

Centralized sanctioning and legitimate authority promote cooperation in humans

Delia Baldassarri^{a,1} and Guy Grossman^b

^aDepartment of Sociology, Princeton University, Princeton, NJ 08544; and ^bDepartment of Political Science, Columbia University, New York, NY 10027

Edited by Michael Hout, University of California, Berkeley, CA, and approved May 26, 2011 (received for review April 5, 2011)

Social sanctioning is widely considered a successful strategy to promote cooperation among humans. In situations in which individual and collective interests are at odds, incentives to free-ride induce individuals to refrain from contributing to public goods provision. Experimental evidence from public goods games shows that when endowed with sanctioning powers, conditional cooperators can discipline defectors, thus leading to greater levels of cooperation. However, extant evidence is based on peer punishment institutions, whereas in complex societies, systems of control are often centralized: for instance, we do not sanction our neighbors for driving too fast, the police do. Here we show the effect of centralized sanctioning and legitimate authority on cooperation. We designed an adaptation of the public goods game in which sanctioning power is given to a single monitor, and we experimentally manipulated the process by which the monitor is chosen. To increase the external validity of the study, we conducted lab-in-the-field experiments involving 1,543 Ugandan farmers from 50 producer cooperatives. This research provides evidence of the effectiveness of centralized sanctioning and demonstrates the causal effect of legitimacy on cooperation: participants are more responsive to the authority of an elected monitor than a randomly chosen monitor. Our essay contributes to the literature on the evolution of cooperation by introducing the idea of role differentiation. In complex societies, cooperative behavior is not only sustained by mechanisms of selection and reciprocity among peers, but also by the legitimacy that certain actors derive from their position in the social hierarchy.

In collective action problems, individuals have to decide whether to contribute to the provision of nonexcludable public goods when incentives to free-ride are pervasive (1). Social sanctioning is widely considered a viable solution to this type of social dilemma (2–5). Experimental evidence shows that, in public goods games (PGGs) (6), participants initially contribute, on average, between 40% and 60% of their endowment.* However, in repeated games, conditional cooperators who wish to avoid being exploited by free-riders gradually refrain from cooperation, thus leading to a drop in contributions in subsequent rounds (7, 8). By contrast, when endowed with sanctioning powers, conditional cooperators can discipline defectors, thus leading to greater overall levels of contribution (2, 9, 10). Peer punishment, as well as reward, provides a possibility of targeted interaction, thus fostering cooperation through mechanisms of direct and indirect reciprocity (11, 12).

Most empirical work on the role sanctioning plays in inducing cooperation has focused on testing evolutionary models of diffused and decentralized punishment (9, 13), and explored the relative efficiency of different peer-sanctioning institutions (14–16). However, scholars have recently begun questioning the ability of spontaneous, uncoordinated, and decentralized peer punishment actions to sustain cooperation in complex societies (17, 18), concluding that “the step from peer punishment to the establishment of sanctioning institutions deserve closer future investigations” (ref. 3, pp. 598–599). Peer sanctioning is only effective under restrictive conditions (3, 19): either individuals in a population are given the option of nonparticipation (20, 21) or the group is sufficiently small (17). In small-size groups, even self-interested contributors may choose to punish defectors at a per-

sonal cost, if interactions are frequent and there are reputation considerations (17). When groups become large and interactions between members infrequent, bilateral punishment is unlikely to sustain cooperation because future gains from punishment cannot be internalized (22).

To overcome this problem, groups tend to develop forms of self-regulation, in which the power to sanction defectors is transferred to a centralized authority (23–25): for instance, villagers in traditional societies turn to their chiefs to adjudicate disputes (26), organized workers developed centralized disciplining institutions to deter defectors (27), and merchants in Medieval Europe created guilds (22).

From an evolutionary perspective, centralized institutions are likely to be more efficient than peer punishment (28) because they are better positioned to overcome coordination failures and free-riding problems (18). To incorporate this phenomenon into theories of public goods provision, we study how group members behave in a context in which a centralized monitor is given a monopoly over sanctioning decisions. Will groups reach high levels of contribution even in a sanctioning regime in which group members do not have the power to decide who should be sanctioned?

Centralized systems of control cannot rely exclusively on the threat of punishment (29). Institutions that are perceived as legitimate only rarely turn to coercion to enforce group norms. In contrast to brute force, the authority of an individual (or organization) depends on the degree of compliance to his commands (30). We expect a leader's authority to depend on his or her perceived level of legitimacy. Our second question is whether the process by which a central authority acquires sanctioning powers is consequential. We hypothesize that involving subjects in the selection of their monitoring authority will confer him greater legitimacy, thereby leading to greater cooperation.

Lab-in-the-Field Experimental Design

To test our hypotheses, we designed a unique adaptation of the PGG. Unlike peer punishment settings in which players are given the ability to reduce each others' payoffs, in our centralized sanctioning setting, this power was given to a single monitor. We played three variants of the PGG. In the baseline condition, players participated in six rounds of a PGG without punishment. In the two

Author contributions: D.B. and G.G. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: dbalda@princeton.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105456108/-DCSupplemental.

*In PGGs, participants anonymously decide how to split an initial endowment between private and public accounts. What players put in the private account remains theirs, and what is contributed to the public account is doubled and redistributed evenly among all group members, regardless of their level of contribution. The most profitable outcome for the group occurs when all players contribute their entire endowment. Nonetheless, regardless of what other people contribute, the most profitable strategy for the individual is to keep the entire endowment in his private account and benefit from what everyone else contributes to the public account. Designed to induce a social dilemma, PGGs capture how players balance self-interest and the well-being of the group.

treatment conditions, after two preliminary rounds of play, one of the players was selected out of the session to become a monitor endowed with sanctioning power. Monitors received the same endowment as the other players [10 monetary units (MUs)] and could spend 1 MU to take away three MUs from players whose contribution level they disapproved of. The monitor's final payoff did not depend on the group's level of cooperation, but only on her sanctioning decisions. In the random monitor condition (T1), the monitor was selected through a random lottery. In the elected monitor condition (T2), the monitor was elected by group members, using a secret ballot (see *Methods and Materials* and *SI Appendix* sections 4 and 5 for additional information and the game script). Comparing the baseline with the random monitor condition allows for an assessment of the effect of a centralized sanctioning regime, whereas a comparison of the random and elected monitor conditions allows an estimate of the independent effect of the process by which the monitor has obtained her sanctioning power.

Laboratory experiments have greatly contributed to our understanding of cooperative behavior, but they are ill-suited to the study of processes in which group identity and cultural factors are expected to be consequential (23, 31, 32). Behavior is influenced by environmental cues. By stripping context away, laboratory experiments open themselves to the charges of undermining the external validity of the findings (33, 34). In addition, repeated interaction in small groups is generally considered the natural condition in which our ancestors had to operate, thus arguably the proper context for studying the emergence of cooperation (11, 12, 35). Incorporating these aspects in a laboratory setting is difficult (36). To overcome these limitations, we carried out our experiment with the members of natural preexisting groups that face collective action problems on a regular basis: our lab-in-the-field experiment was conducted in 50 different producer organizations (farmer cooperatives) in rural Uganda. A total of 1,543 farmers participated in our experiment.

Results

Fig. 1 summarizes our major findings. First, in the presence of a centralized sanctioning system, players significantly increase their contribution to the public good. In the random monitor condition (T1), subjects contributed 15.0% more on average than in the baseline condition ($P = 0.000$, two-sided Mann–Whitney tests). In the elected monitor condition (T2) subjects gave 25.4%

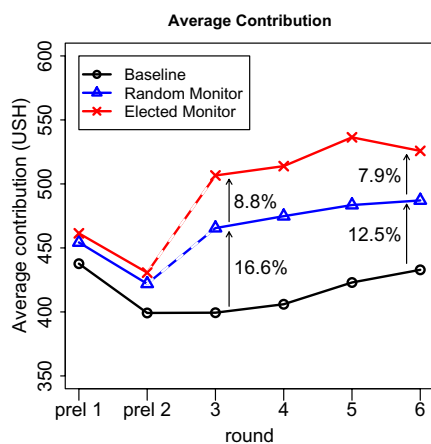


Fig. 1. Average contribution to the public good in the baseline (black), random monitor (blue), and elected monitor (red) conditions. For rounds 3 and 6 we report the percentage increase in contributions comparing the random and baseline conditions, and elected with random [e.g., in round 3, subjects in the random monitor contributed 16.6% more ($P = 0.000$) than in the baseline condition, and subjects in the elected monitor contributed 8.8% more than in the random monitor condition]. $n = 1,446$ (1,543 players – 97 monitors).

more on average than in baseline ($P = 0.000$). Moreover, subjects in both T1 and T2 behaved according to the expectation that monitors would discipline defectors: differences between the baseline and the two monitor conditions are already statistically significant in round 3, before the monitors have acted. In round 3, subjects in T1 contributed to the public account 16.6% more ($P = 0.000$), and subjects in T2 26.8% more ($P = 0.000$), than in the baseline.

Second, the process through which leaders obtain their sanctioning power is consequential: players who participated in the selection of their monitor contributed to the public account 9% more, on average, than players whose monitor was selected at random ($P = 0.005$). Similarly, elected monitors generated greater expectations: in round 3, subjects in T2 gave 8.8% more than subjects in T1 ($P = 0.022$). Statistical models with clustered SEs at the session level that control for preliminary contributions confirm the significance of these results (*SI Appendix, Table S1*).

We explain higher levels of cooperation in T2 as a consequence of the greater legitimacy of elected monitors. We conceive of legitimacy as the monitor's capacity to exercise authority and measure it as the extent to which players defer to the leadership (30). Observational studies have often suggested a positive relationship between the legitimacy of an authority and effectiveness in the exercise of power (29, 37, 38). Our experimental design tests the causal nature of this relationship. We expect players to express greater deference to an authority whose legitimacy has been certified through elections.

To test this hypothesis, we consider, first, players' expectations for the monitor, by looking at the change in their contributions from the second preliminary round to round 3 (Fig. 2A). Second, we consider players' reaction to the monitor's sanctioning behavior, by looking at the change in contributions of the sanctioned players (Fig. 2B). Fig. 2 shows the relative increase in contributions for a representative player in T1 and T2. Parameter estimates come from multilevel models in which we control for individual and group-level predictors (*SI Appendix, Tables S2 and S3*). We find that the introduction of a sanctioning system, even before any manifestation of the monitor's behavior, induced a net increase in contribution of 0.37 MUs in T1 and 0.76 MUs in T2. When considering players' reaction to punishment, evidence of the greater deference players have for elected monitors is even stronger. Having been punished at time $t - 1$ increased players' contribution by 0.57 MUs in T1, and by 1 MU in T2. In both cases, the elected monitor has an impact approximately twice as big as that of a randomly chosen monitor.

To strengthen our claim that the legitimacy of the monitor explains the difference in contribution between T1 and T2, we rule out the possibility that the effect we attribute to legitimacy is due to a "leadership selection" effect—namely, to the fact that players may have elected monitors who are well suited to leadership roles. According to this alternative interpretation, higher levels of cooperation in T2 might be due to the intrinsic qualities of the elected monitor. Indeed, subjects selected monitors with a socially dominant profile—elected monitors were mostly male, and, on average, were wealthier, more educated, and more likely to be born in the village, compared with the pool of eligible monitors (*SI Appendix, Fig. S1*). Nonetheless, none of the monitors' sociodemographic characteristics improved our predictions of players' contributions (*SI Appendix, Table S4*). Moreover, as shown in Fig. 2C, monitors with a socially dominant profile (male, born locally, highly educated, and wealthy) have the same sanctioning effect on contributions as monitors with a nondominant profile. These findings, cumulatively, weaken the possibility of a leadership selection effect. It is possible, however, that elected monitors have certain attributes that induce cooperation, which are unobserved to the researcher but visible to the experimental subjects. Although the possibility of unobserved heterogeneity cannot be ruled out completely, it is not likely to have a decisive role in the experiment.

strategy, monitors' major impact on cooperation was the anticipation and response to sanctioning that they elicited.

Discussion

In recent years, a large cross-disciplinary literature has studied the role that diffused peer sanctioning plays in inducing cooperation. Although this literature has contributed much to our understanding of the evolution of cooperation, large-scale cooperation in complex human groups is not sustainable on the basis of peer punishment alone (39). Turning our attention to the role of centralized authorities opens up new research questions. Our findings suggest that in the presence of a leader, groups can reach higher levels of cooperation, and that monitors, at least in a situation in which their reputation is at stake, are willing to sacrifice part of their welfare to increase cooperation.[‡] These results are qualitatively similar to those obtained using peer sanctioning institutions, with the advantage that a centralized system of monitoring will be more efficient than a decentralized one, because it avoids the problem of uncoordinated over punishment (18). However, the efficiency of a centralized authority decreases as its distance from the monitored group or community increases. In fact, proximity increases monitoring capabilities, reduces information gaps and enforcement errors (41), and facilitates the adaptation of sanctioning practices to local norms (42). Future studies should investigate the relative efficiency of peer vs. centralized sanctioning regimes, distinguishing between centralized local authorities and larger institutional entities, such as the state.

Centralized authorities cannot rely exclusively on coercive punishment (29). Indeed, institutions that are perceived as legitimate only rarely turn to coercive force or power to enforce group norms. In contrast to brute force, the authority of an individual (or organization) depends on the degree of deference to his commands (30). Our study demonstrates that beyond the mere threat of punishment, cooperation is conditional on the perceived level of legitimacy of the sanctioning authority. The fact that subjects who were allowed to select their monitor increased their contribution in subsequent rounds by more than double the increase of subjects who faced an arbitrary authority is indicative of the role legitimacy plays in sustaining cooperation. This research documents experimentally the role of legitimate authority in fostering collective action.[§] This finding is consistent with empirical evidence suggesting that public goods provision is higher in democratic regimes. More work is also needed to understand the relative effectiveness of different sources of legitimacy, considering, for example, more traditional forms of authority.

Role differentiation and dominance hierarchies are organizing principles of most animal groups (43–45). In complex societies,

centralized systems of control can emerge endogenously by virtue of the legitimacy that certain actors derive from their position in the social hierarchy. The literature on the evolution of cooperation has focused mostly on mechanisms of selection and reciprocity among peers. Here, we advance the hypothesis that, under certain conditions, a legitimate authority might constitute an additional mechanism fostering cooperation and thus the construction of new levels of organization. Sources of authority can vary across species and cultures. Position in the pecking order or in the economic structure, religious or traditional power, and political rituals (i.e., elections) are just a few examples of the factors that might enhance authority and facilitate the establishment of leadership roles in humans.

Materials and Methods

To increase the external validity of our findings, we conducted our research with members of farmer associations created as part of one of Uganda's largest recent rural development interventions: the Agriculture Productivity Enhancement Project (APEP). Our lab-in-the-field experiment was conducted in 50 different groups in nine districts across Uganda. In each location ~30 farmers were randomly sampled from a list of all members of a local farmer association, for a total of 1,543 subjects. We used a stratified, random, multistage cluster design to select our sample (see *SI Appendix* section 3 for further details on the sampling strategy).

We played all three variants of the game in each of the 50 sampled farmer associations. Each participant was randomly assigned to one of the variants. The number of participants per session ranged from 8 to 12. This number is higher than in most PGGs (2–4), but was required to protect our subjects' confidentiality. Game participants were endowed, in each round, with 10 coins of 100 Ugandan shillings which corresponds to half a daily wage in rural Uganda, but paid the payoff of only one round, which was randomly selected at the end of the game. In addition, our subjects received a participation fee, travel reimbursement, and additional gains from other behavioral games played during the day.

The voting procedure in the elected monitor condition guaranteed complete anonymity: each player wrote on a piece of paper the identification number of the player she would like to select as a monitor. Subjects could see each other, but were not allowed to talk and were not given any information about the other participants. Subjects were sampled from six different villages and did not know, on average, more than one or two other PGG session participant (i.e., covillagers). They knew, however, that all subjects shared membership in their farmer cooperative. Finally, none of the participants held leadership positions in the farmer associations.

Data were collected between July 2009 and September 2009 by a group of 60 experienced local interviewers, divided into three language teams. The experimenter script was translated and back-translated from English into each of the native languages (Basoga, Luganda, and Ranyankole), and several pilot tests and debriefing sessions were conducted. Interviewers went through a 2-wk training in class and in the field setting, which included training on human subjects issues as well as survey techniques, and they were supervised by team leaders through the entire data collection.

ACKNOWLEDGMENTS. We are grateful to Alexander Barnard, Eliana Horn, Sylvie Hoster, Vivian Lu, and our local team of interviewers for their research assistance. We thank Abigail Barr, Diego Gambetta, Benjamin Goodrich, Michael Hechter, Macartan Humphreys, Daniel Rubenstein, and the anonymous reviewers for helpful comments. D.B. gratefully acknowledges support from National Science Foundation Grant SES(IOS)-0924778 and the Princeton Institute for International and Regional Studies. G.G. gratefully acknowledges support from National Science Foundation Doctoral Dissertation Improvement Grant SES-0921204.

[‡]In the experiment, monitors did not gain from higher levels of cooperation, and they had a monetary disincentive to sanction. Nonetheless, their identity was known by the group, and thus we could not measure the extent to which their sanctioning decisions were affected by reputation considerations.

[§]Barr et al. (40) ran a corruption experiment with random and elected monitors, demonstrating that elected monitors were more willing to invest in exposing corruption.

1. Olson M (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard Univ Press, Cambridge, MA).
2. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
3. Sigmund K (2007) Punish or perish? Retaliation and collaboration among humans. *Trends Ecol Evol* 22:593–600.
4. Henrich J, et al. (2006) Costly punishment across human societies. *Science* 312:1767–1770.
5. Rockenbach B, Milinski M (2006) The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444:718–723.
6. Camerer C (2003) *Behavioral Game Theory: Experiments in Strategic Interaction* (Russell Sage Foundation, New York).
7. Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Econ Lett* 71:397–404.

8. Ostrom E (2000) Collective action and the evolution of social norms. *J Econ Perspect* 14:137–158.
9. Gintis H, Bowles S, Boyd R, Fehr E (2005) *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life* (MIT Press, Cambridge, MA).
10. Fowler JH (2005) Human cooperation: Second-order free-riding problem solved? *Nature* 437:E8, discussion E8–E9.
11. Rand DG, Dreber A, Ellingsen T, Fudenberg D, Nowak MA (2009) Positive interactions promote public cooperation. *Science* 325:1272–1275.
12. Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560–1563.
13. Boyd R, Gintis H, Bowles S, Richerson PJ (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100:3531–3535.
14. Carpenter J, Matthews P (2009) What norms trigger punishment? *Exp Econ* 12:272–288.

15. Masclot D, Noussair C, Tucker S, Villeval MC (2003) Monetary and nonmonetary punishment in the voluntary contributions mechanism. *Am Econ Rev* 93:366–380.
16. Janssen MA, Holahan R, Lee A, Ostrom E (2010) Lab experiments for the study of social-ecological systems. *Science* 328:613–617.
17. Boyd R, Gintis H, Bowles S (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328:617–620.
18. O’Gorman R, Henrich J, Van Vugt M (2009) Constraining free riding in public goods games: Designated solitary punishers can sustain human cooperation. *Proc Biol Sci* 276:323–329.
19. Taylor M (1982) *Community, Anarchy, and Liberty* (Cambridge Univ Press, Cambridge, MA).
20. Fowler JH (2005) Altruistic punishment and the origin of cooperation. *Proc Natl Acad Sci USA* 102:7047–7049.
21. Hauert C, Traulsen A, Brandt H, Nowak M, Sigmund K (2007) Via freedom to coercion: The emergence of costly punishment. *Science* 316(5833):1905–1907.
22. Greif A (1993) Contract enforceability and economic institutions in early trade: The Maghribi traders’ coalition. *Am Econ Rev* 83:525–548.
23. Henrich J, et al. (2010) Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327:1480–1484.
24. Greif A, Milgrom P, Weingast BR (1994) Coordination, commitment, and enforcement: The case of the merchant guild. *J Polit Econ* 102:745–776.
25. Portes A (2010) *Economic Sociology: A Systematic Inquiry* (Princeton Univ Press, Princeton, NJ).
26. Gibson CC, Marks SA (1995) Transforming rural hunters into conservationists: An assessment of community-based wildlife management programs in Africa. *World Dev* 23:941–957.
27. Atleson JB (1969) Union fines and picket lines: The NLRA and union disciplinary power. *UCLA Law Rev* 17:681–757.
28. Guth W, Levati MV, Sutter M, van der Heijden E (2007) Leading by example with and without exclusion power in voluntary contribution experiments. *J Public Econ* 91: 1023–1042.
29. Hechter M (2009) Legitimacy in the modern world. *Am Behav Sci* 53:279–288.
30. Weber M (1922) *Economy and Society* (JCB Mohr, Tübingen, Germany); reprinted (1978) (Univ of California Press, Berkeley).
31. Barr A (2003) Trust and expected trustworthiness: Experimental evidence from Zimbabwean villages. *Econ J* 113:614–630.
32. Habyarimana J, Humphreys M, Posner DN, Weinstein JM (2009) *Coethnicity: Diversity and the Dilemmas of Collective Action* (Russell Sage Foundation, New York).
33. Henrich J, et al. (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford Univ Press, New York).
34. Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J Econ Perspect* 21:153–174.
35. Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. *Science* 319:1362–1367.
36. Jones D (2010) Psychology. A WEIRD view of human nature skews psychologists’ studies. *Science* 328:1627.
37. Tyler TR (1997) The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Pers Soc Psychol Rev* 1:323–345.
38. Levi M, Sacks A (2009) Legitimizing beliefs: Sources and indicators. *Regulation & Governance* 3:311–333.
39. Sigmund K, De Silva H, Traulsen A, Hauert C (2010) Social learning promotes institutions for governing the commons. *Nature* 466:861–863.
40. Barr A, Lindelow M, Sernees P (2009) Corruption in public service delivery: An experimental analysis. *J Econ Behav Organ* 72:225–239.
41. Dickson ES, Gordon SC, Huber GA (2009) Enforcement and compliance in an uncertain world: An experimental investigation. *J Polit* 71:1357–1378.
42. Erikson E, Parent JM (2007) Central authority and order. *Social Theory* 25:245–267.
43. Chase ID, Tovey C, Spangler-Martin D, Manfredonia M (2002) Individual differences versus social dynamics in the formation of animal dominance hierarchies. *Proc Natl Acad Sci USA* 99:5744–5749.
44. Rands SA, Cowlshaw G, Pettifor RA, Rowcliffe JM, Johnstone RA (2003) Spontaneous emergence of leaders and followers in foraging pairs. *Nature* 423:432–434.
45. King AJ, Johnson DDP, Van Vugt M (2009) The origins and evolution of leadership. *Curr Biol* 19:R911–R916.