return for contributing 1 Birr to the project is smaller than one, free riding is the strictly dominant strategy in the game. However, because $2 \times 0.75 > 1$, from the group perspective it is best that each player contributes his entire endowment to the project; players thus face a cooperation dilemma.

Stage 2: Given the contribution decisions of the two players, the leader had the opportunity to assign deduction points to either or both the players. To finance the assignment of deduction points, the leader was given an endowment of 10 Birr. Assigning one deduction point cost the leader 1 Birr and reduced the payoff of the punished member by 3 Birr. Because punishment is not obligatory and assigning punishment is costly to the leader, a purely self-interested leader does not punish. However, other-regarding motives such as efficiency and equality may provide reasons for a leader to punish the players depending on their contribution behavior. We used the strategy method to elicit a leader's propensity to punish the two players as a third party, i.e., for each combination of players' contributions a leader indicated whether and how much he decided to punish each of the two players. Ten out of the 16 possible strategy combinations were shown to the leader (see online Appendix B). In these combinations, player 1 by definition always contributes less than or the same as player 2 $(C_1 \leq C_2)$. We did not elicit the remaining six strategy combinations, in which the contributions of the two players are reversed $(C_1 > C_2)$, because players were anonymous and hence punishment could not be made contingent on the identity of a player.

Once the leader and all players had made their decisions, the leader's punishment decision was matched with the actual contribution decision of the group members to determine final payoffs. Precisely,

$$\text{(2)} \qquad \text{for members:} \quad \Pi_i = \max[\Pi_i^{first} - 3P_i(C_1, C_2), 0]$$

$$\text{(3)} \qquad \text{for the leader:} \quad \Pi_L = 10 - P_1(C_1, C_2) - P_2(C_1, C_2),$$

where $P_i(C_1, C_2)$ denotes the punishment of player i, $i \in \{1, 2\}$, for given contribution decisions C_1 and C_2.[17]

If the leader decided not to punish, he kept his entire endowment. We deliberately gave leaders a higher endowment than group members in order to avoid that leaders punish others merely because they have a lower payoff themselves. All members were aware of the leader's endowment, as well as the one-shot and anonymous nature of the game. Members were assured that the leader would come to know only the contribution decisions taken by them in the game and not their identity. Members and leaders took part in two additional experiments. Because payoffs from all the games were given in sum at the end, members could not determine if the leader punished them or not. This design helped us in ensuring anonymity of the leaders' punishment decisions to rule out opportunities for post-game retaliation by members against their leaders. We also made sure that member decisions were anonymous, in the sense that, while the leader knew that members participated in the game, he could not make out who contributed how much. Thus, neither leaders nor members could acquire a reputation from their behavior in the game.

We implemented a two-player public goods game to keep members' and leaders' decisions simple and to make equality and efficiency concerns easily applicable. In the field, individuals interact of course in groups larger than two. We assume that the main behavioral motives for leader punishment are independent of group size. For instance, if a leader reveals a concern for equality in the two-player game by punishing the player who contributes less, we expect him to behave similarly in groups of size $n > 2$ by punishing those members whose contribution falls below the average contribution of the other group members.[18]

[17] Based on the feedback we got from the pilot study, we made a slight modification with regard to the final payment of the leader. Instead of determining one decision of the leader as payoff relevant for each pair, we incentivized (i.e., paid) each of the ten decisions directly. This was done because the number of pairs who played the public goods game varied from group to group; hence payment based on actual pairs would have introduced extra variation among leaders. Our procedure, which keeps a leader's individual incentives in the game untouched, ensures that all leaders were paid exactly the same number of decisions. Furthermore, the procedure was also easier to explain and to communicate.

[18] The assumption is supported by experimental evidence showing that the punishment of non-cooperators in groups larger than two is determined by the difference between a player's individual contribution and the average contribution of others (e.g., Fehr and Gächter 2000).

B. *Public Goods Game*

In addition to the TPPG, we also used a two-player public goods game to elicit the behavioral type composition in each group with regard to the social norm of conditional cooperation. The norm prescribes cooperation if others cooperate and no-cooperation if others do not cooperate (Fischbacher, Gächter, and Fehr 2001). In the game, individual members had to indicate their contribution conditional on the contribution of an anonymous group member in a one-shot context. We use data from this game to test whether the probability of getting elected as a leader depends on the behavioral type of a leader. Further, based on the result reported in Rustagi, Engel, and Kosfeld (2010) that the share of conditional cooperators in a group has a positive effect on forest outcome, we use this variable as a further control in our empirical estimation.

IV. Punishment Motives and Leader Types

Twenty-two out of 51 leaders (43 percent) punished in the game, spending a total of 186 Birr. The majority of leaders (57 percent) did not punish. On average, a leader spent 0.182 Birr per player per decision. In this section, we first enlist possible motives behind punishment drawn from the experimental and behavioral economics literature, and then test the effect of each motive on observed leader punishment behavior. Subsequently, we use these motives to characterize leaders into distinct types, and also address the robustness of our characterization.

A. *Punishment Motives and Behavior*

Based on previous findings in the literature we identify three main motives for leaders to punish in the third-party punishment game.

(i) *Equality:* Research shows that egalitarian norms play a vital role in human behavior in many societies across the world (Fehr and Schmidt 1999; Dawes et al. 2007; Barr et al. 2009). Leaders in our setting may be motivated to punish a player who contributes less than the other player in the public goods game leading to inequality in terms of contributions and payoffs. We expect punishment to increase in the difference between individual contributions.

(ii) *Efficiency:* Together the two players in the public goods game are best off if they both contribute their full endowment. A second motive behind leader punishment may be the achieved efficiency in terms of joint surplus maximization (Charness and Rabin 2002; Engelmann and Strobel 2004). This implies that leaders punish players who contribute less than their full endowment, independent of the other player's contribution. We expect that punishment increases in the players' deviation from full contribution.

(iii) *Antisociality:* Finally, leaders may also punish irrespective of any deviation from the equality or efficiency norm: in particular, they may punish players who contribute their full endowment to the public good. Herrmann, Thöni,

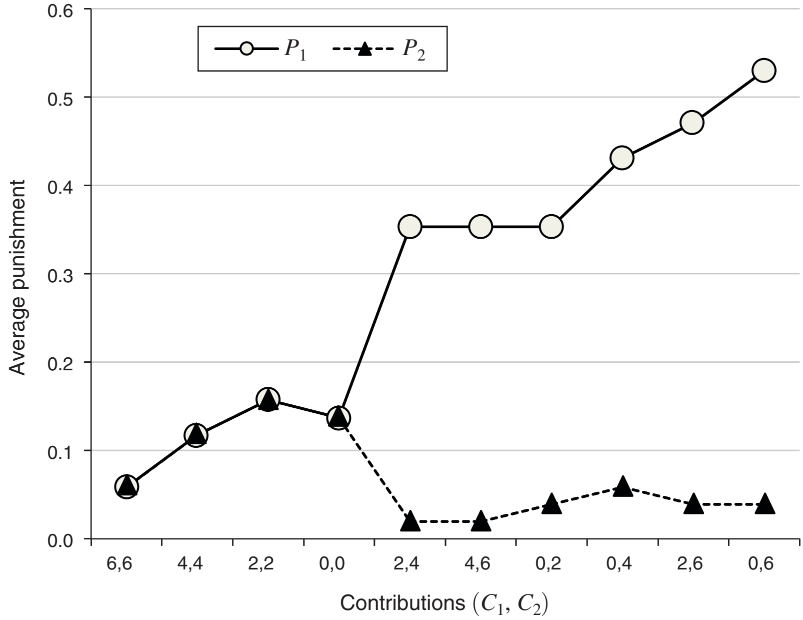

Figure 2. Average Punishment for Different Contributions of Player 1 and 2

Notes: The horizontal axis indicates the contributions of player 1 and 2, respectively. The vertical axis indicates the average punishment in Birr given by leaders. The continuous line with circles indicates average punishment of player 1 (P_1) and the dotted line with triangles indicates average punishment of player 2 (P_2).

and Gächter (2008) label such punishment "antisocial." Evidence suggests that retaliation (Nikiforakis 2008), reduction of relative payoffs (Falk, Fehr, and Fischbacher 2005), and deviation from descriptive norms (Irwin and Horne 2013) play an important role in this punishment. We expect antisocial punishment to be flat with respect to players' contributions.

We test the importance of these motives in explaining the observed leader punishment behavior on the aggregate level. Throughout the paper, we deal separately with punishment of player 1 and 2 because by definition player 1 contributes less than or the same as player 2 but never more in the experiment $(C_1 \leq C_2)$.[19] Further, since most of the punishment occurs in fact on player 1, we focus mainly on this player and discuss results on the punishment of player 2 whenever necessary.

Figure 2 illustrates leader punishment averaged over all leaders in each contribution decision of player 1 and 2. The first four decisions from the left depict average punishment when player 1 and 2 contribute equally, such that deviations from full contribution increase from left to right. The remaining six decisions depict average punishment when player 1 contributes less than player 2, such that the difference in their contribution increases from left to right. Three punishment patterns are noteworthy: (i) player 1 and 2 are punished even when they contribute their full endowment to the public good; (ii) both players are punished equally when they contribute

[19] Thus, while player 1 could violate both efficiency and equality norms, player 2 could violate only the former.

equally, and punishments are higher when players deviate from full contribution; (iii) player 1 is punished much more when he contributes less than player 2, such that punishments seem to increase in the severity of deviation from player 2's contribution. These patterns suggest that antisocial, efficiency, as well as equality motives seem to play a role in leaders' punishment behavior.

Because punishment in the experiment is infrequent, discrete, and count data, we use the following Poisson specification to test whether punishment is significantly different from punishment in the mutual cooperation case as a benchmark:[20]

$$Y_{li} = \exp(\beta \mathbf{K}), \tag{4}$$

where Y_{li} is the expected number of punishments spent by leader l on player i and $\mathbf{K}$ is a vector of dummies representing each of the ten contribution decisions of players 1 and 2. We compute standard errors that are clustered on the leader. Our first result confirms that all three punishment motives are observed on the aggregate level.

RESULT 1: *Deviation from full contribution (efficiency motive) evokes leaders to punish but the extent of this deviation is not statistically significant. Deviation from the other player's contribution (equality motive) also constitutes an important motive; here, the extent of the deviation is significant. Finally, leaders punish players even if both players cooperate (antisocial motive).*

The results are reported in Table 3 and show that punishment in the benchmark category is significantly different from zero, suggesting that leaders are motivated to punish even when players cooperate mutually. Punishments, when both players contribute the same but less than their full endowment (decisions 2 to 4), though jointly significant ($\chi^2_{(3)} = 9.34, p = 0.025$) are not significantly different from each other ($\chi^2_{(2)} = 2.87, p = 0.238$). This implies that the deviation from full contribution constitutes an important motive to punish, but the extent of this deviation is less relevant. In contrast, punishments of player 1 in case both players contribute unequally (decisions 5 to 10) are both jointly significant ($\chi^2_{(6)} = 50.47, p = 0.000$) and significantly different from each other ($\chi^2_{(5)} = 33.16, p = 0.000$), suggesting that both the deviation from the other player's contribution as well as the severity of this deviation evoke leaders to punish.[21]

Figure 2 reveals that player 1 is punished the most in contribution decisions (2, 4) to (0, 6), where he contributes less than player 2. This suggests that the equality motive seems to play the most important role. However, in these decisions player 1 not only deviates from player 2's contribution but also from full contribution. This

[20] Punishment takes few values: 0, 1, and 2. The observed frequencies of these values are 374, 121, and 15, respectively, and are in line with the expected frequency from a Poisson distribution. Moreover, the variance is nearly the same as the mean. Following Cameron and Trivedi (2009), we test the equidispersion assumption of the Poisson model. The test is based on an auxiliary regression of the dependent variable $\left((P_1 - \hat{\mu})^2 - P_1\right)/\hat{\mu}$ on $\hat{\mu}$ without an intercept. If the coefficient of $\hat{\mu}$ is significantly different from zero, it indicates over-dispersion. Results for the specification in Table 3 indicate a borderline case (the coefficient of $\hat{\mu}$ has a p-value of 0.092). We therefore consider also a negative binomial regression. Since the estimates from both regressions are the same, we present the results using the Poisson estimation.

[21] We also estimated equation (4) controlling for leader socioeconomic variables including age, education, and clan. The inclusion of these variables does not change the results reported in Table 3. A Wald test suggests that the effect of leader socioeconomic variables is not significantly different from zero $\left(\chi^2_{(2)} = 1.90, p = 0.593\right)$.

Table 3—Average Punishment of Player 1

	Coefficient	AME
Decision 2 (4, 4)	0.693*	0.059*
	(0.412)	(0.033)
Decision 3 (2, 2)	0.981**	0.098*
	(0.494)	(0.051)
Decision 4 (0, 0)	0.847	0.078
	(0.577)	(0.055)
Decision 5 (2, 4)	1.792***	0.294***
	(0.532)	(0.064)
Decision 6 (4, 6)	1.792***	0.294***
	(0.532)	(0.064)
Decision 7 (0, 2)	1.792***	0.294***
	(0.572)	(0.076)
Decision 8 (0, 4)	1.992***	0.373***
	(0.550)	(0.079)
Decision 9 (2, 6)	2.079***	0.412***
	(0.532)	(0.085)
Decision 10 (0, 6)	2.197***	0.471***
	(0.542)	(0.095)
Constant (6, 6)	−2.833***	
	(0.566)	
Observations	510	510

Notes: Poisson regression with robust standard errors clustered on leaders in parentheses. The dependent variable is the punishment of player 1 (which coincides with punishment of player 2 if both players contribute the same). Punishment in the mutual cooperation case (decision 1) is the benchmark category. AME is average marginal effect.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

potential confound makes it difficult to elicit with precision the effect of the equality motive on punishment. In the online Appendix, we overcome this problem by clustering the ten contribution decisions of players 1 and 2 into four categories and regressing Y_{li} on these categories (online Appendix A.I). The results confirm that the equality motive has indeed the strongest effect on leader punishment behavior, which is more than three times larger than the effect of the efficiency motive.

B. *Leader Punishment Types*

Having shown that antisocial, efficiency, and equality motives are relevant drivers of average leader punishment behavior, we now proceed with a detailed analysis of the punishment behavior of individual leaders. For this purpose, we use the behavior of individual leaders in the TPPG, substantiated by verbal reasons given by the same leader for (not) punishing in the game. The data allow us to categorize the 51 different leaders into four distinct, mutually exclusive types. Table 4 summarizes the definition of types together with their relative frequency.

The first leader type comprises leaders who punish exclusively deviations from the equality norm (L_{EQ}). Altogether, 14 leaders (27.5 percent) are of this type. In the experiment, because it is always player 1 who deviates from the equality norm, L_{EQ} leaders exclusively punish player 1 and never target player 2. The second leader type

Table 4—Leader Types and Punishment Motives

Leader type	Definition	Frequency	Verbal statement (examples)
Equality driven (L_{EQ})	Punishes exclusively deviations from the equality norm	$n = 14$	Make payoffs nearly equal
Equality and efficiency driven (L_{EQEF})	Punishes deviations from equality and efficiency norms	$n = 4$	Make payoffs nearly equal and punish those who contribute less
Antisocial (L_{AS})	Punishes irrespective of deviations, in particular when players contribute their full endowment	$n = 4$	It is so much fun to reduce income; I want to reduce income of those who have
Non-punisher (L_{NP})	Does not punish	$n = 29$	I prefer to have money in my pocket; they get what they have; I do not like to punish

punishes deviations from the equality norm and, in addition, also from the efficiency norm (L_{EQEF}). Four leaders (7.8 percent) fall into this category. These leaders punish player 1 and 2 when they contribute less than 6 Birr but do not punish when either of them contributes his full endowment to the public good. In contrast, the third leader type (L_{AS}) punishes players even when they contribute their full endowment to the public good. Our data show that four leaders are of this type (7.8 percent). Their punishment pattern together with verbally expressed statements clearly reveals an antisocial motive.[22] Finally, 29 leaders (56.9 percent) did not punish in the game and are therefore classified as non-punishers (L_{NP}). The fact that leaders differ in their punishment motivation despite facing an identical experimental environment is in itself startling.

Figure 3 illustrates the punishment behavior of the different leader types separately for player 1 and 2 depending on the contribution decisions of the two players. It shows that leader types differ both in *the way* and *the extent* in which players are punished for different norm violations.

RESULT 2: *L_{EQ} and L_{EQEF} are responsive in their punishment to different norm violations. In contrast, L_{AS} punish all cases alike. L_{EQ} punish deviations from the equality norm the least, while L_{EQEF} punish the same as L_{AS} in this case. Deviations from the efficiency norm are punished more strongly by L_{AS} than by L_{EQEF}.*

Table 5 provides support for Result 2 presenting results from Poisson regressions for the punishment of player 1 both *within* (panel A) and *across* (panel B) leader types.[23] We group contributions into categories C_I to C_{IV} to disentangle equality and efficiency effects. These are: C_I (6, 6), C_{II} (0, 0; 2, 2; 4, 4), C_{III} (0, 2; 0, 4; 2, 4), and C_{IV} (0, 6; 2, 6; 4, 6).[24] Panel A tests whether for a given leader type, punishment of

[22] We are convinced that the behavior of L_{AS} is not due to confusion for the following reasons: (i) all leaders took part in a control test in which they had to answer all questions correctly before they could take part in the experiment; (ii) leaders provided statements to indicate their motivation to punish, while taking punishment decision in the actual game; and (iii) each leader had to put 1 Birr on the payoff of the player whose income he wanted to reduce. After this, he had to remove 3 Birr from the punished player's payoff and hand it over to the experimenter along with 1 Birr from his own punishment endowment. When leaders were done taking punishment decisions, they indicated verbally as well as symbolically that the remaining endowment goes to their pocket. In case a leader did not punish, similar indications were used.

[23] Results on punishment of player 2 are not shown but available upon request.

[24] See online Appendix A.I for further details.

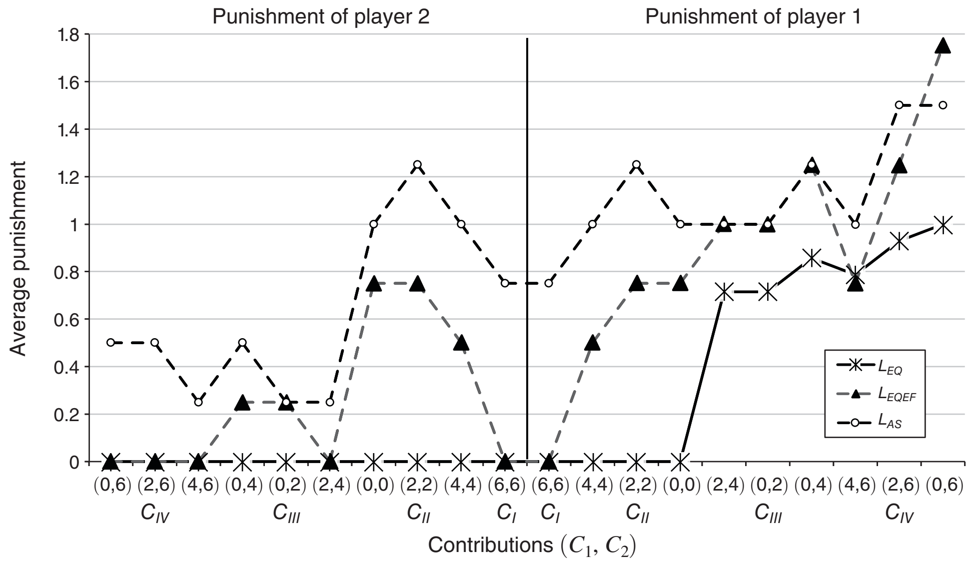

FIGURE 3. AVERAGE PUNISHMENT BY L_{EQ}, L_{EQEF}, AND L_{AS} FOR DIFFERENT CONTRIBUTIONS OF PLAYER 1 AND 2

Notes: The horizontal axis indicates the contributions of player 1 and 2, respectively. Category C_{IV} comprises deviations from equality and efficiency norms by player 1 when player 2 contributes his full endowment; C_{III} comprises deviations from equality and efficiency norms by player 1 when player 2 deviates from the efficiency norm; C_{II} comprises deviations from the efficiency norm by both players 1 and 2. C_I is the mutual cooperation case. The vertical axis indicates the average punishment of player 1 (P_1, right-hand side) and player 2 (P_2, left-hand side) by the different leader types.

player 1 varies across contribution categories with punishment in category C_{IV} serving as the benchmark. Column 1 shows that the punishment of player 1 by L_{EQ} under equality deviations is influenced by player 2's contribution. Column 2 shows that L_{EQEF} punish deviations from the efficiency norm less strongly than deviations from the equality norm. In contrast, though the punishment of player 1 is lower when player 2 does not contribute his full endowment (C_{III}), the difference is not statistically significant. Column 3 shows that L_{AS} do not distinguish between situations where players deviate from any norm and situations where they don't. A Wald test reveals that for L_{AS}, punishment of full contribution is not significantly different from punishment of either efficiency or equality deviations ($\chi^2_{(1)} = 0.50, p = 0.478$). This confirms that L_{AS} punish all cases alike, irrespective of whether a player deviates from a norm or not. Our result mirrors the one of Herrmann, Thöni, and Gächter (2008), who document that in societies where antisocial punishment is prominent, free riders and cooperators are punished with equal likelihood.

Panel B of Table 5 compares differences in punishment behavior across leader types for a given contribution category. L_{EQEF} serves as the benchmark type. Column 1 shows that L_{AS} and L_{EQEF} punish deviations from equality alike (C_{III}), but L_{EQ} punish significantly less, the difference being 0.30 Birr or 13 percent of the daily wage. We get a slightly larger result for category C_{IV} in which player 2 contributes his full endowment (column 2). Column 3 shows that L_{AS} punish deviations from the efficiency norm (C_{II}) significantly more strongly than L_{EQEF}. On average, the difference in punishments is 0.42 Birr or 20 percent of the daily wage, which is quantitatively large.

Table 5—Punishment Patterns within and across Leader Types

	Punishments spent by leaders on player 1 (P_1)		
	(1)	(2)	(3)
Panel A. Punishment within leader types			
Leader type	L_{EQ}	L_{EQEF}	L_{AS}
C_I			−0.575
			(0.421)
C_{II}		−0.629**	−0.208
		(0.283)	(0.149)
C_{III}	−0.172**	−0.143	−0.208
	(0.079)	(0.137)	(0.149)
Constant (C_{IV})	−0.100	0.223*	0.288***
	(0.098)	(0.128)	(0.102)
Observations	84	36	40
Panel B. Punishment across leader types			
Motivation	C_{III}	C_{IV}	C_{II}
L_{EQ}	−0.351***	−0.323**	
	(0.117)	(0.149)	
L_{AS}	−0.000	0.065	0.486*
	(0.185)	(0.145)	(0.260)
Constant (L_{EQEF})	0.080	0.223**	−0.405**
	(0.068)	(0.113)	(0.189)
Observations	66	66	24

Notes: Poisson regression with robust standard errors in parentheses clustered on the leader. In panel A, punishment in C_{IV} is the benchmark category and includes decisions (0, 6) (2, 6) (4, 6). C_I includes (6, 6); C_{II} includes (0, 0) (2, 2) (4, 4); C_{III} includes (0, 2) (0, 4) (2, 4). In panel B, L_{EQEF} is the benchmark type. Coefficients are displayed as semi-elasticity.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

The difference in punishment intensity is also reflected in the average number of punishments spent by the different leader types. On average, L_{AS} punish the most, followed by L_{EQEF} and L_{EQ}, respectively. L_{AS} spent 17.5 Birr on average, of which 8.25 Birr were spent on deviations from the equality norm, 6.5 Birr on deviation from the efficiency norm, and 2.75 Birr on cooperators. In contrast, L_{EQEF} spent on average 11.5 Birr, of which the majority (7.5 Birr) was spent on deviations from the equality norm and the remaining 4 Birr on deviation from the efficiency norm. L_{EQ} punish the least, spending on average 5 Birr exclusively on deviations from the equality norm.

C. *Leader Types and Leader Rating*

One potential concern might be that the characterization of leaders may be an artifact of the experiment. Although this seems doubtful given the procedures we followed and the consistency of leader behavior discussed above, we further dispel this concern with data we collected in the household survey. In the survey, 495 members from all forest user groups were asked to rate their leader on a three-point scale: good, average, and bad. The data show that antisocial leaders have a significantly worse rating relative to the other leader types.

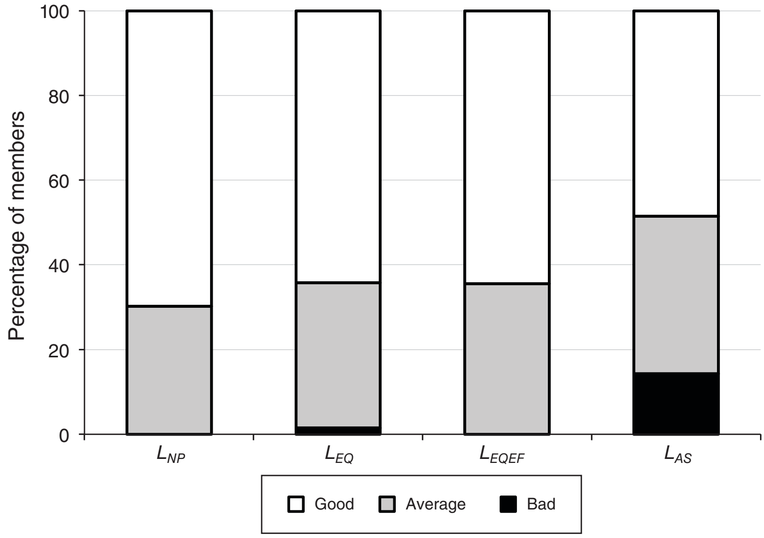

FIGURE 4. LEADER RATING BY GROUP MEMBERS

Notes: Leader rating based on a household survey ($n = 495$), in which members were asked to rate their leaders. The figure shows the percentage of group members who rated their leader into three categories: "bad," "average," and "good." The number of categories was adjusted from five to three because of very few responses in the categories "very good" and "very bad." Category "very good" was merged with "good" and category "very bad" was merged with "bad."

RESULT 3: *In the household survey*, L_{AS} *are significantly more likely to be rated as "bad" by their group members compared to the other leader types.*

Figure 4 illustrates Result 3. Nearly 15 percent of the members interviewed from groups with antisocial leaders rate their leader as bad, in comparison to 1.5 percent for L_{EQ} and none for L_{EQEF} and L_{NP} types. An ordered probit regression of the outcome of the leader rating on a vector of dummies representing a leader's type, controlling for members' and leaders' socioeconomic characteristics, shows that L_{AS} are significantly more likely to be rated as bad than L_{NP} (result not shown). The result establishes an important complementarity between our experimental data and survey measures on leader types. It provides evidence against the argument that punishment behavior of antisocial leaders is merely an artifact of the experiment. Rather, it suggests that in the experiment we pick up an important leader dimension that is related to the negative perception of these leaders.[25]

In the online Appendix, we provide further evidence that our results do not depend on the particular leader characterization by using an alternative approach in which a leader is characterized by the punishment of "less than full contribution" compared to "full contribution." The results confirm our main findings based on the leader characterization above (see online Appendix A.VIII).

[25] It may come as a surprise that not more than 15 percent rate these leaders as bad. However, the finding is consistent with the argument that dominance (rather than prestige) is the pathway through which antisocial leaders attain their leadership position (see the discussion in Section VE).

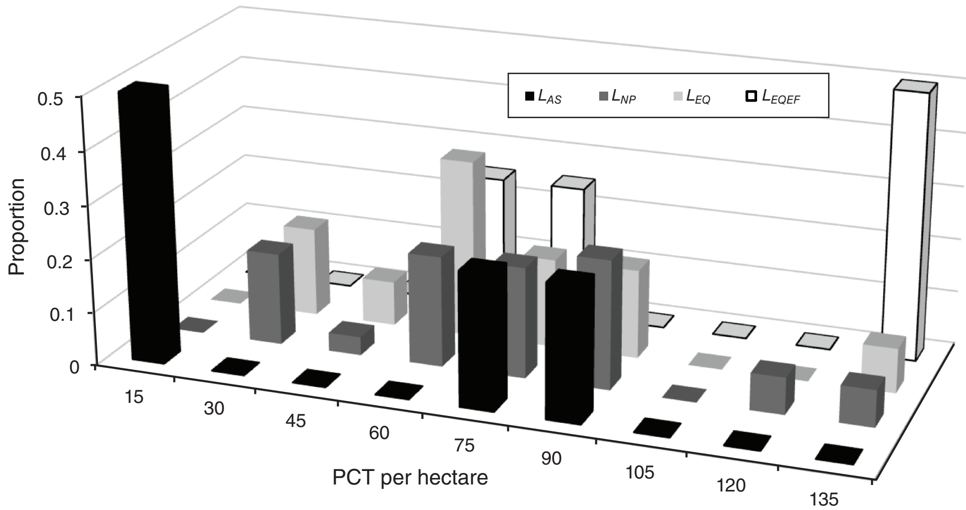

Figure 5. Distribution of Potential Crop Trees (PCT) per Hectare by Leader Type

V. Leader Types and Group Performance

We assess the effect of leader types on group performance using L_{NP} as a benchmark. Given our hypotheses, we expect higher performance in groups led by L_{EQ} and L_{EQEF} and lower performance in groups led by L_{AS}. In addition, we also expect L_{EQEF} groups to perform better than groups led by L_{EQ}. Figure 5 shows the average performance as measured by PCT in groups led by different leader types. It is evident from the figure that there is a large variation in performance across groups led by different leader types. With a mean PCT of 100.12, groups led by L_{EQEF} perform the best, whereas groups with L_{AS} perform the worst, having a mean PCT of 42.84. Groups with L_{NP} and L_{EQ} fall in between and are nearly comparable in performance, each having, on average, about 60 young trees per hectare.

We assess econometrically the effect of variation in a leader's propensity to punish norm violators on group performance. Following previous empirical studies on collective action,[26] we propose the following regression specification augmented by leader variables:

$$(5) \quad PCT_i = \alpha_v + \beta \mathbf{X}_i + \gamma \mathbf{Z}_i + \lambda CC_i + \delta_1 L_{EQ} + \delta_2 L_{EQEF} + \delta_3 L_{AS} + \varepsilon_i,$$

where PCT_i stands for the performance of group i in managing its forest commons as measured by the number of potential crop trees per hectare, α_v are village fixed effects that control for time invariant heterogeneity across villages, $\mathbf{X}_i$ represents a vector of important group-level controls described in Table 1, $\mathbf{Z}_i$ is a vector of leader characteristics, CC_i represents the share of group members with a propensity to cooperate conditional on the cooperation of others, and ε_i is an error term. The remaining three variables in equation (5) capture the effect of equality driven

[26] Cf. Baland and Platteau (1996); Bardhan (2000); Dayton-Johnson (2000); Agrawal (2001); Bandiera, Barankay, and Rasul (2005b); Banerjee, Iyer, and Somanathan (2005); Miguel and Gugerty (2005); Khwaja (2009); and Rustagi, Engel, and Kosfeld (2010).

(L_{EQ}), equality and efficiency driven (L_{EQEF}), and antisocial (L_{AS}) leaders on group performance.

The simplest way to measure the effect of leaders is to model a leader's type as an indicator variable. In this case, the effect of L_{EQ}, L_{EQEF}, and L_{AS} on group performance after controlling for covariates can be expressed in terms of expected values as: $\delta_1 = E[PCT|L_{EQ}] - E[PCT|L_{NP}]$; $\delta_2 = E[PCT|L_{EQEF}] - E[PCT|L_{NP}]$, and $\delta_3 = E[PCT|L_{AS}] - E[PCT|L_{NP}]$. Though simple, the dummy-variable approach has a drawback because it assigns equal weight to leader types within its reference category. However, leaders might differ in their intensity of punishment even within their own reference category. For instance, some leaders might be more equality driven or more antisocial than others. To reflect these differences, we also use the number of punishment points spent by a leader to assign weights to a leader's type. For example, we take for L_{EQ} the total number of punishment points spent by a leader in case the leader punished exclusively deviations from the equality norm, and zero otherwise. For L_{EQEF}, the variable is defined analogously. For L_{AS}, we take the number of punishment points spent by a leader exclusively on players who contribute their full endowment to the public good, and zero otherwise.

A. *Empirical Strategy*

As mentioned in the introduction, there are important concerns, which we need to address before estimating equation (5): reverse causality and omitted variables bias. It is plausible that groups with better outcomes elect good leaders while those with worse outcomes elect bad leaders, leading to reverse causality. However, because all leaders were elected at the beginning of the program, before the outcomes were assessed, this seems unlikely. There is, however, a possibility that unobservable characteristics that are correlated with leader types rather than the leader types themselves are driving the estimated effects. We speculate that this could arise in two forms. First, members may elect leaders along some dimensions, which are correlated with leader types. If these dimensions are related to the outcome, an omitted variable bias could lead to spurious correlation between leader types and performance. Next, members may elect leader types that are similar to theirs. In this case leader types proxy for preference characteristics of the group, and not accounting for them may lead to an upward bias. While, it is difficult to completely rule out endogeneity concerns with data like ours, we briefly present important evidence that alleviates some of the major concerns. All details are discussed in online Appendix A.[27]

First, groups in our setting are similar across many dimensions, including ethnicity, language, religion, occupation, and organization, and nearly identical in geographical conditions and socioeconomic development. These features imply that many factors, which could have otherwise played a role in accounting for

[27] In general, one way to address these concerns is to use appropriate instruments that are correlated with leader types but are orthogonal to performance. However, based on four distinct leader types, this is a very difficult task. Another option is to introduce group-level fixed effects to account for time-invariant omitted variables. This approach requires forest outcome data on all groups for at least two assessments, which we do not have (see Section VD). Even if such data were available, identification of leader effect will be very difficult and possible only if a leader changes his type over time or a leader type turnover occurred. Given that we expect a leader's type to be stable over time, the first option seems unfeasible. The second option is possible only if there is a sufficient gap between leader type turnover and second assessment.

group-level differences in outcomes, are naturally controlled for. Our data confirm this. We show in the online Appendix that groups with different leader types are comparable across observable group-, leader-, and member-level characteristics (Table A3 of online Appendix A). One important mechanism, which causes leader types to be comparable in personal characteristics, is the traditional leader selection process in the Oromo society, which due to shared cultural ties is similar across groups. The Oromo people follow an age-based institution called *Gada*, in which members in the age group of 40–48 years old are considered belonging to the leadership grade. Another characteristic rated important by group members is education because a leader is expected to understand and sign the forest management contract. We show that groups with different leader types do not differ in their pool of individuals from which they could choose their leaders. Further, our data confirm that both education and age are significant predictors of leader election, even after controlling for important other variables such as the leader's wealth or leader's and group members' social types (Table A4 of online Appendix A).

Second, our specifications include village-level fixed effects that absorb any variation in outcomes across groups that are due to village-level differences. Therefore, to the extent that the omitted variables are at the village level, we account for them. Another potential advantage is that groups within a village are relatively homogeneous with regards to characteristics.

Finally, even though we find that groups managed by different leader types are comparable across many dimensions, it is still possible that we may not have included all relevant characteristics, so the possibility of omitted variable bias remains. Following Altonji, Elder, and Taber (2005) and Bellows and Miguel (2009), we test the scope of omitted variable bias by examining the change in the magnitude of leader coefficients in response to the inclusion of control variables (results shown below). If the magnitudes of leader coefficients are robust to the inclusion of powerful and significant controls, it is likely that the effect of leader variables on the outcome is not driven by unobserved variables. In contrast, a substantial attenuation in leader coefficients could imply that the introduction of additional controls may cause further attenuation.

B. *Econometric Results*

We estimate equation (5) using ordinary least squares (OLS) in the cross section first using the weighted measure of a leader's type. Subsequently, we conduct robustness tests by reestimating equation (5) using a leader's type as a dummy variable, different measures of the same control variable, and a host of additional group and leader controls. We also account for possible spatial correlation by clustering standard errors on the village level using wild bootstrap procedures (Cameron, Gelbach, and Miller 2008). Finally, we gauge the importance of how strong the selection on unobservable variables needs to be to explain away the entire effect of leader variables.

RESULT 4: *With respect to non-punishing leaders* (L_{NP}) *as the benchmark, group performance is significantly and positively associated with leaders who enforce both equality and efficiency* (L_{EQEF}), *while the association with leaders who punish* *antisocially (L_{AS}) is negative and highly significant. Performance in groups with leaders who enforce equality only (L_{EQ}) is not significantly different from the benchmark.*

Table 6 shows the association between leader types and group performance. Column 1 includes only the leader variables without any additional controls. With the exception of L_{EQ}, group performance is positively and significantly associated with L_{EQEF} but highly negatively and significantly associated with L_{AS}. A post-regression F-test suggests that the leader variables are both jointly significant ($F(3,47) = 4.82$, $p = 0.005$) and significantly different from each other ($F(2,47) = 7.07$, $p = 0.002$).

In column 2, when we include group-level controls, including the share of conditional cooperators, the coefficients on L_{EQEF} and L_{AS} decline in magnitude but they remain significant. This suggests that the association between group performance and leader types exists over and on top of the association between group performance and conditional cooperation. Other than the leader variables, we also observe a positive association of group performance with the share of conditional cooperators, topographic controls (elevation, historical plantation forest) and time, and a negative association with market distance. The effect of social and economic heterogeneity, as measured by the share of female members and Gini of cattle ownership, is not significantly different from zero. The inclusion of additional controls leads to a jump in the adjusted R^2 from 0.11 in column 1 to 0.74 in column 2, suggesting that control variables are important determinants of group performance. Column 3 introduces village fixed effects to account for village-level differences. This has little effect on the magnitude of leader coefficients and they retain their significance. In contrast, the effect of time disappears and is no longer significant. The village dummies are jointly significant ($p = 0.022$).

In addition to being statistically significant the association between group performance and leader types is also quantitatively important (see column 3). Each punishment point spent by L_{EQEF} leaders on low contributors is associated with an increase in group performance by nearly 2.5 young trees per hectare. Because L_{EQEF} spent on average 11.5 punishment points, a group with a L_{EQEF} leader has on average 28.68 more young trees per hectare than a group with a non-punishing leader (L_{NP}). In contrast, each punishment point spent by L_{AS} leaders on cooperators is associated with a decrease in group performance by over 7.5 young trees per hectare. Since each L_{AS} spent on average 2.75 punishment points on cooperators, this is tantamount to groups with L_{AS} having on average nearly 20.36 young trees less per hectare. These estimates are close to raw differences, which stand at 36.80 trees more in groups with L_{EQEF} and 26.94 trees less in groups with L_{AS}. Given that the average number of young trees per hectare in the sample is 66.79, these effects are very large.

Overall, our model accounts for over 80 percent of the total variation explained in group performance, which is very high for cross-sectional data like ours.

C. *Robustness Analysis*

We adopt a number of approaches to test for the robustness of our results, corroborating the effect of leader punishment on group performance.

Table 6—Leader Types and Group Performance

	No controls	Group level controls	Village fixed effect	Leader type dummies	Vice leaders dropped	Two influential obs. dropped	Leader controls (age, education, clan)	Clan, settlement, and Gini land
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L_{EQ}	−1.186	0.097	−0.638	−6.732	−0.990	0.355	−0.484	0.444
	(1.896)	(1.460)	(1.303)	(7.635)	(1.476)	(1.066)	(1.259)	(1.489)
L_{EQEF}	3.200*	2.349**	2.494***	24.732**	2.739**	2.050***	2.494***	2.932***
	(1.595)	(0.898)	(0.875)	(11.029)	(1.128)	(0.627)	(0.827)	(0.826)
L_{AS}	−9.795***	−6.834***	−7.404***	−19.820**	−8.373**	−8.042***	−8.355***	−9.183***
	(3.315)	(1.809)	(2.396)	(9.105)	(3.032)	(2.495)	(2.329)	(2.500)
CC Share		55.677***	34.624**	27.855*	39.896*	40.032***	38.747***	41.864***
		(13.529)	(15.160)	(15.337)	(21.768)	(13.897)	(13.886)	(14.601)
Elevation		14.179**	21.296**	21.032**	20.343*	23.465***	22.921**	24.689***
		(6.815)	(7.977)	(8.163)	(11.060)	(7.658)	(8.467)	(8.471)
Plantation		21.656***	23.332***	21.554***	23.783***	13.711	22.886***	24.608***
		(4.901)	(6.061)	(6.065)	(6.671)	(15.109)	(5.210)	(6.234)
Group size		0.357	0.522	0.359	0.653	0.870	0.671	1.344
		(0.442)	(0.590)	(0.573)	(0.668)	(0.584)	(0.732)	(0.822)
Market distance		−13.906***	−10.091**	−11.172**	−9.356	−11.180***	−11.473**	−8.599*
		(4.396)	(4.367)	(4.313)	(5.877)	(4.068)	(4.616)	(4.491)
Time		26.334***	2.386	−0.377	5.988	4.965	3.997	4.336
		(5.711)	(10.336)	(10.604)	(12.595)	(9.722)	(10.082)	(9.371)
Social heterogeneity								
Female		15.817	29.286	30.700	23.934	31.208	24.903	10.019
		(18.754)	(22.953)	(22.942)	(27.426)	(25.324)	(27.166)	(26.137)
Clan								3.876
								(15.609)
Settlement								−39.577**
								(19.221)
Economic heterogeneity								
Gini cattle		−40.164	−27.175	−24.128	−8.468	−40.562	−33.110	−44.253
		(33.437)	(31.931)	(33.139)	(37.329)	(31.064)	(34.765)	(32.103)
Gini land								29.767
								(28.609)
Leader characteristics								
Age							0.036	−0.032
							(0.538)	(0.487)
Education							−1.806	−2.316*
							(1.203)	(1.177)
Clan							−4.828	−8.640
							(7.663)	(8.637)
Constant	67.646***	62.692***	89.903***	102.802***	74.607**	79.740***	99.475***	88.530***
	(5.951)	(20.433)	(25.406)	(25.562)	(33.461)	(24.203)	(32.263)	(26.978)
Village fixed effect	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51	51	51	51	44	49	51	51
R^2	0.16	0.79	0.84	0.83	0.85	0.84	0.86	0.88
Adj. R^2	0.11	0.74	0.77	0.75	0.77	0.77	0.78	0.79

Notes: OLS regression with robust standard errors in parentheses. The dependent variable is the average number of young trees per hectare. In models (1–3) and (5–8), leader variables are measured as the actual number of punishments spent in the game. In model (4) leader type is a dummy variable. Group characteristics are as defined in Table 1. Leader age and education are measured in years. Clan is a dummy variable indicating whether the leader belongs to the dominant clan in the group or not. Village fixed effects are dummies for a village to which the group belongs. In column 6, influential observations are assessed using DFITS, which is the difference in fitted values with and without the *i*th observation.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

RESULT 5: *The effect of leader punishment on group performance is robust to alternative specifications, the exclusion of influential observations, and additional leader and group controls. The attenuation in L_{EQEF} and L_{AS} coefficients due to the inclusion of controls is low, suggesting that the selection on unobserved variables needs to be 11 and 15 times stronger than selection on observed variables to explain away the entire effect of leader types on group performance.*

In column 4 of Table 6, we measure leader type as an indicator variable and rerun the model from column 3. The leader variables have the same sign and significance as the results presented in previous columns. Groups with L_{EQEF} leaders have on average 24 more young trees than groups with L_{NP}, whereas groups with L_{AS} have on average 20 young trees less; these are similar to the estimates in column 3. Because the leader-type dummy takes only punishment motives into account, the fact that we arrive at similar estimates to those obtained using punishment motives and punishment intensity together, highlights the importance of punishment motives for the results.

Given the limited number of observations in our sample, we are concerned that vice-leaders or few influential observations might be driving the effects.[28] In column 5, we show that dropping the seven groups in which vice-leaders took part in the experiment has no effect on our results. When we drop two observations[29] that have the most positive and most negative influence (column 6), it has little effect on the magnitude of L_{EQEF} and L_{AS} coefficients. When we drop another two influential observations,[30] this again causes little change in the coefficients on L_{EQEF} and L_{AS} (result not shown). These findings confirm that our results are not due to vice leaders or influential observations.

Next, we test if our results are robust to the inclusion of a variety of additional leader controls. We begin by introducing three important leader variables as covariates in column 7 of Table 6: age, education, and a dummy for whether a leader belongs to the dominant clan in the group or not. As expected, none of the leader variables is significantly associated with group performance, whereas the effect of leader types on group performance remains unchanged. To ensure further that our main results are not due to other leader-level variables, we subsequently control, one by one, for leader characteristics such as ability as rated by group members, demographic characteristics (whether the leader was from the pre-program clan/elder council, number of siblings, birth and wealth order among all children as well as among sons only, number of wives and siblings of wives, and number of children), and physical status (height and chest size). Our results, which are reported in Table A5 of online Appendix A, show that the effects of leader variables remain robust to these additional controls.

Having shown the robustness of our results to a host of leader controls, we turn to additional group-level controls. Thus far, our estimations include only one measure of social and economic heterogeneity at a time. Given the importance of social and economic heterogeneity for cooperation outcomes as suggested by previous

[28] We use DFITS to detect influential observations, which is the difference in fitted values calculated with and without the ith observation. An observation is classified as influential if $|DFIT_i| > 2 \times \sqrt{(k/N)}$, where k is the number of parameters and N is the sample size.

[29] These include one L_{NP} and one L_{EQ}.

[30] These include one L_{EQ} and one L_{EQEF}.

evidence (Alesina and La Ferrara 2000; Baland and Platteau 1996; Dayton-Johnson and Bardhan 2002), we consider additional measures of social heterogeneity (clan and settlement fractionalization) and economic heterogeneity (Gini coefficient of land ownership). Because these measures are not highly correlated with each other, we introduce them at once in column 8. As the results show, this has no effect on the magnitude and significance of leader variables. While the effect of clan, Gini cattle and Gini land is not significantly different from zero, we find a negative and significant effect of settlement heterogeneity, suggesting that the more dispersed a group is the more difficult it is to achieve success in commons management. Our results hold even if we rerun the model with leader type dummies—the resulting coefficients and their respective standard errors on the leader dummies are L_{EQ} (−1.762, SE 8.241), L_{EQEF} (29.247, SE 9.939), and L_{AS} (−25.625, SE 9.922).

We stress-test the sensitivity of our results to further group-level controls and report this in detail in Table A6 of online Appendix A. It is plausible that in addition to clan heterogeneity different clan identities may matter. Therefore, we control for clan differences across groups by introducing, one at a time, dummy variables for each of the three major clans found in the study area. The results show that while the effect of leader types remains robust, the effect of clan dummies is not significantly different from zero. Next, we change the manner in which we account for the effect of conditional cooperators by using a dummy for whether conditional cooperators are in majority in a group and including higher order terms. Again, this has no effect on the significance of either the leader variables or the effect of conditional cooperation. Lastly, we show that our results are robust to additional topographic controls such as the number of perennial and seasonal streams and forest slope.

Because several groups belong to the same village, spatial correlation of standard errors within a village is a possible concern. One option to address this is to cluster standard errors on the village level. Unfortunately, as there are only five villages in our data, the resulting *p*-values will be biased downward leading to a high probability of Type-I error. We therefore follow Cameron, Gelbach, and Miller (2008) and use wild bootstrap procedures, which allow for unbiased estimation even when the cluster size is small. The results show that the effects of L_{EQEF} and L_{AS} remain significant even after taking spatial correlation of errors within a village into account (Table A7 of online Appendix A).

Overall, the robustness of our estimates to alternative specifications and the inclusion of additional controls is remarkable, especially given that the inclusion of additional controls leads to a rise in adjusted R^2 from 0.11 to 0.79 (columns 1 and 8 of Table 6). This means that although the control variables are powerful predictors of group performance, they cause only a small amount of attenuation in the magnitude of leader variables. We use the extent of attenuation in leader variables to gauge the relative importance of omitted variable bias required to explain away the whole leader effect. Comparing leader coefficients in Table 6 without (column 1) and with full controls (column 8) shows that attenuation in L_{EQEF} and L_{AS} coefficients is 0.267 and 0.612, respectively. Given these estimates, the selection on unobserved variables would need to be 11 and 15 times stronger than selection on observed variables to explain away the entire leader effect. Given that the control variables account for a rich set of factors emphasized in the literature on commons management, this seems rather unlikely.

Overall, in the full specification in column 8, leader variables explain 11 percent of the variation in PCT, which is the same as in column 1 without controls.

D. *Additional Results: Second Forest Assessment and Leader Turnover*

So far, our results are based on data from the first forest assessment conducted in 2005 (PCT-I). Up to now, this is the only assessment that is available for all groups. While the second round of forest assessment (PCT-II) had been started, only 25 groups were assessed until January 2013.[31] In this section, we provide preliminary evidence on the association between leadership and group performance for the subsample of groups for which data from the second assessment is available. In addition, we provide information about leader turnover that occurred between our first and second community survey in 2008 and 2013.

Second Forest Assessment.—Of the 25 groups for which PCT-II data is available, 16 are headed by a L_{NP} leader and 9 are headed by a L_{EQ} leader. Unfortunately, no new data is available for groups with either L_{EQEF} or L_{AS} leader. Based on the available data, we test whether there are differences in PCT-II performance between L_{NP} and L_{EQ} leaders, the hypothesis being the same as before, that performance should be higher for L_{EQ} than for L_{NP}. Before doing so, we verify that the subsample of assessed groups is a valid representation of the original sample for a given leader type along important leader- and group-level observables. With the exception of time the program was launched (indicating that older groups were given priority in the second assessment), we find no critical occurrence of selection (Table A8 of online Appendix A).

Results from the second assessment show that while forest outcome has on average declined in L_{NP} groups from 78 trees (PCT-I) to 56 trees (PCT-II), it has increased slightly from 72 to 75 trees in groups with L_{EQ} leaders (Figure 1A of online Appendix A). A nonparametric Wilcoxon matched pairs test confirms that this difference is highly significant for L_{NP} ($p = 0.000$), but not for L_{EQ} ($p = 0.679$). The decline in L_{NP} groups is in line with our hypothesis that non-punishing leaders are less able to uphold cooperation in forest commons management. To test whether differences across the two leader groups are significant, we first reproduce for the subsample of groups our insignificant result from above, i.e., groups with L_{NP} and L_{EQ} leaders do not differ significantly in PCT-I outcomes. Column 1 in Table 7 confirms this. In contrast, column 2 shows that PCT-II outcomes are significantly higher in groups with L_{EQ} leaders. Columns 3–4 present results from a difference-in-differences estimation with group-level fixed effects in which we interact our leader variable with an assessment period dummy. Column 3 shows that L_{EQ} has a positive coefficient that is highly significant. To ensure that this difference is not due to other variables, we control in column 4 for the interaction of the assessment period dummy with important covariates. The results show that the L_{EQ} coefficient decreases slightly but remains statistically significant. Our results

[31] Unfortunately, it is unclear when—and, in fact, whether at all—the remaining groups will be completed. The reason is that the GIZ together with local administrations seem to have changed their strategy from supporting existing groups to implementing new groups (Ameha 2011).

Table 7—Group Performance and L_{EQ} over Time

	PCT-I	PCT-II	DID estimation	
	(1)	(2)	(3)	(4)
L_{EQ}	−0.808	4.209**	5.619***	4.611***
	(1.820)	(1.969)	(1.900)	(1.147)
Constant	89.896**	75.601**	76.027***	75.965***
	(31.593)	(29.002)	(2.243)	(1.786)
Controls	Yes	Yes	No	Yes
Group fixed effect	No	No	Yes	Yes
Observations	25	25	50	50

Notes: Results in columns 1 and 2 are based on OLS regression with robust standard errors in parentheses. Columns 3 and 4 report DID estimates with robust standard errors clustered on the group. Controls include conditional cooperation, group size, market distance, time, female share, and Gini cattle.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

hold even when we use leader dummy as a measure of leader type (not shown). Because L_{EQ} spent on average 4.46 points in the subsample, groups led by L_{EQ} have on average 20.56 more young trees than groups led by L_{NP} in the second assessment. Though this effect is large, note that it falls below the average performance increase of groups led by L_{EQEF} based on PCT-I outcomes (nearly 30 trees). The differential effects highlight, once again, the importance of efficiency motives in leader punishment.

In total, the results from the second forest assessment show that equality-driven leaders are also associated with significant and positive group-performance differences once sufficient time in the assessment of group outcomes is taken into account.

Leader Turnover.—In the 2013 community survey, we gathered information about possible leader turnover across groups. Recall that leaders who participated in our experiment in 2008 had all been elected at the beginning of the program. Our data show that until January 2013, two-thirds of the groups (34 of the 51) had elected a new leader, 11 groups had reelected the same leader, and 6 groups had not held any election. We interviewed each group that had held elections to identify reasons for the election outcome. The answers are listed in detail in Table A9 of online Appendix A. All 11 groups which had reelected the same leader stated that their leader had done a good job. Out of the 34 groups which had elected a new leader, 16 groups explicitly said that their old leader had been dismissed due to "poor performance," e.g., because of no or weak rule enforcement or own rule violation. Six groups had elected a new leader because their old leader had been promoted to a superior position in the village council. Finally, 12 groups explained that their old leader had died, was too old or sick, or quit voluntarily. We summarize these cases as "other reasons."

We hypothesize that reasons for leader turnover are related to a leader's type, with L_{AS} leaders being more likely to be dismissed as a consequence of poor performance and L_{EQEF} and L_{EQ} leaders being more likely to be replaced due to promotion or other reasons. Our data show that this is indeed the case (Figure A2 in online Appendix A). All four L_{AS} leaders were dismissed, three of them due to poor performance, one because he was too old. In contrast, none of the L_{EQEF} leaders were dismissed as a consequence of poor performance. While all these four leaders were replaced as well, this exclusively happened either due to promotion (two leaders) or due to other reasons.[32] For the 7 L_{EQ} leaders and 19 L_{NP} leaders, who were replaced, the picture is somewhat more mixed. Roughly one-half of these leaders were dismissed due to poor performance. The remaining leaders were either promoted or were replaced because of other reasons. Interestingly, the share of promotions is higher among L_{EQ} leaders (29 percent) than among L_{NP} (11 percent).

Overall, the data on leader turnover are remarkably consistent with our results on group performance. While leaders who are associated with positive effects on group performance mostly leave the group due to positive circumstances (i.e., they get promoted), leaders who are associated with negative effects are predominantly replaced due to poor performance. Importantly, when we use leader turnover and turnover reason as proxies for unobserved leader quality, all leader coefficients remain robust and significant while both turnover variables are insignificant (Table A10 of online Appendix A). Together these findings, again, corroborate that our empirical results are not driven by unobserved leader differences but by experimentally revealed heterogeneity in leaders' intrinsic punishment motivation. Interestingly, despite the positive effect of L_{EQEF} and L_{EQ} on group performance and mostly positive reasons behind their turnover, these leaders are not necessarily more likely to stay in power vis-à-vis L_{NP} and L_{AS} types. This brings to mind the importance of path-dependence, the role of other (historical, traditional) factors, and the fact that change does not arise spontaneously (cf. Platteau 2008; Sethi and Somanathan 2008).

E. *Interpreting the Effect of Leader Types*

The empirical results largely support our hypotheses spelled out in Section I. As explained before, one possible reason why L_{EQ} leaders have a relatively weak effect is that these leaders punish indeed *less*, both in terms of number of cases and punishment intensity (cf. Result 2). From the perspective of group members, L_{EQ} leaders thereby turn the social dilemma game into a coordination game, in which every contribution level corresponds to a different Nash equilibrium. In contrast, L_{EQEF} punishment eliminates all Pareto-inferior Nash equilibria and implements full contribution as the unique equilibrium. As long as members in L_{EQ} groups do not have a strong concern for efficiency themselves (which, given that their leaders do not have this, seems rather unlikely), it is actually no surprise that cooperation outcomes are initially close and improve only gradually compared to groups with non-punishing leaders.

Similarly, a possible interpretation for the differential association between L_{EQEF} and L_{AS} leaders and group performance is the manner in which these leaders affect group members' voluntary cooperation, in particular of conditional cooperators. L_{EQEF} leaders, who exclusively and consistently target punishment at low contributors, induce conditional cooperators to cooperate more, while L_{AS} leaders, who punish irrespective of a player's contribution, induce conditional cooperators to

[32] One L_{EQEF} leader was too old, the other quit voluntarily to continue with higher education (cf. Table A9).

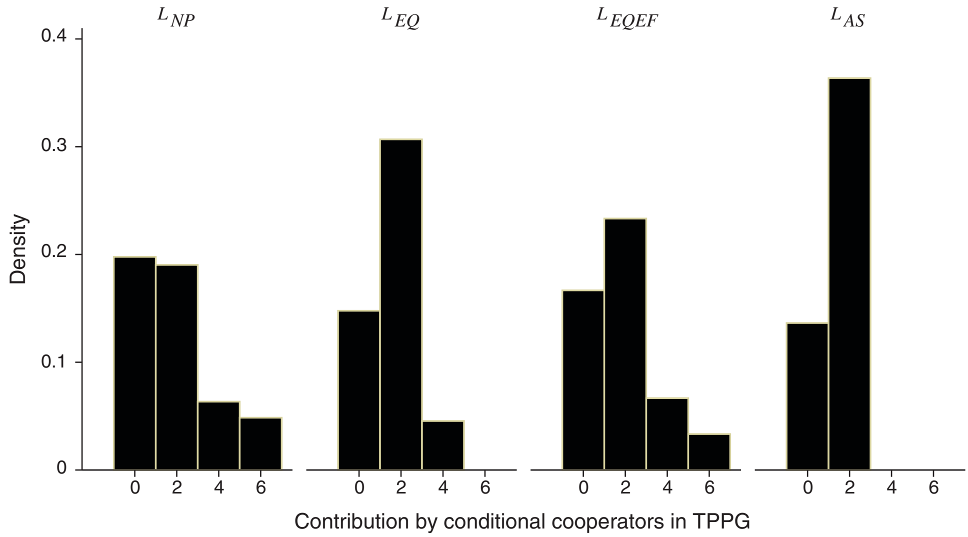

Figure 6. Contribution by Conditional Cooperators in the Third-Party Punishment Game (TPPG)

cooperate less. While it is beyond the scope of this paper to present an in-depth analysis of this mechanism, we can provide preliminary evidence suggesting that it is at work.

Figure 6 shows the histogram of contributions in the TPPG by conditional cooperators from groups with different leader types. As can be seen, a key difference lies in the right tail, which shows contributions above the median (2 Birr). Longer right tails in groups with L_{NP} and L_{EQEF} stand in contrast to the absence of a right tail in groups with L_{AS}, implying that contributions in the latter are significantly lower (Median test, $\chi^2_{(3)} = 9.213$, $p = 0.027$). Thus, conditional cooperators, anticipating that L_{AS} leaders punish players irrespective of their contribution, refrain indeed from contributing higher amounts.

One important question that arises in particular with regard to antisocial leaders in our context is why do groups elect such leaders. Although the data on leader turnover suggests that the problem of antisocial leaders may be limited as no leader is still in power, the large negative association of these leaders with group outcomes nevertheless warrants a deeper analysis of this question. Recent findings in the evolutionary, psychology, and economics literature may help shed some light on this issue (Cheng, Tracy, and Henrich 2010; Goette et al. 2012; Halevy et al. 2012; Hamman, Weber, and Woon 2011). According to this literature, there exist two distinct pathways of attaining leadership in human societies: dominance and prestige. Dominant individuals use intimidation to attain high status and are willing to engage in antisocial behavior to induce fear in their subordinates, paving their way up the dominant hierarchy. Under inter-group competition, dominance is associated with helping in-group members by harming out-group members. For instance, Goette et al. (2012) report that under competition, group membership leads to hostility toward outsiders in the form of antisocial punishment, in which outsiders are punished regardless of whether they cooperate or defect. Hamman, Weber, and Woon (2011) term this behavior "majority tyranny" in which the elected leader from the majority group lets members from the majority group free ride on the contribution of the minority members.

In our field setting, groups comprise multiple clans who may compete for forest resources, providing a possible ground to investigate if perceptions about the punishment behavior of leaders systematically vary with whether an individual belongs to the same clan as the leader or not. In line with the literature above, our data from the household survey in 2008 reveal that the share of members who think that their leader punishes out-clan members more severely than in-clan members is significantly higher for groups with antisocial leaders than for groups with any of the other three leader types ($p < 0.05$), controlling for several leader and member socioeconomic characteristics. Furthermore, in the TPPG conditional cooperators expect to be punished more severely by an antisocial leader if they do not belong to the same clan as the leader ($p < 0.05$), again controlling for socioeconomic characteristics of members and leaders as well as a player's contribution. These experimental results complement the survey findings and suggest that antisocial leaders are likely to punish out-clan members more severely than members of their own clan. Together, the data indicate that antisocial leaders display behaviors typically associated with dominance, which may explain why they were initially elected. Further analysis of these mechanisms represents a key topic for future research.

VI. Conclusion

Cooperation is required for the well functioning of groups and organizations. But, because of cooperation dilemmas, this is often easier said than done. Social scientists advocate that leader punishment can help to solve such dilemmas and improve the prospect of maintaining cooperation. However, despite the prevalence of leader punishment opportunities in many organizations, the variation in group performance is large and persistent.

The aim of this paper is to investigate systematic differences in the propensity of real-world group leaders to enforce cooperation and tie these differences to the performance of groups led by them. We carry out this investigation in the context of a unique field setting in Ethiopia where groups are engaged in the management of common property forests. Members in these groups confront cooperation dilemmas while managing their forest commons and rely on group leaders to punish norm violators. We show that groups are comparable across many group-, member-, and leader-specific dimensions, allowing us to associate differences in group performance to differences in leader enforcement.

Our analysis consists of two main steps. First, we engage leaders in a third-party punishment game in which we control for self-regarding and reputational motives to measure a leader's propensity to enforce cooperation norms due to equality, efficiency, and antisocial motives. Based on this measure, we identify four different leader types: if a leader exclusively punishes members who deviate from the equality norm, we label him equality driven; if a leader punishes deviations from both equality and efficiency norms, we identify him as equality and efficiency driven; in contrast, leaders who punish irrespective of norm deviations are labeled antisocial, while leaders who do not punish are classified as non-punishers. Our experiment shows that real-world leaders make remarkably different decisions in an identical context.

Second, we show that leader types described by the experiment correlate with a range of real-world outcomes—success in forest commons management, perception of the leader, and leader reelection. Our main result shows that while group performance is positively associated with equality and efficiency driven leaders, the association with antisocial leaders is highly negative. The association with pure equality driven leaders is not significantly different from zero if we consider first-round forest assessments, but is positive and significant if we consider second-round assessments. Exploring plausible interpretations behind these differential associations, we show that in groups with antisocial leaders, conditional cooperators contribute significantly less to the public good than in other groups. Additional analyses show that antisocial leaders are more likely to be rated poorly and removed for poor performance by group members.

Our findings have important implications for the design of organizational and development policies. Given a recent trend in decentralization the world over, local leadership has assumed even greater importance. Nowadays, many environment and development programs rely on local institutions (Chattopadhyay and Duflo 2004; Björkman and Svensson 2009) in which leaders could possibly do much more harm than help. Standard practices of eliminating antisocial leaders by holding reelections and by spreading powers over an executive committee may or may not mitigate the detrimental effects of antisocial leadership. In many developing countries, local elections and power sharing are marred by elite capture. One option is to make it much more costly for antisocial leaders to punish norm conformers (see Falk, Fehr, and Fischbacher 2005). Development agencies need to find pioneering approaches to put this into practice. Clearly, this is another important area that future research must address.

REFERENCES

Agrawal, Arun. 2001. "Common Property Institutions and Sustainable Governance of Resources." *World Development* 29 (10): 1649–72.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (5): 777–95.

Alesina, Alberto, and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *Quarterly Journal of Economics* 115 (3): 847–904.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

Ameha, Aklilu. 2011. "Performance of Old PFM Sites in Adaba-Dodola, Chilmo, Borana, and Bonga Sites: Final Report." Addis Ababa: Farm Africa.

Amente, Girma. 2005. "Rehabilitation and Sustainable Use of Degraded Community Forests in the Bale Mountains of Ethiopia." PhD thesis. University of Freiburg, Faculty of Forest and Environmental Sciences.

Baland, Jean-Marie and Jean-Philippe Platteau. 1996. *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford: Oxford University Press.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005a. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics* 120 (3): 917–62.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005b. "Cooperation in Collective Action." *Economics of Transition* 13 (3): 473–98.

Banerjee, Abhijit V., Lakshmi Iyer, and Rohini Somanathan. 2005. "History, Social Divisions and Public Goods in Rural India." *Journal of the European Economic Association* 3 (2–3): 639–47.

Bardhan, Pranab K. 2000. "Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India." *Economic Development and Cultural Change* 48 (4): 847–65.

Barr, Abigail, Chris Wallace, Jean Ensminger, Joseph Henrich, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cárdenas et al. 2009. "Homo Æqualis: A Cross-Society Experimental Analysis of Three Bargaining Games." University of Oxford CSAE Working Paper 2009-02.

Bellows, John, and Edward Miguel. 2009. "War and Local Collective Action in Sierra Leone." *Journal of Public Economics* 93 (11-12): 1144–57.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial Altruism in Humans." *Nature* 442 (7105): 912–15.

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics* 118 (4): 1169–1208.

Bianco, William T., and Robert H. Bates. 1990. "Leadership, Structure, and Collective Dilemmas." *American Political Science Review* 84 (1): 133–47.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124 (2): 735–69.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122 (4): 1351–1408.

Bowles, Samuel, and Sandra Polonía-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature* 50 (2): 368–425.

Camerer, Colin F. Forthcoming. "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Reply to Levitt and List." In *The Methods of Modern Experimental Economics,* edited by Guillaume R. Fréchette and Andrew Schotter. Oxford: Oxford University Press.

Cameron A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–27.

Cameron, A. Colin, and Pravin K. Trivedi. 2009. *Microeconometrics Using Stata*. College Station, TX: Stata Press.

Carpenter, Jeffrey P., and Peter H. Matthews. 2012. "Norm Enforcement: Anger, Indignation or Reciprocity?" *Journal of the European Economic Association* 10 (3): 555–72.

Chanakya. *Arthashastra*. Translated by R. Shamasastry, 1915. Bangalore Government Press.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–69.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–43.

Cheng, Joey T., Jessica L. Tracy, and Joseph Henrich. 2010. "Pride, Personality, and the Evolutionary Foundations of Human Status." *Evolution and Human Behavior* 31 (5): 334–37.

Cinyabuguma, Matthias, Talbot Page, and Louis Putterman. 2006. "Can Second-Order Punishment Deter Perverse Punishment." *Experimental Economics* 9 (3): 265–79.

Dawes, Christopher T., James H. Fowler, Tim Johnson, Richard McElreath, and Oleg Smirnov. 2007. "Egalitarian Motives in Humans." *Nature* 446 (7137): 794–96.

Dayton-Johnson, Jeff. 2000. "The Determinants of Collective Action on the Local Commons: A Model with Evidence from Mexico." *Journal of Development Economics* 62 (1): 181–208.

Dayton-Johnson, Jeff, and Pranab K. Bardhan. 2002. "Inequality And Conservation on the Local Commons: A Theoretical Exercise." *The Economic Journal* 112 (481): 577–602.

Engelmann, Dirk, and Martin Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* 94 (4): 857–69.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2005. "Driving Forces behind Informal Sanctions." *Econometrica* 73 (6): 2017–30.

Falk, Armin, and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952): 535–38.

Fehr, Ernst, and Urs Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25 (2): 63–87.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90 (4): 980–94.

Fehr, Ernst, Karla Hoff, and Mayuresh Kshetramade. 2008. "Spite and Development." *American Economic Review* 98 (2): 494–99.

Fehr, Ernst, and John List. 2004. "The Hidden Costs and Returns of Incentives: Trust and Trustworthiness among CEOs." *Journal of the European Economic Association* 2 (5): 743–71.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–68.

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

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

Gächter, Simon, and Benedikt Herrmann. 2009. "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment." *Philosophical Transactions of the Royal Society: Biological Sciences* 364 (1518): 791–806.

Gächter, Simon, and Benedikt Herrmann. 2011. "The Limits of Self-Governance when Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia." *European Economic Review* 55 (2): 193–210.

Gächter, Simon, and Christian Thöni. 2005. "Social Learning and Voluntary Cooperation among Like-Minded People." *Journal of the European Economic Association* 3 (2-3): 303–14.

Gächter, Simon, and Christian Thöni. 2011. "Micromotives, Microstructure, and Macrobehavior: The Case of Voluntary Cooperation." *Journal of Mathematical Sociology* 35 (1–3): 26–65.

Gittell, Jody H. 2000. "Organizing Work to Support Relational Coordination." *International Journal of Human Resource Management* 11 (3): 517–39.

Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter. 2012 "Competition between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations." *Management Science* 58 (5): 948–960.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312 (5770): 108–11.

Halevy, Nir, Eileen Y. Chou, Taya R. Cohen, and Robert W. Livingston. 2012. "Status Conferral in Intergroup Social Dilemmas: Behavioral Antecedents and Consequences of Prestige and Dominance." *Journal of Personality and Social Psychology* 102 (2): 351–66.

Hamman, John R., Roberto A. Weber, and Jonathan Woon. 2011. "An Experimental Investigation of Electoral Delegation and the Provision of Public Goods." *American Journal of Political Science* 55 (4): 737–51.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cárdenas et al. 2006. "Costly Punishment across Human Societies." *Science* 312 (5781): 1767–70.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment across Societies." *Science* 319 (5868): 1362–67.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics* 13 (2): 324–40.

Irwin, Kyle, and Christine Horne. 2013. "A Normative Explanation of Antisocial Punishment." *Social Science Research* 42 (2): 562–70.

Jones, Benjamin F., and Benjamin A. Olken. 2005. "Do Leaders Matter? National Leadership and Growth since World War II." *Quarterly Journal of Economics* 120 (3): 835–64.

Karlan, Dean S. 2005. "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." *American Economic Review* 95 (5): 1688–99.

Khwaja, Asim I. 2009. "Can Good Projects Succeed in Bad Communities?" *Journal of Public Economics* 93 (7–8): 899–916.

Kidan, Seyoum G., and Tsegaye Tadesse. 2004. "A Model Bylaw for Wajib." Addis Ababa: GIZ.

Kosfeld, Michael, and Devesh Rustagi. 2015. "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.20120700.

Kubsa, Abdurahiman, and Tsegaye Tadesse. 2002. "Granting Exclusive User Rights to the Forest Dwellers in the State-Owned Forest: The Wajib Approach in Ethiopia." Paper presented at the Second International Workshop On Participatory Forestry in Africa, FAO, Arusha, United Republic of Tanzania.

Lazear, Edward P. 2012. "Leadership: A Personnel Economics Approach." *Labour Economics* 19 (1): 92–101.

Levitt, Steve D., and John List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World." *Journal of Economic Perspectives* 21 (2): 153–74.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance* 60 (6): 2661–2700.

Malmendier, Ulrike, and Geoffrey Tate. 2009. "Superstar CEOs." *Quarterly Journal of Economics* 124 (4): 1593–1638.

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review* 93 (1): 366–80.

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.

Nikiforakis, Nikos. 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics* 92 (1–2): 91–112.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: 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.

Platteau, Jean-Philippe, and Erika Seki. 2007. “Heterogeneity, Social Esteem, and Feasibility of Collective Action.” *Journal of Development Economics* 83 (2): 302–25.

Platteau, Jean-Philippe. 2008. “Managing the Commons: The Role of Social Norms and Beliefs.” In *The Contested Commons: Conversations between Economists and Anthropologists*, edited by Pranab K. Bardhan and Isha Ray, 25–45. Oxford: Blackwell.

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld. 2010. “Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management.” *Science* 330 (6006): 961–65.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy.” *Science* 277 (5328): 918–24.

Sethi, Rajiv, and E. Somanathan. 2008. “A Simple Model of Collective Action.” In *The Contested Commons: Conversations between Economists and Anthropologists*, edited by Pranab K. Bardhan and Isha Ray, 125–43. Oxford: Blackwell.

Tang, Shui-Yan. 1992. *Institutions and Collective Action: Self-Governance in Irrigation Systems.* San Francisco: ICS Press.

Trainer, J. 1996. “Forest Inventory of Indigenous Forest of Pilot Area Peasant Associations: Denaba, Bura, Burachele-Adele.” Addis Ababa: Integrated Forest Management Project Adaba-Dodola.

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