ONLINE APPENDIX

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

By MICHAEL KOSFELD AND DEVESH RUSTAGI

APPENDIX A

I. ADDITIONAL RESULTS ON LEADER PUNISHMENT

To disentangle the effect of the equality and the efficiency motive on leader punishment, we cluster the ten contribution decisions of players 1 and 2 into four categories (Table A1).

TABLE A1: MOTIVES FOR PUNISHMENT

Category	Contributions (C_1, C_2)	Player 1 violates EF norm	Player 1 violates EQ norm	Player 2 violates EF norm
C_{I}	(6,6)	No	No	No
C_{II}	(0,0) (2,2) (4,4)	Yes	No	Yes
C_{III}	(0,2) (0,4) (2,4)	Yes	Yes	Yes
C_{IV}	(0,6) (2,6) (4,6)	Yes	Yes	No

Notes. EF stands for efficiency, EQ for equality.

We then regress expected punishment Y_{li} on these categories:

$$Y_{li} = \exp(\beta_1 + \beta_2 C_{II} + \beta_3 C_{III} + \beta_4 C_{IV}). \quad \text{(A1)}$$

The coefficients in equation (A1) can be expressed as follows: (a) β_1 measures the expected value of punishment in the benchmark category, C_I, which is punishment due to anti-social motives; (b) $\beta_2 = E(Y_1 \mid C_{II}) - E(Y_1 \mid C_I)$ measures a change in the expected value of punishment when both players deviate from the efficiency norm; (c) $\beta_3 = E(Y_1 \mid C_{III}) - E(Y_1 \mid C_I)$ measures a change in the expected value of punishment when player 1 violates both efficiency and equality norms, and player 2 violates the efficiency norm, and (d) $\beta_4 = E(Y_1 \mid C_{IV}) - E(Y_1 \mid C_I)$ measures a change in the expected value of punishment when player 1 violates both efficiency and equality norms, but player 2 does not violate the efficiency norm. As before, we cluster standard errors on the leader.

Quantitatively, we can measure the effect of deviating from the equality norm as the difference in the expected value of punishment in categories C_{III} and C_{II}. These categories differ only in two aspects: the extent of deviation by player 1 from the efficiency norm and whether player 1 deviates from the equality norm or not. Because the former has no significant effect on the leader punishment behavior (Result 1), the difference between the two categories must be due to deviation from the equality norm. The difference in the expected value of punishment between C_{III} and C_{IV} allows us to measure the impact of equality motives when player 2 does and does not deviate from the efficiency norm. Thus, $(\beta_3 - \beta_2) = E(Y_1 \mid C_{III}) - E(Y_1 \mid C_{II})$ measures a change in the expected value of punishment due to deviation from the equality norm when player 2 deviates from the efficiency norm, whereas $(\beta_4 - \beta_3) = E(Y_1 \mid C_{IV}) - E(Y_1 \mid C_{III})$ measures a change in the expected value of punishment due to deviation from the equality norm when player 2 does not deviate from the efficiency norm.

Table A2 quantifies the effect of each motive on leader punishment behavior. Columns 1-2 show that average punishment of player 1 increases by 0.078 Birr when he deviates from the efficiency norm and 0.32 Birr when, in addition, he also deviates from the equality norm. The difference between the two coefficients $\left(\hat{\beta}_3 - \hat{\beta}_2\right)$ shows that, on average, a deviation from the equality norm causes an increase in player 1's punishment by 0.24 Birr, when player 2 deviates from the efficiency norm. A Wald test shows that the difference between the two coefficients is highly significant ($\chi^2_{(1)} = 13.14$, $p = 0.000$). Further, $\left(\hat{\beta}_4 - \hat{\beta}_3\right)$ shows that punishment of player 1 increases additionally by 0.072 Birr when player 2 does not deviate from the efficiency norm (Wald test, $\chi^2_{(1)} = 9.05$, $p = 0.003$). These results highlight that quantitatively, deviating from the equality norm has the strongest effect on leader punishment, which is at least 3 times the effect of deviation from the efficiency norm. Columns 3-4 reveal that player 2 is punished in the same way as player 1 in case both players contribute the same amount but somewhat less (although not significantly) if player 1 contributes less than player 2.

TABLE A2: MOTIVES AND LEADER PUNISHMENT

	P_1		P_2	
	(1)	(2)	(3)	(4)
	Coefficient	AME	Coefficient	AME
C_{II}	0.847*	0.078*	0.847*	0.078*
	(0.481)	(0.044)	(0.481)	(0.042)
C_{III}	1.863***	0.320***	-0.405	-0.020
	(0.546)	(0.067)	(0.790)	(0.039)
C_{IV}	2.037***	0.392***	-0.588	-0.026
	(0.532)	(0.076)	(0.680)	(0.031)
Constant (C_I)	-2.833***		-2.833***	
	(0.566)		(0.566)	
Observations	510	510	510	510

Notes. Poisson regression with robust standard errors in parentheses clustered on the leader. The dependent variable is the punishment of player 1 and 2, respectively. C_I (6,6) is the benchmark category; C_{II} includes (0,0) (2,2) (4,4); C_{III} includes (0,2) (0,4) (2,4); C_{IV} includes (0,6) (2,6) (4,6). AME is average marginal effect. *** $p < 0.01$, * $p < 0.1$

II. CONTROL VARIABLES

All the 51 groups included in this study are inhabited by the Bale Oromo people who are linguistically, occupationally, and religiously very similar. 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. Variations in weather conditions occur only across elevation, which ranges roughly from 2600 to 3500 meters. Because weather conditions are correlated with elevation, we expect elevation to be an important factor in mediating the effect of weather on forest growth. Nearly 24 percent of the 51 groups are located above 3000 meters.

Groups also vary in forest type but this, almost exclusively, along elevation. At lower altitudes (up to 2800 m), the forest comprises mainly of *Juniperus excelsa* and *Podocarpus falcatus,* at the middle altitude (2800-3000) it is mostly mixed comprising *Hagenia abyssinica, Hypericum lanceolatum* and *Erica arborea,* with the latter being the most dominant species at altitudes above 3000m. Therefore, the variable *elevation* is a proxy for altitude as well as forest type.

Other than altitude, forest types vary in whether the forest is natural or planted (Eucalyptus and Cyprus trees). The variable 'plantation' captures this and is measured as the ratio of area under historical plantation forest to area under natural forest.

The group size varies from 16 to 30 members, with 40 percent of the groups having 30 members. The criteria for group membership are rigid and include customary right of forest use in the pre-program period as well as long-term residence in an area. As a result of this, members cannot sort themselves into groups of their choice, thus ruling out endogenous sorting. Intra-group migration is possible only through marriage, and because Bale Oromo is a patriarchal society, only women migrate across groups.

In many groups, forest usage rules allow members to sell forest products in weekly markets. Profitability of this trading opportunity, however, depends on the distance of a group from the nearest market. Access to markets is much better in groups located at lower elevations and near the main towns, such as Adaba and Dodola. By contrast, for groups at higher elevations, it takes members nearly four hours to reach the nearest market. This has strong consequences on the extent to which members can earn cash income by selling forest products in the market. Even though market access varies by elevation (Mann-Whitney $z = -2.966$, $p = 0.003$), we do not detect multicollinearity in any of our models. We measure market distance as the average time (number of hours) it takes for a group member to reach the nearest market. It is a weighted measure that takes equine ownership into consideration.

The program was launched in two phases over a period of five years. Accordingly, we treat time as a dummy variable (0 for groups that joined the program between 2003 and 2005, and 1 for groups that joined the program between 2000 and 2002). Half of the groups were established in the first phase of the program.

Socially, groups vary in the share of female members, clan composition, and the number of settlement clusters. The share of female members varies from 0 to 67 percent, with 3 percent of the groups having no female members. Groups also differ in clan composition. While some groups are homogeneous, others have members from different clans. Following previous studies (Alesina and La Ferrara 2000, Miguel and Gugerty 2005), we use an index of fractionalization to measure heterogeneity, which is given by $1 - \Sigma (p_i)^2$, where p_i denotes the proportion of total population from the i^{th} clan. The index varies from 0 to 1 and measures the probability that two persons selected randomly from a group will not be from the same clan. Groups also vary in their settlement fragmentation. Most groups are organized into small hamlets, which may or may not lie adjacent to each other. Settlement heterogeneity is measured analogously to clan heterogeneity, but the two are not significantly correlated ($\rho = 0.035$, $p = 0.806$). The settlement fragmentation is not along clan lines.

Another form of heterogeneity is related to wealth. In a herding society like ours, livestock rearing, particularly cattle ownership is considered an important economic activity and an indicator of wealth. A secondary indicator of wealth is land, which is mainly used for subsistence farming. We use the Gini index of cattle and land ownership to measure inequality; the correlation between the two is positive but insignificant ($\rho = 0.188$, $p = 0.186$).

Table A3 reports summary statistics on observable group-, leader- and member-level characteristics, as well as p-values for a test of the hypothesis that the difference in these characteristics across leader types is zero. With the exception of elevation, family size, cattle, and land holding, there are no differences across groups with different leader types either in the mean or the median.[1] Thus groups with different leader types mostly balance along covariates.

[1] While median cattle and land holding are significant at 10%, elevation and family size are significant at 5%. We verify that our results are robust to controlling for these variables.

TABLE A3: LEADER, GROUP AND MEMBER CHARACTERISTICS BY LEADER TYPES

Variable	L_{NP}				L_{EQ}				L_{EQEF}				L_{AS}				ANOVA	KW
	Obs.	Mean	Med	Std.	Obs.	Mean	Med	Std.	Obs.	Mean	Med	Std.	Obs.	Mean	Med	Std.	(mean)	(median)
Leader (included)																		
Age (years)	29	41.21	40.00	7.36	14	42.29	40.00	6.83	4	35.50	36.00	4.12	4	42.75	43.50	14.77	0.466	0.326
Education (years)	29	4.48	3.00	2.82	14	5.29	5.50	2.92	4	4.250	5.50	3.00	4	4.25	3.00	2.50	0.807	0.772
Dominant clan (%)	29		86.20				86.00				100.00				50		Fisher = 0.289	
Group (included)																		
CC share	29	0.35	0.24	0.23	14	0.29	0.32	0.23	4	0.30	0.36	0.16	4	0.51	0.55	0.10	0.302	0.259
Elevation = 1 (%)	29		37.93		14		0.00		4		25.00		4		0.00		Fisher = 0.017**	
Plantation	29	0.11	0.00	0.44	14	0.02	0.00	0.06	4	0.00	0.00	0.00	4	0.02	0.00	0.04	0.805	0.638
Group size	29	26.66	29.00	4.13	14	26.07	29.50	5.53	4	29.00	30.00	2.00	4	23.75	25	5.50	0.431	0.268
Market distance	29	2.56	2.50	0.87	14	1.92	1.89	0.61	4	1.96	1.75	1.28	4	2.52	2.79	1.00	0.109	0.132
Time = 1 (%)	29		55.17		14		35.71		4		0.75		4		0.25		Fisher = 0.355	
Female share	29	0.22	0.20	0.13	14	0.17	0.16	0.08	4	0.18	0.18	0.08	4	0.19	0.24	0.13	0.724	0.624
Clan	29	0.37	0.50	0.25	14	0.42	0.41	0.23	4	0.27	0.28	0.07	4	0.55	0.52	0.16	0.357	0.342
Settlement	29	0.52	0.57	0.19	14	0.56	0.63	0.19	4	0.64	0.71	0.17	4	0.63	0.62	0.11	0.478	0.251
Gini cattle	29	0.34	0.34	0.09	14	0.32	0.33	0.07	4	0.34	0.36	0.07	4	0.35	0.35	0.05	0.939	0.969
Gini land	29	0.29	0.29	0.07	14	0.31	0.30	0.08	4	0.29	0.28	0.10	4	0.35	0.36	0.05	0.470	0.390
Group (excluded)																		
Free rider share	29	0.13	0.06	0.20	14	0.10	0.03	0.14	4	0.09	0.04	0.13	4	0.05	0.04	0.06	0.811	0.963
Total population	29	176.79	167.00	64.10	14	194.14	183.5	65.07	4	176.5	169.00	65.14	4	177.75	184.50	49.78	0.861	0.783
Forest per member	29	19.46	2.80	10.98	14	13.88	1.74	3.65	4	13.27	2.38	2.02	4	12.43	1.52	0.39	0.130	0.198
Member (excluded)																		
Age	356	35.99	35.00	13.51	170	37.99	35.50	14.58	48	36.63	35.00	12.04	42	37.26	34.00	15.10	0.477	0.544
Education	356	3.25	3.00	3.06	169	3.11	3.00	3.09	46	3.13	3.00	2.94	42	3.69	3.00	3.67	0.747	0.873
Family size	350	8.07	7.00	3.81	169	9.07	8.00	4.47	47	7.62	6.00	4.70	43	8.49	8.00	4.31	0.041**	0.02**
Cattle ownership	356	11.94	9.00	11.40	169	10.34	9.08	8.89	48	8.58	7.00	6.55	43	10.98	9.00	8.38	0.104	0.055*
Horse ownership	357	2.68	2.00	2.12	169	2.45	2.00	2.04	48	2.08	1.50	2.08	43	2.44	2.00	1.58	0.233	0.103
Land ownership	357	1.44	1.00	1.08	170	1.59	1.25	1.18	48	1.33	1.00	1.02	43	1.24	1.00	1.16	0.187	0.070*

Notes. ANOVA is used to test for differences in means and Kruskal-Wallis for differences is medians for key variables across different leader types. Included and excluded refers to whether variables are used in the econometric models in Table 6, main paper. Dominant clan is measured as the share of leaders from within a type who belong to the clan with the largest share in the group led by them. CC share is the share of group members with a propensity to cooperate conditional on the contribution of others in a public goods game (Rustagi et al., 2010). Elevation is a dummy variable, with 1 indicating that the group managed forest is located above 3000 meters, and 0 that 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. 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. Female is the proportion of female members in a group. Clan and settlement are indicators of social heterogeneity measured using a fractionalization index. Gini cattle and Gini land are indicators of economic heterogeneity and are measured using a Gini index. Free rider share is the proportion of members in a group with a propensity to contribute zero in the public goods game (Rustagi et al., 2010). Total population is the population of a group. Forest per capita is size of a forest in hectares/group size. Member characteristics exclude the group leader. ** $p < 0.05$, * $p < 0.1$.

III. LEADER ELECTION

As described in the paper, one mechanism that causes leaders to be comparable in personal characteristics is the traditional leader selection process in the Oromo society in which men in the age group of 40-48 (called *Gada*) are selected as leaders. Another is the need for educated leaders who are able to read, understand, and sign the forest management contract. We first test whether groups with different leader types differ in their pool of individuals from which they could choose leaders. A Kruskal-Wallis test confirms that in the sample comprising only men neither age ($\chi^2_{(3)} = 2.21$, p = 0.53) nor education ($\chi^2_{(3)} = 0.47$, p = 0.92) differ across groups with different leader types. We repeat the above test by restricting our sample further to men from the age category from which leaders are elected (40-48 years) and find similar results for age ($\chi^2_{(3)} = 1.393$, p = 0.71) and education ($\chi^2_{(3)} = 3.032$, p = 0.39). Thus, it seems unlikely that groups led by different leader types are disadvantaged in selecting leaders on key characteristics, such as age and education.

We proceed by showing that age and education are indeed key criteria on which leaders get selected. In Table 2 of the main paper, we see that, on average, leaders are both older and more educated than group members. We now confirm this using a probit model with village fixed effects in which the age-group <40 years is the baseline category. The results, reported in columns 1-2 of Table A4 show that *Gada*-age category and education have a significantly positive effect on the probability of getting elected as a leader, even after controlling for important variables, such as cattle and land ownership, in which leaders differ from members (see Table 2, main paper). The coefficients on age and education remain steady across different specifications.

A related concern is that members might elect leaders of similar types. We gauge the importance of this concern by controlling for the type of an individual in the probit model described above.[2] Because we do not have data on the punishment behavior of members, we measure an individual's type with respect to conditional cooperation, for which we have data on both leaders and members. The assumption is that if leader election is not conditional on an individual's cooperative type, it is also unlikely to be conditional on an individual's punishment type. Adding conditional cooperation to the probit model does not change our results. While the effect of conditional cooperation is not significantly different from zero, age and education still have a significantly positive effect on the likelihood of getting elected as a leader (column 3,

[2] Table 2 (main paper) shows that neither the proportion of conditional cooperators nor the coefficient of conditional cooperation differs across leaders and members.

Table A4). Our results with respect to age and education are robust when we re-run the model separately for groups with higher and lower than the median share of conditional cooperators (columns 4-5). Together these results suggest that traditional criteria that can be seen as exogenous to groups play an important role in the selection of a leader.

TABLE A4: CORRELATES OF LEADER ELECTION

				CC share< Median	CC share > Median
	(1)	(2)	(3)	(4)	(5)
Age: Gada (40-48)	0.844***	0.867***	0.882***	0.964***	0.957***
	(0.164)	(0.178)	(0.179)	(0.263)	(0.243)
Age: above Gada (> 48)	0.364*	0.377	0.369	0.136	0.647**
	(0.213)	(0.233)	(0.236)	(0.371)	(0.320)
Education	0.095***	0.096***	0.092***	0.091**	0.105***
	(0.022)	(0.024)	(0.024)	(0.036)	(0.032)
Cattle holding		0.008	0.010	-0.000	0.020**
		(0.007)	(0.007)	(0.008)	(0.009)
Land holding		0.030	0.016	0.258***	-0.159*
		(0.072)	(0.072)	(0.090)	(0.091)
Conditional cooperator dummy			-0.246	-0.503	-0.151
			(0.172)	(0.387)	(0.219)
Village fixed effect	Yes	Yes	Yes	Yes	Yes
Constant	-2.122***	-2.454***	-2.299***	-2.710***	-2.248***
	(0.173)	(0.223)	(0.226)	(0.319)	(0.290)
Observations	662	657	636	309	327

Notes. Probit regression with robust standard errors clustered on the group in parentheses. The baseline category is less than 40 years.

IV. ROBUSTNESS TESTS

Below we report robustness tests which we conducted using a host of leader- and group-level variables. We control, one at a time, for various leader variables including leader ability, physical status and demographic characteristics. All analyses confirm that it is the leader type and not something correlated with types, which explains differences in group performance.

We assess leader ability using a community survey administered in 2012-13. In the survey, we asked members to rate their leader's ability (in his absence) relative to other group members on a three-point scale (below group members, same as groups members, and higher than group members) in three important economic domains (agriculture, livestock rearing, and forest management). We find that L_{EQEF} perform the best, L_{AS} the worst, and L_{EQ} trail slightly behind L_{NP}. These results are in line with our main result on the effect of leader types on group performance. We conduct a robustness check by explicitly controlling for leader ability as an additional control in table A5. The coefficients on leader variables changes little and they remain significant, but ability has no effect over and on top of leader types.

Next, it could be that there are other confounding leader variables that are driving our main results. We therefore introduce one at a time a host of leader controls including whether the leader was from the pre-program clan/elder council, number of leader siblings, birth order among all children and among sons only, wealth order among all children and sons only, number of wives, number of siblings of wives, number of children, leader's height and chest. Data on these variables was also collected using the community survey administered in 2012-13. Irrespective of the additional controls, we find that the effect of leader types remains robust and significant, whereas none of the additional leader controls are statistically significant.

Finally, we analyze the effect of additional group-level controls. In columns 1-4 of Table A6, we add dummies for the three major clans in the study area (labeled as Clans 1, 2 and 3) in addition to the clan fractionalization index. Clan-1 is a dummy variable with 1 indicating that a group has members from this clan, otherwise zero. Clan-2 and Clan-3 are analogously defined. The results show that while the effect of leader types remains robust, none of the clan dummies is statistically significant. Columns 5-6 analyze different specifications for the influence of conditional cooperators. CC majority is a dummy variable, with 1 indicating that conditional cooperators are in a majority in a group, otherwise zero. CC higher order terms include linear and quadratic terms. Obviously, the effects of leaders and of conditional cooperators do not depend on the specification. Lastly, we consider additional topographic controls. Streams include two variables: number of perennial and seasonal streams in a group-managed forest. Slope is the steepness of the group-managed forest. Both variables were assessed using the community survey in 2012-2013. Columns 7-8 show that our results are robust to the inclusion of these variables.

TABLE A5: FOREST MANAGEMENT OUTCOME, LEADER TYPES, AND ADDITIONAL LEADER CONTROLS

	Leader ability	Council	No. of siblings	Birth-order	Birth-order: sons only	Wealth order	Wealth order: sons only	No. of wives	No. of siblings of wife	No. of children	Height	Chest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
L_{EQ}	0.426	-0.322	0.536	0.439	0.601	0.535	0.560	0.442	0.587	0.462	0.551	0.894
	(1.505)	(1.729)	(1.515)	(1.525)	(1.533)	(1.509)	(1.458)	(1.521)	(1.589)	(1.561)	(1.487)	(1.398)
L_{EQEF}	2.951***	2.705***	2.945***	3.033***	2.907***	2.918***	3.116***	2.933***	2.951***	2.909***	3.132***	3.357***
	(0.882)	(0.862)	(0.836)	(0.872)	(0.815)	(0.768)	(0.853)	(0.834)	(0.828)	(0.860)	(0.961)	(0.817)
L_{AS}	-9.410**	-8.414**	-9.418***	-9.560***	-9.538***	-9.402***	-9.795***	-9.162***	-8.711*	-9.371***	-9.092***	-9.761***
	(3.621)	(3.371)	(2.748)	(2.502)	(2.632)	(2.754)	(2.498)	(2.778)	(5.011)	(2.717)	(3.102)	(2.856)
Conditional cooperation	42.322**	36.774*	45.652**	46.702**	46.296**	44.987**	52.095***	41.815**	46.460***	43.512**	45.590***	46.911***
	(16.105)	(18.698)	(18.098)	(16.825)	(16.781)	(17.993)	(16.059)	(15.418)	(16.215)	(16.061)	(16.217)	(15.126)
Elevation	24.750***	20.234*	26.775**	26.930**	26.847**	26.893**	26.333**	24.728***	26.870**	24.059**	27.071**	28.826***
	(8.923)	(11.083)	(10.170)	(10.361)	(10.223)	(9.915)	(9.872)	(8.501)	(10.293)	(9.051)	(10.593)	(9.843)
Plantation	24.712***	-40.156	24.536***	23.557***	24.840***	24.600***	23.690***	24.585***	24.564***	24.652***	25.102***	24.134***
	(6.411)	(95.508)	(6.780)	(7.562)	(7.193)	(6.769)	(6.293)	(6.583)	(6.921)	(6.624)	(6.967)	(6.681)
Group size	1.344	1.038	1.456	1.353	1.480*	1.438	1.368	1.341	1.502	1.406	1.419	1.658*
	(0.833)	(1.079)	(0.862)	(0.851)	(0.843)	(0.845)	(0.812)	(0.837)	(0.899)	(0.896)	(0.866)	(0.827)
Market distance	-8.669*	-10.148*	-9.304*	-9.805*	-9.277*	-9.284*	-9.051*	-8.581*	-9.367*	-8.624*	-9.466*	-9.851**
	(4.636)	(5.375)	(5.048)	(5.151)	(4.871)	(4.691)	(4.784)	(4.569)	(4.933)	(4.476)	(5.196)	(4.611)
Time	4.415	5.864	7.105	7.330	7.311	6.978	7.185	4.293	7.383	4.067	9.134	7.349
	(9.301)	(11.373)	(10.740)	(10.620)	(10.568)	(10.523)	(10.873)	(9.429)	(10.543)	(9.950)	(15.155)	(9.747)
Female share	10.683	-3.809	3.985	4.569	3.869	3.391	11.574	9.961	3.753	12.092	-3.171	-21.617
	(29.011)	(33.996)	(28.706)	(29.408)	(29.648)	(29.960)	(28.036)	(26.754)	(29.905)	(27.439)	(39.491)	(39.695)
Clan heterogeneity	3.946	11.134	7.401	7.406	7.909	7.249	4.025	3.791	7.443	5.587	7.463	12.350
	(16.164)	(18.928)	(18.827)	(19.901)	(19.703)	(18.407)	(18.420)	(17.205)	(18.946)	(18.207)	(18.291)	(19.319)
Settlement heterogeneity	-39.745*	-34.079	-42.022*	-39.058*	-42.538*	-42.313*	-35.235	-39.673*	-42.296*	-39.135*	-42.975*	-58.572**
	(19.464)	(32.685)	(21.928)	(22.592)	(22.614)	(22.245)	(22.959)	(19.425)	(22.706)	(21.405)	(22.583)	(27.133)
Gini cattle	-44.709	-51.089	-51.072	-53.435	-50.395	-50.428	-62.669	-44.193	-52.092	-42.355	-53.140	-61.202*
	(32.529)	(45.003)	(36.217)	(35.287)	(35.666)	(36.923)	(40.881)	(32.817)	(36.015)	(33.061)	(40.193)	(33.119)
Gini land	31.746	42.829	31.210	34.635	30.056	30.858	41.108	29.513	30.654	30.444	29.566	59.303
	(32.678)	(50.181)	(30.019)	(31.819)	(31.919)	(30.585)	(31.873)	(30.339)	(33.542)	(29.693)	(32.326)	(36.992)
Education	-2.357**	-1.339	-2.427*	-2.325	-2.465*	-2.424*	-2.060	-2.314*	-2.472*	-2.258	-2.539*	-3.117**
	(1.129)	(1.282)	(1.347)	(1.407)	(1.395)	(1.331)	(1.392)	(1.206)	(1.417)	(1.368)	(1.385)	(1.365)
Age	-0.038	-0.115	-0.007	-0.002	-0.005	-0.005	0.115	-0.031	-0.022	-0.044	0.078	0.023
	(0.485)	(0.606)	(0.510)	(0.545)	(0.551)	(0.520)	(0.539)	(0.499)	(0.533)	(0.506)	(0.664)	(0.491)
Clan	-8.892	-3.272	-7.950	-8.017	-7.740	-8.107	-7.489	-8.678	-7.728	-7.932	-7.724	-9.397
	(8.975)	(10.192)	(9.043)	(9.921)	(9.710)	(8.816)	(8.920)	(8.972)	(9.792)	(8.980)	(9.296)	(8.771)

TABLE A5 (CONTINUED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Leader ability	-0.865											
	(9.182)											
From council		-6.633										
		(9.702)										
No. of siblings			0.017									
			(0.429)									
Birth-order				-0.245								
				(0.581)								
Birth-order: sons only					0.102							
					(0.600)							
Wealth order						0.210						
						(1.327)						
Wealth order: sons only							-3.138					
							(3.158)					
No. of wives								-0.152				
								(4.743)				
No. of siblings of wife									0.012			
									(0.378)			
No. of children										0.190		
										(0.442)		
Height											0.222	
											(0.709)	
Chest												0.901
												(0.529)
Constant	89.503***	91.692**	82.730***	87.055**	81.181**	83.670**	79.573***	89.067**	82.256**	81.906**	44.072	1.845
	(27.443)	(40.656)	(29.502)	(31.811)	(31.049)	(31.252)	(28.537)	(33.001)	(31.772)	(36.188)	(141.269)	(67.625)
Village fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51	43	49	48	48	49	49	51	48	50	48	48
R-squared	0.881	0.826	0.883	0.877	0.877	0.883	0.888	0.881	0.877	0.881	0.884	0.896
Adj. R-squared	0.787	0.635	0.784	0.769	0.768	0.784	0.793	0.787	0.768	0.785	0.781	0.804

Notes. OLS regression with robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

TABLE A6: FOREST MANAGEMENT OUTCOME, LEADER TYPES, AND ADDITIONAL GROUP CONTROLS

	Clan-1	Clan-2	Clan-3	Clan 1 and 2 and 3	CC majority	CC higher order terms	Streams	Slope
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L_{EQ}	0.569	0.454	0.981	0.957	0.635	0.586	0.812	1.013
	(1.527)	(1.516)	(1.615)	(1.704)	(1.158)	(1.493)	(1.557)	(1.497)
L_{EQEF}	2.970***	2.967***	3.068***	3.040***	3.339***	3.024***	2.817***	3.063***
	(0.827)	(0.871)	(0.905)	(0.974)	(0.821)	(0.825)	(0.764)	(0.883)
L_{AS}	-9.018***	-9.498***	-9.034***	-8.957**	-12.054***	-9.082***	-8.426***	-9.037***
	(2.486)	(3.327)	(2.462)	(3.419)	(2.761)	(2.566)	(2.619)	(2.589)
CC share	40.537**	42.340**	47.835***	49.100**	25.063***	-1.675	38.564*	49.813**
	(15.135)	(15.554)	(17.069)	(20.685)	(6.487)	(42.811)	(20.079)	(19.213)
CC share^2						59.350		
						(54.848)		
Elevation	24.869***	25.134**	25.864**	25.588**	24.945***	25.734***	21.447*	25.741**
	(8.575)	(9.458)	(9.883)	(10.754)	(7.682)	(8.458)	(10.800)	(9.913)
Plantation	24.593***	24.617***	24.225***	24.146***	22.949***	23.830***	29.725***	24.102***
	(6.519)	(6.320)	(6.869)	(7.008)	(5.333)	(5.950)	(8.295)	(6.425)
Group size	1.146	1.351	0.806	0.856	1.479**	1.537*	0.985	0.906
	(0.799)	(0.838)	(0.920)	(0.902)	(0.687)	(0.776)	(1.007)	(0.909)
Market distance	-8.435*	-8.449*	-9.256*	-9.533*	-5.801	-8.577*	-7.480	-9.598*
	(4.630)	(4.444)	(4.809)	(5.095)	(4.315)	(4.530)	(5.021)	(5.187)
Time	4.303	3.713	7.191	7.845	1.241	6.787	6.344	8.293
	(9.546)	(10.424)	(10.228)	(11.721)	(7.928)	(9.333)	(10.060)	(11.198)
Female share	10.290	10.951	8.369	7.623	7.237	9.638	10.349	6.937
	(26.540)	(28.226)	(27.149)	(29.527)	(24.634)	(26.212)	(23.414)	(29.089)
Clan heterogeneity	2.507	3.762	3.690	4.487	5.669	5.066	2.102	5.006
	(17.076)	(15.625)	(17.592)	(19.148)	(13.574)	(15.499)	(17.685)	(18.258)
Settlement heterogeneity	-35.760*	-39.761*	-33.645*	-34.979	-41.783**	-46.963**	-49.473**	-36.757*
	(20.325)	(19.453)	(19.331)	(21.667)	(16.741)	(21.266)	(22.221)	(20.018)
Gini cattle	-46.214	-43.714	-42.219	-40.867	-37.268	-47.425	-48.115	-43.177
	(33.996)	(33.050)	(31.381)	(35.944)	(31.053)	(32.296)	(35.211)	(34.432)
Gini land	34.990	28.526	42.320	41.421	27.745	30.026	40.005	38.243
	(30.252)	(29.819)	(30.437)	(33.126)	(25.611)	(28.005)	(34.887)	(32.813)
Leader education	-2.181*	-2.382*	-2.076	-2.090	-1.912*	-2.559**	-2.509**	-2.158*
	(1.265)	(1.292)	(1.228)	(1.422)	(1.044)	(1.226)	(1.122)	(1.251)
Leader age	0.039	-0.039	0.038	0.006	-0.158	-0.144	-0.089	0.004
	(0.509)	(0.507)	(0.479)	(0.541)	(0.518)	(0.531)	(0.460)	(0.507)

TABLE A6 (CONTINUED)

Leader clan	-8.868	-8.596	-9.657	-9.571	-9.110	-9.552	-9.982	-9.726
	(8.975)	(8.873)	(9.903)	(10.672)	(6.149)	(7.551)	(9.094)	(9.462)
Clan-1	8.767			-5.778				
	(20.815)			(21.593)				
Clan-2		-2.012		1.095				
		(11.299)		(12.208)				
Clan-3			-0.094	-0.433				
			(8.032)	(8.490)				
Perennial streams							2.241	
							(4.919)	
Seasonal streams							7.095	
							(6.525)	
Slope								-0.899
								(4.058)
Constant	82.619**	89.224***	92.568***	96.721***	88.230***	88.254***	91.300***	92.615***
	(31.335)	(28.086)	(29.350)	(34.293)	(26.862)	(26.957)	(26.858)	(30.973)
Village fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51	51	50	50	51	51	50	49
R-squared	0.882	0.881	0.888	0.888	0.901	0.885	0.890	0.888
Adj. R-squared	0.789	0.787	0.796	0.780	0.829	0.795	0.793	0.792

Notes. OLS regression with robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

V. SPATIAL CORRELATION OF ERRORS

Because several groups belong to the same village, spatial correlation of standard errors within a village is a possible concern. A key issues here is that we have just five villages in our dataset. With such few clusters, the standard errors and corresponding p-values are likely to be biased downwards. As a result, hypothesis testing with asymptotic standard normal critical values tends to over-reject the null. Given this concern, we follow the suggestions laid down by Cameron, Gelbach and Miller (2008) and use wild bootstrap procedure, which does not generate high type-I error rates despite a small number of clusters. Column 3 in Table A7 reports the p-values obtained using wild bootstrap-t procedures with 400 replications. We find that the effect of L_{EQEF} and L_{AS} coefficients is significant at 5 and 10 percent, respectively. These results are robust even if we increase the number of repetitions to 800 and 1200, at which the p-values seem to become stable (columns 4-5).

TABLE A7: PCT AND LEADER TYPES

	PCT per hectare				
	Robust	Cluster	Wild bootstrap		
			(400)	(800)	(1200)
	(1)	(2)	(3)	(4)	(5)
L_{EQ}	0.768	0.739	0.539	0.557	0.560
L_{EQEF}	0.001	0.004	0.045	0.027	0.029
L_{AS}	0.001	0.019	0.092	0.094	0.095

Notes. Results based on OLS regression of PCT on full control list and village fixed effects (cf. column (9) in Table 6, main paper). Controls include CC share, elevation, plantation, group size, market distance, time, female, clan, settlement, Gini cattle, Gini land, leader's age, leader's education, and leader's clan. P-values are based on standard errors clustered on the village obtained using CGM wild bootstrap procedures implemented using bootwildct command. The number of observations is 51.

VI. SECOND FOREST MANAGEMENT ASSESSMENT (PCT-II)

Until January 2013, 25 out of 51 groups considered in this study were assessed for the second time. This makes the sub-sample analysis prone to a selection problem. In Table A8, we test the null hypotheses that observables across the full and the sub-sample within a leader's type are not significantly different from zero. We consider only a selective list of observables because there is not much variation in the excluded ones, such as elevation and historical plantation share. Except for time, which is significant at 10 percent, none of the covariates are significantly different. Using the Bonferroni correction[3], a joint null that these differences are not significantly different from zero cannot be rejected at 10 percent.

[3] The revised p-value using the Bonferroni correction is 0.1/11 = 0.009.

TABLE A8: SUMMARY STATISTICS BY FULL AND SUB-SAMPLES OF LEADER TYPES

Variable	L_{NP} original sample	L_{NP} sub-sample	Difference	L_{EQ} original sample	L_{EQ} sub-sample	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
PCT I	68.38	78.35	9.97	60.82	71.90	11.08
	(33.18)	(21.56)	(8.18)	(29.30)	(27.96)	(12.12)
CC share	0.35	0.40	0.05	0.29	0.33	0.04
	(0.23)	(0.24)	(0.07)	(0.21)	(0.22)	(0.09)
Group size	26.66	26.81	0.16	26.07	26.22	0.15
	(4.13)	(4.31)	(1.32)	(5.53)	(5.93)	(2.45)
Market	2.56	2.49	-0.08	1.92	1.83	-0.09
	(0.87)	(0.83)	(0.26)	(0.61)	(0.70)	(0.28)
Time	0.55	0.81	0.26*	0.36	0.44	0.09
	(0.51)	(0.40)	(0.14)	(0.05)	(0.53)	(0.22)
Female share	0.22	0.24	0.02	0.17	0.18	0.01
	(0.13)	(0.15)	(0.05)	(0.08)	(0.10)	(0.04)
Gini cattle	0.34	0.35	0.01	0.32	0.34	0.01
	(0.09)	(0.11)	(0.03)	(0.07)	(0.07)	(0.03)
Gini land	0.29	0.31	0.02	0.31	0.30	-0.00
	(0.07)	(0.05)	(0.02)	(0.08)	(0.10)	(0.04)
Leader age	41.21	40.63	-0.58	42.29	42.11	-0.17
	(7.36)	(6.09)	(2.16)	(6.83)	(7.61)	(3.05)
Leader education	4.48	4.56	0.08	5.29	5.22	-0.06
	(2.82)	(2.39)	(0.84)	(2.92)	(2.64)	(1.20)
Leader clan	0.86	0.88	0.01	0.86	0.78	-0.08
	(0.35)	(0.34)	(0.11)	(0.36)	(0.44)	(0.17)
Observations	29	16		14	9	

Notes. Columns 1, 2, 4, and 5 report mean with standard deviation in parentheses. Columns 3 and 6 report difference with robust standard errors in parenthesis. Difference represents the average difference in covariates between the full and the sub-sample within a leader's type, computed as the regression coefficient on a dummy (1=group covered in the second assessment, 0 otherwise).

Figure A1 shows the difference in the performance of groups led by L_{NP} and L_{EQ} over time using data from PCT-I and PCT-II.

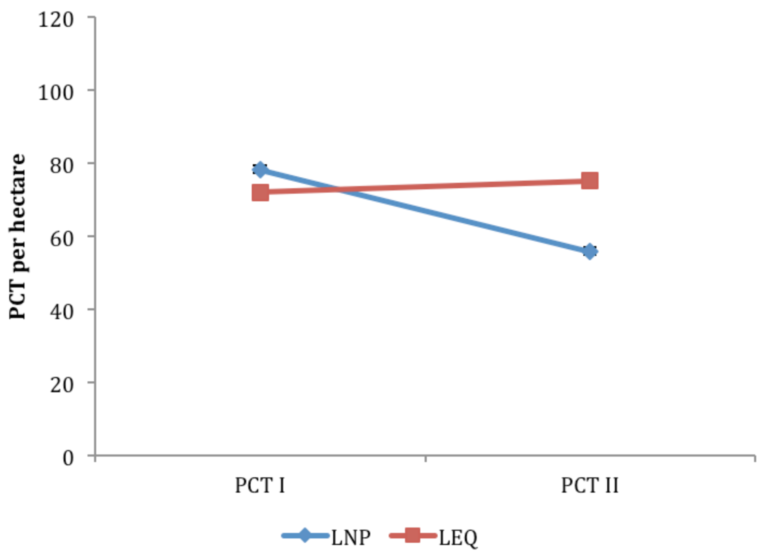

FIGURE A1: FOREST MANAGEMENT OUTCOME FOR L_{NP} AND L_{EQ} LEADERS BASED ON FIRST (PCT-I) AND SECOND (PCT-II) FOREST ASSESSMENT

VII. LEADER TURNOVER

Table A9 reports the leader type, turnover status, as well as the reason behind turnover for all 51 groups. Figure A2 shows the percentage of reasons behind turnover within each leader type.

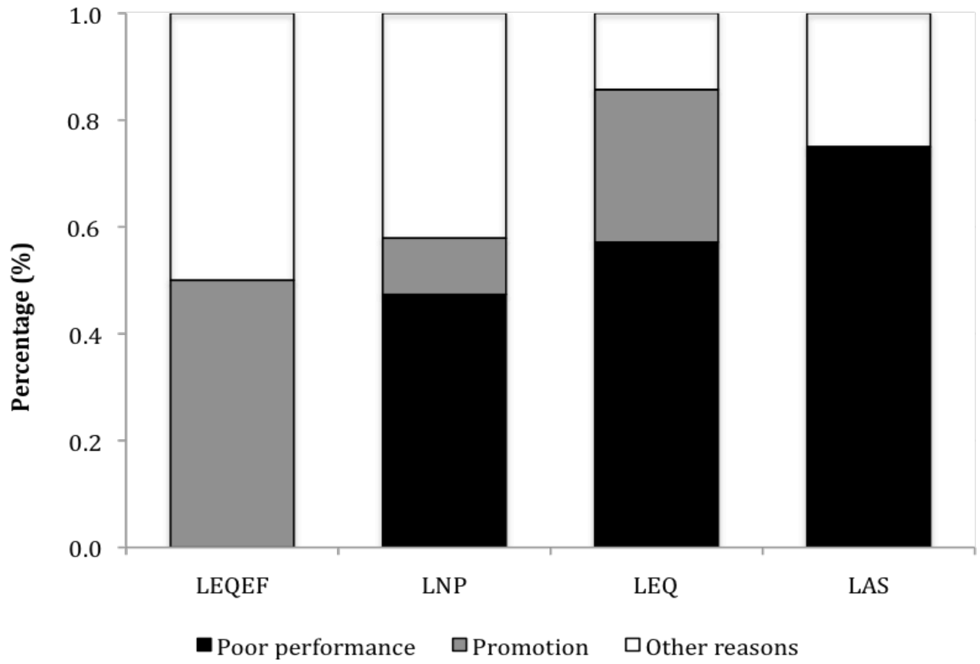

FIGURE A2: REASONS BEHIND LEADER TURNOVER BY LEADER TYPE

To ensure that leader types are not masking for leader quality as reflected by leader turnover, we control for leader turnover. Table A10 reports the results and begins with the baseline model without turnover. Columns 2-3 show that the effect of leader types is robust to controlling for both leader turnover and the reason behind leader turnover. Column 4 further shows that our results hold even when we simultaneously control for both leader quality (proxied by turnover) and leader ability (cf. Table A5 above).

TABLE A9: LEADER TURNOVER BY LEADER TYPE

Group code	Leader type	Leader turnover	Reason	Classification
1	L_{NP}	Re-elected	Did a good job in forest management	Good job
2	L_{NP}	Removed	To improve forest management	Poor
3	L_{EQ}	Re-elected	Did a good job in forest management	Good job
4	L_{EQ}	Removed	To ensure improvement in forest condition	Poor
5	L_{EQ}	Re-elected	Did a good job in forest management	Good job
6	L_{NP}	Removed	Weak and careless	Poor
7	L_{EQ}	Removed	Elected as a leader to the village council	Promotion
8	L_{AS}	Removed	Did not work properly	Poor
9	L_{EQEF}	Removed	Elected as a leader to the village council	Promotion
10	L_{EQEF}	Removed	Elected as a leader to the village council	Promotion
11	L_{NP}	Re-elected	Did a good job in forest management	Good job
12	L_{NP}	Removed	Negligent and bad tempered	Poor
13	L_{NP}	Removed	Rule violation by family members	Poor
14	L_{NP}	Removed	Did not improve forest management	Poor
15	L_{NP}	Removed	To ensure that people cooperate better	Poor
16	L_{AS}	Removed	The leader himself did not follow the rules	Poor
17	L_{EQ}	Re-elected	Did a good job in forest management	Good job
18	L_{EQ}	Removed	Did not pay attention to forest management	Poor
19	L_{EQ}	Removed	Poor performance and rule implementation	Poor
20	L_{NP}	Removed	Old age	Other
21	L_{NP}	Removed	Sickness	Other
22	L_{NP}	Removed	Quit voluntarily	Other
23	L_{NP}	Re-elected	Did a good job in forest management	Good job
24	L_{NP}	Removed	Elected as a leader to the village council	Promotion
25	L_{NP}	Removed	Old age	Other
26	L_{NP}	Removed	Old age	Other
27	L_{NP}	Removed	Sickness	Other
28	L_{NP}	Removed	Not strict with enforcing rules	Poor
29	L_{NP}	Re-elected	Did a good job in forest management	Good job
30	L_{NP}	Removed	Elected as a leader to the village council	Promotion
31	L_{EQ}	Re-elected	Did a good job in forest management	Good job
32	L_{EQEF}	Removed	Quit to continue with education	Other
33	L_{NP}	Removed	Hot tempered and negligent	Poor
34	L_{AS}	Removed	Got old	Other
35	L_{NP}	Removed	Old and not active anymore	Other
36	L_{NP}	Re-elected	Did a good job in forest management	Good job
37	L_{EQ}	Removed	Elected as a leader to the village council	Promotion
38	L_{EQ}	Re-elected	Did a good job in forest management	Good job
39	L_{AS}	Removed	Did not do his job well, loose character	Poor
40	L_{NP}	No election		
41	L_{NP}	No election		
42	L_{NP}	Removed	Did a good job in forest management	Poor
43	L_{EQ}	No election		
44	L_{NP}	Removed	Died	Other
45	L_{NP}	No election		
46	L_{NP}	No election		
47	L_{EQ}	No election		
48	L_{NP}	Re-elected	Did a good job in forest management	Good job
49	L_{EQEF}	Removed	Old age	Other
51	L_{EQ}	Removed	Did not want to be a leader	Other
53	L_{EQ}	Removed	Did not perform well	Poor

Notes: Outcome data for group 50 and 52 was not available.

TABLE A10: GROUP PERFORMANCE, LEADER TYPES, AND LEADER TURNOVER

	(1) Baseline results	(2) Leader turnover	(3) Turnover reason	(4) Leader turnover and ability
L_{EQ}	0.444	0.341	0.210	0.246
	(1.489)	(1.535)	(1.521)	(1.564)
L_{EQEF}	2.932***	3.126***	3.136***	3.250***
	(0.826)	(0.967)	(0.882)	(1.154)
L_{AS}	-9.183***	-8.997***	-8.447***	-9.744**
	(2.500)	(2.687)	(2.891)	(3.706)
Controls	Yes	Yes	Yes	Yes
Village fixed effect	Yes	Yes	Yes	Yes
Observations	51	51	51	51
Adj. R-squared	0.794	0.789	0.793	0.782

Notes. OLS regression with robust standard errors in parentheses. Controls include share of conditional cooperators, elevation, plantation, group size, market distance, time, female share, Gini of cattle and land ownership, settlement and clan heterogeneity, and leader characteristics (age, education and clan).

VIII. ALTERNATIVE LEADER CLASSIFICATION

We also consider an alternative classification, in which we create a new variable L_{PUN}, which is defined as:

$$L_{PUN} = 1+\overline{P}(C<6)/1+\overline{P}(C=6),$$

where C is a player's contribution, $\overline{P}(C<6)$ is a leader's average punishment of a player, who contributes less than six Birr, and $\overline{P}(C=6)$ is a leader's average punishment of a player, who contributes six Birr. Intuitively, the variable L_{PUN} measures a leader's punishment of 'less than full contribution' as a proportion of punishment of 'full contribution'. We add one to both the numerator and the denominator because the latter is zero for leaders, who never punish full contribution. Adding one to the denominator only would result in a distortion as leaders who punish $C_i < 6$ and $C_i = 6$ equally but with different intensity (across leaders) would receive different values. For example, a leader who punishes both cases equally with 1, would receive the value 1/2, while a leader who punishes both cases with 2, would receive 2/3. In the present specification, both leaders receive the value 1.

The advantage of the alternative measure is that it is continuous and we thus exploit the full range of punishing and non-punishing leaders. The disadvantage is that it is difficult to distinguish between non-punishing and anti-socially punishing types. Both leaders are characterized by a flat punishment pattern, as they do not punish less contribution more than full contribution: non-punishing leaders by definition (because they do not punish at all), anti-socially punishing leaders because they punish all cases alike (cf. Result 2). Yet, the two kind of leader behavior clearly differ, both with respect to their underlying motivation and with respect to their expected effect on cooperation outcomes.

When we compare L_{PUN} across leader types, we find that L_{EQEF} (1.792, s.d. 0.198) have the highest average value, followed by L_{EQ} (1.417, s.d. 0.139) and then L_{AS} (1.352, s.d. 0.338). Because L_{NP} do not punish, their value is 1. A Kruskal-Wallis test shows that differences in L_{PUN} by leader types are highly significant ($p = 0.000$).

We regress PCT on L_{PUN} using the full specification as in column 8 of Table 6 in the main paper. The results are reported in Table A11 below. Column 1 shows that L_{PUN} has a positive effect on the forest management outcome as measured by the first assessment (PCT-I). The higher L_{PUN}, i.e., the more a leader punishes less contribution more than full contribution, the better are forest management outcomes. The effect is robust to the inclusion of village fixed effects (column 2). One standard deviation increase in L_{PUN} (0.72 points) increases the forest management outcome by 16.12 trees, which is quantitatively large.

Because it is difficult to discriminate between non-punishing and anti-socially punishing leaders, we re-run the regression in column 3 with punishing leaders alone, i.e., we exclude L_{NP}. As expected, the coefficient of L_{PUN} increases and is significant. In other words, the positive effect of leaders who score high on L_{PUN} (i.e., L_{EQ} and L_{EQEF}) is even larger among punishing leaders than among all leaders. This nicely mirrors the negative association of L_{AS} leaders compared to L_{NP} as reported in the paper. Finally, column 4 shows that the results also hold when we consider data from the second forest assessment (PCT-II). Overall the results corroborate our main findings using the classification of leaders into four different types.

TABLE A11: LEADER CLASSIFICATION AND PCT PER HECTARE

	PCT-I			PCT-II
	(1)	(2)	(3)	(4)
L_{PUN}	22.303*	22.291*	42.109*	50.504**
	(12.401)	(12.400)	(21.325)	(23.630)
Controls	Yes	Yes	Yes	Yes
Village fixed effects	No	Yes	No	No
Observations	51	51	22	25
Adj. R-squared	0.705	0.712	0.721	0.736

Notes. OLS regression with robust standard errors in parentheses. In columns (1-2) controls include share of conditional cooperators, elevation, plantation, group size, market distance, time, female, settlement, clan, Gini cattle, Gini land, leader's age, leader's education, and leader's clan. In the restricted samples in columns (3-4), we include only a limited set of control variables because there is not much variation in the excluded variables. These include share of conditional cooperators, group size, market distance, time, female share, and Gini cattle. In column (3), groups led by L_{NP} are dropped.

APPENDIX B

GAME INSTRUCTIONS AND EXPERIMENTAL PROCEDURES

I. THIRD-PARTY PUNISHMENT GAME

We will now proceed with the second game. In this game there are three players: Two FUS members and one FUS leader. This game is similar to the game, which we just played. In this game, you just have to take one decision, which is like decision 1 in the first game. For your benefit, we will now repeat the instructions.

Instructions for the FUS Members (make sure that the leader is present)

1. There are two players in a group. Again, their identity is kept anonymous.

2. Each player receives an initial endowment of 6 ETB.

3. The players have to decide how much they would like to put in a project. In this game, you can put only 0, or 2, or 4, or 6 ETB in a project (*ask the group to repeat this)*. Can you put 1 ETB in a project? Can you put 3 ETB in a project? Can you put 5 ETB in a project? Please remember that you cannot put 1 or 3 or 5 ETB in a project.

4. Any amount in a project will be multiplied by 1.5 times and then divided equally between you and your partner player in the game (*just like in the previous game*).

5. After you have made this decision, we will call the leader of your FUS and show him the amount you put in the project as well as your income. But the leader will not come to know of your identity.

6. We will give the leader 10 ETB (*do this in front of the actual leader so that he understands the game as well*). After the leader has seen the amount that you put in the project, he has an option to reduce the income of either any one or both the players.

7. If the leader decides to reduce the income of a player, he has to spend one ETB, but the player loses 3 ETB. However, if the leader decides not to reduce, he gets to keep all 10 ETB in his pocket and the players don't lose any income either.

8. Your final income from the game will be calculated after subtracting any amount that was reduced by the leader.

9. Please note that you will not come to know if the leader reduced your income or not. We will not tell you this (*for we do not want you to beat the leader in the evening – optional and as humour*).

10. We will now give you two examples to show how this works (*Give two examples: leader takes a reduction decision and leader does not take a reduction decision. Select randomly two players and give them 6 ETB)*.

Let us assume that player 1 puts 0 ETB and player 2 also puts 0 ETB in a project. Can the players put 1 or 3 or 5 in the project (*ask one by one*)? How much money is there in the project? Does the project money increase? How much money does each player has in his pocket? What is each player's income? Up to here the game is similar to the one we played before.

Now let us assume that Devesh is the leader of your FUS. (*Display the contributions and payoffs of each player and call the actual FUS leader to stand nearby. Let the main assistant explain Devesh and the actual leader how the display is organized*).

This is what player 1 puts in the project and this is his income. This is what player 2 puts in the project and this is his income. How much did player 1 put in the project? What is his income? How much did player 2 put in the project? What is his income? Do you understand this? (*Devesh: while responding please count bills as local people do*).

I will now give you 10 ETB and you may use this money to reduce either or both player's income (*display10 ETB clearly to both members and the leader*). Would you like to reduce any or both player's income? (*Devesh: say No*). Ask the group: What happens if the leader says he does not want to reduce? The money goes to his pocket.

Let us now see what happens if the leader reduces players' income. Would you like to reduce the income of any one or both the players? (*Devesh: say Yes and 'punish' each player by one Birr*). Please note that the leader decided to reduce the income of both the players by 1 ETB. How much did the leader spend on reducing the income of each player? How much does each player lose? Because of the reduction by the leader, we will reduce each player's income by 3 ETB (*remove 3 ETB from each player's income in front of all the members*). How much does each player earn? How much does the leader earn? Is the leader obliged to reduce income in the game?

11. Have you understood the game? Do you need more examples?

12. One very important thing: before you put any money in the project, please think if the leader will reduce your income because of the money you put in the project or not. We will ask you on this.

Instructions for the FUS Leader

We hope that you have understood the game. In this game you have to take a decision on whether or not you would like to reduce any one or both player's income. This decision is entirely yours and we will not help you on this. You are not obliged to reduce. We will not inform anyone of the decision you take.

You will receive 10 ETB from us, which you may use to reduce a player's income. In case you decide to reduce a player's income, you will have to spend 1 ETB for each reduction, but the player's income will be reduced by 3 ETB.

Have you understood the game? Would you like to play a mock round? (*Encourage the leader to play a mock round. Use 4,6 combination for this. Make sure that the leader learns that he has to spend one Birr to reduce, that the 'punished' player loses 3 ETB, and that reduction is not mandatory*).

Now, we would like to show you the contribution decisions of two members from your FUS (*show each contribution set separately by using one ETB bills so that the leader is not confused. The table is for your own understanding and is not to be shown to the leader*).

Player contributions		**Punishment by Leader**		**Player contributions**		**Punishment by Leader**	
A	B	A	B	A	B	A	B
0	0			2	4		
0	2			2	6		
0	4			4	4		
0	6			4	6		
2	2			6	6		

Display from 2,4 onwards. Ask the leader: Can you tell me where are the contributions of player 1 and 2 on this display? Where are the payoffs of player 1 and 2 on this display? How much did player 1 put in the project? What is his income? How much did player 2 put in the project? What is his income?

Good. Here is your endowment of 10 ETB. You may now take a reduction decision (*record each decision in the table*).

Once all the decisions have been taken, ask each leader why he punished (or did not punish) in the game.

II. EXPERIMENTAL PROCEDURES

We briefly describe the experimental procedures we adopted for conducting the experiments. The procedures are described in detail in Rustagi et al. (2010). We employed community mobilizers to recruit group members on our behalf. These mobilizers knew nothing about the experiments and were asked to invite 15-25 members on an appointed date at a designated venue. Overall 670 members (including leaders) from 49 groups participated in the experiment, covering over 50 percent of the member population in these groups.

Conducting experiments in the field with a non-standard subject pool creates many challenges, which we resolved in several steps. First, we used fixed written instructions, set of examples, and control test questions as our guidelines for conducting the experiments. Second, these instructions were tested in a pilot study in a linguistically similar but different region of Ethiopia, as well as in three forest user groups. Based on the feedback we got, the instructions were fine-tuned to meet the limited literacy of our subject pool. Third, we tested each player for her game comprehension using a fixed set of control questions and allowed only those players to participate in the experiment who answered these questions correctly. Fourth, in line with the members' familiarity with small currency bills, we used actual bills of one Birr while delivering instructions, examples, testing, and for playing the actual game; this made it easier for the participants to make simple calculations. Fifth, we took great care to ensure members that their decisions will be anonymous to each other and to the field assistants. We assigned each member an identity card bearing the names of Swiss Cantons or German States. Members were aware that once the experiments are over and the cards are handed in, the experimenter would have no chance of identifying them. These practices not only ensured uniformity in procedures in all groups but also that subjects understood the games well enough to participate in them.

We are aware that conducting field experiments in developing countries is fraught with contagion and contamination risks. We adopted a two-fold approach to assuage these risks: First, we focused on groups from one village at a time, and second, wherever possible, we invited two neighboring groups at a time. In addition, we ran experiments every day, without a break, till all groups from a village had taken part in all the experiments. Once this was achieved, we moved to the next village and followed similar procedures there. This meant that members who took part in the experiments were left with less time to interact with members from groups yet to take part in the experiments. These practices minimized the scope of contagion or contamination.

Lastly, all experiments were hand-run and were conducted by one of the authors together with the main assistant. Overall, members took part in three experiments, of which the experiment discussed here, was the second. We paid members their earnings from all the games in the end. The stake size was high. On average, each player earned more than four times the daily wage in the study area.