Essays on Leadership Selection and Public Goods Provision in Self-Help Organizations

Guy Grossman

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Essays on Leadership Selection and Public Goods Provision in Self-Help Organizations

Guy Grossman

This dissertation examines the relationship between leadership selection and public goods provision in self-help organizations. Leadership selection is defined as the rules for selecting leaders, as well as the factors that determine the quality of the leadership class. Self-help organizations are defined as relatively small-size voluntary groups that are created to provide goods and services to members and that select their leader via democratic procedures. Examples include micro-lending and micro-insurance groups, common-pool resource groups, women and artisan cooperatives and farmer associations.

The dissertation focuses on Ugandas recent largest development project: the Agriculture Productivity Enhancement Project (APEP). USAID funded, APEPs stated goal is to expand rural economic opportunities by supporting the transition of smallholder producers into commercial farming. During the projects lifespan (2005-2009), APEP helped about 60,000 small-scale producers to organize into over 200 farmer associations (i.e. cooperatives). Importantly, the success of these farmer associations in overcoming social dilemmas and in providing goods and services to their members, varied tremendously. *Why are some groups more successful than others in overcoming the social dilemmas inherent in public goods production?* To explain this variation, the dissertation uses a range of disciplinary perspectives — drawn mainly from political science, economics, social psychology and sociology — as well as a diverse set of methodological tools.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to sincerely thank my dissertation advisor and mentor Macartan Humphreys for his outstanding guidance, caring, patience, and for providing me with an excellent atmosphere for doing research. His contribution to this project extends beyond what may be revealed in the text.

I am especially indebted to Timothy Frye and Kimuli Kasara for their close readings, valuable suggestions and for supporting a dissertation topic that all but conforms to the mainstream. Special thanks goes to David Laitin and David Stasavage, who were willing to participate in my defense committee at a short notice.

I am also in awe to my teachers Vicky Murillo, Isabela Mares, Sheri Berman, Lucy Goodhart and Mona el-Ghobashy whose passion to the discipline was inspiring and contagious, and to John Huber, Robert Shapiro, Andy Gelman and Greg Wawro who taught me the tools that made this project possible.

I wish to extend a special thanks to Delia Baldassarri, whose encouragement, support and belief in this research project from the preliminary to the concluding level were vital for its successful completion.

I am grateful to the many faculty and colleagues who helped me along the way: Akbar Noman, Bill Mcallister, Kate Baldwin, Alex Scacco, Mat Winters, Bernd Beber, Benjamin Goodrich, Mike

Ting, Alessandra Casella, Navin Kartik, Massimo Morelli, Virginia Oliveros, Narayani Lasala, Boliang Zhu, Stefano Recchia, Peter van der Windt, Pierce O'Reilly, Neelanjan Sircar, Kelly Rader, Ram Fishman, Jordan Kyle, Kristin Michelitch, the participants at the Comparative Workshop at Columbia University and at the Contemporary African Political Economy Seminar.

For generously supporting my research, I thank Columbia University, The Mellon Interdisciplinary Graduate Fellows Program, Columbia's Center for the Study of Development Strategies, Columbias multidisciplinary Program on International Development and Globalization (IGERT) and the National Science Foundation.

I thank Sylvie Hoster, Alex Barnard, Eliana Horn, Vivian Lu and my Ugandan research team for their excellent research assistance.

I am heartily thankful to Milan Vaishnav, Laura Paler, Cyrus Samii, Janet Lewis, Walker Hanlon and Elizabeth Sperber who provided not only endless streams of insight and support, but who never once failed to be what we hope to find in friends.

I thank my mother, my siblings and my daughter Maya for their love, support and understanding.

Finally, I thank Advah, my love, for everything else. Her endless caring, support and encouragement is woven into every page of this dissertation.

I dedicate this dissertation to my late father Emerich Grossman.

Part I

Dissertation Chapters

Chapter 1

Overview

1.1 Motivation

State withdrawal and increasing democratization of public life in many developing countries have encouraged the rapid proliferation of local, self-help groups — autonomous, voluntary-based, grassroots organizations, which are created to provide collective goods and services to their members. Empirical evidence supports this claim: in Senegal 10% of sampled villages reported having at least one self-help group in 1982; by 2002 this figure was 65%. In Burkina Faso the figures were 22% for 1982 and 91% in 2002 (Bernard et al., 2008). The rise of these groups — which include farmer associations, artisan cooperatives, common-pool resource groups, saving and micro-credit groups and micro-insurance — is also driven by the active support of the aid community. In the past decade the World Bank, bilateral donor governments and INGOs have increasingly funded development interventions that promote self-help groups, as part of a paradigm that stresses the positive effects of participatory, community-driven development.

However, the evidence at hand suggests that the effectiveness of self-help organizations in alleviating poverty has been very mixed (Easterly, 2001, Hellin, Lundy and Meijer, 2009). To provide their members with high value goods and services, self-help groups must overcome classic collective action problems (Olson, 1965)[1]. *Why are some groups more successful than others in overcoming the social dilemmas inherent in public goods production*[2]?

1.2 Theoretical Framework

This dissertation addresses this question by drawing on three, distinct literatures: (i) the sociological and economics research on the effectiveness of self-help groups, (ii) the political economy literature on political selection, and (iii) the political science and political economy work on public

[1] In this study I focus exclusively on existing self-help groups, which have already solved the collective action problem inherent in group formation. This focus allows me to view those groups as small communities, while sidelining free-riding tensions between members and non-members. The social dilemmas addressed in this study are therefore part of a broader class of obstacles to cooperation between group members in the production of club goods and services.

[2] The term social dilemma refers to situations in which group and individual incentives are at odds (Heckathorn, 1996). Throughout the dissertation I use the terms social dilemma and collective action problem interchangeably.

goods provision. In doing so, I seek to provide new insights into an organizational form — self-help groups — that is playing an increasingly important role in shaping welfare outcomes in poor areas. But I also seek to do so in a way that contributes to each of these three larger theoretical literatures. The subsections that follow briefly summarize the dissertations contributions to each of these literatures.

Literature on the Effectiveness of Self-help Groups

The study of voluntary self-help organizations performance, with few exceptions, has been dominated by sociologists and development economists. The sociological approach generally emphasizes the strength of social ties between group members, through the *social capital* paradigm (Woolcock and Narayan, 2000) or via *social network analysis* (Simmons and Birchall, 2008). Here, the strength of informal relations has been shown to influence the spontaneous creation of protection systems (Krishna, 2002), facilitate the diffusion of information (Fafchamps and Minten, 2001) and of innovation adoption (Isham, 2002), increase the success of microcredit programs (Cassar, Crowley and Wydick, 2007, Karlan, 2007), and foster civic engagement conducive to cooperation (Gittell and Vidal, 1998). The economic approach, by contrast, commonly focuses on factors such as *group attributes* (e.g., group size (Agrawal and Goyal, 2001); income, religious or ethnic heterogeneity (Alesina and Ferrara, 2000)), *peer sanctioning* (Ostrom, 1990), and various factors external to the group (e.g., terrain, rainfall, soil quality, or market conditions (Berdegué, 2001)).

Though this body of work has contributed much to our current understanding of voluntary self-help groups, it suffers from a fundamental limitation: it tends to treat self-help groups as non-hierarchical organizations, ignoring the key role that social differentiation processes play in solving social dilemmas[3]. Specifically, this literature overlooks the importance of leaders in inducing cooperation in small group settings. Integrating a political science perspective — one that takes power relations and hierarchies as its starting point — into the study of self-help groups' effectiveness, is the first theoretical contribution of this study.

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

The Political Selection Literature

In contrast, the political economy literature on political selection (Besley, 2005, Kotakorpi and Poutvaara, 2010), the second body of work that informs this study, does assume an hierarchical world. I build on the key insight of this literature — throughout the dissertation but most directly in Chapter 3 — that the profile (e.g., quality) of the political class has welfare and distributional implications (Bianco and Bates, 1990, Caselli and Morelli, 2004). However, as far as my knowledge goes, the political selection literature has yet to explore its implications beyond the selection of professional career politicians (Mattozzi and Merlo, 2008). The second major theoretical contribution of this study is to expand the scope of this literature by offering, in Chapter 3, a formal model adapted specifically to the small group setting.

Public Goods Provision Literature

Finally, the dissertation builds on the political science and political economy literatures on the relation between governance institutions and public goods provision (Funk and Gathmann, 2009, Persson, Roland and Tabellini, 2000). Specifically, I apply a 'new institutionalist' perspective in exploring how institutional design matters for self-help organizations. There are, however, several limitations to the public goods literature, which this project seeks to address.

First, with few exceptions, this body of work focuses almost exclusively on the distributive roles of local (Besley and Case, 2003) and national governments (Persson and Tabellini, 2003), overlooking the importance of self-help groups in public goods provision. As a result, the current political science literature is quiet about the possibility that institutional design matters in intimate and information rich environments (*cf.* Olken (2010)). The third theoretical contributions of this research project is, therefore, *testing whether institutional rules 'have a bite' in small-group settings*. I explore this question directly in chapter 2 and chapter 4.

Second, following a 'new institutionalist' approach, the political science work on public goods provision begins from the assumption that individuals' expectations and behavior are shaped by

incentives embedded within formal and informal rules (Levitsky and Murillo, 2009). Yet with the exception of a small literature on legitimacy (Levi, Sacks and Tyler, 2009), this literature does not consider the possibility that individuals' behavior is also a function of the *process by which institutions were put into place*. Another theoretical contribution of this research project is to examine this possibility directly: I ask, *do the procedural rules for selecting leaders impact individuals' behavior in ways that cannot be reduced to some incentives scheme?*

The third limitation of the public goods provision literature lies in its difficulty to demonstrate its causal claims. This limitation stems from the fact that (i) national political institutions have many elements that are bundled together and that likely change at the same time (Acemoglu and Johnson, 2005), (ii) the adoption of governance institutions tends to be a function of highly endogenous factors (Acemoglu, Johnson and Robinson, 2001), and (iii) empirical analysis face the serious problem of not having a valid micro foundation, because the theory's predictions are commonly tested with aggregated national data. This final limitation is important: new institutionalism is ultimately a micro foundational theory: individuals' behavior is explained as a function of incentives embedded within institutional arrangements (North, 1990). I seek to address these problems through the dissertation's research design.

1.3 Research Design and Methodology

In contrast to national and state-level institutions, establishing causal effects of political institutions and tracing micro-foundational behaviors can be more feasible at the local level. In this section, I briefly introduce the Ugandan farmer associations that are the focus of the empirical parts of the study. I then describe the collective action problem that hinders their performance. Finally, I turn to review the methodology I use for studying the impact of leadership selection on public goods production[4].

[4]More accurately, most of the goods that self-help groups produce should be classified as 'club goods', a subtype of public goods that are *excludable* but *non-rivalrous*. I prefer the use of the term public goods, in order to relate this study more easily to existing literatures.

A Very Short History of Farmer Associations in Africa

Farmer associations' raison d'étre is to improve the performance of their members' farms as economic units engaged in market transactions. Among self-help groups, farmer associations occupy a central role in developing countries' poverty alleviation strategies (Narayan-Parker, 2002*a*). In poor countries such as Uganda, the majority of citizens derive their lion share of income from agriculture. For example, according to the 2002 census (UBOS, 2002), over 88% of Ugandans live in areas that are regarded rural. Most of those living in rural areas are smallholder producers that do not participate actively in the monetized economy. For this reason, policies for integrating smallholder, subsistence farmers into markets have been among the most salient issues for both colonial and post-colonial governments (Bates and Lofchie, 1980). Yet despite numerous efforts in recent decades to collectivize production, and later crop marketing, many initiatives have failed, not least because in many developing countries, colonial and national governments sought to control and exploit farmer organizations (Bates, 1981, Lele, 1981).

In the early 1990s, as a result of structural adjustment liberalization policies, almost all government controlled farmer cooperatives had become insolvent (Ponte, 2002). This process triggered an immediate renewal in the interest of the development community in supporting farmer groups (Hussi et al., 1993). In recent years, development agencies have been increasing their support for farmer organizations, which are still believed to be the most efficient means for integrating small-scale producers into markets (Birchall, 2003). Here too, however, studies assessing the contribution of aid-supported farmer groups to raising income and farm output report very mixed results (Ashraf, Gine and Karlan, 2009, Berdegué, 2001).

The Research Site: APEP Groups

All of the farmer associations surveyed for this project were created as part of one of Uganda's largest recent development projects: the Agriculture Productivity Enhancement Project (APEP)[5].

[5]APEP was funded by USAID, and implemented by Chemonics International, an International Development consulting firm based in Washington DC.

APEP's goal was to support small-holder farmers' transition into commercial farming. Between 2004 and 2008, it helped organize about 60,000 small producers into more than 2,500 village-level groups, known as producer organizations (POs), which were further organized into more than 200 farmer associations across Uganda, known as Depot Committees (DCs). Serving, on average, 200 members from ten neighboring village-level groups, the farmer associations were designed to exploit economies of scale and to bargain for better prices based on quality and volume[6].

Social Dilemmas in Farmer Associations: A Running Example

Consider the social dilemma that hinders the effectiveness of farmer associations in providing the most important service to their members: securing higher output prices through collective marketing. This social dilemma will serve as our running example throughout this dissertation.

Because of the high costs of transportation and market information in many developing countries, dispersed small-holder farmers are restricted to sell their crops through local middlemen. These agents often exploit asymmetries in information and bargaining powers, offering dispersed farmers below-market prices. By contrast, organized farmers can avoid operating through these agents. Instead, they can sell their crops via their association in bulk and obtain higher prices—their size increases their bargaining power and reduces buyers' transaction costs (Staatz, 1987).

Yet, once a farmer group is founded, middlemen raise their prices offered to individual farmers to remain competitive. Since middlemen, unlike most farmer groups, collect the crops at the farm-gate and pay cash-on-delivery, members have a private interest in selling their produce to middlemen. The private gain of selling to middlemen ('defecting'), however, is conditional on a sufficient number of *other members* selling via the group ('cooperating'). This is because the price offered by middlemen depends on the price that the farmer group secures, which depends crucially on volume. If too many members defect, collective marketing collapses and all individuals typically receive a profit lower than they would, had they cooperated with the group. As Figure 1.1 makes clear, some groups manage to overcome this tension between private and group inter-

[6]For a discussion of reasons for choosing to focus on the APEP groups, see Appendix A.1.

ests. Many others fail[7]. Explaining this variation in the effectiveness of self-help groups is the central goal of this study.

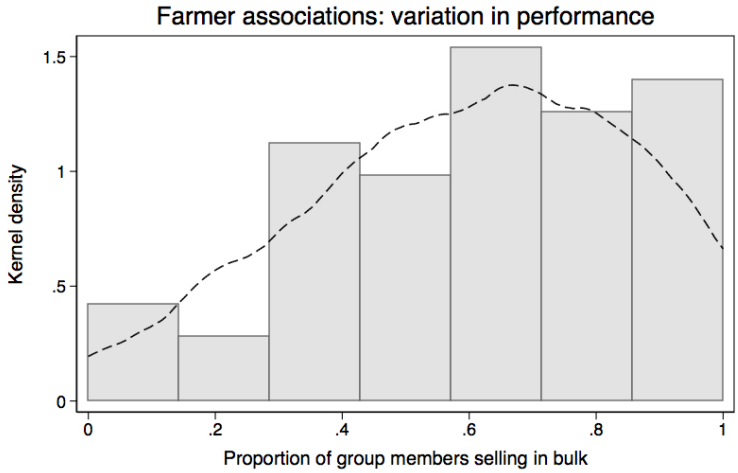

Figure 1.1: Variation in performance. Figure displays the distribution of the proportion of members in each of the 50 sampled Ugandan farmer association who sold their coffee via the group, at least once, in the past season.

Methodological Approach

As mentioned above, research at the local level can apply a wide set of methodological tools that are not typically available for national-level research. In this section, I briefly discuss the methodological approach I take for testing the study's hypotheses regarding the impact of leader selection rules on public goods production.

First, the empirical analysis is informed by a year of fieldwork in Uganda, in which I applied ethnographic methods. These include conducting, in person, dozens of semi-structured interviews with APEP staff and with members and leaders of the APEP groups, as well as reviewing a large number of documents and reports produced by the project administrators. The ethnographic work proved invaluable. For one, it provided the basis for constructing the survey instruments. Secondly, it allowed me to develop the intuition I use for introducing a theoretical model on the

[7] In Chapter 3 I demonstrate that farmer groups that are able to overcome this social dilemma contribute significantly to their members' welfare (Grossman, 2011*a*).

determinants of leader quality in self-help groups in chapter 3. Third, the familiarity I gained with the administrative apparatus of APEP allowed me to identify natural conditions that resulted in (plausibly) exogenous variation in leader selection rules. This is the basis for the encouragement research design (Angrist and Krueger, 2001) I use for studying the impact of leader selection rules on leader accountability in chapter 4.

Second, I base the study's identification strategy on taking a large set of behavioral experiments typically conducted in a laboratory environment to a field setting, following (Habyarimana et al., 2009) . These experiments, which I adapted specifically to address the issues at hand, allow me to further uncover causal relations between leader selection rules and public goods provision [Chapters 2 and 4][8].

Why not conduct the experiments in a laboratory at Columbia University in New York City? First, recent evidence suggests that students' behavior in behavioral experiments might not be representative of the larger society (Jones, 2010). Secondly, there is growing evidence suggesting that in the context of behavioral experiments that measure social-preferences, subjects from developing countries behave differently than their counterparts in the developed world (Cardenas and Carpenter, 2008). Third, the dissertation's strategy for increasing the external validity is based on conducting experiments with, and collected observational data on, members of pre-existing groups that face collective action problems on a regular basis.

For this purpose I conducted more than 3,100 individual-level surveys with a random sample of members of those groups. In addition, I collected unique social network data in each of the 50 sampled farmer associations studied herein. These original data allow me to corroborate the experimental results, contributing to the external validity of the study's key findings. One of the major contributions of this study is, therefore, methodological: combining original survey data, social network analysis, formal modeling, behavioral games and a encouragement research design within the framework of a single study.

[8]See (Habyarimana et al., 2009, pp. 13–19) for a comprehensive exposition of the advantages of conducting behavioral experiments in a field-setting.

1.4 Chapter Summaries

Chapter 2: Internal Centralized-Sanctioning and Cooperation in Self-Help Groups

This chapter contributes to a debate on the role sanctioning plays in fostering cooperation in social dilemmas. Sanctioning is widely considered to be a viable solution to the collective action problem inherent in public goods production (see Sigmund (2007) for a review). Scholars have generally focused on two forms of sanctioning solutions. In the first solution, *centralized authorities* that are *external* to the group — such as the state (Scholz and Gray, 1997) — are the locus of coordination and enforcement of cooperative efforts. Most recent empirical and theoretical work has focused, instead, on a second solution, in which cooperation emerges from diffused and decentralized punishment (Boyd et al., 2003, Gintis et al., 2005)

However, scholars have recently begun questioning the ability of spontaneous, decentralized and uncoordinated *peer-punishment* actions to sustain cooperation in complex organizational structures (Boyd, Gintis and Bowles, 2010, O'Gorman, Henrich and Van Vugt, 2009), concluding that "the step from peer punishment to the establishment of sanctioning institutions deserve closer future investigations" (Sigmund, 2007, p. 598-9). Peer-sanctioning, I argue, is only effective under very restrictive conditions: namely, it can only sustain cooperation in small-size groups, where the cost of punishment is likely to be recuperated. In such groups, self-interested contributors may choose to punish defectors at a personal cost, as long as they have reasons to believe that punishment will increase future contributions. When groups become large and interactions between members infrequent, bilateral punishment is unlikely to sustain cooperation because future gains from punishment cannot be internalized.

To overcome this problem, groups commonly develop forms of self-regulation, in which the power to sanction defectors is transferred to a centralized authority, internal to the group (Greif, Milgrom and Weingast, 1994). Internal centralized sanctioning institutions are likely to be more efficient than peer-punishment (Guth et al., 2007), since they are better positioned to overcome coordination failures (O'Gorman, Henrich and Van Vugt, 2009). To incorporate these ideas into

theories of public goods provision, I study how group members behave in a context in which a centralized monitor is given a monopoly over sanctioning decisions. I ask, will groups reach high levels of contribution, even if their members do not have the power to decide who should be sanctioned? One of the contributions of this paper is testing and documenting the effectiveness of centralized authorities, *internal to the group*, in fostering cooperation[9].

The chapter addresses an additional open question: whether the *political process* through which internal centralized authorities obtain their sanctioning powers is consequential for cooperation. Differently from peer-sanctioning systems, in which the right to punish defectors comes hand-in-hand with group membership, in a centralized-sanctioning regime it is important to distinguish between the impact of sanctioning and the impact of the way in which sanctioning powers are granted. Specifically, I test whether elections have a positive impact on group members' contributions to public goods production. When analyzing the causal mechanisms that possibly tie elections and leader's effectiveness in fostering cooperation, I test the hypothesis that elections increase cooperation through a "legitimacy effect". Namely, it examines whether individuals are more likely to commit to a leader's authority if they participate in her selection.

To investigate these hypotheses, I combine "lab-in-the-field" behavioral experiments with observational data on 1,543 producers from a sample of 50 Ugandan farmer associations. I developed a novel adaptation of the public goods game, which is the conventional behavioral experiment used to study the conditions under which groups can overcome individual incentives to defect (Camerer, 2003). The experimental setup allows me to attest the impact of centralized-sanctioning institutions on cooperative behavior as well as to demonstrate that the size of this effect depends on the process by which these institutions are established. To assess the external validity of those findings, I relate the farmers' behavior in the experiment to their level of cooperation in the farmer organization and show that farmers' deference to authority in the controlled setting predicts cooperative behavior in their natural environment, in which they face a similar social dilemma.

[9]See also Frye (2000) that examines the conditions under which private groups govern themselves without turning to an outside agent for enforcement.

Chapter 3: A Theory of Leadership Selection in 'Small' Self-Help Groups

This chapter continues my examination of the factors affecting leaders' effectiveness in public goods production, by focusing on the determinants of the *quality* of internal centralized authorities. Leader quality is thought of, here, as a combination of the leader's ability and the amount of effort that the leader exerts while working for the group.

While many factors may affect the quality of leadership obtained by a small group, this chapter focuses specifically on two. First is the ability of a group to monitor its leader in order to incentivize effort, which depends on the monitoring institutions available to the group. Second is the availability of private income opportunities outside of the group, which depends on local economic conditions. While both of these factors are likely to evolve over time, neither can be changed rapidly from one election cycle to the next[10]. The main goal of this chapter is to present and test a model that uncovers the conditions under which monitoring institutions can *worsen* the quality of the candidacy pool.

The model I develop is applicable to groups in which leaders, who play a decisive role in producing a group public good, are chosen endogenously from the set of members through democratic procedures. This is a feature shared by many small groups, including the farmer associations studied herein. The starting point for the model is, therefore, a citizen-candidate framework (Besley and Coate, 1997), which a number of recent studies have used to investigate issues of leadership quality (Gagliarducci, Nannicini and Naticchioni, 2010, Messner and Polborn, 2004).

This chapter departs from existing theories in that it reflects the particular features of self-help groups, which differ from larger political units in a number of ways. First, unlike large political units, in small groups members generally know each other well. This means that incomplete information plays a smaller role in determining outcomes in small group settings. It also means that small groups have an advantage over large political units in offering 'high-powered' incentive

[10]From a group's perspective, there are good reasons to ensure relative 'stickiness' of monitoring institutions. One reason is to protect against attempts by incumbents to alter monitoring institutions to increase incumbency advantage. I thus treat monitoring as a fixed parameter for any given elections and consider the quality of leadership that a group obtains, given the monitoring institutions and local private income opportunities present at any specific time.

schemes that condition remuneration on effort or outcome. Second, small groups are often formed with a specific purpose in mind, so that members' goals are generally more closely aligned than in larger political units. Third, participating in small groups, even as the leader, is generally a part-time affair. Rarely do small groups have the resources to employ full-time leaders, as is common in larger political units. The result is that leaders must choose how to allocate their time between producing the group public good and generating private income. Fourth, the leader receives significant benefits from the public good that she produces. This has an important effect on members' incentives to seek leadership positions.

The model's principal theoretical argument is that groups may face a trade-off between the ability of the elected leader and the amount of effort that the leader exerts. This trade-off is driven by two effects. First, an increase in monitoring of the leader will incentivize the leader to exert more effort. However, a higher level of monitoring may also cause high ability members to self-select out of the candidate pool, resulting in leaders with lower ability being elected. When both of these effects are operating, the result is a rough inverted U-shaped relationship between the level of monitoring and the value of the public good produced. At low levels of monitoring, high ability leaders are elected, but the exert little effort, leading to a low public good value. An increase in the level of monitoring causes the leader to exert more effort, increasing the value of the public good. However, if the level of monitoring continues to increase eventually high ability members will start to self-select out of the candidate pool, reducing the value of the public good. Importantly, this trade-off exists *only when private income opportunities are sufficiently high*. If private income opportunities are low, then high ability members have little reason to opt-out of candidacy. Thus, when there are few private income opportunities outside of the group activities, groups will be able to obtain high ability leaders who also exert a relatively high level of effort.

The model's predictions are tested using the original data I gathered from an extensive survey of Ugandan farmer associations. The data support the predictions of the model, while also providing evidence for one of the central arguments of this dissertation: that the quality of leadership in small groups affects the value of the public good produced and, thereby, members' welfare.

Chapter 4: Do Leader Selection Rules Impact Accountability?

In chapter 2, I use a set of behavioral experiments to study the impact of elections on cooperation. In that chapter, the main focus of the analysis was the *cooperative behavior of group members* in a social dilemma, under different experimentally-induced political institutions. In chapter 4, I return to explore the impact of leader selection rules. In this chapter, however, my identification strategy is based on natural conditions that resulted in exogenous variation in the rules for selecting the senior managers of the APEP associations. Specifically, I examine the causal impact of moving from an appointment-based rule to popular direct elections on *leader accountability*[11].

A survey of the literature on the impact of the rules for selecting officials – such as judges, public regulators, school board members and the CEOs of traded companies – suggests that the debate over the relative virtues of elections and appointments is far from settled. On the one hand, there are good theoretical reasons to assume that different leadership selection methods will have varying effects on political and economic outcomes of interest. The empirical findings, however, are ambiguous, not least because the adoption of governance institutions tends to be a function of other group-specific factors, which makes the identification of causal impact nearly impossible (Acemoglu, Johnson and Robinson, 2001).

This chapter addresses the identification problem by exploiting plausibly exogenous variation in the rules for selecting the senior managers (henceforth leaders) of the APEP groups. The paper's identification strategy — explored in length in Section 4.4 — rests upon the fact that (i) APEP field-facilitators played a foundational role in establishing the farmer associations, (ii) almost all groups adopted the field facilitator recommendation when choosing between an election and an appointment rule for selecting their manager (iii) the idiosyncratic preferences of individual field-facilitators that informed their recommendations are plausibly orthogonal to other characteristics of the groups they were hired to work with, and (iv) the deployment of field facilitators to districts was orthogonal to the characteristics of the groups.

[11] Appointment denotes a rule by which the association's board of directors, in which the farmer groups that make-ep the association are equally represented, select the DC manager. Direct denotes a rule by all registered members from each of the village-level group that make up the association can cast their vote.

Using the facilitators' recommendation as an instrumental variable, I test the impact of moving from appointments by board directors to direct elections along two dimensions[12]. First, I examine the impact of leader selection rules on the extent to which leaders are monitored by, and accountable to, group members. I find that direct-voting results in leaders who are significantly more accountable to group members. Compared to appointment-based groups, farmer associations that use direct, popular elections are more likely to employ internal and external auditors, to hold selection procedures on a more regular basis, and to ensure that members receive receipts when selling crops through the association.

To explore possible causal mechanisms, I use a set of behavioral experiments, such as dictator and third-party-punishment games. The experimental findings suggest that elections trigger a stronger sense of commitment by local leaders, reciprocating their election by the members of the group. This sort of reciprocity likely operates independently of reelection considerations. In addition, analysis of social network data suggests evidence of an 'affinity effect,' whereby appointment corresponds to affinity between appointees and appointed officials inimical to monitoring and accountability

Secondly, I examine the impact of leader selection rules on the profile of the chosen leaders. Here, I test the hypothesis that appointments yield more competent and/or trustworthy leaders, as compared to popularly elected leaders. The intuition here, developed more extensively in Section 4.3, is that, compared to average group members, boards of directors are more informed about candidates and the current 'state of the world', have longer time-horizons, and are more likely to base their vote on the welfare of the group rather than narrow self-interest. I do not find evidence to support this argument. Selection methods do not appear to impact the profiles of leaders with respect to (i) ability (i.e. marketable skills) or other socio-demographic attributes, (ii) centrality of network position, or (iii) other-regarding preferences.

[12]See Appendix A, Fig. A.2 for a chart of the organizational structure of the APEP associations.

1.5 Lessons Learned

Leader Selection Processes Matter. In the past two decades, economists, political scientists and political economists working within a 'new institutionalist' framework, have focused on demonstrating how individuals' behavior is shaped by *incentives* embedded within institutions. Importantly, the various expansions of North's basic approach — which include distinguishing between formal and informal, weak and strong, or exogenous vs. endogenous institutions — all continue to place incentives at the center of the theory's micro-foundation. This dissertation complements and enriches the 'new-institutionalist' framework by pointing to the causal impact of the *process by which institutions were put into place*. In this study I limit my focus to leader selection rules. For example, in chapter 2, I demonstrate that when leaders acquire their regulatory and sanctioning powers through elections, they are able to exercise their sanctioning and monitoring roles more effectively. In chapter 4, I demonstrate that compared to appointed leaders, elected leaders are more accountable and responsive to their constituents. Future work should test the extent to which these findings apply to other institutional settings.

Pecuniary Considerations. The second key lesson learned from this research project is that process matters *independently from pecuniary considerations*. In chapter 2, I demonstrate that the procedure for selecting internal centralized authorities impact cooperative behavior *independently from the threat of sanctioning*, i.e. independently from monetary considerations. Instead, I provide evidence supporting the idea that elections legitimate the use of power: individuals are more likely to confer authority to leaders whom they have directly elected. These findings have implications that far exceed the study of small self-help groups. Past research has demonstrated that generalized trust and norms of reciprocity (Henrich et al., 2004) can "trump" monetary incentives, challenging some of the basic tenets of economics. This dissertation adds to this body of literature by demonstrating how leader selection procedures can impact behavior in ways that trump income maximization strategies.

Power Considerations. The third lesson is that process matters *independently from power considerations*. In chapter 4, I demonstrate not merely that office holders' responsiveness to their constituents is a function of leader selection processes, but also that the mechanism that ties elections to leaders' responsiveness is based on reciprocal expectations, which *cannot be reduced to reelection considerations*. This finding challenges the conventional wisdom in political science that tends to explain the behavior of politicians instrumentally, as a function of power calculations (i.e. winning elections) or of rent extraction opportunities.

Leader Quality Matters The forth lesson is that leader quality matters. In chapter 3, I demonstrate that the ability of leaders and, especially, the effort they exert while working for the group have a positive and significant impact on public goods production and, thereby, members' welfare. This finding sheds new light on the effectiveness of voluntary self-help organizations since prior work, as I have argued, tends to overlook the importance of hierarchical structures.

Analysis of Punishment in Public Goods Games. In chapter 2 I introduce a new method for analyzing sanctioning behavior in public goods games, where the experimenter does not control the contributions to the public goods that monitors face. The method — matching on distributions of contributions — improves on the current practice that analyzes punishment behavior in public goods games as if it was experimentally induced. This contribution should extend beyond the boundaries of our discipline.

Chapter 2

The Role of Internal Centralized-Sanctioning in Public Goods Provision in Self Help Groups

2.1 Introduction

Self-help groups are small to medium size organizations that provide members with valuable public goods. Such groups come in many varieties, from farmer associations in Uganda and micro-credit groups in Bangladesh to common-pool resource communities in Japan (Ostrom, 1990). As these few examples suggest, self-help groups are present in many facets of political and social life, and in countries of all income levels. The importance of these groups, especially in developing countries, has grown in recent years, as larger political units have sought to democratize, decentralize and liberalize their economies[1]. The ubiquity of those groups, together with their potential to affect their members' welfare, calls for a better understanding of the factors that determine their effectiveness in public goods provision.

To provide their members with public goods, self-help groups must overcome collective action problems (Olson, 1965). Consider the social dilemma that hinders the effectiveness of many Ugandan farmer associations – the group considered herein[2]. Farmer associations exist to provide members with group public goods, the most important of which is securing higher output prices through collective marketing[3]. Because of the high costs of transportation and of market information in many developing countries, dispersed small-holder farmers are restricted to sell their crops through local agents. These agents, or middlemen, likely exploit asymmetries in information and in bargaining powers, offering dispersed farmers below-market prices. By contrast, organized farmers who sell via their association (in bulk), can obtain higher prices by increasing their bargaining powers and by reducing buyers' transaction costs (Staatz, 1987).

Yet, collective marketing is subject to a social dilemma: once a farmer association is in place, middlemen tend to raise their prices to remain competitive. Since middlemen, unlike most farmer groups, collect the crops at the farmers' gate and pay cash-on-delivery, group members have a pri-

[1]For example, in Senegal 10% of sampled villages reported to have, at least, one self-help group in 1982; by 2002 this figure was 65%. In Burkina Faso the figures were 22% for 1982 and 91% in 2002 (Bernard et al., 2008).

[2]The term social dilemma refers to situations in which group and individual incentives are at odds (Heckathorn, 1996). In the following I use the terms social dilemma and collective action problem interchangeably.

[3]The literature on the potential of farmer organizations as engines of growth is large. This literature suggests that farmer associations can play an important role in generating development and reducing poverty, especially in the context of the developing world. See, among others, Narayan-Parker (2002*a*) and (Bosc et al., 2002).

vate interest in selling their produce to middlemen. The private gain of side-selling to middlemen ('defecting'), however, is conditional on a sufficient number of *other members* selling their crops via the farmer group ('cooperating'). This is because the price offered by middlemen depends on the price that the farmer group secures ('yardstick effect'), which depends crucially on volume. If too many members defect, collective marketing collapses. Some groups manage to overcome this tension between private and group interests. Many others fail[4].

Sanctioning plays a key role in fostering cooperation in social dilemmas (Fehr and Gächter, 2002). In the farmer associations I study, more than 70% of the sampled members acknowledged, when surveyed, that side-selling to middleman is forbidden and punishable, and more than a quarter of those have reported having received warnings or punishment for being caught side-selling. A full 75% of the group leaders I surveyed claimed to invest efforts in discovering whether members side-sell, and more than 50% claimed that they had been approached by members reporting on other members' defection. In these associations, the enforcement of the norm of cooperation (i.e., 'thy shall not sell to middlemen') involves peer monitoring, but crucially depends on the coordination and centralization of sanctioning powers in the hands of a few group leaders.

Most generally, scholars have identified two forms of sanctioning solutions to public goods provision. In the first solution, *central authorities* that are *external* to the group are the locus of coordination and enforcement of cooperative efforts. Examples of these solutions include theories of the state (Scholz and Gray, 1997). In the second solution, cooperation emerges from uncoordinated and *decentralized* punishment efforts that are *internal* to the group. Examples of these solutions include peer-sanctioning regimes (Ostrom, 1990)[5].

These two solutions, however, do not account for all relevant situations, as our example of the Ugandan farmer associations demonstrates. In fact, even small-size groups and communities are characterized by some level of social differentiation and hierarchy from which *internal centralized*

[4] In chapter 3 I show that farmer groups that are able to overcome this social dilemma make a significant contribution to their members' welfare (Grossman, 2011*a*).

[5] This literature begins from the premise that cooperation can be sustained only if, at minimum, some uncoordinated group members are willing to bear the cost of punishing defectors (Gintis et al., 2005). Decentralized peer punishment actions include (i) social sanctioning such as shaming and gossiping, (ii) monetary punishment such as severing of commercial ties, and (iii) applying force or physical constraint.

authorities emerge (King, Johnson and Van Vugt, 2009)[6]. To adjudicate disputes, villagers in traditional societies turn to their chiefs (Gibson and Marks, 1995), and merchants in Medieval Europe created guilds (Greif, Milgrom and Weingast, 1994). Similarly, union leaders punish workers who cross the picket line (Atleson, 1969), and school committees sanction parents who fail to contribute to local public goods (Miguel and Gugerty, 2005). The first contribution of this paper is to document the effectiveness of internal centralized-sanctioning institutions in fostering cooperation[7].

In the past two decades, formal and experimental research have focused almost exclusively on peer-sanctioning institutions. Within this framework, scholars have demonstrated that the threat of sanctioning induces greater cooperation by changing individuals payoff functions (Boyd et al., 2003, Fehr and Gächter, 2002, Gintis et al., 2005)[8]. If we consider, however, centralized-sanctioning institutions (Guth et al., 2007), their effectiveness may not rely only on the threat of punishment, but also on the extent to which they are perceived as legitimate (Dickson, Gordon and Huber, 2009, Eckel, Fatas and Wilson, 2010). Legitimacy induces greater cooperation by eliciting a stronger (moral) obligation to centralized authorities (Greif, 2006, Levi, 1989). Legitimacy, understood here as the popular acceptance of an authority's right to govern (Rawls, 1971), can be derived from different sources of authority, such as charismatic, traditional, or rational-legal authority (Weber, 1922). Focusing on the latter, the second contribution of this paper is to demonstrate that the *political process* by which an authority originally acquires its sanctioning powers is consequential for cooperation.

Studying the causal effect of punishment institutions on behavior, in a natural setting, is constrained by several methodological challanges. First is the problem of reverse causality: punishment affects cooperative behavior, yet at the same time, levels of cooperation affect the use of punishment. Secondly, there is endogeneity between the type of sanctioning regime (i.e., peer *vs.* centralized sanctioning) and the nature of the group under consideration. Similar problems con-

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

[7]See also Frye (2000) that examines the conditions under which private groups govern themselves without turning to an outside agent for enforcement.

[8]The main focus of this strand of literature is in identifying conditions for overcoming the second-order collective action problem inherent in costly decentralized punishment.

strain the study of the impact of different selection mechanisms for leaders (Grossman, 2011*c*). In most natural settings, the factors that determine the institutions or processes for selecting leaders endowed with sanctioning powers are endogenous (Acemoglu, Johnson and Robinson, 2001). These identification challenges have impeded the study of the causal effects of institutional change on cooperative behavior (Dal Bo, Foster and Putterman, 2010).

To investigate these aspects while overcoming the identification problems, I adopt a methodological framework that combines "lab-in-the-field" behavioral experiments with observational data on 1,543 producers from 50 Ugandan farmer associations. I developed a novel adaptation of the public goods game (PGG), which is the conventional behavioral experiment used to study the conditions under which groups can overcome individual incentives to defect (Camerer, 2003). The experimental setup allows me to attest the positive impact of centralized-sanctioning institutions on cooperative behavior as well as to demonstrate that the size of this effect depends on the process by which these institutions are established. To assess the *external validity* of my findings, I then relate subjects' behavior in the PGG to their behavior in the natural setting, in which they face a similar social dilemma[9].

The paper unfolds as follows. After discussing the theoretical framework and stating my hypotheses, I describe the research design and provide some background information on the Ugandan associations that are the subject of my study. Following a description of the experimental manipulation, I present a first set of results. To study the impact of a centralized-sanctioning authority, I compare the cooperative behavior of subjects that do not face a threat of punishment with the behavior of subjects in a context in which sanctioning powers are given to a single monitor. I find that the introduction of a centralized-sanctioning authority has a strong positive impact on cooperation. To study whether the political process through which leaders acquire their powers is consequential, I compare the cooperative behavior of subjects who elected their monitor with the behavior of subjects who were assigned a monitor selected at random. I find that the way in which centralized authorities obtain their sanctioning powers is consequential for cooperation

[9] A similar strategy has been taken in the context of gift exchange experiments (List, 2006) and donation experiments (Benz and Meier, 2008). Whether individuals pro-social behavior in experiments correlates with their behavior in natural settings is largely an open question.

above and beyond the threat of punishment: participants are more responsive to the authority of an elected rather than randomly chosen monitor.

Following a brief summary of the observational data and empirical measures of legitimacy, I present a second set of results showing that the relationship found in the controlled experiment also exists in the participants' natural setting. First, I find a strong association between the perceived legitimacy of the managers of the farmer associations and their members' level of cooperation. Second, I show a positive relationship between farmers' behavior in the PGGs and their level of cooperation in the farmer associations. I conclude that the experimental setting captures institutional conditions and group dynamics that are relevant for determining levels of cooperation in the subjects' natural setting.

2.2 Theoretical Framework and Hypotheses

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

In PGGs, participants initially contribute, on average, between 40 and 60% of their endowment. However, in repeated games, 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 (Fischbacher, Gächter and Fehr, 2001, Ostrom, 2000). By contrast, when participants are allowed to punish other subjects, conditional cooperators can discipline defectors, thus lead-

ing to greater overall levels of contribution (Fehr and Gächter, 2002). Peer punishment provides a possibility of targeted interaction, thus fostering cooperation through mechanisms of direct and indirect reciprocity (Lubell and Scholz, 2001, Rand et al., 2009).

Peer sanctioning, however, is only effective under very restrictive conditions (Sigmund, 2007, Taylor, 1982). It can only sustain cooperation in small-size groups, where the cost of punishment is likely to be recuperated (Olson, 1965). In such groups, self-interested contributors may choose to punish defectors at a personal cost, as long as they have reasons to believe that punishment will increase the future contributions of the "targets." This, in turn, depends on the frequency of interaction between members (Boyd, Gintis and Bowles, 2010). As the number of members increases and interactions become infrequent, bilateral punishment becomes unlikely to sustain cooperation because future gains from punishment cannot be internalized (Greif, 1993).

The limited scope of peer sanctioning induces groups, organizations and communities to delegate sanctioning powers to internal centralized authorities. These institutions are likely to be more efficient than peer punishment (Guth et al., 2007, O'Gorman, Henrich and Van Vugt, 2009), since they are better positioned to overcome coordination failures and free-riding problems (Lake, 2009), although they might experience flaws in information, thus leading to enforcement errors (Dickson, Gordon and Huber, 2009). The centralization of sanctioning is the likely outcome of an *endogenous* process of social differentiation: virtually all social groups, even those characterized by low levels of complexity, experience elementary forms of division of labor that lead to the emergence of hierarchical structures and leadership roles. Sanctioning, of course, is only one way in which leaders impact cooperative behavior. Other means include persuasion (Henrich and Gil-White, 2001), coordination (Wilson and Rhodes, 1997), provision of information (Dewan and Myatt, 2008), and leading by example (Levati, Sutter and van der Heijden, 2007). To incorporate these intuitions into theories of public goods provision, I study how group members behave when a single individual is given a monopoly over sanctioning. Specifically I test the following:

Hypothesis 1 *Internal centralized authorities that are given a monopoly over sanctioning decisions will punish defectors at a personal cost.*

Hypothesis 2 *Centralized-sanctioning induces greater contribution toward public goods production, even if participants do not have any control over sanctioning decisions.*

My second research question is whether and how the political process through which internal centralized authorities obtain their sanctioning powers is consequential for cooperation[10]. Differently from peer-sanctioning systems, in which the right to punish defectors comes hand-in-hand with group membership, in a centralized-sanctioning regime it is important to distinguish between the effect of sanctioning and the effect of the way in which sanctioning powers are granted. Internal centralized authorities might derive their legitimacy from different sources: tradition (e.g., village chiefs and the pope), charisma, (e.g. Gandhi and Martin Luther King, Jr.), or rational-legal procedures (Weber, 1922). In this paper I focus on the latter source of legitimacy and am interested in testing whether elections have a positive impact on subjects' contributions to public goods production. Importantly, my focus on the method for selecting leaders addresses a major gap in the current literature on the impact of leaders on cooperation: the tendency to treat the emergence of leadership institutions as wholly exogenous (Ahlquist and Levi, 2011)[11].

There are several plausible complementary mechanisms through which the participation of group members in the selection of a sanctioning authority would induce greater cooperation. In this paper I focus on testing whether the *procedure* of elections can increase subjects' cooperation toward public goods provision via its impact on the legitimacy of the centralized authority (Tyler, 2005). Defining legitimacy as the extent to which people feel *morally obliged* to follow an authority (Greif, 2006) and accept its right to govern (Rawls, 1971), I test whether subjects are more likely to defer to a leader and accept her punishment decisions, if allowed to participate in her selection.

There are several reasons why legitimacy should mediate the relation between elections and subjects' cooperative behavior. First, elections may have a ritualistic or symbolic value that confers

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

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

on leaders greater authority. According to Beetham (1991, p. 69), the exercise of coercive power is justified only if it is based upon a "common framework of belief between the dominant and the subordinate in any power relationship." In many groups, communities, and societies, *explicit consent* via elections provides such a "common framework"[12]. Second, elections can contribute to the legitimacy of centralized authorities through a "peer effect" (Zelditch, 2001). In short, people are influenced by what others think about authorities. Election results signal to subjects the worthiness of the chosen authority, affecting the way subjects encode information from leaders, and the importance they give to their messages. Finally, a core argument of a large political science and social psychology literature is that the fairness of the procedure through which authorities gain power and/or exercise power shapes the willingness of subjects to defer to their authority[13]. This claim is widespread in legal (Gibson, 1989), political (Levi and Sacks, 2009), and managerial settings (Hoffmann, 2005).

There are two other notable mechanisms that might explain the relationship between elections and cooperative behavior: a leadership selection effect and accountability. First, direct elections may enable group members to select "better" leaders, namely, leaders whose *socio-demographic profile, status, or characteristics* make them more effective in triggering cooperation (Eckel, Fatas and Wilson, 2010, Grossman, 2011*a*). Second, periodic elections may induce centralized authorities to enforce cooperation more stringently to increase the likelihood of reelection (Gordon and Huber, 2002), thereby increasing cooperative behavior. Since these mechanisms may confound the effect of legitimacy, I designed the experiments such that leader selection effects are minimized and controlled for in the analysis, while accountability effects are eliminated[14].

To sum, by legitimizing their use of coercive powers, elections are expected to *increase the effectiveness of leaders*, thereby affecting public goods provision. Building on this framework, the experiment was designed to test the following hypotheses:

[12] The idea that consent is needed to legitimize authority is usually accredited to Locke (1990).

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

[14] Accountability effects are eliminated by having the monitor selected 'once and for all' subsequent rounds. Leaders' selection effects were minimized by randomly sampling the experimental subjects from six different villages and from a pool of members that did not hold any leadership roles in their associations. I return to these points in the research design and analysis sections.

Hypothesis 3 ***Outcome:*** *the process by which a centralized authority acquires its sanctioning power is consequential for cooperation: contribution levels to public goods provision are higher when authorities are elected rather than selected at random.*

Hypothesis 4 ***Mechanism:*** *elections increase cooperation through a "legitimacy effect": individuals are more likely to commit to a leader's authority if they participate in his/her selection.*

My third research question concerns the external validity of my results. Extant empirical evidence has documented the positive relationship between the perceived legitimacy of an authority and citizens' compliance (Hechter, 2009). Although there have been some valuable attempts to determine the causal relationship between leadership legitimacy and collective outcomes using panel data and cross-country comparisons (Levi and Sacks, 2009, Murphy, 2005, Tyler, 1997), it is often quite difficult to draw definitive and general conclusions relying exclusively on observational data. In contrast, the experimental condition allows me to assess the existence of a causal relationship between centralized authority and cooperation in PGGs. To attest the external validity of my findings, I relate group members' performance in the PGGs to their behavior in a natural setting in which they face similar collective action dilemmas. I, therefore, test the following hypotheses:

Hypothesis 5 *The relationship between leaders' legitimacy and levels of cooperation observed in the experimental setting also exists in the natural setting.*

Hypothesis 6 *Group members' cooperation in the experimental settings would resemble their level of cooperation in the natural setting, especially when the experimental conditions reproduce key features of the natural setting.*

2.3 Research Site, Sampling, and Experimental Design

To study these hypotheses I took a behavioral experiment with well-established properties, typically conducted in a laboratory environment, to rural Uganda. Behavioral experiments need not be preoccupied with attempts to mirror naturalistic conditions – a task that such experiments are ill-suited for (Berkowitz and Donnerstein, 1982, p.247–8). Instead, to increase the confidence that the experiment captures decision-making in similar real-world situations, I conducted the experiment with, and collected observational data on, members of pre-existing groups that face an analogous collective action problem on a regular basis. I conducted our experiment with members of farmer associations, since they repeatedly face in their natural setting – as explained above – a similar social dilemma to the one studied by PGGs. Comparing the cooperative behavior of the Ugandan farmers in their real-world setting to their behavior in the controlled 'lab-in the field' environment, is among the most important methodological contributions of this study[15].

From an external validity perspective, there are several other benefits in studying farmers from Uganda, which is among the world's least developed countries. PGGs are assumed to capture individuals' behavior in real-world social dilemmas. Yet, the universal applicability of these experiments has been limited by the fact that almost all past studies have relied on students from western universities[16]. Here my research design offers an advantage over existing studies. First, recent evidence suggests that students' behavior in PGGs might not be representative of the larger society (Jones, 2010). Secondly, there is growing evidence suggesting that in the context of behavioral experiments that measure social-preferences, subjects from developing countries behave differently than their counterparts in the developed world (Cardenas and Carpenter, 2008).

That the study's subjects are members of pre-existing groups also contributes to its external validity. It has been argued that laboratories are a poor setting for testing how subjects balance the tension inherent in social dilemmas (Burnham and Johnson, 2005). On one hand, on the basis

[15] A similar strategy has been taken in the context of gift exchange experiments (List, 2006) and donation experiments (Benz and Meier, 2008). Previous work suggests that whether individuals pro-social behavior in experiments correlates with their pro-social behavior in the field is still largely open.

[16] The Dictator and Ultimatum games have been widely used in field settings (Henrich et al., 2004). By contrast, PGGs in field settings are rare; *cf.* Habyarimana et al. (2007) and Fearon, Humphreys and Weinstein (2009).

of PGGs that took place in laboratories and that use ad-hoc groups, researchers have argued that *group selection* affects cooperation through its impact on pro-social preferences (Boyd et al., 2003). On the other hand, laboratories strip individuals from context, and are limited in their ability to replicate the mutual trust, past experience, shared norms and group identity which are central for balancing tension between private and public interests (de Rooij, Green and Gerber, 2009, Henrich et al., 2004, Levitt and List, 2007). This consideration has led researchers to argue that *pre-existing groups* are the most natural setting to test theories about the emergence of cooperative behavior (Herrmann, Thóni and Gächter, 2008, Nowak, 2006, Rand et al., 2009).

The Research Site: APEP Farmer Groups

The farmer associations I surveyed were created as part of one of Uganda's largest recent rural development interventions: the Agriculture Productivity Enhancement Project (APEP)[17]. APEP's goal was to support subsistence farmers' transition into commercial farming. Between 2004 and 2008 it helped organize over 60,000 small-holder farmers into about 2,500 village-level groups (known as producer organizations, or POs), which were further organized into more than 200 farmer associations. Serving, on average, 200 members from ten neighboring village-level groups, the farmer associations (known as Depot Committees, or DCs) were designed to exploit economies of scale and to bargain for better prices based on quality and volume.

Studying the APEP groups presents many advantages. First, the project's scope and size allow me to conduct a large-scale quantitative study within the boundaries of a single nation, thus securing the homogeneity of the political and legal environments, as well as many project-related factors. Moreover, the process of group formation occurred under the lead of a few project field-trainers. As a consequence, APEP groups have similar governance structures and leadership positions whose roles and functions are comparable across sites. Each farmer association has an executive committee – responsible for making key decisions at the association level – comprised of a manager, chairperson, secretary, and treasurer. Operationally, the manager is the leader of the

[17] APEP was funded by USAID, and implemented by Chemonics, a Washington D.C. Consultancy.

association. His most important responsibilities include organizing crop collection, searching for buyers, and negotiating output and input prices. Additionally, managers are involved in coordinating training activities, facilitating the diffusion of information, and sanctioning members who do not follow the association's rules and bylaws.

The Sample

I used a stratified, random, multistage cluster design to select my sample[18]. The use of random samples is not common in behavioral experiments. This is because the main goal of behavioral experiments is testing the accuracy of general causal statements (Berkowitz and Donnerstein, 1982), and not to determine the probability that a certain event will occur in a particular population. Drawing a representative sample from the population of APEP members, nonetheless, had two benefits. First, it helped subjects to form consistent beliefs about the behaviors of the individuals with whom they were playing (Habyarimana et al., 2007, p.712). Second, it allowed me to make inferences from the behavior of my sample to the groups from which our subjects were drawn. Representative samples are especially valued when researchers wish to use experiments as a measurement device that allows them make inferences about groups, and not only individual behavior (Levitt and List, 2007).

From each of the 50 sampled association, different types of data were collected. I interviewed the DC executives and the members of the DC board to gather information at the cooperative level. Data on the DCs' economic activities were also assembled from the associations' books and records. In each association, I sampled six producer organizations (or POs), for a total of 287[19]. An interview with the leaders of the sampled POs allowed to collect additional data at that level. I also collected individual-level data. From each sampled PO, I further sampled, in average, six members for a total of 36 members per association. Sampled members were surveyed in person by trained interviewers in the respondents' language, for a total of 1,781 surveys[20].

[18] An elaborate description of the sampling scheme can be found in the Supplementary Appendix A

[19] When a farmer association had fewer than seven POs, I selected all its village-level groups.

[20] Though I surveyed 1,781 out of 1,800 sampled members, only 1,543 of those participated in the experiments. This gap is due to the fact that the experiment were conducted in a central location. To reduce attrition, interviewers returned

Though the empirical analysis uses data from the individual and group-level surveys, the construction of those instruments and the meanings I derived from them, relied heavily on more than a year of field work, during which I conducted dozens of interviews and meetings with group members and leaders, APEP-staff and Chemonics staff in Uganda and Washington D.C.

Experimental Design

To test hypotheses 1 through 4, I designed an adaptation of the PGG. In each round of the experiment, subjects received an endowment of 10 coins of 100 USH – 10 monetary units (MUs) – which is the equivalent of about half a daily wage in rural Uganda. Subjects then had to decide, anonymously, how to split this endowment between a private and a public account. What subjects put in the private account remained theirs, while what was contributed to the public account was doubled and redistributed evenly among all group members.

I played three variants of the PGG: baseline, random monitor, and elected monitor. Subjects assigned to the baseline participated in six rounds of a PGG *without* sanctioning. In the two monitoring treatments, I introduced a centralized-sanctioning institution. Differently from peer-sanctioning settings, in which subjects may punish each other, I gave sanctioning powers to a single authority. Namely, after two preliminary rounds of play, one of the subjects was assigned the role of a monitor. Monitors' – whose identity was public knowledge – received the same endowment as the other subjects, but could not contribute to the PGG, nor receive part of the public account. Instead, monitors were able to spend 1 MU to take away 3 MUs from subjects whose contribution level they disapproved[21]. Monitors' payoff did not depend on the group's level of cooperation, but only on their sanctioning decisions. A monitor's payoff in round t is, therefore, 10 (MUs) minus the number of subjects sanctioned in that round. Subjects' payoff is calculated as $\pi_{it} = (10 - x_{it}) + \frac{2\sum x_{it}}{n} - P_{it} \times 3$, where $x_{it} \in \{0, 1, \ldots, 10\}$ is the contribution to the public account, and $P \in \{0, 1\}$ indicates whether player i was punished at round t.

to sampled villages several times to locate sampled members who did not show up on that day, for whatever reason.

[21] This study follows the convention in PGGs, according to which the threat of punishment is credible and substantial: sanctioned subjects have no option but to comply with their punishment.

The two sanctioning treatments differed only in the procedure for selecting the monitor. In the random-monitor treatment (T1), the monitor was selected through a *lottery*. Comparing the baseline with the random-monitor treatment allows for an assessment of the causal effect of instituting a centralized-sanctioning regime. In the elected-monitor treatment (T2), participants *elected* their monitor using a secret ballot[22]. A comparison of the random and elected-monitor treatments allows an estimate of the independent effect of the process by which the monitor has obtained her sanctioning powers. I played all three variants in each of the 50 sampled associations. Participants were randomly assigned to only one variant[23]. Table 2.1 summarizes the experimental design.

Game Variant	Centralized Sanctioning	Elections	Sessions	Avg subjects per Session	Rounds per Session
Baseline			50	10	6
Random Monitor	X		48	10	2 prelim + 4
Elected Monitor	X	X	49	10	2 prelim + 4

Table 2.1: Experimental Design

[22]The voting procedure guaranteed anonymity: each player wrote on a piece of paper the ID 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 subjects. For more details on the selection process, see the game scripts in the Appendix.

[23]The number of subjects per session ranged from 8 to 12. This is higher than in most PGGs (2 to 4), and was required in order to protect our subjects' anonymity.

- **Stage 1: Contribution to the PG.** Players decide *anonymously* how to divide 10 MUs between a private and a public account. To ensure anonymity, players made their allocation decisions behind 3-sided cardboard screens.
- **Stage 2: Contributions become common knowledge.** RAs display publicly *all* the contributions to the public account (in USH), from the lowest contribution to the highest. Players are unable to match between contributions and players' identity.
- **Stage 3: Payoffs.** RAs calculate *publicly* the mean contribution to the public account and the size of the social return. Using this information, RAs display the actual payoff (private + social returns) for each of the contributions displayed on the board.
- In the control group, play repeats for six rounds without threat of punishment. In T1 and T2, monitors are selected after preliminary round 2: Round 3 is, thus, the first round in which players are subjected to sanctioning. Stages 4 and 5 are only played in T1 and T2, the two sanctioning treatments.
- **Stage 4: Punishment decisions.** Monitors, whose identity is known to all players, decide whether to assign 'reduction points'. Because it is not possible to infer the identity of players from contributions, monitors sanction contribution levels, *not individual players*. Because monitors made their sanctioning decisions publicly, players knew what levels of contribution were punished, but could not match between punishments to the identity of the sanctioned players.
- **Stage 5: Payoffs recalculated.** Following the monitors' sanctioning decision, RAs reduce the payoffs of sanctioned contributions by 3 MUs (300 USH). Play repeats for four rounds under a sanctioning regime.

2.4 Experimental Findings

Figure 2.1 summarizes the main results of the lab-in-the-field experiment (see also Appendix, Table 2.5). First, corroborating hypothesis 2, in the presence of a centralized-sanctioning authority subjects significantly increased their contribution to the public good[24]. In the random-monitor condition (T1), subjects contributed 15.0% more, on average, than in the baseline ($P = 0.00$)[25]. In the elected-monitor condition (T2), subjects contributed 25.4% more than in the baseline ($P = 0.00$)[26]. Moreover participants acted under the expectation that monitors would punish defectors: significant differences in contributions between baseline, on the one hand, and both T1 and T2, on the other hand, are already visible in round 3, thus *before observing monitors' behavior*. In round 3, subjects in T1 contributed to the public account 16.6% ($P = 0.00$) and in T2 24.4% more ($P = 0.00$) than subjects in the baseline treatment.

Second, confirming hypothesis 3, the (political) process through which monitors obtain their sanctioning power is consequential. Subjects in T2 contributed to the public account, on average, 9% more than subjects in T1 ($P = 0.005$). The experiment, thus, provides evidence that elections have a positive impact on subjects' level of cooperation (see also Appendix, Table 2.6). I now turn to explore the mechanisms that might be responsible for such a positive effect.

According to hypothesis 4, we expect greater contributions in the elected monitor to be due to a "legitimacy effect": subjects should express greater deference to monitors whose legitimacy has been certified through elections. I have defined legitimacy as the extent to which subjects feel obliged to follow an authority and accept its right to govern and measure it as follows. First, I consider subjects' expectations, by looking at changes in contributions from the second preliminary round to round 3, before monitors' decisions over sanctioning take place. While in the preliminary rounds average contributions to the public account in T1 and T2 are similar, in round 3 subjects

[24]This finding is equivalent to other lab-in-the-field PGGs, where peer sanctioning increased cooperation relative to the baseline (see Barr (2001) and Carpenter (2004)).

[25]Two-sided Mann–Whitney test. The dependent variable is the mean contribution in rounds 3 through 6.

[26]The modest decline in baseline is consistent with findings from PGG studies in non-western settings. Whereas cooperation declines significantly with college-aged participants in the United States, cooperation rates remain higher and are sustained longer with African and Asian subjects (Cardenas and Carpenter, 2008, p. 313).

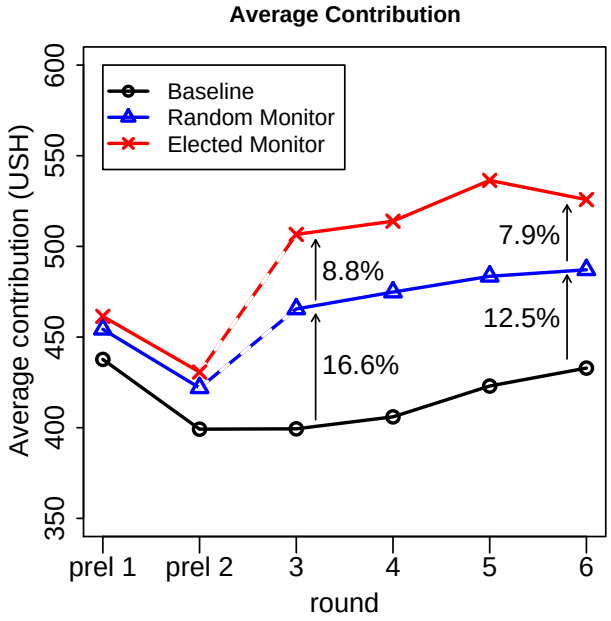

Figure 2.1: **Average contribution to the public good by treatment.** For rounds 3 and 6, the graph reports the percentage increase in contributions comparing random with baseline condition, and elected with random (e.g., in round 3, subjects in the random monitor contributed 16.6% more than in the baseline condition).

in T2 contributed significantly more (+8.8%) than subjects in T1 (P = 0.022). Regression models that control for monitors' characteristics confirm the significance of these results (Appendix, Table 2.7)[27]. Figure 2.2.A shows the marginal estimated effect of the type of monitor on a representative subject: the introduction of a centralized-sanctioning institution in round 3 induced a net increase in contributions of 0.37 MUs in T1, and 0.76 MUs in T2.

Second, I consider players reactions to punishment as further evidence of the greater deference to elected monitors. Parameter estimates come from a multilevel regression, in which I model the change in player i's contribution from $t-1$ to t as a function of whether player i has been sanctioned at $t-1$ as well as the type of monitor, controlling for the monitor's socio-demographic characteristics, the monitor's sanctioning behavior at time $t-1$, and player i's preliminary contributions. On average, having been punished at time $t-1$ increases subjects' contribution at time t by 0.57 MUs in T1, and by almost 1 MU in T2 (see Figure 2.2.B and Appendix, Table 2.8).

[27] Results from other model specifications produce similar results, and can be provided upon request.

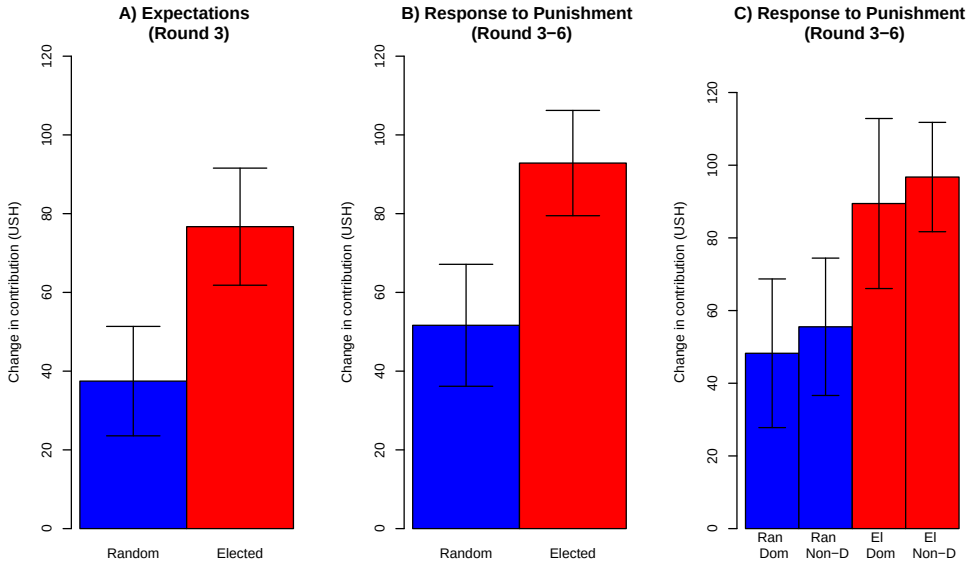

Figure 2.2: **Predicted change in contribution in the two sanctioning treatments.** Plot of the estimated variation in contributions from $t-1$ to t (A) for all subjects in round 3; (B) for sanctioned player in rounds 3 to 6; and (C) for sanctioned subjects in rounds 3 to 6 distinguishing between monitors with a dominant (male, wealthiest quintile, highest level of education, born in the village) or non-dominant profile. Continuous variables are held at their mean values, while categorical variables at their median values.

Having found evidence for the "legitimacy effect" hypothesis, I now turn to rule out the possibility of a confounding impact due to leadership selection. Namely, it is possible that higher levels of cooperation are due to the characteristics of the elected monitors. I find, on one hand, that subjects elected monitors with socially dominant profiles – elected monitors were more likely to be male, wealthier, more educated, and more likely to have been born in the village, compared to the pool of eligible monitors (Appendix, Fig. 2.5). On the other hand, knowing a monitor's profile – namely his gender, education, age, wealth, place of birth, and church attendance – does not improve our capacity to predict subjects' contributions: when regressing subjects' contribution on monitors' profile in T1, the characteristics of monitors, whether tested separately or jointly, do not have a significant effect on contributions (Appendix, Table 5).

In addition, the socio-demographic profile of monitors does not affect subjects' change in contributions from preliminary round 2 to round 3 (Appendix, Table 2.7), nor their reaction to sanctioning (Appendix, Table 2.8). Figure 2.2.C demonstrates that monitors with a socially dominant

profile have the same impact on sanctioned subjects' subsequent contributions as monitors with a non-dominant profile.

These findings, cumulatively, weaken the possibility of a leadership selection effect[28]. It is possible, however, that elected monitors have certain attributes that induce cooperation, which are unobserved to the research team but visible to the experimental subjects. Though the possibility of unobserved heterogeneity cannot be ruled out, it does not seem to play a decisive role in this experiment. For one, elected monitors are not more *public-spirited* than random monitors, at least as this is reflected in their contributions to the public good in the preliminary rounds[29]. Second, as I demonstrate below, I do not find much evidence suggesting that elected monitors were enforcing cooperation more ardently than random monitors. Third, elected monitors' *religiosity* is not higher than random monitors, at least as this is reflected in church attendance[30]. Finally, recall that the experimental subjects were drawn from six neighboring villages, none held leadership positions in the farmer association, and that they were not allowed to talk through the entire course of the experiment. These design features further reduce the possibility that participants had private knowledge of how well other subjects would perform as monitors.

In sum, the experiment's findings support the hypothesis that subjects' greater deference to elected monitors is due to a "legitimacy effect": elections impact cooperation by conferring greater authority to leaders. In the next section I analyze monitors' sanctioning behavior and test whether the different criteria for monitor selection may have affected monitor's sanctioning strategies, and thus, indirectly, cooperation.

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

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

[30]In Uganda, religiosity is considered an important attribute for those seeking public office.

2.5 Monitors' Sanctioning Behavior

Table 2.2 compares the sanctioning behavior of the two monitors by round of play. Though elected and random monitors sanctioned, on average, the *same number of players* per round (from an average of about 2.5 players sanctioned in round 3, to about 1.5 players in round 6), they varied in the *amount of contribution sanctioned*. In round 3, the first round under a sanctioning regime, random and elected monitors punished similar levels of contribution. In subsequent rounds, the average maximum contribution sanctioned by random monitors gradually declined, while it increased for elected monitors. By round 6, random monitors did not punish, on average, subjects who contributed more than 24.6% of their endowment. By contrast, elected monitors kept punishing, on average, players that contributed up to 37.1% of their endowment ($P = 0.022$). Similarly, whereas random monitors, on average, allowed players who contributed more that 31.4% of their endowment to go unpunished, elected monitors did not allow players who contributed less than 38.5% of their endowment to go unpunished (see bottom panel, Table 2.2).

Frequency of Punishment (No. players punished in round t)				
	Round 3	Round 4	Round 5	Round 6
Random Monitor	2.47	1.86	1.78	1.49
	(2.05)	(1.55)	(1.39)	(1.68)
Elected Monitor	2.40	2.08	1.71	1.58
	(1.95)	(1.60)	(1.13)	(1.35)
Maximum Punishment (in coins / Monetary Units)				
	Round 3	Round 4	Round 5	Round 6
Random Monitor	3.09	2.67	2.76	2.46
	(2.76)	(2.75)	(3.00)	(2.93)
Elected Monitor	3.20	3.60	3.29	3.71
	(2.63)	(2.88)	(2.92)	(3.05)
Minimum Contribution Not Punished (in coins / Monetary Units)				
	Round 3	Round 4	Round 5	Round 6
Random Monitor	2.96	2.98	3.14	3.10
	(1.80)	(1.73)	(1.94)	(1.96)
Elected Monitor	3.96	3.79	3.85	3.50
	(1.91)	(1.93)	(2.27)	(2.59)

Table 2.2: Frequency, Max Punishment and Min Contribution Not Punished in the Random (T1) and Elected (T2) monitor conditions. N=97 (49 Random and 48 Elected monitors). Standard deviation in parentheses.

These findings seem to indicate that elected monitors used punishment to enforce cooperation more ardently. However, because both types of monitors faced, in all rounds, different distributions of contributions (reported in Table 2.5 in the Appendix) the data is also consistent with an

alternative explanation: both monitors may have followed a similar strategy of sanctioning a few contributors at the bottom of the distribution[31].

To overcome this problem, I used the Kullback–Leibler (K–L) divergence measure to match each of the *distributions of contributions* to the public pot that an elected monitor faced with the closest distribution of contributions that a random monitor faced[32]. Using the matched pairs, I was then able to test the extent to which elected and random monitors, facing similar distributions of contributions, acted in a similar manner[33]. Previewing the results, I find no evidence that elected monitors punished more frequently, and only weak evidence suggesting that elected monitors used a higher threshold to signal an accepted level of cooperation.

In Table 2.3 and in Figure 2.8 in the Appendix, I demonstrate the improvement in the divergence between distributions of contributions achieved by matching. On average, matching reduced the (K–L) distance between the distributions that the different monitors faced, by a factor of about 4.

	Matched Pairs					All T2–T1 Pairs	
	Mean	sd	Min	Max	N Pairs	Mean	N pairs
Round 3	0.091	0.050	0.027	0.228	48	0.328	2,352
Round 4	0.079	0.032	0.030	0.182	48	0.306	2,352
Round 5	0.088	0.041	0.025	0.178	48	0.336	2,352
Round 6	0.074	0.044	0.010	0.183	48	0.310	2,352

Table 2.3: K–L Divergence Measure for both Matched Pairs and for all possible T2–T1 pairs, by round.

I use three measures to test whether elected monitors enforce norms of cooperation more ardently than random monitors *facing similar contributions*: (i) the number of players being punished in round t, (ii) maximum contribution being punished in round t, and (iii) minimum contribution not punished in round t. Results are presented graphically in the Appendix. Figure 2.10 graphs the number of players punished by elected monitors, as a function of the number of players pun-

[31] The difference in distributions that the two types of monitors faced is a product of the experimental design: contributions levels were selected by the players endogenously and were not under the control of the research team.

[32] A detailed description of the Kullback–Leibler (K–L) divergence measure is provided in the Appendix 2.C.

[33] Within blocks defined by rounds, I used a nearest neighbor with replacement matching algorithm.

ished by random monitors, for all matched pairs. Jittered dots that are above (below) the 45% line are matched pairs in which the elected monitor has punished more (less) players than the matched random monitor. Similarly, Figure 2.11 and Figure 2.12 in the Appendix, graph the maximum contribution punished and the minimum contribution not punished in round t by elected monitors, as a function of the behavior of random monitors, for all matched pairs.

Differences in mean behavior of the matched pairs, including p-values derived from a Wilcoxon matched-pairs signed-ranks test, are provided in Table 2.4. Starting with the frequency of punishment, elected monitor punished slightly more players in rounds 4-6 than random monitors facing a similar distribution of contributions, however, those differences are not significant. Turning to the maximum contribution sanctioned, using matched pairs (Table 2.4) instead of the entire sample (Table 2.2), the difference between the two types of monitors is now smaller and non-signifcant. Finally, there is some evidence that the minimum contribution not punished under elected monitors is higher than under random monitors, at least in rounds 3-5[34].

	N Punish			Max Punish			Min Not Punish		
	T2	T1	P-value	T2	T1	P-value	T2	T1	P-value
Round 3	2.396 (1.954)	2.417 (2.030)	0.955	3.205 (2.629)	2.891 (2.759)	0.669	3.958 (1.913)	2.896 (2.106)	0.010
Round 4	2.083 (1.596)	1.688 (1.355)	0.129	3.600 (2.879)	4.250 (3.992)	0.670	3.792 (1.935)	3.062 (1.827)	0.041
Round 5	1.708 (1.129)	1.583 (1.200)	0.124	3.289 (2.920)	3.255 (3.590)	0.460	3.854 (2.269)	3.312 (1.870)	0.065
Round 6	1.583 (1.350)	1.396 (1.807)	0.097	3.711 (3.048)	3.378 (3.752)	0.355	3.500 (2.593)	3.438 (2.072)	0.807
N pairs			48			48			48

Table 2.4: Table compares the behavior of the matched pairs of elected and random monitors along three dimesnions: Number of players punished, maximum contribution punished and minimum contribution *not* punsihed, in round t. Columns 5-6 and 8-9 report results in Monetary Units (coins). Columns 4, 7, 10 report p-values from a Wilcoxon matched-pairs signed-ranks test. Standard deviations in parenthesis.

[34]Note that averaging the maximum contribution punished is not optimal for comparing the behavior of the two monitors. First, it ignores the fact that punishing very high contributions (7 coins and above), may actually suppress contributions in subsequent rounds. Second, the measure fails to incorporate cases where no players was sanctioned (missing value in the dataset). For these reasons, minimum contribution not punished seems as a more reliable measure.

Matching the distributions that the two types of monitors faced, I find only weak evidence that elected monitors are enforcing norms of cooperation more forcefully. These findings further weaken Hypothesis 4 and the likelihood of unobserved heterogeneity in monitors' profiles. More so, the findings of this exercise – together with the fact that there is no discernible relation between the frequency of punishment in round t and the average group contribution in that round – suggests that both types of monitors were using a heuristic rule of generally punishing the lowest contributors. Put differently, monitors do not seem to punish according to some predefined acceptable level of contribution, but rather they consider players' contribution relative to the contribution of others in the group. Monitors' motivation aside, I find that the number of players sanctioned and the maximum amount sanctioned do not affect changes in players' contributions (see Table 2.8). The commutative evidence thus suggests that monitors' major impact on contributions is derived from the response to sanctioning that they elicit.

In recent years a large number of studies have demonstrated experimentally that subjecting individuals to a threat of sanctioning increases contribution to public goods production. The above experiment builds on that intuition, but expands the literature in two directions. First, consistent with Hypothesis 2, the experiment demonstrates that local centralized sanctioning authorities have a positive impact on cooperation. Consistent with Hypothesis 1, both elected and random monitors punished defectors, even though they did not derive monetary benefits from the resulting increase in contribution and even though punishment was costly. Interestingly enough, players in both T1 and T2 anticipated this, thereby increasing their contributions immediately after the selection process, even before observing the monitors' punishment behavior.

Secondly, the experiment further expands the literature by suggesting that beyond the mere threat of punishment, individual's cooperation is also conditional on the process by which leaders acquire their authority (Hypothesis 3). Players who participated in electing a centralized authority contributed, on average, more towards a public good than players who were not granted such a right. Analyzing possible mechanisms to explain these results, I do not find much evidence supporting a leader-selection effect (Hypothesis 4). By contrast, I find relatively strong evidence

that elections impacted cooperation via a legitimation process. Consistent with hypothesis 5, subjects were more likely to confer authority to, and justify punishment from, monitors whom they have directly elected. These findings call for a closer examination of the roles that leaders' selection rules and legitimacy play in local public goods provision. In the following section I turn to addressing issues related to the experiment's external validity.

2.6 Analysis of Observational Data

The experimental findings point to the relevance of the *procedure* through which the leadership is selected as the base for the legitimacy of the centralized authority[35]. In other words, it has demonstrated that legitimacy likely mediates the causal relation between a certain (selection) procedure and members' cooperation. In this section, I analyze observational data to attest the external validity of the experimental findings. I do so by relating group members' cooperative behavior in the PGGs to their cooperative behavior in the farmer associations, a natural setting in which they face a similar social dilemma.

In the attempt at relating experimental findings to the ordinary life of farmer associations, I focus the empirical analysis on a set of measures of procedural legitimacy. In general, measuring legitimacy is a difficult task. Following previous scholarship, I rely on a variety of survey questions that capture different aspects that might affect the respondent's perception of the legitimacy of the group leader. I consider a total of six proxy measures of legitimacy. Two are attitudinal and rather abstract: they are members' assessment of (a) whether the DC manager is monitored; and (b) whether the DC manager is accountable. Three other indicators are more specific; they report: (c) how often does a member receive receipts when selling through the association; (d) whether the member knows the name of the manager; and (e) whether the member knows the method for electing the DC manager. Finally, using principal component analysis, I also combined these responses into (f) an index of "perceived legitimacy."

[35]See Rothstein (2009) for a fruitful discussion on the distinction between 'procedural' and 'performance'-based legitimacy.

The key outcome variable is members' level of participation in public goods production. Collective marketing is the major activity of farmer associations, and, as explained before, it is the major collective action problem that farmer groups must overcome. I therefore measure cooperation by looking at the marketing decisions of group members. A high level of cooperation exists when members sell a large fraction of their crops via their farmer group[36].

Two *self-reported* measures of members marketing decisions are used to measure cooperation: (i) a dichotomous variable capturing whether a member sold his crops via the association, at least once, in the past season; and (ii) the proportion of a member's total seasonal coffee yield that was sold via the farmer group in the past season. Moreover, to further check the robustness of the self-reported measures of collective marketing, I gathered similar information from other sources. First, the leaders of each of the 287 sampled village-level POs were asked to provide information about their group members, including whether members sold their coffee via the group in the past season. Members' self-report and the PO leaders' report were consistent for over 70% of the sample. Finally, I used information provided by the DC executives to construct a variable capturing the proportion of members selling in bulk in the past season. Results are similar using either measure of cooperation. Here, I report results using self-reports and relying on the dichotomous measure of cooperation.

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

$$logit\{Pr(y_{ij}|X_{ij},\zeta_j)\} = \beta_0 + \beta_1 L_{ij} + X_{ij}\Gamma_1 + F_j\Gamma_2 + \zeta_j \tag{2.1}$$

with $\zeta_j|X_{ij} \sim N(0,\psi)$ and ζ_j independent across groups j. The dependent variable y_{ij} is an

[36]This measure is superior than alternative measures of public goods production such as price. For one, it directly measures members' cooperative behavior. Second, price may be affected by a host of factors outside the control of the association, such as altitude, rainfall, etc.

indicator that takes the value of one if individual i in group j sold coffee via her farmer group, at least once, in the past season. Our variable of interest is L_{ij}, which measures individual $i's$ perception of the manager's legitimacy. X_{ij} is a vector of individual-level controls, including variables such as gender, age, education, log seasonal yield, years since joining the farmer groups, the rating of the local middleman's honesty, richness of associational life, and frequency of attendance to religious services. F_j is a vector of group-level controls, including variables such as the age of the association, membership size, and manager's effort level. To relax the assumption of conditional independence among members of the same farmer group, we include a farmer association-specific random intercept ζ_j in the linear predictor[37].

In Figure 2.3, I report the predicted change on the probability that an 'average' member will cooperate, as a function of each legitimacy measure[38]. For example, controlling for individual- and group-level variables, knowing how the manager has been elected increases the likelihood of cooperation by 12.4% (5.7–19.2%, 95% CI). Similarly, receiving a receipt from the DC manager when selling through the group increases the likelihood of cooperating by more than 13% (4.6–21.4%), compared to those who do not get a receipt. In sum, across a wide range of legitimacy proxy measures, the change in the estimated probability of cooperation is positive and substantial.

Comparing Behavior 'Inside' and 'Outside' the Lab

In the last part of the analysis, I turn to compare the subjects' cooperative behavior in the controlled experiment directly with their behavior in their natural environment, as members of farmer groups. To my knowledge, this is the first study to make such comparison in the context of PGGs[39].

[37] I also tested model specifications that (1) nested farmer associations with strata (three-level model), and (2) that allowed the key independent variable (legitimacy) to vary across groups (random-intercept model). A likelihood ratio test rejected those specifications in favor of the two-level random intercept model, which is more parsimonious. Regression results can be found in Table 2.12 in the Appendix, while summary statistics of the variables used in this analysis are presented in Table 2.11 in the Appendix.

[38] Predicted estimates are calculated by setting continuous control variables to their mean values and categorical variables to median values. Standard errors are calculated using the delta method. For the legitimacy index, predicted change is calculated as one standard deviatio'n change from the grand mean.

[39] Laury and Taylor (2008) compare behavior in a PGG to behavior in "a naturally occurring public good." However, in that study, cooperation in the 'natural' setting was measured as the amount subjects were willing to donate from their experiment's earnings to a local NGO. Since this behavior is not part of the subjects' natural activity, their results should be treated with some care.

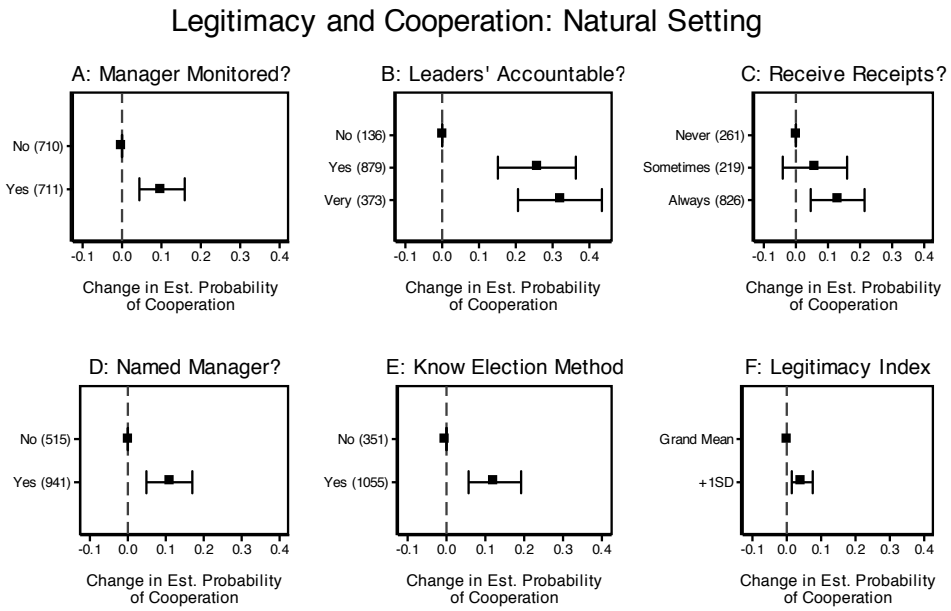

Figure 2.3: **Relation between legitimate authority and the likelihood of cooperating to public goods provision.** Estimates are calculated by setting continuous control variables at their mean levels and categorical variables at their medians. Standard errors are calculated using the delta method.

In particular, I expect that cooperative behavior in the elected-monitor treatment, which I interpret as capturing individuals' deference to legitimate authority, will predict behavior in the farmers' natural setting, where I have shown a positive relation between procedural legitimacy and cooperation. While this would not be a direct proof of the causal impact of a leader's legitimacy on cooperation in the natural settings, a positive correlation can be considered as an indirect validation of such a causal statement. In other words, what works under the "Petri dish" of a controlled experiment might be at work in the natural setting as well.

To test these hypotheses, I run the following Generalized Linear Model (GLM) for each of the three treatment conditions:[40]

[40]Since the contribution to the public account is a proportion of an initial endowment, i.e., bounded between 0 and 1, I implemented the GLM estimator proposed by Papke and Wooldridge (1996). The results I present are robust to other specifications: e.g., running the GLM with and without controls; running equivalent OLS regressions, treating mean contribution to the public account as a continuous variable; and using other specifications of cooperation in the natural setting.

$$\bar{Y}_{ij} = \beta_0 + \beta_1 C_{ij} + \Gamma X_{ij} + \epsilon_{ij} \tag{2.2}$$

where $\bar{Y}_{ij}$ is the mean contribution to the group account in rounds 3 to 6 of member i from group j, C_{ij} is an indicator of whether that subject cooperated with the farmer group by participating in collective marketing and X_{ij} is a vector of individual-level controls, such as sex, age, education, associational life, and contribution in the preliminary rounds. To account for dependence between members of the same group, I cluster standard errors at the farmer association level[41].

Regression results are presented in Figure 2.4 (see also Appendix, Table 2.13). For each treatment, I graph the expected values of the mean contribution to the public account ($\bar{Y}_{ij}$), for both cooperators (i.e., those participants who sold their crops via the group) and defectors (i.e., those who sold only to middlemen), including 95% confidence intervals. I find that in the elected-monitor condition cooperative behavior in the PGG is positively and significantly related to behavior in the farmer group[42].

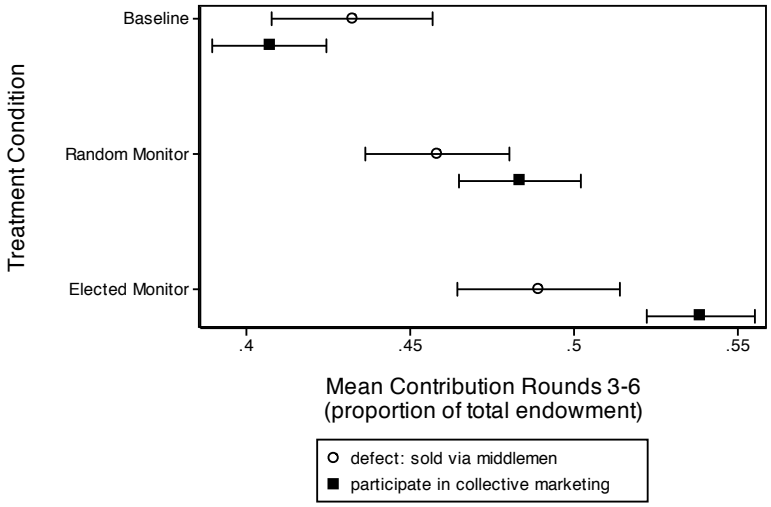

Figure 2.4: **Relation between subjects' cooperative behavior in the PGG and as natural setting.** The plot describes the expected value of a member's contribution to the public account, averaged across the sanctioning rounds (3 through 6), as a function of her participation in collective marketing. Caps represent 95% confidence intervals.

[41] The proportion of cooperators in the natural setting was equally balanced across the three treatments.

[42] Similar results are obtained with group-level data and with different measures of cooperation in the farmer group.

I interpret these findings to reflect the fact that the elected-monitor treatment approximates the process of legitimation of centralized authorities that occurs in farmer groups. Farmers who elected their monitor were able to draw on their past experiences and group norms. As a consequence, their game behavior reflects real-life patterns of behavior in a manner that is not visible in the other treatment conditions. These findings are consistent with a conventional premise in social psychology: individuals generalize their experiences from one situation to the other, making the same causal attributions in seemingly disparate settings (Dweck, Goetz and Strauss, 1980).

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

My findings suggest that the legitimacy of the centralized authority is an important factor in determining the success of farmer groups, and likely of other similar organizations. These findings underscore the merit of behavioral experiments, insofar as results suggest that the PGGs used in the current study were able to isolate causal mechanisms that are likely to be at work in both the experimental and real-world settings: in this case, a mechanism that increases cooperation in a social dilemma, in a context where subjects are habituated to elections. The correlation between the behavior in the lab-in-the-field setting (when exposed to the legitimacy stimulus/cue) and in the natural environment (where the stimulus is experienced and reported by the subjects) guarantees that there is some continuity between the two contexts. This, indirectly, provides support to the idea that the mechanism I tested in the lab might be at work in the real world.

2.7 Conclusion

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

Centralized-sanctioning systems, however, cannot rely exclusively on coercive punishment. Indeed, institutions that are perceived as legitimate only rarely turn to brute force to enforce group norms. To minimize the use of brute force, which is costly, centralized authorities need to be recognized as legitimate by the ruled (Ahlquist and Levi, 2011). I conceive of legitimacy as the capacity of a centralized power to exercise authority, and measure it as the extent to which subjects comply with its directives. The fact that sanctioned subjects who elected their monitor increased their contribution in subsequent rounds by more than double the sanctioned subjects who faced an arbitrary authority is indicative of the role legitimacy plays in sustaining cooperation.

Demonstrating how legitimacy mediates the relationship between political processes and the effectiveness of leaders is the second contribution of this paper. I have shown that beyond the mere threat of punishment, the political process through which leaders acquire their authority is consequential. When leaders acquire their regulatory and sanctioning powers through elections, they are able to exercise their sanctioning and monitoring roles more effectively. These findings are consistent with recent evidence of a positive impact that democratic elections at the village level in rural China have on public goods provision (Martinez-Bravo et al., 2010). As such, my findings have important implications to our understanding of the development of local forms of democratic rule. This study, therefore, calls for a closer examination of the role that leader selection rules and procedural legitimacy play in public goods production (see also (Grossman, 2011*c*).

The paper also contributes to the study of legitimacy, which is a central concept in normative theory and applied research (Hechter, 2009). Until the late 1980s only little attention was paid to empirical investigation of legitimacy (McEwen and Maiman, 1986). In recent years, however, scholars have increasingly sought to document the political and social effects of legitimacy. For example, Lieberman (2007) uses legitimacy to explain compliance with health regulations during an epidemic, Levi (1997) used it to explain people's support of war efforts, and Linz and Stepan (1996) to explain democratic transitions. Similarly, Gibson, Caldeira and Spence (2005), who focus on courts, and Murphy (2005), who focuses on tax agencies, find that political institutions can gain acceptance for unpopular decisions and policies, when legitimate. As these few examples suggest, past research focused almost exclusively on whether legitimacy encourages deference to laws and norms. This study expands this literature by examining the role of legitimacy in securing cooperation in public goods provision.

Turning our attention to the role of internal centralized authorities opens up a new set of questions for future research. I have found that in the presence of a centralized-sanctioning authority, groups can reach higher levels of cooperation and that monitors, at least in a situation in which their reputation is at stake, are willing to bear the cost of punishing in order to increase cooperation[43]. These results are qualitatively similar to those obtained using peer-punishment institutions, with the possible advantage that a centralized system of monitoring will be more efficient than a decentralized one. Future studies should investigate the relative efficiency of decentralized (i.e., peer) *vs.* centralized-sanctioning regimes. In addition, more work is needed in order to understand the motivation behind the observed behavior of both regular subjects and monitors and explore the relative effectiveness of different legitimation processes, for example democratic legitimacy *vs.* more traditional forms of authority.

Finally, the paper makes a contribution to the debate regarding the utility of using behavioral experiments in the social sciences[44]. Behavioral experiments are traditionally used to identify

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

[44]For a recent review of the debate concerning the external validity of behavioral experiments see Levitt and List

general patterns in human behavior. In our research design, which combines a lab-in-the-field experiment and observational data, behavioral games also have been deployed to reveal differences between individuals and groups. Namely, I study the impact of centralized authority on public goods provision using a behavioral experiment and then relate subjects' performance in the experiment to their real-life outcomes.

The study offers three core findings: (i) in an experimental setting, individuals are responsive to legitimacy cues: participants contributed more to public goods production when their monitor obtained its sanctioning powers through elections; (ii) the relation between legitimate authority and cooperation exists also in the participants' natural environment; and (iii) greater cooperation in participants' natural environment translates into a greater response to the legitimacy cue: group members who participate in collective marketing contribute more, on average, in the elected-monitor condition than 'defectors' who sell their coffee to local middlemen. Taken together, my findings suggest that the legitimacy of internal centralized authorities is an important factor in determining the success of farmer associations, and likely of other similar organizations. Combining a lab-in-the-field experiment with corresponding observational data, I am able to isolate one of the elements that makes group members cooperate in real life.

(2007). See also Ahlquist and Levi (2011) for a recent critique of lab experiments in political science for their inadequate attention to external validity considerations.

Appendix

Appendix 2.A Analysis of contributions

Table 2.5 reports the average contribution to the public good in the baseline, random monitor, and elected monitor conditions; it complements Figure 1 in the text.

	Prelimin Contr 1	Prelimin Contr 2	Round 1	Round 2	Round 3	Round 4	N
Baseline	437.7	399.2	399.4	405.9	423	432.8	518
Random Monitor	455.1	422.0	465.5	474.8	483.5	487.1	474
Elected Monitor	462.3	431.5	506.6	513.9	536.4	525.8	454

Table 2.5: Mean contribution to the public good in the baseline, random monitor, and elected monitor conditions. N=1446 (1543 players - 97 monitors).

In the analysis of contributions (see two-sided Mann-Whitney tests in the text) I argue that the difference between the sanctioning treatments and baseline (random *vs.* baseline; elected *vs.* baseline), as well as the difference between random and elected treatments are significant. Additional support for this finding comes from three statistical models in which I control for preliminary contributions and cluster the standard errors at the farmer cooperative level (Table 2.6). Contributions to the public account in public goods games are treated as proportions or fractions of the initial endowment (in our case a fraction of 1,000 USH). To handle this type of data appropriately, one must take into account the bounded nature of the response variable. Using a proportion in linear regression models will generally yield out-of-range predictions for extreme values of the regressors. I follow Papke and Wooldridge (1996) strategy for handling proportions in which zeros and ones may appear, as well as intermediate values. I apply a generalized linear model (GLM), using a logit transformation of the dependent variable and the binomial distribution. Such a technique allows me to generate predictions from the model which can easily be transformed back into units of the response variable. This is an improvement over past tradition of using censored normal regression techniques, such as Tobit models, on proportions that include zeros and ones. That approach is not completely appropriate, since the observed data is not really censored: values outside the interval [0, 1] are simply not possible for proportional data. Note, however, that results using Tobit models are very similar.

Analysis of Players' Contributions

	BL-T1		BL-T2		T1-T2	
	Coef.	s.e	Coef.	s.e	Coef.	s.e
ATE	0.223**	(0.086)	0.397***	(0.085)	0.172*	(0.081)
Prelim Contribution	2.552***	(0.143)	2.238***	(0.145)	2.204***	(0.153)
Constant	-1.385***	(0.084)	-1.256***	(0.084)	-1.018***	(0.081)
Observations	990		970		926	
Log Likelihood	-449.04		-447.36		-423.52	

Standard errors clustered at the farmer cooperative level in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: **DV: Mean Contribution rounds 3-6.** The DV, which is measured as proportions of the initial endowment, is modeled as a function of treatment assignment, controlling for preliminary contributions. BL-T1 refers to the Average Treatment Effect (ATE) of random monitor (T1) compared to baseline (BL) condition. BL-T2, is the ATE of elected monitor (T2) compared to baseline (BL), and T1-T2, is the ATE of elected (T2) compared to random monitor (T1). Coefficients from the GLM regressions are reported following logit transformation.

Appendix 2.B Analysis of expectations and response to punishment

Table 2.7 complements Figure 2.A in the text.

Players' Expectations of Monitors' Behavior

	Coef.	s.e.
Elected monitor	37.897*	(18.124)
Monitor gender	6.009	(24.644)
Monitor born local	-6.039	(19.758)
Monitor age	0.944	(0.725)
Monitor church attendance	6.154	(24.533)
Monitor education	-1.466	(3.571)
Monitor wealth	-1.804	(3.349)
Constant	19.766	(81.546)
σ_u		
Constant	23.325	(17.532)
σ_e		
Constant	237.806***	(5.850)
Observations	882	
Log likelihood	-6081.230	

Standard errors clustered at the farmer cooperative level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.7: **DV: Change in contribution from (preliminary) round 2 to round 3.** The DV is modeled as a function of the monitor type, controlling for the monitor's sociodemographic characteristics.

Table 2.8 complements Figures 2.B in the text.

Response to Punishment		
	Coef.	s.e.
Elected monitor condition	-18.282	(12.45)
Sanctioned at $t-1$	57.293***	(16.04)
Sanctioned by Elected	60.840**	(21.47)
Preliminary contribution	-0.009	(0.02)
Number of Players in Session	-1.575	(4.37)
Number of players sanctioned at $t-1$	3.043	(3.67)
Maximum contribution sanctioned at $t-1$	-0.046	(0.03)
Minimum contribution sanctioned at $t-1$	-0.043	(0.03)
Monitor gender	-2.589	(14.70)
Monitor age	-0.193	(0.47)
Monitor education	-1.039	(2.23)
Monitor wealth	1.774	(2.03)
Monitor church attendance	0.676	(15.02)
Monitor born in village	2.561	(11.91)
Intercept	34.018	(50.19)
Observations	2448	
N Sessions	91	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.8: **DV: Change in individual contribution from $t-1$ to t.** The DV is modeled as a function of whether player i has been sanctioned at $t-1$ and the type of monitor, controlling for the monitor's sociodemographic characteristics, monitor's sanctioning behavior at time $t-1$, and player i's preliminary contributions. Standard errors clustered at the session level.

Leader Selection Effect

In the text I consider the possibility of a leader selection effect. Figure 2.5 below, shows the average sociodemographic characteristics of the monitors in the random and elected monitor conditions. Elected monitors are disproportionally male, locally born, more educated and richer. There are, however, no differences with respect to age and church attendance.

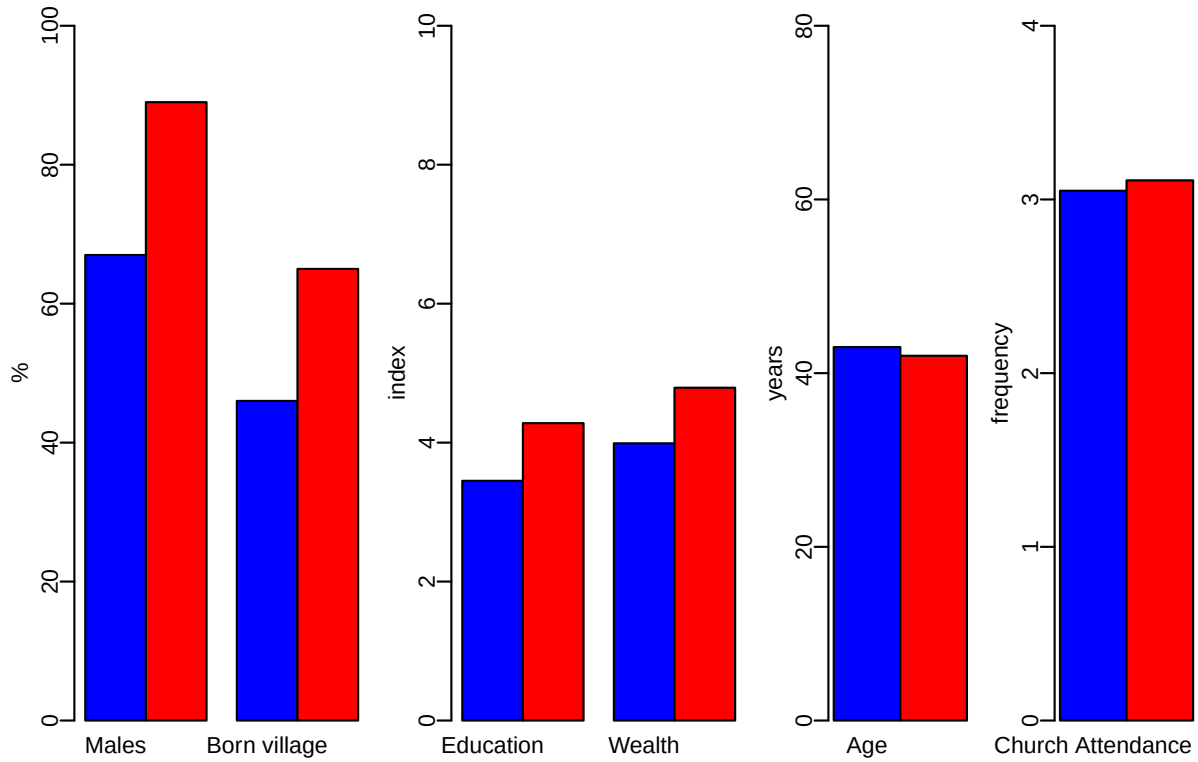

Figure 2.5: **Monitors' sociodemographic characteristics.** Mean values for the random (blue) and elected monitor (red) conditions.

Table 2.9 shows that a monitor's profile does not affect contributions. I run two models to test the significance of the monitor's sociodemographic profile. The first is an OLS model and the second is a GLM model. In both cases, the dependent variable is players' mean contribution in the sanctioning rounds 3–6. In the GLM model, the mean contribution is calculated as a proportion of the maximum contribution (bounded by 0 and 1). In both models, the mean contribution is modeled as function of six characteristics of monitors (gender, local birth, age, church attendance, education and wealth). I report results from regressions run on data from the random treatment only. Since in the random treatment the monitors were chosen by means of a lottery, we can exclude the possibility of spurious correlations, and confidently assess the impact of monitors'

sociodemographics on contributions. Moreover, since some of these sociodemographic characteristics are strongly correlated with each other (i.e., males are, on average, richer and more educated than women), I compute a joint significance test for the OLS model and consider two statistics. According to the R^2 statistics, the model explains a mere 5% of the variance in contributions. Similarly, the value for the F statistic is 1.34, which is much lower than the critical value needed for significance ($P = 0.2598$).

Significance of Monitors' Profile

	OLS		GLM	
	Coef.	s.e.	Coef.	s.e.
Monitor gender	8.074	(42.063)	0.032	(0.169)
Monitor born in village	-27.473	(40.722)	-0.111	(0.164)
Monitor age	-0.983	(1.154)	-0.004	(0.005)
Monitor church attendance	88.567	(62.126)	0.359	(0.253)
Monitor education	-12.016	(6.748)	-0.048	(0.027)
Monitor wealth	-2.109	(5.129)	-0.009	(0.021)
Intercept	200.13***	(199.500)	-0.110	(0.811)
Observations	446		446	
R^2	0.049			
Joint Significance	1.34			
Degrees of Freedom	6.00		6.00	

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table 2.9: **Impact of monitors' profile on contributions.** The DV, mean contribution in rounds 3-6, is modeled, using OLS and GLM, as function of the monitors' sociodemographic characteristics (random monitor treatment). Standard errors are clustered at the farmer cooperative level.

Appendix 2.C Analysis of Monitors' Behavior

Descriptive Analysis

1. Elected and random monitors sanction, on average, the same number of players, from 2.5 players sanctioned in round 1 to 1.5 in round 4 (See top panel, Table 2.2 in the main text).

2. The two types of monitors differ in the range of contributions sanctioned (See bottom panels, Table 2.2 in the main text)

3. The data suggests that both types of monitors have similar sanctioning patterns. Fig. 2.6 shows, for each possible contribution, the proportion of contributors that were sanctioned. The likelihood of being sanctioned, under both sanctioning treatments, drops sharply between 0 to about 400 USH, before flattening for higher levels of contributions. There is some evidence suggesting that elected monitors sanction a greater proportion of those contributing between 200 to 400 USH. This difference, however, is not significant. The goal of the matching exercise is to test whether this difference is the result of the different distribution of contribution that both monitors face, from round 3 onwards.

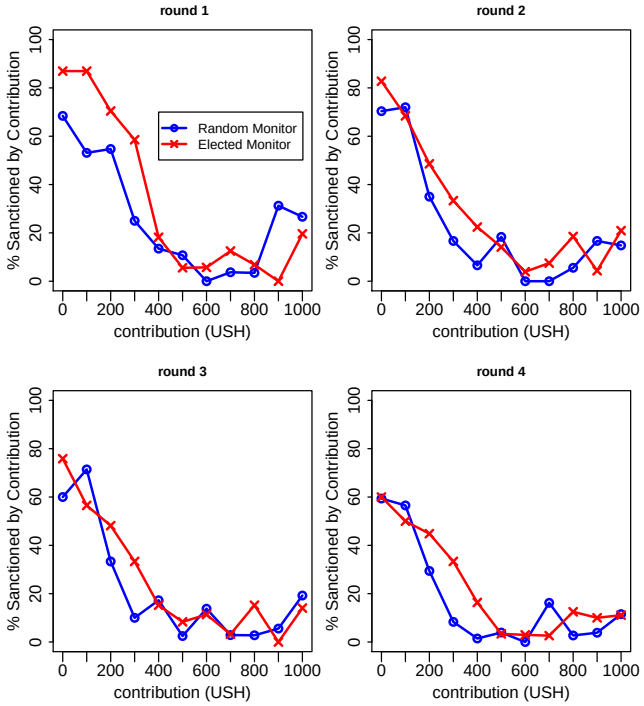

Figure 2.6: **Proportion of contributions sanctioned by level of contribution:** Random monitor (blue) and elected monitor (red) conditions.

Matching Analysis based on the Distributions of Contributions

The Kullback–Leibler (K–L) divergence is a non-symmetric measure, commonly used to calculate the difference or distance between two probability distributions. For probability distributions P and Q of a discrete random variable, their K–L divergence is defined to be:

$$D_{KL} = (P||Q) = \sum P(i) log \frac{P(i)}{Q(i)} \tag{2.3}$$

The K–L divergence is, therefore, the average of the logarithmic difference between the probabilities P and Q, where the average is taken using the probabilities P. The K–L divergence is only defined if (A.) $\sum P = 1$, (B.) $\sum Q = 1$, and (C.) if $Q(i) > 0$ for any i such that $P(i) > 0$[45]. To calculate the K–L divergence between two distributions of contributions to the public good, I first calculated, for each session j, in each round t, the probability that the number of coins i was donated to the PG (See example below, Table 2.10). Given a frequency distribution, it is pretty straightforward to then calculate P_i, Q_i, and D_{KL} (see example in Equation 2.4).

Distribution of contributions in treatment d, for session j in Round t											
Contribution-level	0	1	2	3	4	5	6	7	8	9	10
Frequency P_{d_2jt}	2	1	1	3	1	1	4	2	3	1	1
P_i	0.1	0.05	0.05	0.15	0.05	0.05	0.2	0.1	0.15	0.05	0.05
Frequency Q_{d_1jt}	3	2	1	1	2	1	2	3	1	2	2
Q_i	0.15	0.1	0.05	0.05	0.1	0.05	0.1	0.15	0.05	0.1	0.1

Table 2.10: The table provides data for demonstrating how to calculate the K–L divergence measure of two possible distributions of contributions to the public pot, for elected d_2 and random monitor conditions d_1.

$$D_{KL} = 0.1 log \frac{0.1}{0.15} + 0.05 log \frac{0.05}{0.1} + \cdots + 0.05 log \frac{0.05}{0.1} \tag{2.4}$$

[45]To overcome the problem that not all contribution-levels (coins) were realized in some session j in round t, I smoothed the probability distribution by assigning a small positive probability when $P(i) = 0$ or when $Q(i) = 0$. After assigning ϵ_{ij} I readjusted the probabilities in each session j by dividing $P(ij)$ by $1 + \sum \epsilon_{ij}$ such that $\sum P = 1$ and $\sum Q = 1$. The value of epsilon is the probability that a given contribution level was realized across all sessions, blocked by treatment and round

In Figure 2.7 I demonstrate how matching helps ensure that we compare the behavior of monitors that face similar contributions:

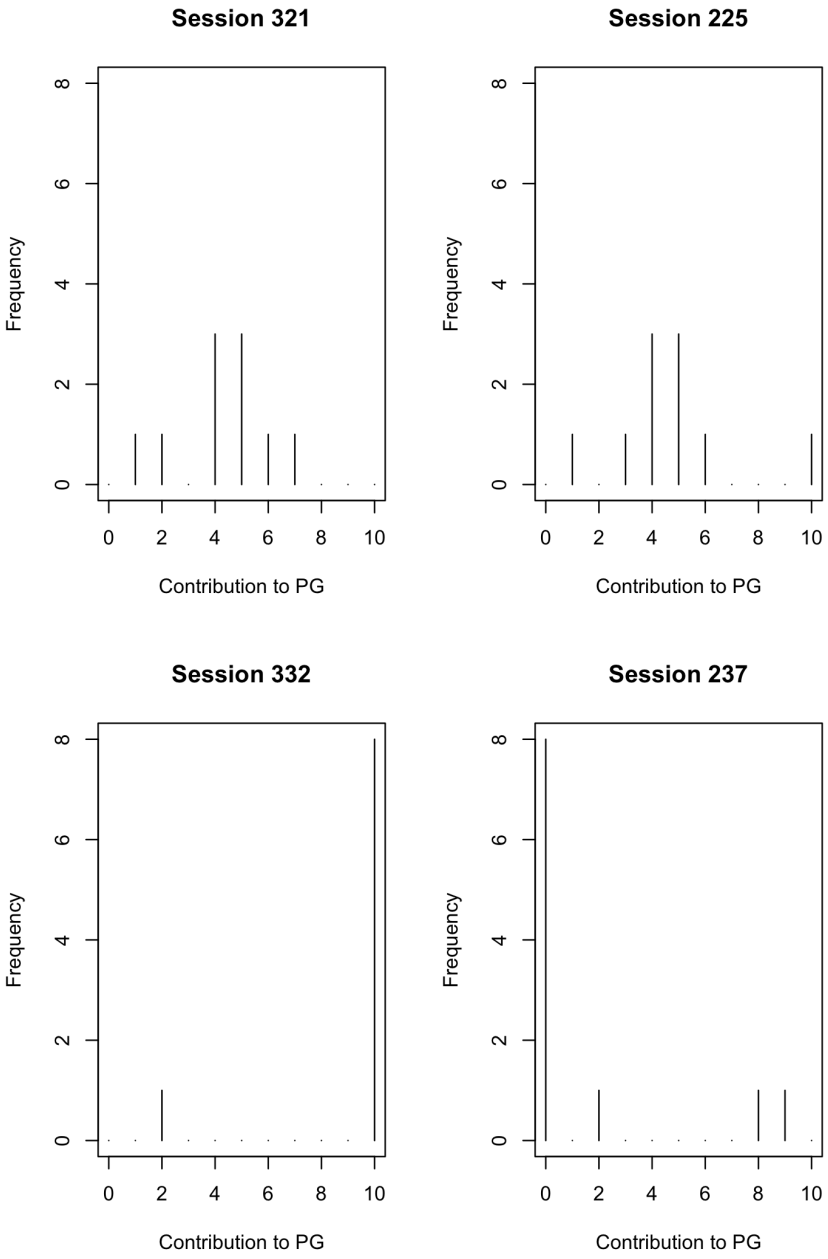

Figure 2.7: In the top panel I graph the pair of distributions of contributions with the lowest K–L divergence value ($D_{KL} = 0.01$), in round 6. The bottom panel graphs the pair of distributions, in that round, with the highest K–L divergence ($D_{KL} = 1.17$). Numbers refer to session IDs. Sessions 321 and 332 were drawn from T2 whereas sessions 225 and 237 are drawn from T1. Contribution (x-axis) are reported in monetary units (i.e., No. of coins).

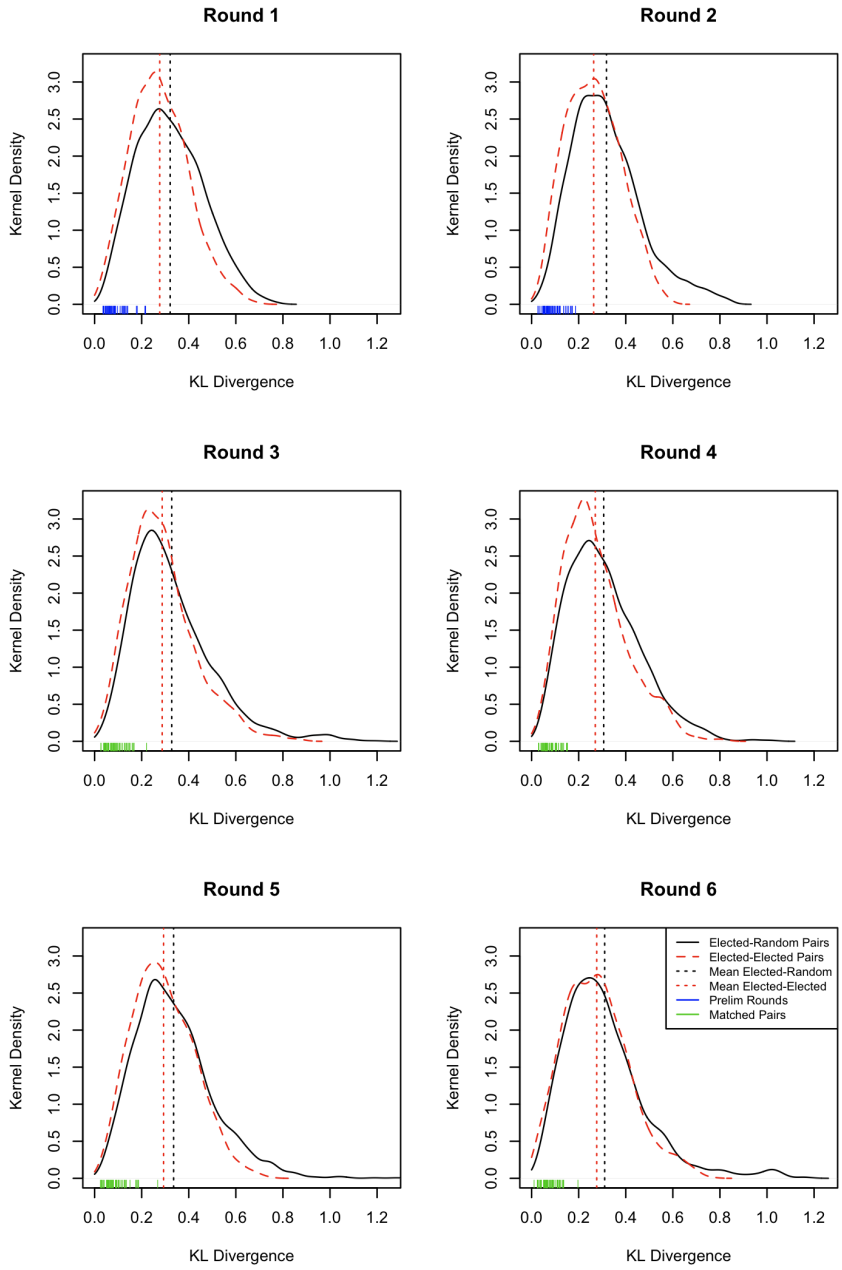

Figure 2.8: **K-L divergence measure of all possible pairings:** Eelected-Random monitors (solid line), and of of Elected-Elected monitors (red dashed line). Rugged lines represent the K–L divergence measure of the matched Elected-Random pairs.

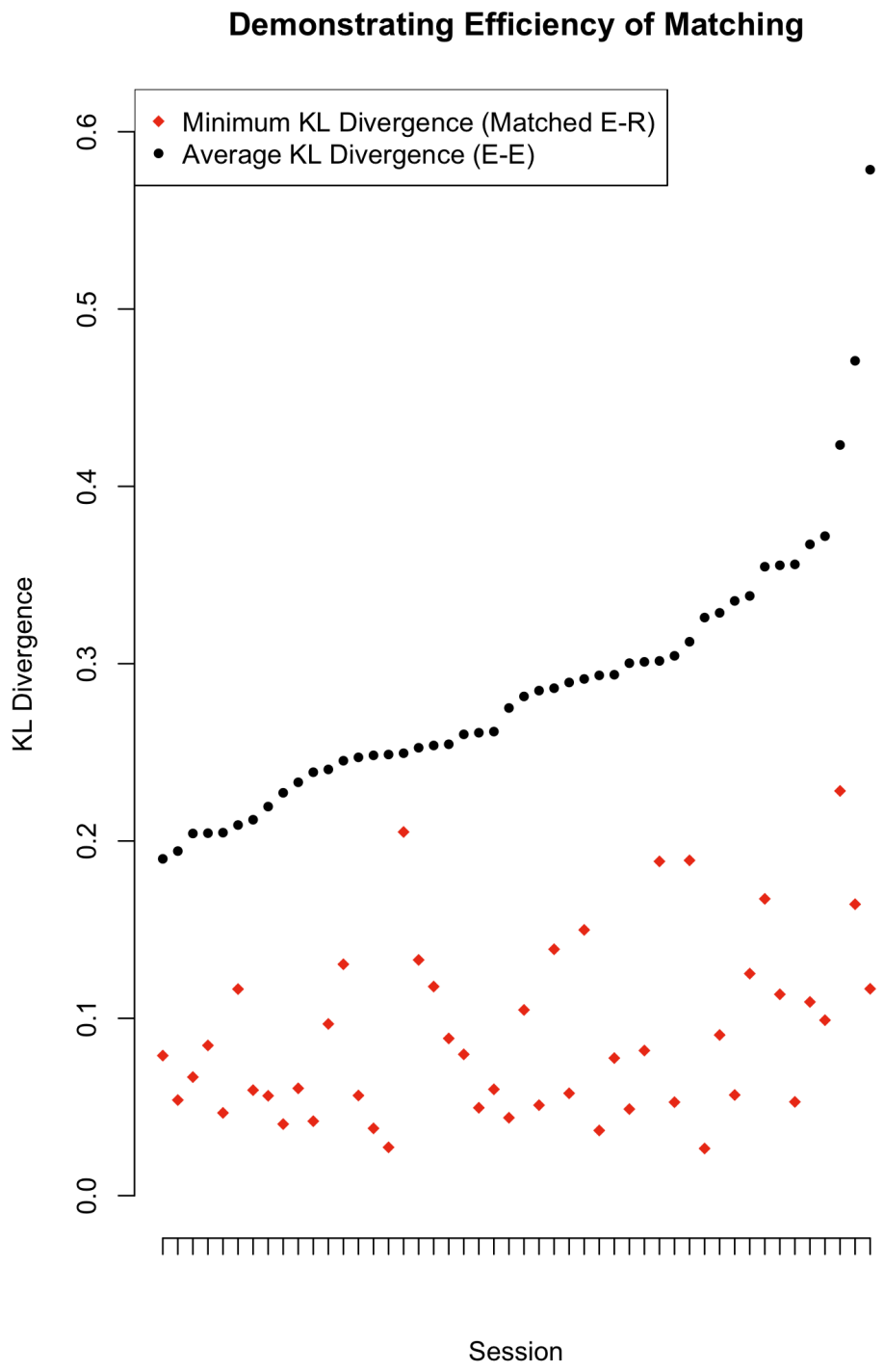

Figure 2.9: This figure demonstrates how matching on distributions may improve our ability to compare the behavior of elected and random monitors. The figure graphs, for each session of elected monitor in round 5, the mean KL divergence distance to (a) *all other elected monitor sessions* (black circle) and to (b) its matched pair (red diamond). For each of the elected monitor sessions, the divergence between its distribution and the distribution of its matched pair, is significantly smaller than its divergence to another elected monitor session picked by random.

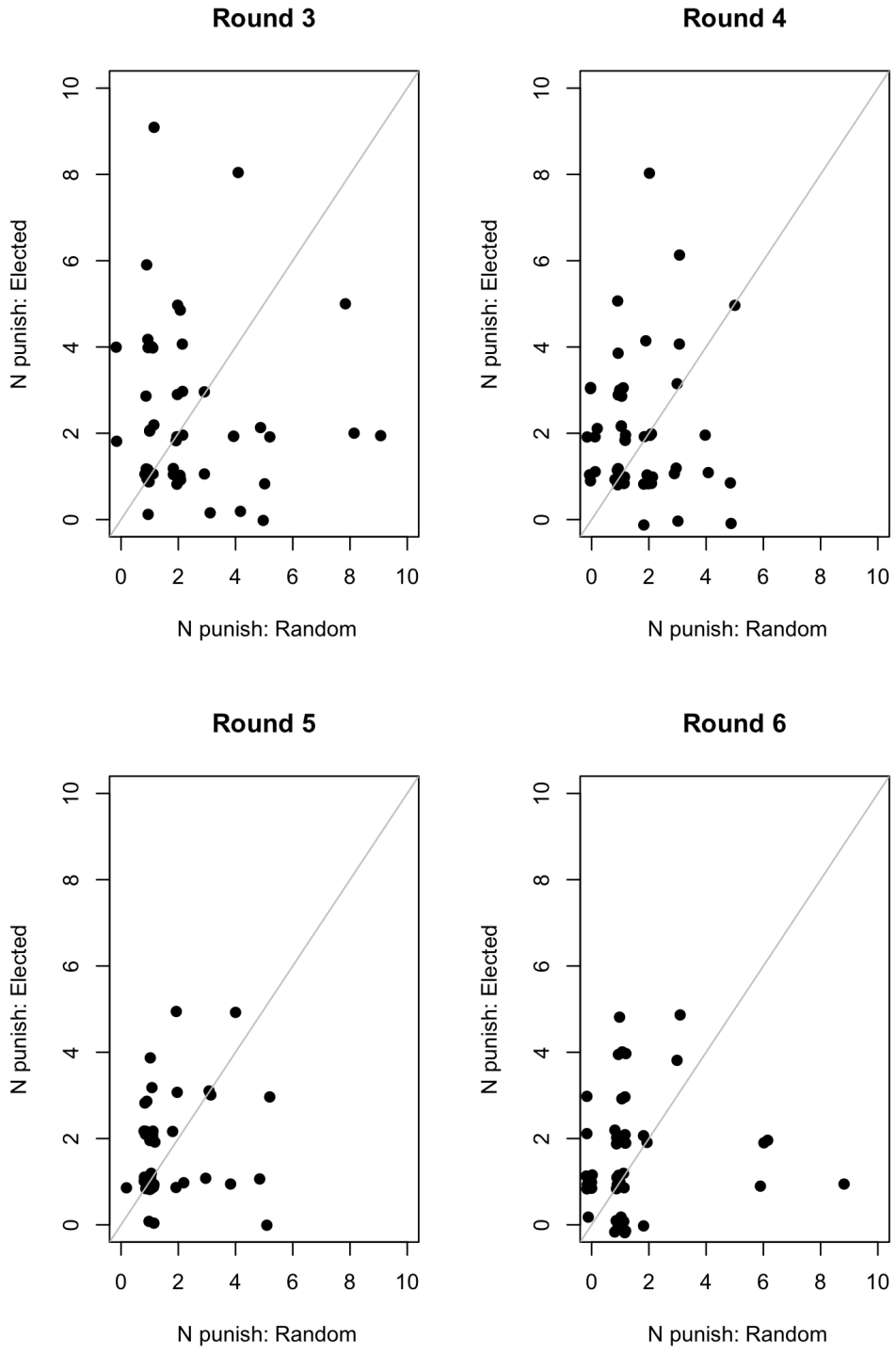

Figure 2.10: Number of players punished in round t by elected monitors as a function of the number of players punished by the random monitors, for all matched pairs.

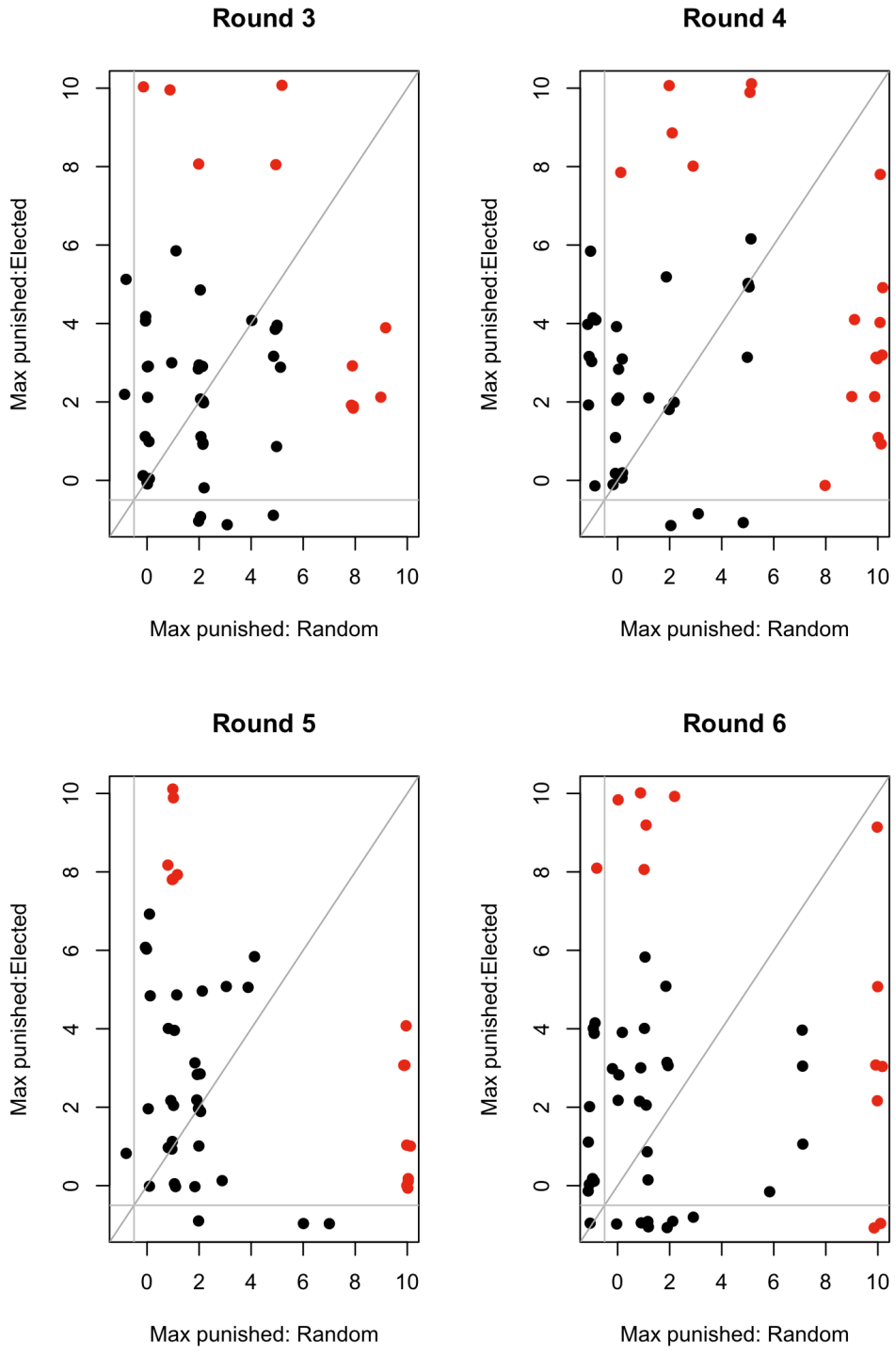

Figure 2.11: Maximum contribution level punished in round t by elected monitors, as a function of the maximum contribution level punished by the Random monitors for all matched pairs. Red dots signal that the maximum contribution punished in that round was equal or above 8 coins (MUs). Note that -1 on the X (Y) axis refer to sessions in which the random (elected) monitor has not punished any player (more accurately, any level of contribution.) Punishment decisions (x-axis) are reported in MUs (coins).

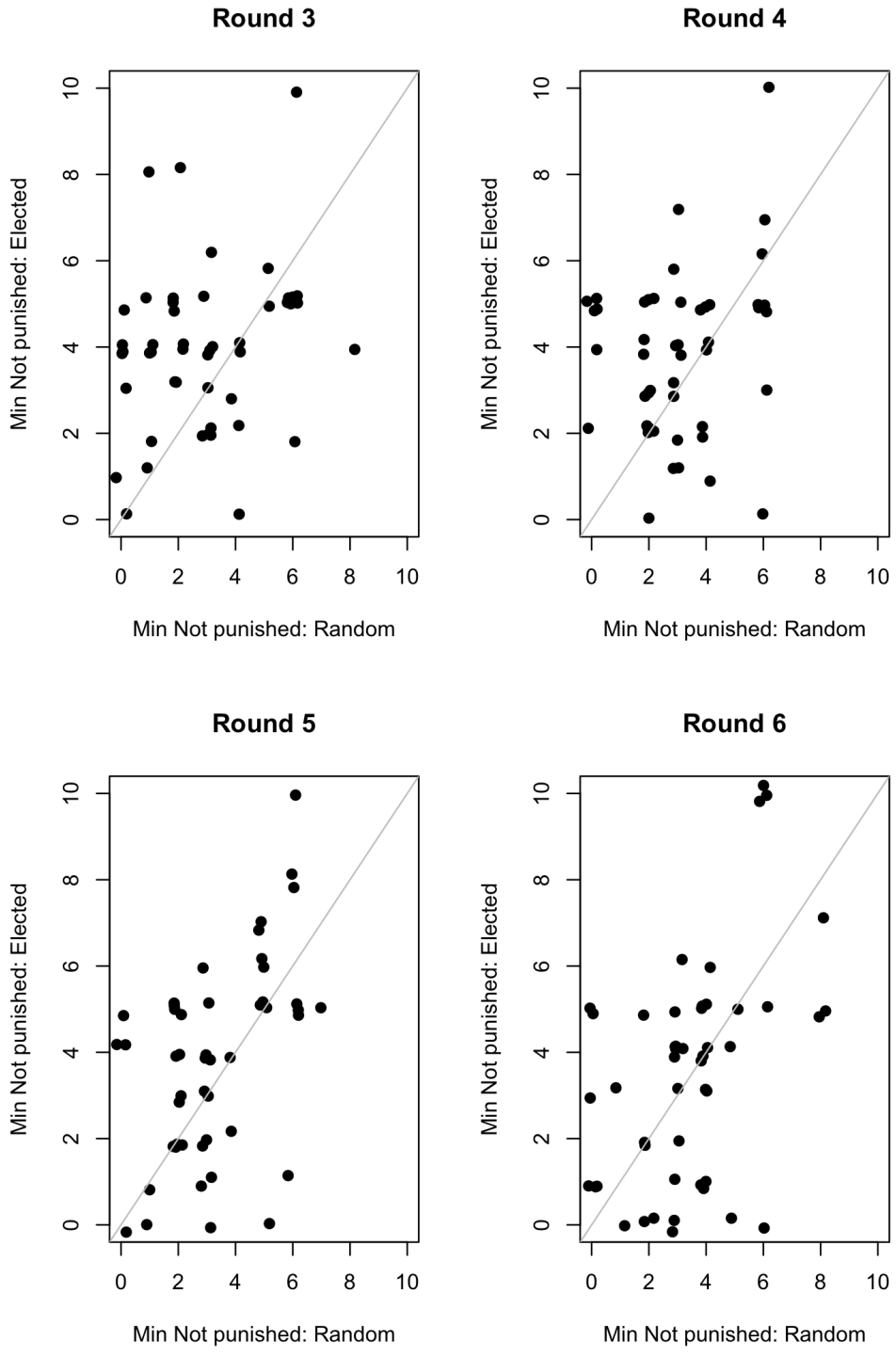

Figure 2.12: Minimum contribution that was *not* punished in round t by elected monitors as a function of the minimum contribution that was *not* punished by random monitors, for all matched pairs. Punishment decisions (x-axis) are reported in MUs (coins).

Appendix 2.D Analysis of Observational Data

In this subsection I provide additional information that supplements the analysis of the relation between legitimate authority and cooperation in the farmers' natural setting.

Legitimate Authority and Cooperation (Natural Setting)

Table 2.11 provides the descriptive statistics of the variables used for analyzing the observational data. Regression results from the six multilevel random intercept models are provided in Table 2.12 below.

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Dependent Variable					
Sell in Bulk (cooperate)	0.61	(0.49)	0	1	1746
Legitimacy Proxy Variables					
Leader monitored	0.49	(0.5)	0	1	1723
Leaders Accountable	1.16	(0.58)	0	2	1670
Receipt given	1.41	(0.82)	0	2	1544
Name manager	0.63	(0.48)	0	1	1781
Know selection rule	0.73	(0.45)	0	1	1697
Legitimacy index (std.)	0	(1)	-2.53	1.08	1476
Individual Controls					
Male	0.68	(0.47)	0	1	1781
Age (units of 10)	4.56	(1.44)	1.4	9.5	1781
Formal education	0.83	(0.37)	0	1	1773
Log Seasonal Yield	5.41	(1.13)	0	9.21	1683
Years since Joining Group	3.71	(1.76)	1	8	1769
Middleman honesty	0.91	(0.28)	0	1	1698
Associational-life (std.)	0	(1)	-3.14	2.85	1633
Church attendance	2.99	(0.53)	1	4	1781
Group Controls					
N members (units of 50)	4.29	(2.21)	1.22	10.72	1781
Mean seasonal yield	3.8	(1.78)	0.86	7.76	1781
Leader's Effort (std.)	0	(1)	-2.69	2.53	1781

Table 2.11: **Summary Statistics:** variables used to analyze the relation between legitimate authority and cooperation in the farmers' natural environment.

Relation Between Legitimacy and Cooperation

	A	B	C	D	E	F
Leader Monitored	0.458***					
(Yes)	(0.132)					
Leaders Accountable		1.063***				
(Yes)		(0.22)				
Leaders Accountable		1.367***				
(Very)		(0.25)				
Receipt Given?			0.259			
(Sometimes)			(0.22)			
Receipt given?			0.602**			
(Always)			(0.19)			
Name Manager				0.489***		
(Yes)				(0.13)		
Know Voting Rule					0.541***	
(Yes)					(0.15)	
legitimacy index (std.)						0.212**
						(0.07)
Ind and DC Controls	*X*	*X*	*X*	*X*	*X*	*X*
Constant	-2.746***	-3.257***	-1.831**	-2.814***	-2.782***	-1.529*
	(0.66)	(0.70)	(0.68)	(0.65)	(0.66)	(0.69)
Constant	-0.521**	-0.491**	-0.584**	-0.594***	-0.583***	-0.591**
(Random Intercept)	(0.17)	(0.17)	(0.19)	(0.17)	(0.18)	(0.19)
Observations	1421	1388	1306	1456	1406	1258

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.12: **Legitimacy and cooperation: natural setting.** DV: Selling coffee at least once via the farmer association in the past season ('cooperation'). The dependent variable is modeled as a function of one of six legitimacy proxy measures, using mulit-level regression models. All models control for individual and for group-level predictors. To relax the assumption of conditional independence among members of the same farmer group, we include a farmer association-specific random intercept

Cooperative Behavior in the PGG and in the Natural Setting

	Baseline	Random	Elected
Sold in Bulk	0.110	-0.056	0.280***
	(0.15)	(0.09)	(0.07)
Sex	-0.185	-0.170*	-0.045
	(0.13)	(0.07)	(0.10)
Age (units of 10)	0.089*	0.007	-0.004
	(0.04)	(0.03)	(0.03)
Attendance at religious services	-0.148	-0.101	-0.070
	(0.11)	(0.07)	(0.08)
Years since Joining Group	0.007	0.016	-0.019
	(0.03)	(0.02)	(0.02)
Associational-life (std.)	0.015	-0.015	0.069
	(0.05)	(0.04)	(0.05)
Education (std.)	-0.041	0.015	0.023
	(0.07)	(0.04)	(0.05)
Contribution prelim 2 (Prop.)	0.760***	0.748***	0.551***
	(0.07)	(0.06)	(0.06)
Constant	-0.282	0.279	0.236
	(0.39)	(0.26)	(0.29)
Observations	373	364	337

Standard errors clustered at the session level in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.13: **Relation between cooperation in PGG and natural setting.** DV: subject's mean contribution to the public good in the punishment rounds (3-6). The dependent variable is modeled a function of the subjects cooperative behavior as a member of her farmer group (i.e. sold coffee in bulk), controlling for the subjects' socio-demographic characteristics and for contributions in the preliminary round.

Appendix 2.E Assessing Balance Across Treatments

Covariate	Baseline Mean	Random Diff	Elected Diff
Male	.67	-0.002	-0.027
Age	45.24	0.174	0.888
Church attendance	2.95	0.050	0.052
Born in village	.468	0.035	0.059
Read	.718	0.007	-0.001
Write	.688	0.015	0.006
Advanced English	.212	-0.008	-0.055
Education	5.45	0.065	-0.257
Wealth	5.54	.0178	-0.303
Observations	493	449	442

Table 2.14: **Sample Characteristics:** column 1 reports means in baseline, whereas columns 2 and 3, report the difference from that mean for Random and Elected monitor conditions. Male, Born in village, Read, Write and Advanced English are dummies. Church attendance is a categorical variable measuring the frequency of attending religious services. Education and Wealth are composite measures constructed via principal component analysis, broken down to deciles.

	BL-T1 $t-test$ $p-value$	BL-T1 K-S $p-value$	BL-T2 $t-test$ $p-value$	BL-T2 K-S $p-value$	T1-T2 $t-test$ $p-value$	T1-T2 K-S $p-value$
Male	0.964	1.000	0.539	0.994	0.539	0.998
Age	0.865	0.616	0.376	0.720	0.510	0.089
Church attendance	0.249	0.965	0.200	0.999	0.957	1.000
Born in village	0.443	0.916	0.222	0.367	0.604	0.999
Read	0.807	1.000	0.967	1.000	0.780	1.000
Write	0.641	1.000	0.830	1.000	0.784	1.000
Advanced English	0.800	1.000	0.059	0.460	0.093	0.675
Education	0.748	0.846	0.171	0.620	0.111	0.380
Wealth	0.944	0.676	0.236	0.117	0.213	0.351
Observations	955	955	945	945	900	900

Table 2.15: **Formal Tests of Covariats Balance.** Columns 1, 3, and 5, report $p-values$ from regressing each covariate on dummies representing treatment comparisons. Column 1 reports the significance test of the comparison between those sampled into baseline (BL) and those sampled into Random monitor (T1); column 3 compares those sampled into baseline against Elected monitor (T2); and column 5 compares Random (T1) to Elected (T2). Columns 2, 4, and 6 report $p-values$ from two-sample Kolmogorov-Smirnov test for equality of distribution functions. Note that none of the t-tests of means nor the permutation test of distribution is statistically significant.

Chapter 3

A Theory of Leadership Selection in Small Groups - With Evidence from Ugandan Farmer Associations

3.1 Introduction

Self-help organizations are relatively small, voluntary groups that provide goods and services to members. Examples include micro-finance groups, Chambers of Commerce, PTAs, common-pool resource groups, artisan cooperatives and farmer associations. Such organizations are present in many facets of economic and social life, and in countries of all income levels. Their importance has arguably increased in recent years, particularly in developing countries, where larger political units have democratized, decentralized and liberalized[1]. Thus, the ubiquity of self-help organizations and their growing impact on social welfare calls for a better understanding of the factors that determine their effectiveness, in particular.

This study considers one factor affecting the success of self-help groups: the quality of the group leader. Leadership quality is thought of as a combination of the leader's ability and the amount of effort that the leader exerts while working for the group. While many factors may affect the quality of leadership obtained by a small group, this chapter focuses specifically on two. First, the ability of a group to monitor its leader in order to incentivize effort, which depends on the monitoring institutions available to the group. Second, the availability of private income opportunities outside of the group, which depends on local economic conditions. While both of these factors are likely to evolve over time, neither can be changed rapidly from one election cycle to the next[2]. Thus, I will treat these as fixed parameters for any given elections and consider the quality of leadership that a group obtains, given the monitoring institutions and local private income opportunities present at any specific time.

The chapter begins by introducing a theoretical model designed to elucidate some of the factors affecting leadership quality in small group settings[3]. The model is applicable to self-help groups in which leaders, who play a decisive role in producing a group public good, are chosen

[1] Empirical evidence supports this claim: In Senegal 10% of sampled villages reported having at least one self-help group in 1982; by 2002 this figure was 65%. In Burkina Faso the figures were 22% for 1982 and 91% in 2002 (Bernard et al., 2008).

[2] From a group's perspective, there are good reasons to ensure relative 'stickiness' of monitoring institutions. One reason is to protect against attempts by incumbents to alter monitoring institutions to increase incumbency advantage.

[3] Throughout the chapter I use the terms 'small groups' and 'self-help groups' interchangeably.

endogenously from the set of available members through democratic procedures. This is a feature shared by many small groups, including those described above. The model does not apply, however, to other "small groups" in which leaders are selected from a set of non-members or are appointed by non-members[4]. The starting point for this chapter is, therefore, a citizen-candidate framework (Besley and Coate, 1997, Osborne and Slivinski, 1996), which a number of recent studies have used to investigate issues of leadership quality (Caselli and Morelli (2004), Messner and Polborn (2004), Gagliarducci, Nannicini and Naticchioni (2010))[5].

The model departs from existing theories in that it reflects the particular features of self-help groups, which differ from larger political units (regions, nations, etc.) in a number of ways. First, unlike large political units where each citizen knows only a few of her fellows, in small groups members generally know each other well. This means that incomplete information plays a smaller role in determining outcomes in this settings. It also means that small groups have an advantage over large political units in offering incentive schemes that condition remuneration on effort[6]. Second, small groups are often formed with a specific purpose in mind, so that members' goals are generally more closely aligned than in larger political units[7]. Third, participating in small groups, even as the leader, is generally a part-time affair. Rarely do small groups have the resources to employ full time leaders, as is common in larger political units. The result is that leaders must choose how to allocate their time between producing the group public good and generating private income. Fourth, the leader receives significant benefits from the public good that she produces. In contrast, in larger political units, the value that leaders derive from the public good they produce is often small relative to the amount of effort they exert or the overall value of the public good. This has an important effect on members' incentives to seek leadership positions.

[4]In the first example, a board of directors may choose a CEO from outside a company. In the second case, leaders of military units are commonly appointed by higher authorities, external to the group/unit. Such groups do not fit the definition of "small groups" used in this study.

[5]Other important contributions to this literature include Barro (1973), Ferejohn (1986) and Persson, Roland and Tabellini (1997). For a recent survey of the literature, see Besley (2005).

[6]See Besley (2004, 197-198) which provides a thoughtful discussion on the problems that large political units face in trying to devise high-powered incentives for politicians.

[7]Note that the assumption that society is comprised of citizens that have competing interests, together with incomplete information about candidates who cannot credibly commit to voters, forms the basis of the canonical citizen-candidate model. See, among others, Besley and Coate (1997).

The study's principal theoretical finding is that groups may face a trade-off between the ability of the elected leader and the amount of effort that the leader exerts. This trade-off is driven by two effects. First, an increase in monitoring of the leader will incentivize the leader to exert more effort, which I call the *discipline effect*[8]. However, a higher level of monitoring may also cause high ability members to self-select out of the candidate pool, resulting in leaders with lower ability being elected, which I term the *self-selection effect*. When both of these effects are operating, the result is a rough inverted U-shaped relationship between the level of monitoring and the value of the public good produced. At low levels of monitoring, high ability leaders are elected, but the exert little effort, leading to a low public good value. An increase in the level of monitoring causes the leader to exert more effort, increasing the value of the public good. However, if the level of monitoring continues to increase eventually high ability members will start to self-select out of the candidate pool, reducing the value of the public good. Importantly, this trade-off exists *only when private income opportunities are sufficiently high*. If private income opportunities are low, then high ability members have little reason to opt-out of candidacy. Thus, when there are few private income opportunities outside of the group activities, groups will be able to obtain high ability leaders who also exert a relatively high level of effort.

These results extend existing theoretical work on leadership quality building on the citizen-candidate framework, most of which has focused on larger political units. The self-selection effect appearing in the model was suggested by Caselli and Morelli (2004), who pointed out that high ability citizens may opt-out of the candidate pool if the rewards from holding office are set too low. Similar results appear in Messner and Polborn (2004), who offer a model that moves towards the small group setting by having only one leader, as opposed to the set of leaders in Caselli and Morelli (2004), and allowing the leader to benefit from the value of the public good that she produces. Another important contribution is Gagliarducci, Nannicini and Naticchioni (2010), which introduces a model in which leaders can allocate their time to generating private income while in office. They show that under these circumstances, high ability citizens will prefer to run, but may exert little effort while in office. One contribution of my theory is to bring together these

[8]This terminology follows Besley (2006).

ideas in a framework that allows me to evaluate the welfare implications of the trade-off between leader ability and effort[9]. The present theory also differs from existing work in that it is adapted to the particular features of small groups. Finally, I pay more attention to the role of private income opportunities in determining whether the trade-off between leader ability and effort exists.

To test these results, I use original data collected through an extensive survey of associations of coffee farmers in Uganda. These associations, recently established through a USAID funded intervention, provide a good context for testing the model because we are able to look across a large number of groups, all with relatively similar structures, and all formed around the same time for the same purpose. However, while the surveyed associations have similar governance structures, there exists a significant amount of variation across associations in the availability of private income opportunities and in the level of monitoring undertaken by the groups, which depend on their having the necessary monitoring institutions. This variation results from a number of factors, including the identity of the facilitator who helped the farmers set up their associations, local economic conditions, and variation in the cohesiveness of different localities, as manifested in the strength of their social networks (Grossman, 2011*c*). Exploiting variations in monitoring institutions, and in local private income opportunities, allows us to assess the capacity of the model to explain the determinants of leader quality in self-help groups.

The Ugandan farmer associations that I study fit the features of small groups, described in the theoretical portion of the model, well. The associations are made up of farmers from several nearby villages. All group members share a common main goal, obtaining higher prices for their outputs, with secondary goals including obtaining lower input prices and learning about better farming practices. The group leaders spend only part of their time working for the group, with the rest devoted to farming their land or working at other off-farm jobs. When leaders negotiate higher prices for their crops, they benefit directly and significantly from the group public good through the higher price that they receive for their own crops.

[9]In contrast, Gagliarducci, Nannicini and Naticchioni (2010) predict that when incumbents are allowed to split their time between private and public activities, citizens will be forced to elect either high ability leaders who do not exert any effort towards generating group goods, *or* low ability leaders who do. Unlike my model, which makes clear welfare predictions, Gagliarducci, Nannicini and Naticchioni (2010) cannot assess which of those two corner solutions is preferable.

The data collection effort involved conducting extensive surveys of over 3,000 members and leaders drawn from a sample of 50 farmer associations. These data allow me to construct a broader set of variables than were available in previous studies, including information on the value of the public good, members' ability, leaders' effort, groups' monitoring institutions, local private income opportunities, and changes in the welfare of group members since joining their group.

These variables were measures using survey responses as well as other sources. For example, the value of the public good is measured as the share of group members selling collectively through the group, or the average share of members' output sold collectively. The effort measure depends on members' survey responses rating the leader's effort level, as well as the number of collectively selling events organized by the leader. Ability is measured using information on members' education, literacy, English proficiency, and scores from cognitive tests that I administered. The monitoring measure takes into account institutions such as a whether there is a committee responsible for monitoring the leader, whether external audits were used, and whether the group specified how much the leader should work. Local private income opportunities are measured as the share of group members holding off-farm jobs. Measure of members' change in welfare conditions is based on change in the ownership of twelve different durable assets, such as livestock and bicycles, since joining the group. Whenever possible, these measures were checked against information from other sources, such as interviews with the leaders, or the groups' books.

To summarize the empirical results, I find support for both the discipline and self-selection effects, as well as the trade-off between them represented by the inverted U-shaped relationship between the level of monitoring and the value of the public good. Furthermore, my results suggest that the level of private income opportunities plays an important role in determining whether these effects coexist. Finally, I find evidence that higher quality leaders produce a higher public good value, and that the value of the public good is associated with positive welfare effects.

Several existing studies look for evidence of a self-selection effect, including (Ferraz and Finan, 2010, Gagliarducci and Nannicini, 2010, Gehlbach, Sonin and Zhuravskaya, 2010). These studies find evidence supporting the self-selction effect, though the effect is often modest. A particular

interesting study by Kotakorpi and Poutvaara (2010), which uses data on MPs in Finland, finds evidence of the self-selection effect for women, but not for men. The authors speculate, but do not test, that this difference may be due to differences in the private income opportunities available to women relative to men.

This chapter makes two advances over existing tests of the self-selection effect. First, my theory suggests that the self-selection effect exists only under certain conditions: when the level of monitoring is high and private income opportunities are available. I take these conditions into account when testing for the existence of the self-selection effect. Second, I have individual-level data on the *entire pool of potential candidates*. In contrast, existing studies observe only those individuals who choose to become candidates. This allows me to look at individual choices when assessing the self-selection effect, and control for individual characteristics. As predicted by the model, the empirical findings suggest that high-ability member are less likely to become the leader when the level of monitoring is high and there are sufficient private income opportunities.

This study is also related to existing work on the relationship between the costs and rewards of office and the effort exerted by incumbents. In a recent paper, Ferraz and Finan (2010) find that an increase in leader remuneration leads to an increase in the effort of legislators in Brazilian municipalities. One of the key testable predictions of my model is the discipline effect, in which an increase in the level of monitoring causes leaders to exert more effort. These are complementary approaches because both remuneration and monitoring affect the costs and benefits of holding office and the returns to exerting effort[10]. My data allow me to measure both effort and the level of monitoring institutions available to each group. The results suggest that an increase in monitoring levels is associated with a significant increase in the effort exerted by the leader, consistent with the theoretical predictions. Thus, my findings regarding the role of monitoring are similar to the results of Ferraz and Finan (2010) regrading remuneration.

This research is related to two other sets of existing literature. First, my findings lend support

[10]Remuneration affects the returns to exerting effort by changing the benefits of winning reelection. Monitoring affect these benefits in slightly different way: by tying the rewards and costs to the amount of effort exerted.

to the growing literature on the impact of leaders' characteristics on welfare outcomes[11]. Second, the chapter is related to the study of farmer organizations as engines of growth. This literature suggests that farmer associations can play an important role in generating development and reducing poverty[12], yet the success of interventions of this type have been mixed[13]. The impact of leadership has been identified as one factor that could be important in determining the success of these programs[14]. I contribute to this literature by showing how the governance structure of these associations, and the economic environment in which they are embedded, can affect the quality of leadership that they obtain and, thus, their effectiveness.

In the next section of the paper I present the theoretical model and derive several testable predictions. Section 3.3 describes the Ugandan farmer associations that are used to test the model, while Section 3.4 describes the data collection procedure. Section 3.5 presents a brief study of associations from two districts, Iganga and Masaka. The empirical analysis is contained in Section 3.6, while Section 3.7 concludes.

3.2 Theory

The model introduced in this section builds on a growing literature investigating factors that determine the quality of leaders obtained through democratic political institutions utilizing a citizen-candidate framework. There are two main differences between my model and previous theories. First, I incorporate a number of the features found in previous studies into a single theoretical framework. In particular, the theory includes both the discipline and self-selection effects, and group members are heterogeneous in their ability. These features allow me to make welfare

[11]See, among others,Chattopadhyay and Duflo (2004) who study the effect of political reservations for women on the type of public goods provided by Indian Village Councils, Jones and Olken (2005) on the effect of leaders' quality on growth, and Kasara (2007) on the relation between leaders' origin and taxation policy.

[12]In the context of the developing world, see, among others, Narayan-Parker (2002*b*). In the context of the developed world see Staatz (1987), Sykuta and Cook (2001), and Sexton (1990).

[13]See, e.g., (Hellin, Lundy and Meijer, 2009).

[14]See Bingen, Serrano and Howard (2003). Other factors that have been identified include (i) the legal and policy environment (Hussi et al., 1993), (ii) project-design components (Bingen, Serrano and Howard, 2003) and (Shepherd, 2007), (iii) the nature of the links between producers and buyers (Shepherd, 2007), (iv) group-specific factors, such as size, membership homogeneity, internal cohesion and trust (Stringfellow et al., 1997) and (Agrawal and Goyal, 2001), and (v) market conditions (Hellin, Lundy and Meijer, 2009) and(Berdegué, 2001) .

statements regarding the trade-off between these two effects, which was not possible in previous frameworks[15]. Second, I present a theory that has been adapted to the small-group setting[16]. The characteristics of small groups that I have described lead me to specify a model in which, (i) members have perfect information about each others abilities, group monitoring institutions, and the availability of private income opportunities, (ii) leaders are sanctioned (or rewarded) depending on the amount of effort they exert, (iii) group members' preferences are perfectly aligned, (iv) leaders divide their time between producing a group public good and earning private income, and (v) leaders benefit significantly from the public good that they produce.

Model setup

The model considers a group of N members, indexed by $i \in (1, ...N)$, which is formed in order to produce a group public good. The members elect a leader who is responsible for producing the good. The value of the public good produced depends on the effort exerted by the leader and the leader's ability. Members other than the leader do not participate in public goods production[17]. Members are heterogeneous in their ability $A_i \in (0, \bar{A})$, with no two having the same ability level, so that group members can be strictly ordered according to their ability. Each member is endowed with one unit of effort that can be allocated between generating private income and public goods production. However, since only the leader produces the public good, all other members allocate all of their effort towards generating a private income.

Two parameters play key roles in the model. The first is the level of monitoring of the leader undertaken by the group, represented by $m \geq 0$. The monitoring level is an exogenous parameter which depends on the institutional monitoring technology available to the group, such as the

[15] E.g., Gagliarducci, Nannicini and Naticchioni (2010).

[16] Note that 'small' self-help groups are not defined with respect to some particular size. In effect, the size of a "small group" is context-specific. Groups with many members may display these characteristics, though as group size grows it seems increasingly unlikely that most or all of these characteristics will be sustained. Moreover, not every group with few members will satisfy these criteria. Notwithstanding these caveats, I argue that these features are present in a great many groups with few members.

[17] It would not change my results if all group members were forced to put a fixed amount of effort towards public good production. The more complex possibility that there may be complementarities between leader quality and the amount of effort that members devote to public goods production is interesting, but beyond the scope of this paper.

existence of a committee responsible for overseeing the leader[18]. Treating monitoring levels as an exogenous parameter fits the empirical setting that I study well, as discussed in Section 3.3. It also matches the existing literature on this topic, which generally takes the costs and rewards of office as exogenously given[19]. The level of monitoring influences both the amount of effort that the leader devotes to public goods production and group members' candidacy choices.

The second important parameter is the value of the private income opportunities available to group members, relative to the value public good, represented by $\alpha \in (0, 1)$. I refer to this parameter as "private income opportunities". In practice, this relative value may depend on both the availability of private income opportunities, or on factors affecting the potential value of the public good, though I focus primarily on how α is affected by the availability of private income opportunities. It is important to note that this parameter represents income opportunities that are outside of the group and are not affected by the level of the group public good[20]. Also, α is a group-level parameter, which applies to all group members; individual-level variation in private income opportunities is captured instead by each individual's ability. The level of private income opportunities will affect the leader's allocation of effort, and each member's candidacy choice.

Group members derive utility from their income, Y_i according to an increasing and concave utility function $U_i = U(Y_i)$. The income of a member i who does not become a candidate (nor the leader) is given by Equation 3.1, where the leader is some individual l.

$$Y_i = I(A_i, 1, \eta_I)\alpha + P(A_l, e_l, \eta_P)(1 - \alpha) \tag{3.1}$$

The $I(A_i, 1, \eta_I)$ term represents member's private income from sources outside the group activities, which depends on the member's ability, the member's effort devoted to generating private

[18]While it is possible that these monitoring institutions are adjusted over time, these adjustments are likely to take place slowly. This is likely do to the fact that changes to monitoring institutions must be made by incumbent leaders, so that allowing monitoring institutions to adjust rapidly would open them to manipulation by politicians. Caselli and Morelli (2004) describes the negative results that can occur when incumbent politicians are able to adjust the costs and rewards of holding office.

[19]See, e.g. Messner and Polborn (2004) and Gagliarducci, Nannicini and Naticchioni (2010). An exception is Caselli and Morelli (2004) who suggest that incumbent politicians can reduce the benefits of holding office in order to increase their chances of reelection, with negative effects.

[20]For example, in the Ugandan farmer associations empirical setting, working for a local NGO represents a private income opportunity, while farm work does not.

income, which members who are not the leader will always set to one, and a random noise term η_I. The $P(A_l, e_l, \eta_P)$ term represents the value that a member receives from the public good, which depends on the *leader's* ability A_l, the amount of effort the leader devotes to generating the public good, e_l, and a random noise term η_P. The income for an individual who becomes a candidate but is not elected the leader is the same as Equation 3.1, less some additional cost of candidacy $\phi > 0$[21]. The income of the member, $i = l$, who becomes the leader is given by Equation 3.2.

$$U_l = U(I(A_l, 1 - e_l, \eta_I)\alpha + P(A_l, e_l, \eta_P)(1 - \alpha) - C(m, e_l) - \phi) \tag{3.2}$$

The leader's value from the public good depends on her ability and the amount of effort she devotes to producing the public good, $e_l \in [0, 1]$, where $1 - e_l$ is the amount of effort the leader devotes to generating private income. The leader may also receive sanctions (or benefits) from the group, based on the amount of effort exerted. These sanctions depend on both the amount of effort exerted by the leader, as well as the amount of monitoring undertaken by the group, m, and are represented by the $C(m, e_l)$ term, which may also incorporate some fixed remuneration paid to the leader. The ability to sanction or reward leaders based on their effort is one result of the high level of information available in the small group setting. Finally, to be elected, the leader must have also paid candidacy cost ϕ.

The fact that the sanctions or rewards that leaders receive for their service depend on effort, rather than the public good value, is important. A rewards scheme based on effort is preferred when there is sufficient uncertainty in the public good value, represented in the model by η_P, which is a normally distributed independent random variable with mean zero. This variable is not observed by group members, so that the ultimate value of the public good is a weak signal of leader ability and effort. Remunerating the group leader based on this signal introduced a high level of uncertainty into their income, reducing the benefits of remuneration to risk averse individuals. High uncertainty in the value of the public good fits the empirical setting well, where the

[21] This candidacy cost may be monetary or social. Previous studies, such as Caselli and Morelli (2004), have highlighted the role of social rewards, or "ego rents", in rewarding politicians for holding office. It seems likely that there are similar social costs and benefits related to winning or losing elections.

prices negotiated by the leader depend on external forces – such as volatile world coffee prices, the exchange rate, or changes in the structure of local competition – and members have great difficulty in obtaining information about these market conditions. In contrast, effort is more accurately observed in the small group setting. These considerations underline my decision to focus on cases in which remuneration is based on leader's effort, rather than the overall public good value. I discuss this issue more formally in the Appendix. The η_I term represents similar uncertainty in private income. This variable is independent with mean zero, and has the same probability distribution function as η_P.

The functions $I(A_i, 1 - e_i, \eta_I)$ and $P(A_i, e_i, \eta_P)$ are increasing in the effort and ability arguments, concave in the $1 - e_i$ and e_i terms, respectively, and twice differentiable in the ability and effort terms. The random terms η_I and η_P can be thought of as additive noise induced by external forces, so for example, $I(A_i, 1 - e_i, \eta_I) = I(A_i, 1 - e_i) + \eta_I$. When no member chooses to become a candidate, and no leader is elected, the public good value is $P(0, 0, \eta_P) = \eta_P$. I assume that Inada conditions hold in both private income generation and public goods consumption as $1 - e_i \to 0$ and $e_i \to 0$, respectively, and that there is a complementarity between ability and effort in either task: $\partial^2 I(A_i, 1 - e_i, \eta_I)/\partial A_i \partial(1 - e_i) > 0$ and $\partial^2 P(A_i, e_i, \eta_P)/\partial A_i \partial e_i > 0$[22].

Monitoring institutions are necessary since effort is not perfectly observable by the group, and cannot be inferred based on the public good value due to the unobserved random noise term η_P. Monitoring can be thought of as a mechanism that detects whether the leader is failing to perfectly perform some leadership tasks. The more effort exerted by the leader, the less likely it is that shirking will be detected. On the other hand, the more monitoring undertaken, the more likely it is that the group will detect some task that the leader failed to properly perform. The

[22]The complementarity between ability and effort is an important assumption. The intuition behind this formulation is that effort (including time) is the means through which ability is translated into results. For example, a highly able individual that exerts no time at a task will achieve no results, but will achieve good results if they devote an hour to the task. Thus, the product of their ability depends on the effort exerted. Similarly, a low ability individual who spends an hour on a task may make a mess of things, while a high ability individual that spends the same hour will achieve good results. So the payoff to an hour of effort also depends on ability. However, note that all of the results, except Prop. 5, continue to hold if the complementarity between ability and effort in public goods production is set to zero, as long as there is still complementarity in generating private income. This is an important point because the empirical results provide no strong evidence that ability and effort are complements in public goods production in the farmer associations I study.

monitoring mechanism is explored in more detail in the Appendix. The $C(m, e_i)$ function will be increasing and weakly convex in m, decreasing in e_i, and twice differentiable. Also, the greater is the level of monitoring, the greater is the benefit of increasing effort: $\partial^2 C(m, e_i)/\partial m \partial e_i < 0$.

The model has three stages. In the first stage, members decide whether to offer themselves as a candidate for the leadership position. Members base this decision on a comparison of payoffs from being the leader to their payoffs from being a regular group member. Next, members vote in order to choose a leader out of the pool of available candidates. In the final stage, the elected leader decides how much effort to devote to producing the public good, knowing that devoting effort to producing the public good reduces the amount of effort that can be put towards generating private income. Once the leader's effort is chosen, the values of η_I and η_P are realized, the public good is produced, members receive their payoffs, and the game ends.

Members begin the model with perfect information on the ability of other group members, the group's level of monitoring, and the availability of private income opportunities. In contrast, the amount of effort exerted by the leader is not perfectly observed. The ability of the group to assess the leader's effort level will depend on the available monitoring institutions. The values of the random variables η_I and η_P are also unobserved by group members.

To solve the model, I work backwards, starting with determining the effort that each member would exert if they are the leader. These expected effort levels are used by members to determine who to elect in the second stage, given each potential set of candidates. Moving back another step, the expected election outcomes are used in members' candidacy choices.

Leader effort

If member i is the leader, she will decide how to allocate effort between public goods production and generating private income by solving the optimization problem below. For simplicity, I will abuse notation slightly by writing the expected value of the I() and P() functions as follows: $I(A_i, 1 - e_i) = E(I(A_i, 1 - e_i, \eta_I))$ and $P(A_i, e_i) = E(P(A_i, e_i, \eta_P))$.

$$\max_{e_i} I(A_i, 1-e_i)\alpha + P(A_i, e_i)(1-\alpha) - C(m, e_i) - \phi$$

The optimal effort level, denoted e_i^*, is the solution to the first order condition[23].

$$-\frac{\partial I(A_i, 1-e_i)\alpha}{\partial e_i} + \frac{\partial P(A_i, e_i)(1-\alpha)}{\partial e_i} - \frac{\partial C(m, e_i)}{\partial e_i} = 0 \tag{3.3}$$

One implication of allowing the leader to divide effort between public goods production and generating private income is that there is the possibility that higher ability members may make worse leaders. This will occur if higher ability members, when leaders, substitute so much effort away from public goods production that the reduction in effort offsets the benefits of their ability. While this is an interesting possibility, in this chapter I will consider only situations in which high ability members are better leaders. In other words, I consider only situations in which $dP(A_i, e_i^*)/dA_i > 0$ for all possible parameter values. To do so I will make Assumption 1, which amounts to placing restrictions on the complementarity of ability and effort in generating private income relative to public goods production.

Assumption 1 *The public good value produced in equilibrium is increasing in the ability of the group leader.*

$$dP(A_i, e_i^*)/dA_i = \frac{\partial P(A_i, e_i^*)}{\partial A_i} - \frac{\partial P(A_i, e_i^*)}{\partial e_i^*}\left[\frac{\frac{\partial^2 I(A_i, 1-e_i^*)\alpha}{\partial A_i \partial e_i^*} - \frac{\partial^2 P(A_i, e_i^*)(1-\alpha)}{\partial A_i \partial e_i^*}}{\frac{\partial^2 I(A_i, 1-e_i^*)\alpha}{\partial e_i^{*2}} - \frac{\partial^2 P(A_i, e_i^*)(1-\alpha)}{\partial e_i^{*2}} - \frac{\partial^2 C(m, e_i^*)}{\partial e_i^{*2}}}\right] > 0$$

According to this expression, the change in the public good value due to the direct effect of higher ability, $\partial P(A_i, e_i^*)/\partial A_i$, is greater than the change due to the indirect effect of effort, $(\partial P(A_i, e_i^*)/\partial e_i^*)$ multiplied by the change in the leader's optimal effort level induced by the higher ability, $\partial e_i^*/\partial A_i$. There are three motivations for imposing this assumption. First, I assume that under most circumstances, higher ability members will make better leaders. The empirical evidence confirms that this is the case, at least, in the empirical setting that I investigate. Second, eliminating this additional complexity will make it easier to focus on the mechanisms that I am-

[23] An interior solution is ensured by the functional form assumptions.

most interested in. Third, this assumption is consistent with most of the existing literature on this topic, making it easier to compare my work to previous results[24].

Election

Given a set of candidates, group members will choose the leader based on the value of the public good that they are expected to produce. Because individuals know the ability of all other group members they are able to calculate the effort that each candidate would exert if elected, e_i^*, and the expected value of the public good that they would produce. Members can then rank the available candidates according to $P(A_i, e_i^*)$[25]. Each member has one vote, which must be used to vote for one candidate, if any are available. If no candidates are available, no vote takes place, and no leader is elected. I consider only strategies that are not weakly dominated[26]. In equilibrium, each member will always either vote for the candidate delivering the highest public good value or themselves (if the rewards from holding office are great). The candidate delivering the highest public good value will be elected[27].

Candidacy choice

Each member's candidacy choice will depend on a comparison between her expected utility from being the leader and her utility from not being the leader. The key trade-off is that, as the leader, the member benefits from the public good she produces, but producing the public good requires substituting effort away from generating private income.

Candidacy choice is a game played simultaneously by all group members. I will look for Nash Equilibrium solutions to this game in pure strategies. Each group member will choose between

[24] See Caselli and Morelli (2004), and Messner and Polborn (2004).

[25] This is possible given Assumption 1, which ensures that, since no two members have the same ability, and public good production is strictly increasing in ability, no two members will deliver the same public good value.

[26] This rules out weakly dominated strategies in which members vote for a candidate other than their preferred candidate, but no one has incentive to change their vote because none of them represent the decisive vote.

[27] There is the possibility that, if the rewards from holding office embodied by the $C(m, e_i)$ function are very large, then all members may choose to run and vote for themselves. In this case there will be a tie vote, and I assume that the members must vote again for a different candidate, at which point the best available candidate will be elected.

two strategies: {*Run, Not Run*}. We will see that, under most circumstances, multiple equilibria exist. This occurs because higher ability leaders (those delivering higher public good values) may choose Run if they believe that lower ability members will choose Not Run, in which case it is optimal for low ability members to choose Not Run. On the other hand, lower ability members may choose Run if they believe that higher ability members will choose Not Run. This will occur if higher ability members prefer to free ride on a lower quality leader rather than to run themselves. The following four conditions are necessary and sufficient for equilibrium existence.

EC 1 *There is at most one member who chooses Run in each equilibrium.*

This condition must hold because no member i would choose Run, given that another member j with $P(A_j, e_j^*) > P(A_i, e_i^*)$ also chooses Run, since member i would never be elected under these conditions but would still have to pay the cost of candidacy.

EC 2 *If a member i chooses Run, that member must have a non-negative payoff from choosing Run relative to a situation in which no leader is chosen, i.e.,* $CP_i \geq 0$ *where,*

$$CP_i(A_i, \alpha, m) = I(A_i, 1 - e_i^*)\alpha + P(A_i, e_i^*)(1 - \alpha) - C(m, e_i^*) - \phi - I(A_i, 1)\alpha \tag{3.4}$$

This must hold because member i will never choose Run if she would be better off with no public good.

EC 3 *If some member i chooses Run, then no other member j, who would deliver a higher public good value than i (*$P(A_j, e_j^*) > P(A_i, e_i^*)$*), has a positive payoff from choosing Run given that member i chooses Run. I.e,* $CP_j - P(A_i, e_i^*) \leq 0$ *where* CP_j *is as in Equation 3.4.*

This must hold because, in an equilibrium in which i chooses Run, it cannot pay for a better potential leader j to also prefer Run, or else j would run, and i would not.

EC 4 *If no member chooses Run, then it must be the case that no member has a positive payoff from choosing Run relative to a case in which no one runs, i.e.,* $CP_i \leq 0$ *for all i.*

If no member chooses Run, then it must be the case that no member has a positive payoff from choosing Run relative to a situation in which no leader is elected and no public good is produced.

Candidacy Incentives

The relationship between a member's ability and his incentive to become a candidate is central to the model. To address this relationship, I first provide a more rigorous definition of what I mean by candidacy incentives.

Def. 1 *High ability members have* **greater candidacy incentives** *relative to low ability members when* $dCP_i/dA_i > 0$. *Low ability members have relatively greater candidacy incentives when* $dCP_i/dA_i < 0$.

Candidacy incentives are driven by a trade-off, faced by leaders, between having less time to spend producing private income, and producing and benefiting from a higher value public good. Low ability members will have greater candidacy incentives if the benefits of being the leader fall, for higher ability members, because the higher public good value they produce does not compensate them for the foregone private income. This is the case in Caselli and Morelli (2004), where the benefit that leaders derive from the public good they produce is set to zero, so low ability candidates will always have greater candidacy incentives. However, in the smaller group setting considered here, leaders benefit from the public good they produce, which opens up the possibility that higher ability individuals may have greater candidacy incentives[28]. In the upcoming analysis, I clearly separate results which hold only when low ability members have greater candidacy incentives relative to high ability members, which I will call Condition 1.

Condition 1 *High ability members have less incentive to be the leader than low ability members, i.e.,* $dCP_i/dA_i < 0$.

This condition plays two roles in the upcoming analysis. First, I examine how the model behaves when Condition 1 holds. Second, I will look for parameter values under which Condition 1 holds.

[28] See also Messner and Polborn (2004).

The Model's Predictions

Next I derive the predictions of the model which will later be taken to the data. I first consider how the leader's effort is affected by the parameters of the model, then consider how the parameters work through member's candidacy decision to affect the ability of the elected leader. Lastly, I consider how the sum of these effects determines the value of the public good produced.

Discipline effect

I begin by showing the discipline effect, i.e., that holding the identity of the leader constant, an increase in monitoring increases the leader's optimal effort level and therefore the value of the public good. It can also be shown that an increase in private income opportunities reduces the leader's optimal effort level.

Proposition 1 *Holding the identity of the leader constant, the amount of effort allocated to producing the public good is increasing in the level of monitoring, m, and decreasing in the level of private income opportunities, α. I.e., $de_i^*/dm > 0$ and $de_i^*/d\alpha < 0$.*

This proposition can be easily derived by applying the implicit function theorem to Equation 3.3.

Self-selection effect

Here I present results that describe how high private income opportunities and monitoring can work together to cause high ability individuals to self-select out of candidacy. The argument is divided into three propositions. To begin, I show that an increase in monitoring reduces a group member's payoff from choosing Run, and can lead him to always prefer Not Run in equilibrium.

Proposition 2 *Consider an equilibrium with monitoring level m in which member i chooses Run, which implies $CP_i(A_i, \alpha, m) > 0$. There exists a cutoff monitoring level $\bar{m}_i > m$ for member i such that $CP_i(A_i, \alpha, \bar{m}) = 0$. For any $m' > \bar{m}_i$, $CP_i(A_i, \alpha, m') < 0$ and member i does not choose Run in equilibrium.*

The intuition is that an increase in monitoring increases the leader's expected sanctions (or decreases the expected rewards), thus reducing the attractiveness of holding office. Thus, for each member, there will exist some monitoring level $\bar{m}_i$ at which she is indifferent between choosing Run and Not Run given that no other member runs, and for any monitoring level greater than $\bar{m}_i$, she will choose Not Run. A formal proof is available in Appendix 3.A. Next, I show that when Condition 1 holds, the cutoff monitoring level $\bar{m}_i$ is lower for higher ability members.

Proposition 3 *Suppose that Condition 1 holds, so that low ability members have greater candidacy incentives, and* $A_i > A_j$. *Then* $\bar{m}_i < \bar{m}_j$.

Under Condition 1, a higher ability member will always have lower candidacy incentives. This implies that $CP_i(A_i, \alpha, m) < CP_j(A_j, \alpha, m)$ when $A_i > A_j$. Thus, individual i will become indifferent between Run and Not Run given that no other member runs ($CP_i(A_i, \alpha, m) = 0$) at a lower monitoring level than individual j. A formal proof is available in Appendix 3.A. Given the results above, it is important to know the parameter values under which low ability members have greater candidacy incentives (Condition 1 holds). These will be the conditions under which, in equilibrium, higher monitoring levels will cause high ability individuals to opt out of the candidate pool before lower ability individuals. The following proposition shows that Condition 1 holds for high levels of private income opportunities.

Proposition 4 *There exists a level of private income opportunities* $\bar{\alpha} < 1$ *such that for all* $\alpha > \bar{\alpha}$, *Condition 1 holds, i.e.,* $dCP_i(A_i, \alpha, m)/dA_i < 0$.

The intuition here is that an increase in private income opportunities decreases the candidacy incentives of high ability individuals more, because the private income gains that they forgo if they become the leader are larger than for a lower ability individual due to the complementarity between effort and ability. A formal proof is available in Appendix 3.A. Putting Propositions 2 - 4 together, Corollary 1 is obtained.

Corollary 1 *When private income opportunities are sufficiently high ($\alpha \geq \bar{\alpha}$), lower ability individuals will have relatively greater candidacy incentives (Condition 1 will hold). When Condition 1 holds, high monitoring levels will cause high ability individuals to choose Not Run in equilibrium.*

Corollary 1 is the main empirical result related to the self-selection effect. It shows that private income opportunities and monitoring can work together to drive high ability individuals out of the candidate pool. It is this three-way relationship that is taken to the data in Section 3.6. Finally, I derive one more result showing that, when there are sufficient private income opportunities, higher ability leaders exert less effort, implying a negative correlation between leader ability and leader effort.

Proposition 5 *For every monitoring level m, there exists a level of private income opportunities $\alpha^* < 1$ such that for all $\alpha > \alpha^*$, leader effort is decreasing in leader ability, i.e., $de_l^*/dA_l < 0$. Similarly, there exists a level of private income opportunities $\hat{\alpha}$ such that for all $\alpha < \hat{\alpha}$, $de_l^*/dA_l > 0$.*

The intuition here is that the complementarity between effort and ability causes high ability individuals to concentrate more of their effort on tasks where the returns are higher. Thus, as private income opportunities grow, high ability leaders will substitute effort away from public good production more rapidly than low ability members. On the other hand, when private income opportunities are low, high ability individuals will substitute more effort towards public goods production. A proof is available in Appendix 3.A.

Combined effects

This section explores how changes in monitoring levels affect the public good output when both the discipline and self-selection effects are operating. Simulation results – rather than analytical results – are used because, with a finite number of group members, the relationship between monitoring and the public good value will not be smooth, a feature that makes deriving analytical results difficult. Results are generated by assuming an initial distribution of abilities from which

the ability of N group members are drawn at random. The model is then used to derive the candidate pool, identify the leader, and calculate the public good value obtained by each group. Repeating this procedure many times for each set of parameter values, allows us to discern how changes in parameter values affect the outcomes of the model. Simulations are run on groups with 10 members who's abilities are drawn form a uniform [0,1] distribution. Results are obtained by repeating the exercise 200 times for each set of parameter values. When there are multiple equilibria, I focus on the equilibrium delivering the highest possible public good value[29].

Particular functional forms and parameter values were selected such that they are consistent with the model's assumptions and to allow displaying a range of possible scenarios. For example, functional forms and parameter values were chosen such that at low values of m and α the incentives for individuals to be the leader are high and the candidate pool is large, while at high values of m and α there are few incentives for individuals to be the leader and the candidate pool is small. This ensures that the simulations cover the range of interesting scenarios. The details of the functional forms and parameter values used are presented in Appendix 3.D.

Figure 3.1 shows the average levels of leader effort (left) and ability (right) as a function of monitoring for various levels of private income opportunities. The left panel demonstrates the discipline effect: an increase in monitoring increases the amount of effort exerted by the group leader. The right panel demonstrates the self-selection effect: as monitoring increases, the expected ability of the leader decreases. This effect binds earlier when there are more private income opportunities. Additional results, available in Appendix 3.D, suggest that increasing m also reduces the ability rank, relative to other group members, of the leader.

Figure 3.2 shows the result of these combined effects on the value of the public good. There is a clear inverted U-shaped relationship present for higher levels of private income opportunities. In these cases, the discipline effect dominates at lower monitoring levels and the self-selection effect dominates at higher levels. The higher are private income opportunities, the earlier this inflection point is reached. However, at low levels of private income opportunities, the self-selection effect

[29]Other papers in this literature also focus on the best equilibrium. See, e.g., Bernheim and Kartik (2010). Similar results are obtained by taking the averages over all possible equilibria (see Appendix 3.D).

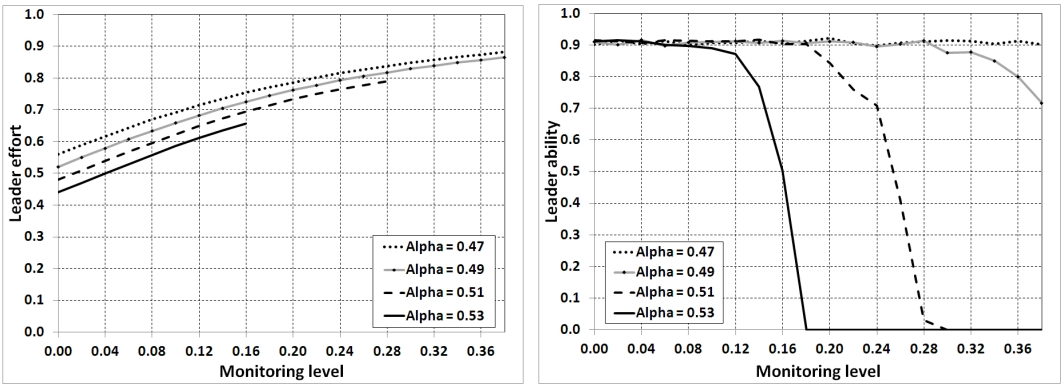

Figure 3.1: Simulated Leader Effort and Ability

disappears, and thus, there is no inverted U-shape. In this case, high ability members prefer to run and they exert a high level of effort once elected. I now move to compare the model's predictions to real-world data from one set of small self-help groups.

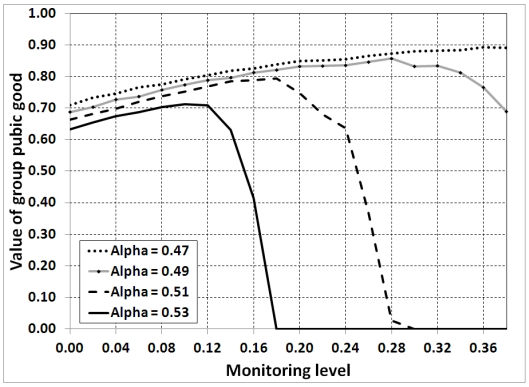

Figure 3.2: Simulated Public Good Values

3.3 Empirical Setting

Farmer associations' raison d'étre is to improve the performance of their members' farms as economic units engaged in market transactions. Relatively small and self-governed, farmer associations provide members, who join voluntarily, with a group public good. The preferences of

members of farmer groups with respect to their group's services — the most important of which is securing higher outputs prices through collective marketing — are closely aligned[30]. Because of the high costs of transportation and of market information, dispersed small-holder farmers across the developing world have little options but to sell to trading partners who are able to exploit asymmetries in information and in bargaining powers. In contrast, organized farmers who sell their cash crops via their association (in bulk), can obtain higher prices by increasing their bargaining powers and by reducing buyers' transaction costs (Staatz, 1987). Also, members of farmer associations, which cover relatively small geographical units, tend to have a high level of information about other members (Grossman, 2011*c*). These factors are consistent with the small group features described in the theory.

APEP: The development project

All the farmer associations I surveyed were created as part of one of Uganda's largest recent development projects: the Agriculture Productivity Enhancement Project (APEP)[31]. APEP's goal was to support small-holder farmers' transition into commercial farming. Between 2004 and 2008 it helped organize over 60,000 small producers into more than 2,500 village-level farmer groups, known as producer organizations (POs), which were further organized into more than 200 farmer associations across Uganda, known as Depot Committees (DCs). Serving, on average, 200 members from ten neighboring village-level groups, the farmer associations were designed to exploit economies of scale and to bargain for better prices based on quality and volume.

Studying the APEP groups presents several advantages. First, the project's scope and size allow me to conduct a large-scale quantitative study within the boundaries of a single nation, thus securing the homogeneity of the political and legal environments, as well as many project-related factors. Second, APEP groups have similar governance structures and leadership positions whose roles and functions are comparable across sites. Each farmer association is led by

[30] Other services that farmer groups may provide include securing lower input prices and training in agriculture.

[31] APEP was funded by USAID, and implemented by Chemonics International, an International Development consulting firm based in Washington DC.

an executive committee (management team), which is part of the DC board of directors. The executive committee is comprised of a DC manager, the chairman of the board, a secretary and a treasurer. Operationally, the DC manager is the principal leader of the association (henceforth, leader). The leader's most important responsibilities include searching for buyers, negotiating input and output prices, and organizing the collection of crops (including hiring and supervising employees). Additionally, leaders help coordinate group activities and facilitate the flow of information throughout the association.

The association leader is chosen from the set of board directors. In other words, the board directors — also referred to as representatives — form the pool of potential candidates for the senior leadership position[32]. Responsibilities of the board of directors, which is comprised of two elected representatives from each of the village-level POs, include monitoring the work of the DC executives (including the leader), representing the opinions of PO members at the associational level, and aiding in the transmission of information. Thus, there exist three types of members: ordinary group members, DC representatives, and the DC leader[33].

Third, while all of the associations adopted governance structures that are generally similar, there exists a significant amount of variation in the monitoring structures adopted by the groups. Early on in the intervention, APEP's management reached the conclusion that village-level POs were too small to be viable economic players. As a result, APEP strongly encouraged neighboring POs to form a single association with a federal structure. A uniform template was then used by APEP field-officers to facilitate the process of association formation[34]. The identity of the facilitator assigned to each association, the information group members had about each other prior to the association formation, and representatives' previous experience in participating in other self-help groups, generated variation in the monitoring institutions that groups adopted[35]. Once

[32] At the village-level, each producer organization (PO) elects its own local leadership which is responsible for implementing decisions taken by the DC leadership.

[33] A graphical representation of the association governance structure is provided in Appendix A, Fig. A.2.

[34] The process of establishing the DC governance structures took place in three workshops, led by APEP facilitators. These workshops took place between 12 to 24 months after creation of the village-level POs. For additional information see Chapter 4, Section4.4. The group formation manual can be provided by the author upon request.

[35] Groups that were created by the same facilitator share more similar structures and rules than groups created by different facilitators. For a more detailed description and rigorous assessment of the exogeneity of the associations'

established, groups, by and large, retained these institutions, usually enshrined in constitutions[36]. Leaders' compensation can serve as a good example for the resilience of the DCs' governance institutions. When established, APEP facilitators strongly encouraged new groups to keep monetary remuneration to leaders as low as possible. The data confirms that 3-5 years after their creation, only one association paid its manager any regular salary.

3.4 Data and sampling scheme

This section briefly describes the data used in this chapter and how it was collected. To reduce crop-related variability, I first limited the study's target population to associations marketing coffee, the most common cash crop sold by the APEP groups. I then sampled 50 associations from 5 district-areas (strata) using a stratified, random, multistage cluster design[37].

Quantitative data for the empirical analysis was collected between July and September 2009 by a team of 60 local interviewers. Within each association, several different types of data were collected. One key data set is based on individual-level surveys of ordinary group members. For each association, I sampled six POs, for a total of 287 village-level groups[38]. From each sampled PO, I surveyed, on average, six members, for a total of 36 members per association[39]. Sampled members were surveyed in person by trained interviewers in the respondents' local language, for a total of 1,781 surveys. I refer to this data source as the "members' survey".

A second key data set is based on individual-level surveys of *all DC representatives*, for a total of 1,316 interviews. These surveys cover the complete pool of potential candidates for the DC leadership position. These "representatives' surveys" only partially overlap with the "members' surveys", as they were tailored to capture the representatives' roles and responsibilities within

level of monitoring, see Grossman (2011*c*).

[36]For example, the vast majority of the constitutions I examined had both quorum and super-majority rules for making constitutional amendments.

[37]Additional information on the sampling scheme, including a map locating DCs in the five strata, can be found in Appendix A, Section A.2.

[38]In few cases, when a farmer association had fewer than seven groups, I selected all of its POs.

[39]The number of sampled members from each of the six *sampled* village-level groups was proportional to the size of the PO, to ensure that the sample is self-weighted.

the association governance structure. I collected additional data at the associational-level using questionnaires completed by the DC senior leadership, and at the PO-level using questionnaires completed by the PO leaders. Data on the DCs' economic activities were also collected from the associations' books.

Measurement of key variables

In this section I will walk through the information and procedures employed to construct each of the variables used in the subsequent analysis. One advantage of the farmer associations I study is that one of the key variables in the model, the value of the public good, is relatively straightforward to measure. Since farmer associations' central activity is collective marketing, it is reasonable to relate the value of the public goods directly to the marketing decisions of members. A high value public good exists when members sell a large fraction of their crops via their farmer association[40]. In the analysis, I use two measures of member's marketing decision to proxy the value the public good: (i) an indicator variable capturing whether a member sold his crops via the association, at least once, in the past season, and (ii) the share of a member's total seasonal coffee yield that was sold via the farmer group in the past season[41].

Measuring the remaining key variables in the model is a more complex task. For some of the variables — e.g., members' ability, leader effort and group monitoring — a number of questions were asked related to different aspects of the variable, and responses were collected from a variety of sources. These values were then collapsed into single measures using principal component analysis[42]. A similar approach was recently used by Bjorkman and Svensson (2009) to construct a single composite index of monitoring using six monitoring proxy variables.

[40] I use individual marketing decisions rather than crop prices, since the former is a much less noisy signal of the performance of the leader. This is because coffee prices depend, to large extent, on exogenous factors.

[41] I check my results against two alternative measures of collective marketing, gathered from interviews with group leaders and the associations' books and records. These alternative measures deliver similar results.

[42] This technique is used to re-express multiple variables as one (or more) variables that explain as much of the variation in the original variables as possible. In technical terms, the first principal component of a set of variables $X_1, X_2, \ldots X_n$ is the linear combination of these variables that exhibits maximum variance. A good source for more information about principal component analysis is Lattin, Carroll and Green (2003). For an example of the use of principal component analysis in practice see McKinzie (2005).

To construct individual's ability measure I used information on respondents' literacy level, educational attainment and English proficiency[43]. Respondents also completed two cognitive tests[44]. I combined these variables into a single summary index using principal component analysis. All of the above variables are positively correlated, and the first principal component was able to explain more than 61% of the variance. Several checks increase our confidence in the composite ability measure. First, members who hold high-skilled off-farm jobs have significantly higher ability scores than those who do not (Appendix A, Fig. A.5, Panels A-C). Second, (mean) ability scores are increasing with the leadership role in the association (Appendix A, Fig. A.5, Panel D). Third, the composite ability measure is highly correlated with wealth (Appendix A, Fig. A.6).

To measure the effort leaders spend producing the public good, I combined effort ratings from sampled members and from the DC representatives. I also used information on the number of times the leader organized collective marketing in the past season – the associations' most important activity. All of these variables were positively and highly correlated, with the first principal component explaining 45% of the variables' variance. To check the reasonableness of the effort measure, I compare it with the leader's self-assessment of their effort. I find that leaders who have high effort scores also report working longer hours and have a better sense of whether members are following the association's rules and by-laws.

Table 3.1 provides information on the variables I used to measure groups' monitoring level. The variables, which are derived from both the members and the representatives' surveys, were combined, using principal component analysis, to obtain a single summary index of monitoring measure for each DC. All of the variables are positively correlated, often strongly, with the first principal component explaining more than 46% of the variance.

The availability of private income opportunities – a group-level variable – also plays a key role in the model. I used respondents' employment status to construct this measure[45]. The primary

[43]Because no local language is spoken by more than 20% of Ugandans, English is the lingua franca of the business and political class. English proficiency allows individuals to communicate with potential trading partners outside their small geographic areas.

[44]The cognitive tests included two assignnments: solving a simple maze in less than two minutes and solving a raven test comprised of 12 questions in two minutes.

[45]In the model, α measures the value of private income opportunities in relation to the potential value of the public

Behavioral Measures	Type	Survey
Whether there is a rule regarding the manager's time commitment	dummy	Reps
The extent to which external auditors are used	categorical	Reps
Whether respondents have asked to review the DC's books	dummy	Reps
Whether someone is responsible for monitoring the DC manager	dummy	Reps
Attitudinal/ Perceptional Measures		
Whether the manager is transparent	dummy	Reps
The extent to which the manager is accountable	categorical	Members
Whether the manager is monitored	dummy	Members
The extent to which the manager is transparent	categorical	Members

Table 3.1: Variables used to measure group monitoring level (m_j)

measure of a group's private income opportunities is the fraction of representatives in the association holding off-farm jobs (except as unskilled laborers). Note that only data from the complete sample of board directors were used to construct this measure, since this is the relevant candidate pool. Alternative measures were also tested and produced similar results.

In order to test whether higher values of the group public good have positive welfare effects, I construct a measure of the change in a member's welfare since joining his or her farmer group. The measure was constructed using questions about ownership of 12 different assets that reflect the purchasing power of farmers, such as bicycles and livestock[46]. For each asset type, respondents were asked to provide information on the number of items they currently have and the number of items they had in the year before joining the group. Measurement errors, typical in survey-based recall questions, are reduced given that (i) the median member joined her group merely three years ago, (ii) the creation of the farmer group is considered a major milestone to the majority of members, and (iii) the included assets are central to households in rural Uganda.

good. Because I am not able to accurately measure the *potential* value of the public good, I proxy (α) through a measure of private income opportunities.

[46]Using asset ownership to measure the welfare of poor households is a commonly used technique in poor developing countries where monetary measures of income and wealth are problematic. See, e.g., Filmer and Pritchett (2001).

3.5 Case studies: Iganga district vs. Masaka district

This section describes the experience of farmer associations from two district-areas, Masaka and Iganga, before I test the model predictions more rigorously in Section 3.6. Masaka is a relatively well-off district compared to Iganga; group members in Masaka are more educated, have more available land, and are wealthier. Whereas the median member of a farmer association in Iganga grows coffee on less than one acre, producing a seasonal median yield of 250 KGs, the median group member in Masaka grows coffee on 1.5 acres, yielding 363 KGs per season. Given these facts, one might expect that the farmer associations in Masaka would function more effectively than in Iganga. Yet, I find that farmer groups in Iganga have higher levels of public goods production. For example, 85% of the members of the farmer associations in Iganga report selling at least once via their association in the past season, compared with 49% of the members of groups from Masaka (Appendix A, Fig. A.3). In addition, members in Iganga sell 69% of their seasonal yield via their farmer groups. In Masaka the figure is 31% (Appendix A, Fig. A.4).

The model suggests that variation in the value of the public good depend on leaders' ability and effort. I find that the mean ability scores of DC managers in Iganga is 0.61 standard deviations above the mean ability scores of DC managers in Masaka (*p-value*=0.09)[47]. Furthermore, when I examine the entire network of DC representatives, the mean ability of Iganga board directors is, on average, 0.29 standard deviations higher than those in Masaka (*p-value*$= 0.00$). This contrast sharply with the findings from the members' survey, in which the average *member* in Masaka is significantly more educated than members in Iganga (*p-value*$= 0.06$). In other words, high-ability group members in Masaka appear to be less willing to take on leadership positions than those in Iganga. Turning to the leader's effort, I find that the average effort that leaders in Masaka exert, e_i, is 2 standard deviations *lower* than the effort exerted by Igangan leaders. This occurs even though, in Masaka, the mean monitoring level is 1.5 standard deviations *higher* than in Iganga.

Given that leaders in Iganga have higher ability and spend more time in producing the public good, it is not surprising that farmer groups in that district are more effective. The question then,

[47] P-values in this section are based on a one-tail, two-group, mean-comparison test with unequal variances.

is why groups in Masaka, which are endowed with better educated members on average, end up with lower quality leadership than groups in Iganga. The model provides an explanation to this result, by pointing to the important role of private income opportunities. According to the model, in areas that have high private income opportunities, both the discipline and self-selection effects are present, and groups face a trade-off between leaders' effort and ability. This causes higher-ability members to drop-out of the candidacy pool at relatively lower levels of monitoring, and causes elected leaders to exert less effort. In contrast, when private income opportunities are low, my framework suggests that this trade-off is not present. Under these conditions, high ability members will choose to be candidates and will exert high levels of effort if they are the leader, *even without the incentives created by enhanced monitoring technology*.

Turning to the data, I find that groups in Masaka have, on average, the highest level of private income opportunities, with a mean across DCs that is 0.66 standard deviations above the mean for the entire sample. Groups in Iganga, on the other hand, have the lowest mean score of private income opportunities for representatives, at 0.95 standard deviations below the mean for the entire sample. This difference is both substantial and significant (*p-value*$= 0.01$). These findings support the predictions of the model. In particular, they suggest that, because of high local private income opportunities, groups in Masaka are forced to trade-off between leader effort and ability. In contrast, the low levels of local private income opportunities in Iganga allow groups do avoid this trade-off, and therefore obtain leaders with high ability, who are also willing to exert more effort in public goods production, resulting in a higher public good value. In the next section I test the model's predictions on the entire sample.

3.6 Empirical analysis

In this section I look for evidence of the main predictions of the theoretical model: the discipline effect, the self-selection effect, and the existence of an inverted U-shaped relationship between group monitoring level and the value of the public good produced.

Discipline effect

The model makes two key predictions with respect to the manager's effort. The first prediction – formalized in Theorem 1 – is that an increase in the level of monitoring (m_j), increases the amount of effort exerted by the group leader (e_j). The second prediction is that, when private income opportunities (α_j) are sufficiently high, high-ability leaders choose to exert less effort, all else equal (Theorem 5). To explore these predictions I regress the standardized score of the leader's effort of group j on the group's level of monitoring, private income opportunities, the ability of the association's leader (A_j) and the interactions between the key variables. OLS regression results are shown in Table 3.2. The first specification includes only group monitoring level and strata fixed effects, while the second model adds DC-level controls for the age and size of the association, the mean size of members' seasonal coffee yield, the association's density of associational life and measures of ethnic and religious fractionalization among the association's board directors[48]. The third model adds a variable measuring private income opportunities and its interaction with monitoring, and the fourth specification adds the ability of the associations' manager and its interaction with private income opportunities and group monitoring.

In accordance with the discipline effect, I find a positive, substantial, and significant relationship between groups' level of monitoring and the amount of effort exerted by leaders. Turning to the second prediction, tested in model D, I find some evidence that the relation between manager's ability and effort is conditional on the level of outside income opportunities (Appendix, Fig. 3.8). When there are few private income opportunities (two standard deviation below the grand mean), one unit change in leader ability is associated with a 0.40 standard deviation increase in leader effort. In contrast, when private income opportunities are available (two standard deviations above the grand mean), a one unit increase in leader ability is associated with a 0.11 standard deviation *decrease* in leader effort, as predicted in Theorem 5, though these results are not statistically significant at conventional levels.

[48]Ethnic heterogeneity makes solving collective action problems more challenging, in Uganda (Björkman and Svensson, 2010, Habyarimana et al., 2007). Ethnic fractionalization index was constructed using a Herfindahl concentration index: $ELF = 1 - \sum_{i=1}^{n} s_j^2$ where s_j is the share of group j, and $(j = 1 \ldots n)$. To measure the density of associational-life I calculate the group mean of the number of voluntary associations in which DC representatives are regular participants.

DV: LEADER'S REALIZED EFFORT IN PUBLIC GOODS PRODUCTION

	Model A	Model B	Model C	Model D
Monitoring (std.)	0.772***	0.709***	0.731***	0.690***
	(0.118)	(0.121)	(0.129)	(0.235)
Monitoring × PIO			-0.079	0.058
			(0.117)	(0.125)
Leader's ability (std.)				0.151
				(0.175)
Monitoring × Leader's ability				-0.092
				(0.093)
Leader's ability × PIO				-0.128
				(0.128)
Group Controls		X	X	X
Strata FEs	X	X	X	X
Observations	50	50	50	42
r^2	0.720	0.781	0.791	0.800

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: **Relation Between the association's monitoring level and the leaders' realized effort.** Results from OLS regressions using group-level data. PIO stands for private income opportunities.

An increase in the leader's effort level is only valuable if it increases the value of the group public good. A significant variation in group public good levels is observed across associations 1.1 and district-areas. For example, the share of member selling through the association in the past season varies from a low of 49% in Masaka to a high of 85% in Iganga (Appendix A, Fig. A.3). Similarly, the share of member's yield sold through the group varies from 31% in Masaka to 69% in Iganga (Appendix A, Fig. A.4). To explore how leader's effort and ability translates into the value of the public good, I run a series of logistic regressions, using the following specification:

$$y_{ij} = I[\tilde{y}_{ij} > 0]$$

$$\tilde{y}_{ij} = \beta_0 + \beta_1 A_j + \beta_2 \bar{e}_j + \beta_3 (A_j \times \bar{e}_j) + X_{ij}\Gamma_1 + J_j\Gamma_2 + \zeta_j + \epsilon_{ij} \tag{3.5}$$

Here the dependent variable y_{ij} indicates whether member i sold his coffee through his farmer association j during the past season (rather than selling it all to a local middleman). This measure is derived from the reporting of the producer organizations' leaders[49]. The key independent

[49]Similar results are obtained by using members' self-report as an alternative measure of the dependent variable.

variables are the ability of the DC leader A_j and effort e_j, and the interaction between the leader's ability and effort. In model B I add a set of individual level controls X_{ij}, and in model C I further add group-level controls J_j. All models include strata fixed effects and standard errors are clustered at the farmer association level.

Results, which are displayed below in Table 3.3, suggest that both leader's effort and ability are positively related to the value of the public good, though only the coefficient on the leader's effort is significant. Using the full specification of model C, moving from one standard deviation below the mean effort to one standard deviation above the grand mean (a range that covers 38% of associations) is associated with a 63% points increase in the predicted probability that a member sells his coffee through his farmer association (*p-value*$= 0.00$), holding the control variables at meaningful values (means or medians for categorical variables)[50].

DV: SELLING IN BULK VIA THE DC IN THE PAST SEASON

	Model A	Model B	Model C
Leader's effort (std.)	0.716***	0.752***	0.653***
	(0.191)	(0.198)	(0.189)
Leader's ability (std.)	0.124	0.129	0.090
	(0.180)	(0.178)	(0.196)
Leader's ability × Leader's effort	-0.073	-0.086	-0.013
	(0.136)	(0.128)	(0.140)
Individual Controls		X	X
Group Controls			X
Strata FEs	X	X	X
Observations	1456	1233	1180
Log Likelihood	-877.344	-719.194	-670.097

Standard errors clustered at the farmer association level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: **Leader quality and marketing decision.** Members' decision to sell via their group is modeled as a function of the leader's ability and effort. Model A includes only the key independent variables, Model B includes, in addition, a series of individual-level controls, Model C includes DC-level controls.

These results are reported in the Appendix.

[50]The signs of the control variables seem reasonable. Women and older members are more likely to sell in bulk. The probability of selling via the group is positively related to the total size of coffee yield, the time since joining the group, the dishonesty of local middlemen, and with a member's richness of associational-life. The further a member lives from the DC's crop collection point, the less likely she is to bulk. Selling in bulk is positively correlated with distance to the nearest trading center. This distance likely affects the availability of information about market prices and of buyers 'outside' the group, so as the distance to the local trading center increases, farmers are likely to have greater information asymmetry and less bargaining power. Under such conditions, farmer groups are well-positioned to increase the value of the public good for members. A full set of results can be found in Appendix, Table 3.9.

Self-selection effect

The next central prediction of the model, presented formally in Corollary 1, is that an increase in monitoring level decreases the likelihood that high ability members will be candidates (and thus the probability that they become the group leader), *but only in areas with sufficiently high private income opportunities*. In particular, the model predicts that when there are ample private income opportunities and monitoring levels are high, high-ability individuals will opt out of candidacy, so lower ability leaders will be elected.

I explore this prediction by looking at how the identity of the elected leader is affected by both monitoring levels and private income opportunities. I run an individual-level logistic regression on the complete sample of DC representatives, where the dependent variable indicates whether the individual is the group leader[51]. The key independent variables are board director's ability, group's monitoring level and private income opportunities, and the interactions between these variables. The regression equation is given below, where $\tilde{y}_{ij}$ is an unobserved latent variable that determines whether a board director becomes the DC leader, y_{ij} is an indicator variable that takes the value one if individual i in group j is the group leader, A_{ij} represents ability, α_j represents private income opportunities, m_j represents group monitoring level, X_{ij} is a set of individual level controls and F_j is a set of DC-level controls. To account for correlation in the error terms of members of the same DC, in all models I cluster standard errors at the farmer association level.

$$y_{ij} = I[\tilde{y}_{ij} > 0]$$

$$\begin{aligned} \tilde{y}_{ij} = \quad & \beta_0 + \beta_1 A_{ij} + \beta_2 \alpha_j + \beta_3 m_j + \beta_3 (A_{ij} \times \alpha_j) + \beta_4 (A_{ij} \times m_j) + \beta_5 (\alpha_j \times m_j) + \\ & \beta_6 (A_{ij} \times \alpha_j \times m_j) + X_{ij}\Gamma_1 + F_j \Gamma_2 + \epsilon_{ij} \end{aligned} \tag{3.6}$$

The main coefficient of interest in this analysis is β_6, the coefficient on the interaction be-

[51]The model's predictions are most clear for the identity of the elected leader. The model's predictions are less clear about the identity of the losing candidates.

tween ability, monitoring, and private income opportunities. Based on the model's predictions we should expect a negative coefficient, since increase in monitoring should *decrease* the probability that a higher ability individual is the leader when there are more private income opportunities.

Results obtained from the regressions are displayed in Table 3.4. These results support the main prediction of the model, as we observe a negative and significant coefficient on the interaction term between ability, monitoring level and private income opportunities. Other coefficients take reasonable values . As expected, individual's ability, wealth, health and the size of their coffee yield all have a positive influence on an individual's likelihood of being the leader. Interestingly, individuals who are active participants in other organizations, such as church groups or village committees, are significantly less likely to become farmer association leaders (see Appendix, Table 3.11). This suggests that group members face time constraints which influence their candidacy decisions, in ways that are consistent with my theory.

DV: IDENTITY OF THE ASSOCIATION'S LEADER

	Model A	Model B	Model C	Model D
Rep ability (std.)	1.128***	1.147***	0.874***	0.877***
	(0.257)	(0.207)	(0.199)	(0.203)
Monitoring	0.042	0.235	0.158	-0.006
	(0.215)	(0.150)	(0.207)	(0.222)
Rep ability × Monitoring	-0.066	-0.050	0.139	0.121
	(0.228)	(0.147)	(0.192)	(0.199)
Private Income Opportunities (std.)	0.073	0.089	-0.018	-0.140
	(0.240)	(0.141)	(0.147)	(0.154)
Rep ability×PIO	-0.218	-0.245	-0.157	-0.132
	(0.265)	(0.167)	(0.175)	(0.171)
Monitoring ×PIO	0.264	0.286***	0.142	0.150
	(0.179)	(0.088)	(0.113)	(0.117)
Rep ability×Monitoring ×PIO	-0.349*	-0.361***	-0.259**	-0.246**
	(0.191)	(0.081)	(0.105)	(0.103)
DC Controls		X	X	X
Reps Controls			X	X
Strata FEs				X
Observations	1058	1058	903	903
Log Likelihood	-163.284	-162.101	-125.116	-124.294

Standard errors clustered at the farmer association level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.4: **Test of the self-selection effect:** relationship between board director's ability, group's monitoring level and private income oportunities (PIO).

In order to get a sense of the magnitude of the relationships found in this regression, I use a graphical representation. In Figure 3.3 I plot the predicted change in the likelihood that any member of the association's "potential candidates" is the actual leader, as group's move from low to high monitoring environment and private income opportunities changes in it's entire range, while setting members' ability constant[52]. In the right panel, members' ability is set to two standard deviations above the grand mean, while in the left panel it is set to two standard deviations below the mean. As Fig. 3.3 makes clear, when members' ability is *low*, the marginal effect of the level of monitoring on the likelihood of being the group's leader is negative, when group's private income opportunities are low. However, as private income opportunities become more abundant, moving from low to a high monitoring level has a *positive* effect on the likelihood that low-ability members become leaders. This change reflects the fact that high-ability members begin dropping out of the candidacy pool. By contrast, when members' ability is *high*, the effect of change in monitoring levels on the likelihood of being the leader is positive when private income opportunities are sparse. As the model predicts, the marginal effect of monitoring turns negative once private income opportunities become abundant (about one standard deviations above the grand mean).

Inverted U-shaped relation: Monitoring and the value of the Public Good

This subsection explores the third main prediction of the model: the existence of an inverted U-shaped relationship between groups' value of the public good and monitoring. To test this prediction, I fit a locally weighted regression of the value of the public good (group-means) on the group's monitoring level[53]. Figure 3.4 graphs the results using the share of members who have sold their crops at least once in the past season via the farmer groups, as the measure of the public goods value. Though understandably crude, a rough inverted-U shape is apparent.

[52]Note that the regression coefficients represent unconditional effects, whereas marginal changes of each of the independent variables is conditional on the values of the other two variables.

[53]This approach was inspired by Urquiola and Verhoogen (2009).

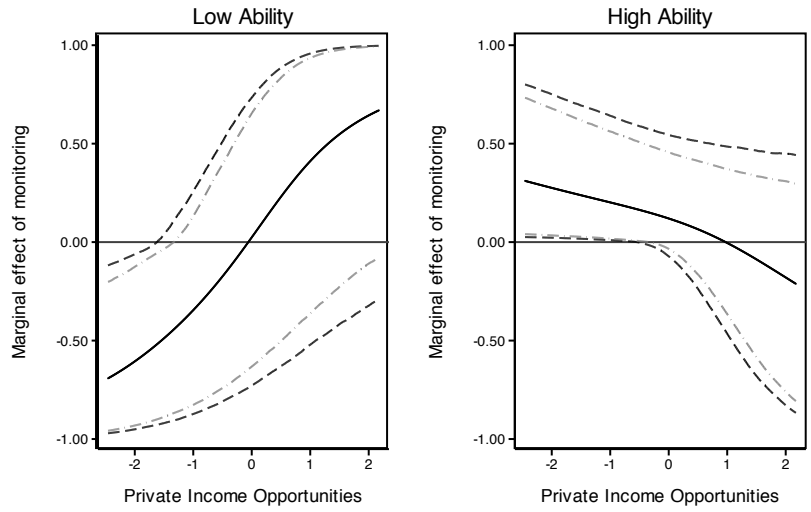

Figure 3.3: **Marginal effect of monitoring on the likelihood of being elected as leader as private income opportunities increases in it's entire range, for high-ability and low-ability representatives.** Inner and outer dashed lines represent 90% and 95% confidence intervals, based on bootstrap SEs (20,000 iterations). Marginal effect of monitoring is measured from two standard deviations below the grand mean to two standard deviations above.

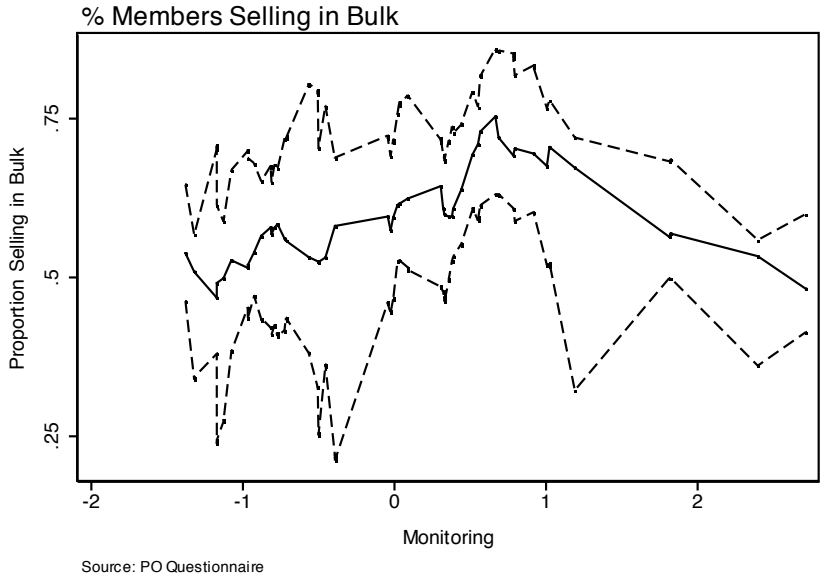

Figure 3.4: **Monitoring and the value of the Public Good.** In each panel, the lines plot fitted values of locally weighted regressions (using Stata's *lowess* command and a bandwidth of 0.3) of group-level means of the value of the public good on monitoring. 95% confidence intervals are based on bootstrap simulations with 1,000 replications.

Welfare effects

In the last piece of the empirical analysis, I look at the relationship between leadership quality, the value of the public good and members' welfare. In groups with a high public good value, where many members sell via their association, we expect a larger increase in welfare since members had joined the group.

To test this prediction, I fit three models. First, I run an individual-level OLS regression, in which a standardized measure of member's welfare change since joining the group is modeled as a function of the amount of collective marketing undertaken by the group and the member's own marketing decision. The amount of collective marketing undertaken by the group is measured as the fraction of members selling via the group in the past season, whereas individual's marketing decisions are measured as the share of coffee yield that was sold via the association. Second, I fit a reduced form OLS, in which member's welfare change is modeled as a function of the leader's ability, effort and their interaction. Third, I model the standardized measure of member's welfare change as a function of the amount of collective marketing undertaken by the group, which itself is instrumented by the leader's ability and effort[54]. All key variables were standardized to allow for easier comparison between regression coefficients. I include individual level controls for sex, age, year of joining the group, health status and level of education, as well as association level controls for group age and membership size. In all models I include strata fixed effects and cluster standard errors at the association level.

Table 3.5 presents the regression results. First, I find that increase in members' welfare since joining the farmer group is positively and significantly related to the value of the public good (Model A). Second, I find that leader effort is positively and significantly related to the value of the public good but decreasing in leader ability (Model B). I find no significant effect of leader ability on welfare. Third, I find a strong significant association between members' welfare and the value of the public good, when the latter is instrumented by leader's ability and effort. Turning

[54]The first stage of this regression is not shown here because it is nearly equivalent to the results appearing in Table 3.3. I assume that leader's effort and ability impact DC members' welfare only via collective marketing activity, which is the central activity of the association.

to an individual member's decision to sell through the association, I find that controlling for the value of the public good, the larger the share of crops a member sells via the farmer association the larger is the welfare change.

DV: CHANGE IN A MEMBER'S WELFARE SINCE JOINING THE FAMER GROUP

	A: Direct	B: Reduced	C: IV
Fraction of members selling in bulk (std.)	0.072*		0.229**
	(0.037)		(0.096)
Leader's ability (std.)		-0.020	
		(0.040)	
Leader's effort (std.)		0.079*	
		(0.041)	
Leader's ability × Leader's effort		-0.056**	
		(0.023)	
Share of Yield sold in bulk (std.)	0.117***	0.100***	0.063*
	(0.028)	(0.030)	(0.036)
DC Controls	X	X	X
Members Controls	X	X	X
Strata FEs	X	X	X
Observations	1448	1229	1229
r^2	0.113	0.123	0.102
Log Likelihood	-1908.469	-1569.030	-1583.633

Standard errors clustered at the farmer association level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: **Relation Between Value of the PG and Welfare.** In model A, the standardized dependent variable is a function of the value of the public goods measured as the fraction of members selling in bulk via the assocaition. In model B, the dependent variable is a function of leader's ability and effort, whereas in the model C, the key independent variable – the value of the public good – is instrumented by the leader's effort and ability.

3.7 Conclusions

I have argued in this chapter that self-help groups may face a trade-off between the ability of their leader and the amount of effort the leader exerts, with consequences for the quality of leadership obtained, the value of the public good produced, and the welfare of group members. To a large degree, this trade-off is driven by the group's monitoring institutions. High levels of monitoring incentivize leaders to exert more effort, but may also drive higher ability members to self-select out of the candidate pool. Whether this trade-off exists depends on the availability of private income opportunities. Evidence from a sample of Ugandan farmer associations support these predictions.

One surprising implication of the results found in this chapter is that more and better monitoring institutions do not necessarily lead to better outcomes. My results suggest that introducing new monitoring institutions may result in higher ability group members opting out of candidacy through the self-selection effect, with negative consequences for the value of the public good. The theory suggests that these negative effects can be avoided, however, if increased monitoring of the group leaders is balanced by an appropriate increase in leader rewards.

From a policy perspective, this study suggests that the quality of group leaders should be considered, at least partially, to be endogenous. Thus, small groups should be structured in ways that take into account how these structures will affect the quality of leadership obtained. My results also suggest that small group structures should take into account the private income opportunities available to group members. For example, it may not be appropriate to apply broad standards to small groups that exist in different economic settings, as was done when APEP recommended that farmer associations not offer monetary remuneration to leaders regardless of local economic conditions. This also has implications for interventions that affect the set of private income opportunities available to group members. According to the theory presented in this paper, an intervention that increases private income opportunities may have unintended effects on the success of small groups operating in that area. Ideally one should take into account these externalities when assessing the costs and benefits of development interventions.

My results tend to support the predictions of existing models designed for larger political units. They also demonstrate, however, that previous results must be modified in order to accommodate the small group setting. On the other hand, my results may also help inform our understanding of leadership selection dynamics in larger political units. For example, Kotakorpi and Poutvaara (2010) have speculated that differences in available private income opportunities may explain their finding of a self-selection effect for women, but not men. I present evidence that this intuition is correct, at least for the case of small groups. Thus, while small groups are worthy of study in themselves, research on small groups can inform, and be informed by, research on larger political units.

There remain a number of outstanding questions related to leadership quality in small groups. One interesting set of questions centers on the ability of self-help groups to adjust the costs and rewards of holding office in order to obtain better leadership. In the case I study, groups were constrained by institutional structures imposed when the groups were formed, a time when they likely lacked the information and experience to choose optimal monitoring institutions. The theory suggests that there can be substantial gains if groups are able to more easily adjust monitoring institutions (and leader remuneration) in order to obtain better leaders. However, in practice we often observe institutional structures that are rigid or only change slowly over time. One reason may be that allowing flexible institutional structures increases the ability of incumbent leaders to make changes to benefit themselves. Understanding the trade-offs surrounding the flexibility of institutional structures is likely to be an interesting avenue for future research.

Appendix

Appendix 3.A Proofs

Proof of Prop. 2

Two conditions must hold to prove this Proposition[55]. First, we need. $dCP_i(A_i, \alpha, m)/dm < 0$. Taking this derivative, we have the following.

$$\frac{dCP_i(A_i, \alpha, m)}{dm} = -\frac{\partial C(m, e_i^*)}{\partial m} < 0$$

Second, we need $\lim_{m \to +\infty} CP_i(A_i, \alpha, m) < 0$. This must hold given the result above and the weak convexity of $C(m, e)$ with respect to m. These two results, together with the continuity of the $CP_i(A_i, \alpha, m)$ function imply that as m increases eventually a monitoring level $\bar{m}_i$ is reached at which $CP_i(A_i, \alpha, \bar{m}_i) = 0$ and any further increases in m result in $CP_i(A_i, \alpha, m) < 0$.

Proof of Prop. 3

We have two members with $A_i > A_j$. Under Condition 1, this implies that $CP_i(A_i, \alpha, m) < CP_j(A_j, \alpha, m)$ Suppose that $m = \bar{m}_i$ is such that $CP_i(A_i, \alpha, m) = 0$. This implies that $CP_j(A_j, \alpha, m) > 0$. Given Prop. 2, it must be the case that $\bar{m}_i < \bar{m}_j$.

Proof of Prop. 4

This proof requires two steps. First, we need to show that dCP_i/dA_i is decreasing in α. To do this, we show that $d^2CP_i/dA_i d\alpha < 0$. The second step involves showing that there exists some $\bar{\alpha} \in (0, 1)$ such that $dCP_i(A_i, \bar{\alpha}, m)/dA_i \leq 0$ for all m.

Step 1

$$\frac{dCP_i}{dA_i} = \frac{\partial I(A_i, 1 - e_i^*)\alpha}{\partial A_i} + \frac{\partial P(A_i, e_i^*)(1-\alpha)}{\partial A_i} - \frac{\partial I(A_i, 1)\alpha}{\partial A_i} + \left(\frac{\partial I(A_i, 1 - e_i^*)\alpha}{\partial e_i^*} + \frac{\partial P(A_i, e_i^*)(1-\alpha)}{\partial e_i^*} - \frac{\partial C(m, e_i^*)}{\partial e_i^*} \right)$$

Using Equation 3.3, this simplifies.

[55] As mentioned, the model was formalized together with W. Walker Hanlon. I use the plural form in this section to signal that the proofs are the result of a joint effort.

$$\frac{dCP_i}{dA_i} = \frac{\partial I(A_i, 1-e_i^*)\alpha}{\partial A_i} + \frac{\partial P(A_i, e_i^*)(1-\alpha)}{\partial A_i} - \frac{\partial I(A_i,1)\alpha}{\partial A_i}$$

Taking the derivative with respect to α, we obtain the following.

$$\frac{d^2CP_i}{dA_i d\alpha} = \left[\frac{\partial I(A_i, 1-e_i^*)}{\partial A_i} - \frac{\partial I(A_i,1)}{\partial A_i}\right] + \left[\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial e_i^* \, \partial A_i}\alpha + \frac{\partial^2 P(A_i, e_i^*)}{\partial e_i^* \, \partial A_i}(1-\alpha)\right]\frac{de_i^*}{d\alpha}$$

Note that $de_i^*/d\alpha < 0$, $\partial^2 I(A_i, 1-e_i^*)/\partial A_i \partial e_i^* < 0$, and $\partial^2 P(A_i, e_i^*)/\partial A_i \partial e_i^* > 0$. Thus, all terms are negative except $(\partial^2 I(A_i, 1-e_i^*)/\partial A_i \partial e_i^*)(de_i^*/d\alpha) > 0$. Denote $-de_i^*/d\alpha = \Delta > 0$. We rewrite the equation above by splitting the first term into two.

$$\begin{aligned}\frac{d^2CP_i}{dA_i d\alpha} &= \left[\frac{\partial I(A_i, 1-e_i^*)}{\partial A_i} - \frac{\partial I(A_i, 1-e_i^*+\Delta)}{\partial A_i}\right] + \left[\frac{\partial I(A_i, 1-e_i^*+\Delta)}{\partial A_i} - \frac{\partial I(A_i,1)}{\partial A_i}\right] \\ &- \left[\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial e_i^* \, \partial A_i}\alpha + \frac{\partial^2 P(A_i, e_i^*)}{\partial e_i^* \, \partial A_i}(1-\alpha)\right]\Delta\end{aligned}$$

Next, we take a linear approximation of the first term on the right-hand side.

$$\begin{aligned}\frac{d^2CP_i}{dA_i d\alpha} &\approx \Delta\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial A_i \partial e_i^*} + \left[\frac{\partial I(A_i, 1-e_i^*+\Delta)}{\partial A_i} - \frac{\partial I(A_i,1)}{\partial A_i}\right] \\ &- \left[\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial e_i^* \, \partial A_i}\alpha + \frac{\partial^2 P(A_i, e_i^*)}{\partial e_i^* \, \partial A_i}(1-\alpha)\right]\Delta\end{aligned}$$

Rewriting,

$$\begin{aligned}\frac{d^2CP_i}{dA_i d\alpha} &\approx \left[\Delta\left(\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial A_i \partial e_i^*}\right) - \Delta\left(\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial A_i \partial e_i^*}\right)\right]\alpha \\ &+ \left[\Delta\left(\frac{\partial^2 I(A_i, 1-e_i^*)}{\partial A_i \partial e_i^*}\right) - \Delta\left(\frac{\partial^2 P(A_i, e_i^*)}{\partial e_i^* \, \partial A_i}\right)\right](1-\alpha) \\ &+ \left[\frac{\partial I(A_i, 1-e_i^*+\Delta)}{\partial A_i} - \frac{\partial I(A_i,1)}{\partial A_i}\right] < 0\end{aligned}$$

Step 2

We are looking for an $\bar{\alpha}$ such that $dCP_i/dA_i \leq 0$ for all values of m. To begin, set $dCP_i/dA_i = 0$

to get a cutoff $\bar{\alpha}(m)$ given a particular monitoring level m.

$$\bar{\alpha}(m)\left(\frac{\partial I(A_i, 1-e_i^*)}{\partial A_i} - \frac{\partial I(A_i,1)}{\partial A_i}\right) + \left(\frac{\partial P(A_i, e_i^*)}{\partial A_i}\right)(1-\bar{\alpha}(m)) = 0$$

Applying the implicit function theorem to this formula, we can show that $d\bar{\alpha}(m)/dm > 0$. Thus, we need to find $\bar{\alpha} = lim_{m\to+\infty}\bar{\alpha}(m)$. As $m \to +\infty$, $e_i^* \to 1$. Thus, the $\bar{\alpha} \in (0,1)$ which satisfies $dCP_i/dA_i \leq 0$ for all m is implicitly defined by the following equation.

$$\bar{\alpha}\left(\frac{\partial I(A_i,1)}{\partial A_i}\right) = \left(\frac{\partial P(A_i,1)}{\partial A_i}\right)(1-\bar{\alpha})$$

Proof of Prop. 5

Applying the implicit function theorem to Equation 3.3, we obtain,

$$\frac{de_i^*}{dA_i} = -\frac{\frac{\partial^2 I(A_i,1-e_i^*)}{\partial A_i \partial e_i^*}\alpha + \frac{\partial^2 P(A_i,e_i^*)}{\partial A_i \partial e_i^*}(1-\alpha)}{\frac{\partial^2 I(A_i,1-e_i^*)\alpha}{(\partial e_i^*)^2} + \frac{\partial^2 P(A_i,e_i^*)(1-\alpha)}{(\partial e_i^*)^2} - \frac{\partial^2 C(m,e_i^*)}{(\partial e_i^*)^2}}$$

The denominator of this expression is negative, so the sign depends on the numerator. When $\alpha = 1$ the numerator will be $\partial^2 I(A_i, 1-e_i^*)/\partial A_i \partial e_i^* < 0$. Therefore, since this is a continuous function, we must have $de_i^*/dA_i < 0$ for sufficiently high α. Similarly, when $\alpha = 0$, the numerator will be $\partial^2 P(A_i, e_i^*)/\partial A_i \partial e_i^* > 0$ and we must have $de_i^*/dA_i > 0$ for sufficiently low α.

Appendix 3.B Incentive mechanism

In this section we substantiate our argument that when there is a great deal of uncertainty regarding the value of the public good, an incentive scheme based on leader's effort will be preferred to one based on outcome. The driving force behind this theoretical result is that members are risk averse. Compensating the leader based on the public good value forces her to accept additional uncertainty in her utility function, reducing the utility that she derives from being the leader. Compensation based on effort, which is more easily observed, avoids the disutility generated by this additional uncertainty. This advantage must be weighed agains the fact that compensation based on the value of the public good allows higher ability members to exert less effort than lower ability member and still receive the same costs or benefits from holding office. This reduces the opportunity cost of holding office for high ability members, relative to low ability members, increasing the chances of high ability members choosing to become candidates. These issues have

been analyzed previously (Baker, 2000, Lazear, 1986), and so we only revisit them briefly here.

Consider two incentive schemes. The first is based on effort and denoted $C_1(m, e_i)$. The second is based on the public good value and denoted $C(P(A_i, e_i, \eta_P))$. The utility of a manager under an effort-based, and output-based performance scheme, respectively, are given below.

$$U_1 = U[I(A_l, 1 - e_l, \eta_I)\alpha + P(A_l, e_l, \eta_P)(1 - \alpha) - C_1(m, e_l) - \phi]$$

$$U_2 = U[I(A_l, 1 - e_l, \eta_I)\alpha + P(A_l, e_l, \eta_P)(1 - \alpha) - C_2(P(A_l, e_l, \eta_P)) - \phi]$$

Now consider the effect of an increase in the variance of η_P (with a commensurate increase in the variance of η_I). The effect on the utility through the first and second terms inside the U() function will be equivalent regardless of the incentive scheme, ceteris paribus. However, an increase in the variance of η_P will also act through $C_2(P(A_l, e_l, \eta_P))$ under the output-based incentive scheme, which will cause U_2 to decrease more rapidly in η_P than U_1. Thus, if there is sufficient uncertainty in the value of the public good, the returns to being a leader may be lower under an output-based incentive scheme than under an effort-based incentive scheme. This would motivate groups to offer effort-based incentive schemes, such as the scheme posited in this paper.

Appendix 3.C Monitoring mechanism

The monitoring technology in this theory can be motivated as follows. Suppose that there are a set of T tasks that the group leader should fulfill. If the leader allocates an effort level $e_l^* \in [0, 1]$ to the leadership position, then a fraction e_l^* of these tasks are completed, while $1 - e_l^*$ of them remain undone. The monitoring technology $m \in [0, T]$ allows the group to look at m of these tasks and observe whether they were completed. The leader is then punished based on the number of tasks identified which are incomplete. If we consider a large number of tasks then this is equivalent to a variable with a binomial distribution, where the probability of identifying an incomplete task is $1 - e_l^*$ and m gives the number of observations obtained. The expected number of incomplete tasks observed is then $m(1 - e_l^*)$ with variance $me_l^*(1 - e_l^*)$.

In order to simplify the model it is helpful to modify the framework above to eliminate the uncertainty. One way to do this is to assume that the expected value is always achieved, i.e., that the group always observes $m(1-e_l^*)$ incomplete tasks given an effort allocation e_l^* and monitoring level m. This simplification allows us to ignore the impact of the member's risk aversion on their effort allocation, greatly simplifying the analysis. Note that the resulting function $C(m, e_l^*)$ satisfies the model assumptions.

Appendix 3.D Simulations

The following functional forms are used in the simulation exercise.

$$I = A_i^{\beta}(1-e_i)^{1-\beta} \qquad P = p * A_i^{\beta} e_i^{1-\beta} \qquad C = m(1-e_i)$$

The parameter values used for the simulations are $N = 10$, $\beta = .5$, $\phi = .1$, and $p = .1$. The simulations are run for values of m from 0 to 0.4 by steps of .02 and for $\alpha = \{0.47, 0.49, 0.51, 0.53\}$.

Figure 3.D presents additional results from the simulation exercise. The data are constructed by ranking the individuals in each group by their ability, with 10 being the highest and 1 being the lowest ability member. It shows the ranking of the member that ultimately becomes the leader. The main point here is that not only is leader ability falling, but that it is falling even though higher ability members are available.

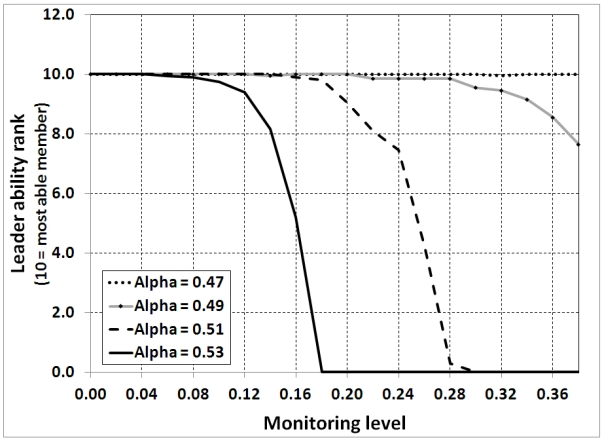

Figure 3.5: Simulated Leader Ability Rank

We also calculate results where we take average values over all equilibria, rather than just the equilibrium that delivers the best manager, for each set of parameter values. The average leader effort, leader ability, and public good value produced are displayed in Figures 3.6 and 3.7 below. These show the same results as obtained when we focused only on the best available equilibrium. Note that in some cases there will be only one equilibrium and that under these conditions these results will be exactly the same as those displayed in Figures 3.1 and 3.2.

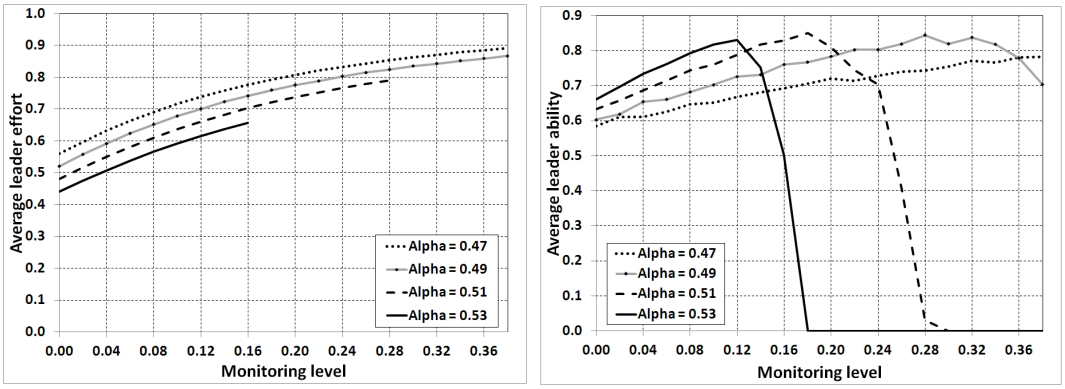

Figure 3.6: Simulated Leader Effort and Ability Averaged Over All Equilibria

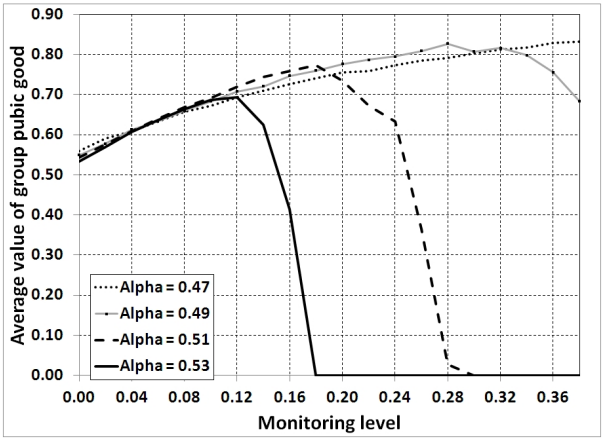

Figure 3.7: Simulated Public Good Value Averaged Over All Equilibria

Appendix 3.E Empirics

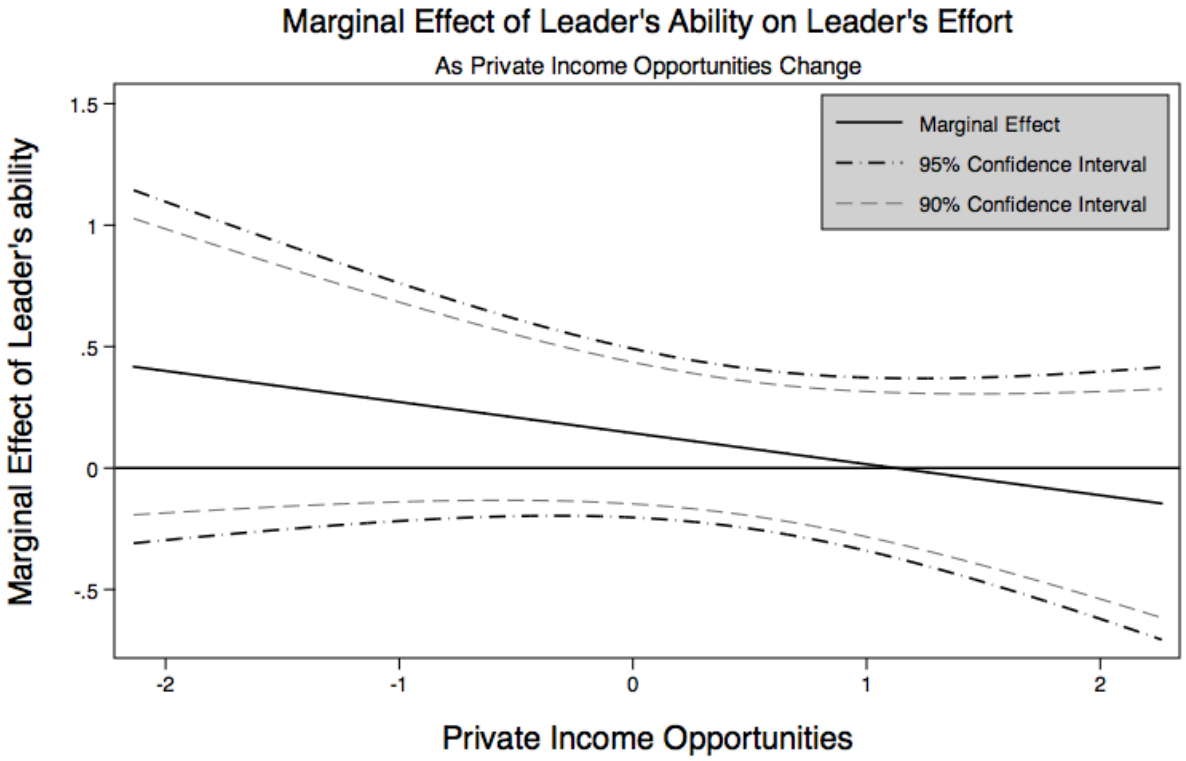

Figure 3.8: **Marginal effect of leader ability on effort as a function of the private income opportunities.** Marginal effects are estimated by holding continuous variables at mean values and categorical variables at median values.

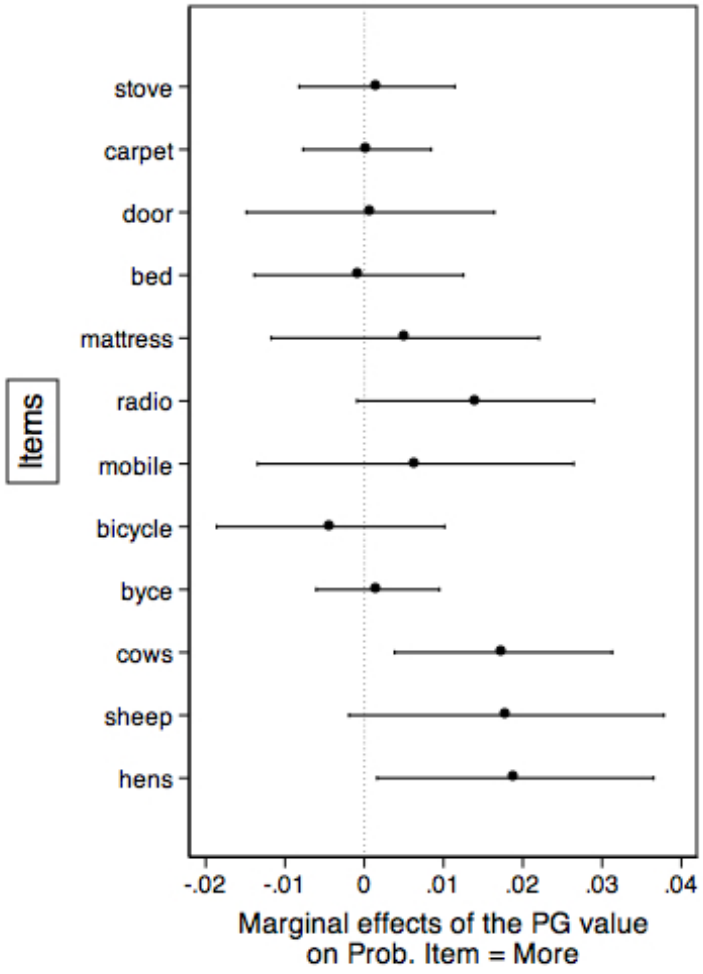

Figure 3.9: **Marginal effect of collective marketing on assets-base.** This figure shows the relation between collective marketing and and the presence of items used to construct the welfare index. It provides information on the marginal effect of the value of the public goods measured as the fraction of members selling in bulk on the likelihood that member i from DC j has *more* of item k_s since joining the farmer group. Marginal effects are calculated for an increase of 10% in the fraction of members selling in bulk from the grand mean, holding continuous variables at their mean levels and categorical variables at their median. The Ordered logit model include strata fixed effects and clustered standard errors at the association's level. The figure suggests that the primary driver of the increase in measures welfare was an increase in livestock (cows, sheep, hens).

DV: LEADER'S REALIZED EFFORT IN PUBLIC GOODS PRODUCTION

	Model A	Model B	Model C	Model D
Monitoring (std.)	0.772***	0.709***	0.731***	0.690***
	(0.118)	(0.121)	(0.129)	(0.235)
Private Income Opportunities			0.087	0.014
			(0.132)	(0.168)
Monitoring × PIO			-0.079	0.058
			(0.117)	(0.125)
Leader's ability (std.)				0.151
				(0.175)
Monitoring × Leader's ability				-0.092
				(0.093)
Leader's ability × PIO				-0.128
				(0.128)
Age of DC		-0.108	-0.115	-0.165*
		(0.082)	(0.085)	(0.093)
N members (units of 50)		0.069*	0.062	0.031
		(0.040)	(0.044)	(0.077)
Mean Seasonal Yield (units of 100)		0.103**	0.083	0.056
		(0.050)	(0.060)	(0.089)
Ethnic fractionalization		-0.113	-0.195	-0.706
		(0.527)	(0.563)	(1.003)
Religious fractionalization		-0.625	-0.489	-0.481
		(0.632)	(0.644)	(0.770)
Associational-life (std.)		-0.288**	-0.330**	-0.444**
		(0.125)	(0.140)	(0.161)
Middleman honesty		1.050	1.376	1.639
		(1.248)	(1.421)	(1.985)
Constant	0.483***	-0.274	-0.536	-0.385
	(0.123)	(1.101)	(1.222)	(1.906)
Strata FEs	X	X	X	X
Observations	50	50	50	42
r^2	0.720	0.781	0.791	0.800

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: **Relation Between the association's monitoring level and the leaders' realized effort.** Results from OLS regressions using group-level data. This table replicates Table 3.2 in the main text, while also providing the full set of coefficients for the control variables. PIO stands for private income opportunities.

DV: LEADER'S REALIZED EFFORT IN PUBLIC GOODS PRODUCTION

	Model A	Model B	Model C	Model D
Monitoring (std.)	0.733***	0.662***	0.682***	0.601***
	(0.116)	(0.107)	(0.107)	(0.143)
Private Income Opportunities			0.066	-0.070
			(0.104)	(0.127)
Monitoring × PIO			-0.087	0.094
			(0.083)	(0.114)
Leader's ability (std.)				0.246**
				(0.116)
Monitoring × Leader's ability				-0.152
				(0.095)
Leader's ability × PIO				-0.122
				(0.109)
Age of DC		-0.137	-0.144*	-0.253***
		(0.088)	(0.087)	(0.098)
N members (units of 50)		0.051	0.045	-0.010
		(0.044)	(0.044)	(0.053)
Mean Seasonal Yield (units of 100)		0.108**	0.090*	0.062
		(0.052)	(0.053)	(0.060)
Ethnic fractionalization		-0.610	-0.649	-1.586**
		(0.607)	(0.599)	(0.698)
Religious fractionalization		-0.798	-0.691	-1.079
		(0.764)	(0.781)	(0.879)
Associational-life (std.)		-0.371***	-0.410***	-0.498***
		(0.137)	(0.138)	(0.149)
Middleman honesty		2.023	2.240	3.299**
		(1.369)	(1.371)	(1.448)
$\sigma_{regions}$	1.245	0.792	0.817	0.353
Observations	50	50	50	42
Log likelihood	-51.035	-43.960	-43.062	-34.637

Standard errors clustered at the strata (region) level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: **Relation Between the association's monitoring level and the leaders' realized effort.** This Table complements Table 3.2 in the text, by providing results from random intercept mixed-effects regression models. PIO stands for private income opportunities.

DV: LEADER'S REALIZED EFFORT IN PUBLIC GOODS PRODUCTION

	Measure A	Measure B	Measure C	Measure D
Monitoring (std.	0.427*	0.690***	0.617*	1.451
	(0.219)	(0.235)	(0.311)	(0.994)
Private Income Opportunities	-0.063	0.014	-0.056	-0.095
	(0.171)	(0.168)	(0.158)	(0.159)
Monitoring × PIO	0.008	0.058	0.006	-0.010
	(0.156)	(0.125)	(0.136)	(0.163)
Leader's ability (std.)	0.099	0.151	0.111	0.089
	(0.171)	(0.175)	(0.181)	(0.156)
Monitoring × Leader's ability	-0.166	-0.092	-0.119	-0.186
	(0.118)	(0.093)	(0.124)	(0.142)
Leader's ability × PIO	-0.127	-0.128	-0.144	-0.120
	(0.139)	(0.128)	(0.127)	(0.136)
Age of DC	-0.203*	-0.165*	-0.163	-0.215*
	(0.114)	(0.093)	(0.099)	(0.122)
N members (units of 50)	0.045	0.031	0.030	0.052
	(0.073)	(0.077)	(0.095)	(0.075)
Mean Seasonal Yield (units of 100)	0.117	0.056	0.085	0.153
	(0.095)	(0.089)	(0.101)	(0.091)
Ethnic fractionalization	-0.801	-0.706	-0.776	-0.945
	(1.112)	(1.003)	(1.002)	(1.083)
Religious fractionalization	-0.893	-0.481	-0.690	-1.108
	(0.803)	(0.770)	(0.762)	(0.912)
Associational-life (std.)	-0.459*	-0.444**	-0.437**	-0.428
	(0.232)	(0.161)	(0.192)	(0.258)
Middleman honesty	3.015	1.639	1.200	3.218
	(2.561)	(1.985)	(2.399)	(2.650)
Constant	-1.339	-0.385	-0.080	-2.587
	(2.292)	(1.906)	(2.252)	(2.267)
Strata FEs	X	X	X	X
Observations	42	42	42	42
r^2	0.730	0.800	0.760	0.719

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: **Robustness Check:** this table provides a robustness check for the composite monitoring variable. It compares the relation between the association's monitoring level and the leaders' realized effort (Table 3.2 in the main text), by juxtaposing four different measure of monitoring. Measure A is a composite index constructed using only perceptional/attiudinal information (e.g. whether the manager is accountable); Measure B is constructed by adding behavioral information and is our preferred measure used throughout the paper (e.g. whether external auditors are used); Measure C is constructed by adding information on whether leaders provide receipts to members who sell in bulk; Measure D is simply whether there is a rule specifying an effort requirement. Results from OLS regressions using group-level data. PIO stands for private income opportunities.

DV: SELLING IN BULK VIA THE DC IN THE PAST SEASON			
	Model A	Model B	Model C
Leader's effort (std.)	0.716***	0.752***	0.653***
	(0.191)	(0.198)	(0.189)
Leader's ability (std.)	0.124	0.129	0.090
	(0.180)	(0.178)	(0.196)
Leader's ability × Leader's effort	-0.073	-0.086	-0.013
	(0.136)	(0.128)	(0.140)
Male		-0.156	-0.178
		(0.167)	(0.177)
Age (units of 10)		0.060	0.073
		(0.056)	(0.058)
Seasonal Coffee Yield (units of 100)		0.045	0.041
		(0.032)	(0.032)
Years since Joining Group		0.111**	0.124**
		(0.054)	(0.057)
Middleman honesty		0.393**	0.354*
		(0.198)	(0.194)
Education (Std.)		-0.043	-0.007
		(0.072)	(0.076)
Leader's co-villager		0.387	0.455
		(0.396)	(0.421)
Associational-life (std.)		0.028	0.051
		(0.071)	(0.075)
Community leader		-0.011	0.001
		(0.141)	(0.145)
Farming Primary Occupation			0.391
			(0.264)
Village distance to District Capital			0.009
			(0.007)
Village Distance to Trading Center			0.115*
			(0.063)
Village Distance to DC Crop Collection			-0.091**
			(0.042)
N members (units of 50)			0.158*
			(0.094)
Age of DC			0.043
			(0.137)
Constant	0.603	-0.405	-1.448*
	(0.382)	(0.489)	(0.776)
Strata FEs	X	X	X
Observations	1456	1233	1180
Log Likelihood	-877.344	-719.194	-670.097

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: **Members' decision to sell in bulk as a function of the leader's ability and effort.** This table replicates Table 3.3 in the main text, while also providing the full set of coefficients for the control variables. Model A includes only the key independent variables, Model B includes, in addition, a series of individual-level controls, Model C includes DC-level controls (C). In all models standard errors are clustered at the farmer association level.

DV: SELLING IN BULK VIA THE DC IN THE PAST SEASON

	Model A	Model B	Model C
Leader's effort (std.)	0.841***	0.836***	0.802***
	(0.226)	(0.225)	(0.245)
Leader's ability (std.)	0.110	0.126	0.134
	(0.210)	(0.208)	(0.228)
Leader's ability × Leader's effort	-0.021	-0.053	0.065
	(0.198)	(0.194)	(0.207)
Male		-0.081	-0.067
		(0.160)	(0.168)
Age (units of 10)		0.112**	0.128**
		(0.051)	(0.053)
Seasonal Coffee Yield (units of 100)		0.029	0.024
		(0.018)	(0.018)
Years since Joining Group		0.103**	0.097**
		(0.045)	(0.047)
Middleman honesty		0.348	0.340
		(0.257)	(0.264)
Education (Std.)		-0.063	-0.032
		(0.077)	(0.081)
Leader's co-villager		0.567**	0.615**
		(0.254)	(0.271)
Associational-life (std.)		0.059	0.089
		(0.083)	(0.087)
Community leader		0.090	0.049
		(0.167)	(0.175)
Farming Primary Occupation			0.245
			(0.163)
Village distance to District Capital			0.006
			(0.007)
Village Distance to Trading Center			0.098***
			(0.035)
Village Distance to DC Crop Collection			-0.113***
			(0.033)
N members (units of 50)			0.128
			(0.108)
Age of DC			0.091
			(0.200)
Constant	0.453	-0.870*	-1.947**
	(0.280)	(0.453)	(0.850)
$\sigma_{regions}$	.409	0.425	0557
σ_{groups}	1.040	1.009	1.050
Observations	1456	1233	1180
Log Likelihood	-824.826	-683.420	-638.064

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: This table complements Table 3.3 in the main text. Here I fit three-levels random intercept logistic regressions, in which members' decision to cooperate is modeled as a function of the leader's ability and effort. Model A includes only the key independent variables, Model B adds a series of individual-level controls, Model C includes also DC-level controls (C). In all models, members are nested within farmer groups which are nested within strata.

DV: IDENTITY OF THE ASSOCIATION'S LEADER

	Model A	Model B	Model C	Model D
Rep ability (std.)	1.128***	1.147***	0.874***	0.877***
	(0.257)	(0.207)	(0.199)	(0.203)
Monitoring	0.042	0.235	0.158	-0.006
	(0.215)	(0.150)	(0.207)	(0.222)
Rep ability × Monitoring	-0.066	-0.050	0.139	0.121
	(0.228)	(0.147)	(0.192)	(0.199)
Private Income Opportunities (std.)	0.073	0.089	-0.018	-0.140
	(0.240)	(0.141)	(0.147)	(0.154)
Rep ability×PIO	-0.218	-0.245	-0.157	-0.132
	(0.265)	(0.167)	(0.175)	(0.171)
Monitoring ×PIO	0.264	0.286***	0.142	0.150
	(0.179)	(0.088)	(0.113)	(0.117)
Rep ability×Monitoring ×PIO	-0.349*	-0.361***	-0.259**	-0.246**
	(0.191)	(0.081)	(0.105)	(0.103)
N members (units of 50)		-0.098***	-0.187**	-0.212**
		(0.037)	(0.076)	(0.084)
Ethnic fractionalization		-0.711	-1.005	-2.175**
		(0.533)	(0.714)	(0.996)
Community leader			0.163	0.209
			(0.467)	(0.457)
Wealth (std.)			0.380***	0.376***
			(0.131)	(0.139)
Associational-life (std.)			-0.370**	-0.401**
			(0.173)	(0.186)
Seasonal Coffee Yield (units of 100)			0.063***	0.064***
			(0.016)	(0.017)
Good health			1.260	1.280
			(0.866)	(0.883)
Born in village			0.423	0.463
			(0.368)	(0.369)
Constant	-3.642***	-3.032***	-4.574***	-4.529***
	(0.238)	(0.252)	(1.038)	(1.095)
Strata FEs				X
Observations	1058	1058	903	903
Log Likelihood	-163.284	-162.101	-125.116	-124.294

Standard errors clustered at the association level in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.11: **Relation between reps' ability and group's monitoring level and private income opportunities.** This table replicates Table 3.4 in the main text, while also providing the full set of coefficients for the control variables. PIO stands for private income opportunities.

DV: CHANGE IN A MEMBER'S WELFARE SINCE JOINING THE FAMER GROUP

	A: Direct	B: Reduced	C: IV
Fraction of members selling in bulk	0.072*		0.229**
	(0.037)		(0.096)
Leader's ability (std.)		-0.020	
		(0.040)	
Leader's effort (std.)		0.079*	
		(0.041)	
Leader's ability × Leader's effort		-0.056**	
		(0.023)	
Share of Yield sold in bulk (std.)	0.117***	0.100***	0.063*
	(0.028)	(0.030)	(0.036)
Seasonal Yield (units of 100)	0.011**	0.007	0.006
	(0.005)	(0.006)	(0.006)
Years since Joining Group	0.059***	0.057***	0.052***
	(0.013)	(0.015)	(0.015)
Education attainment	0.013	0.017	0.014
	(0.022)	(0.025)	(0.025)
Associational-life (std.)	0.097***	0.119***	0.124***
	(0.031)	(0.031)	(0.031)
Community leader	0.010	0.015	0.014
	(0.021)	(0.020)	(0.020)
Leader's co-villager	0.116	0.163	0.171*
	(0.088)	(0.100)	(0.104)
Male	0.163***	0.169***	0.145***
	(0.048)	(0.051)	(0.049)
Age (units of 10)	-0.106***	-0.095***	-0.095***
	(0.017)	(0.018)	(0.017)
Good health	0.105**	0.107*	0.089*
	(0.052)	(0.056)	(0.051)
Shop owner	0.159	0.205	0.228
	(0.133)	(0.139)	(0.142)
Age of DC	-0.027	-0.027	-0.064
	(0.028)	(0.028)	(0.041)
N members (units of 50)	-0.002	0.010	-0.031
	(0.021)	(0.021)	(0.036)
Constant	-0.073	-0.161	0.098
	(0.190)	(0.165)	(0.247)
Strata FEs	X	X	X
Observations	1448	1229	1229
r^2	0.113	0.123	0.102
Log Likelihood	-1908.469	-1569.030	-1583.633

Standard errors clustered at the farmer association level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12: **Relation Between Value of the PG and Welfare.** This table replicates Table 3.5 in the main text, while also providing the full set of coefficients for the control variables. In model A, the standardized dependent variable is a function of the value of the public goods measured as the fraction of members selling in bulk via the assocaition. In model B, the dependent variable is a function of leader's ability and effort, whereas in the model C, the key independent variable – the value of the public good – is instrumented by the leader's effort and ability.

Chapter 4

Do Leader Selection Rules Impact Accountability? Exploiting Exogenous Variation in the Formation of Ugandan Farmer Associations

4.1 Introduction

Social scientists have long argued that governance institutions affect economic and political outcomes. A sizable literature focuses on leadership selection rules, debating the possible trade-offs between the election and appointment of office holders. The question of leadership selection is arguably the most salient issue in judicial studies (Hanssen, 1999)[1], for instance, and has also received increased attention from scholars of corporate governance following the recent financial crisis (Frey and Benz, 2005).

Yet, a survey of the literature on the impact of the rules for selecting officials – such as judges, public regulators, school board members and the CEOs of traded companies – suggests that the ongoing debate over the relative virtues of elections and appointments is far from settled. On the one hand, there are good theoretical reasons to assume that different leadership selection methods will have varying effects on political and economic outcomes of interest. The empirical findings, however, are ambiguous.

One reason for the mixed results likely stems from the question of context: the appointment of a CEO by a company's board of directors may be only weakly comparable to the appointment of a utility regulator by an elected state governor, for example. Yet, even within studies of the same types of institutions, research in this field is plagued by endogeneity problems. The fact that the adoption of governance institutions tends to be a function of other group-specific factors makes assessment of causal impact nearly impossible (Acemoglu, Johnson and Robinson, 2001). In other words, there may be unobserved characteristics that correlate both with the adoption of certain governance institutions and the socio-political or economic outcomes of interest (Dal Bo, Foster and Putterman, 2010). Finally, past work is mostly based on aggregated cross-national or intra-state data. The structure of such data seriously limits our ability to gain insight into the micro-foundational processes associated with different selection methods[2].

[1] "It is fairly certain that no single subject has consumed as many pages in law reviews over the past 50 years as the subject of judicial selection"– the election versus the appointment of state judges" (quoted in Hanssen (1999, p. 205)).

[2] Causal inference is further frustrated by the fact that institutions are bundled within states and also change slowly over time (Acemoglu and Johnson, 2005).

This paper contributes to the debate by exploiting plausibly exogenous variation in leadership selection rules within the context of local self-help organizations. Self-help organizations are relatively small, voluntary groups that provide goods and services to members. Examples include micro-finance groups, Chambers of Commerce, PTAs, common-pool resource groups, artisan cooperatives and farmer associations. Such organizations are present in many facets of economic and social life, and in countries of all income levels. Their importance has arguably increased in recent years, particularly in developing countries, where larger political units have democratized, decentralized and liberalized[3]. Thus, the ubiquity of self-help organizations and their growing impact on social welfare (Grossman, 2011*a*) calls for a better understanding of the factors that determine their effectiveness, in particular.

More specifically, this paper studies Ugandan farmer associations that were created as part of the Agriculture Productivity Enhancement Project (APEP), 2004-2009[4]. APEP is among the largest recent rural development projects in East Africa. This intervention targeted over 60,000 smallholder farmers from 30 districts throughout Uganda in an effort to support their transition to commercial farming. Smallholder producers were organized into farmer groups of about 20-25 members known as producer organizations, or POs. POs operated at the village level. To exploit economies of scale and to increase the farmers' bargaining power, APEP subsequently organized an average of ten neighboring POs into larger "associations," known as Depot Committees.

The APEP associations share virtually identical organizational structures. This is because most aspects of group structure were specified in an extensive training manual provided to all the project field staff that oversaw the formation of larger associations. As per the APEP manual, each farmer association is led by an executive committee (management team), which is part of the DC board of directors. The executive committee is comprised of a DC manager, the chairman of the board, a secretary and a treasurer. Operationally, the DC manager is the principal leader of the

[3]Empirical evidence supports this claim: In Senegal 10% of sampled villages reported having at least one self-help group in 1982; by 2002 this figure was 65%. In Burkina Faso the figures were 22% for 1982 and 91% in 2002 (Bernard et al., 2008).

[4]APEP was funded by USAID and implemented by Chemonics International, an International Development consulting firm based in Washington D.C.

association (henceforth, leader)[5]. The leader's most important responsibilities include negotiating input and output prices, organizing the collection of crops (including hiring and supervising employees), and searching for buyers. Additionally, leaders help coordinate group activities and facilitate the flow of information throughout the association. Responsibilities of the board of directors, which is comprised of two elected representatives from each of the village-level POs, include monitoring the work of the DC executives (including the leader), representing the opinions of PO members at the associational level, and aiding in the transmission of information.

Crucially, though, the APEP field manual failed to specify the rule that the new associations should adopt for selecting managers. Instead, the local field facilitators were unguided in their advocacy of either appointment or election rules. In this respect, the idiosyncratic preferences of APEP facilitators played an influential role in the process of group formation.

In the event, APEP facilitators recommended appointment-based selection in approximately 60% of cases. Specifically, these facilitators recommended that the DC board of directors appoint the DC manager[6]. A large number (78%) of DCs that received a recommendation to appoint leaders followed the suggestion. By contrast, APEP facilitators in the remaining 40% of cases suggested that PO members use popular, direct voting to select the DC manager[7]. Attesting to the centrality of the APEP facilitators in the process of group formation, 94% of associations followed this recommendation.

This chapter uses the facilitator recommendation as an instrumental variable with which to test the impact of varying leadership selection rules. I consider possible impact along two dimensions[8]. First, I employ behavioral measures to examine the impact of leader selection rules on monitoring and accountability within groups. Findings suggest that popular direct elections yield significantly higher levels of monitoring and accountability, substantially increasing the (i)

[5]The DC manager is equivalent to a CEO in a traded company. In about a quarter of the sampled DCs, the DC manager also serves as the chairman of the board.

[6]The term appointment here loosely lumps up slightly different ways in which DC board of directors, which is comprised equally of representation from constituent POs, select the manager. These included, for instance, a show of hands among the members of the board of directors, or deliberated consensus.

[7]In direct vote associations, all registered members from each of the village-level group can cast their vote. The direct-vote method, however, also loosely lumps up some variations in quorum requirement and election technology.

[8]I discuss the validity of this instrument below, in Section 4.4

frequency of selection processes (i.e. elections or (re-)appointments), (ii) the likelihood that an association hires external auditors to examine its books and records; and (iii) the likelihood that the association conducts internal audits. In addition, members in direct election associations are significantly more likely to receive receipts when selling their crops through the association[9].

Consistent with these behavioral results, I find that both ordinary members and board directors in direct-vote associations are more likely to perceive the DC leader as being monitored and accountable. Among other things, this suggests that, contra some arguments for appointment-based systems, members of farmer associations are sophisticated voters, capable of constructing realistic judgments regarding a leader's actions. Further, the leaders of direct vote associations themselves testify to higher levels of restraint placed on the leadership of direct-election associations: directly elected leaders are *more than two times* more likely than appointed leaders to report that other board directors warned them they were not doing their job well or were not exerting enough effort.

Secondly, I examine the impact of leader selection rules on the profile of the chosen leaders. Here, I test the hypothesis that appointments yield more competent and/or trustworthy leaders, as compared to popularly elected leaders. The intuition here, developed more extensively in Section 4.3, is that, compared to average group members, board directors are more informed about candidates and the current 'state of the world', have longer time-horizons, and are more likely to base their vote on the welfare of the group rather than narrow self-interest. I do not find evidence to support this argument. Selection methods do not appear to impact the profiles of leaders with respect to (i) ability (i.e. marketable skills) or other socio-demographic attributes, (ii) centrality of network position, or (iii) other-regarding preferences.

After reporting these main results, I turn to examine mechanisms linking elections and accountability. To do so, I conducted a set of well-known behavioral experiments with members of the farmer associations, which I adapted for the question in hand. These experiments are impor-

[9]In Uganda, most coffee farmer associations are unable to pay members cash-on-delivery. Instead farmers receive a receipt when delivering their crops to the association, which specifies the volume delivered (in kilograms). Only after the association collects the crops from all its members, the manager can finalize a deal with potential buyers. Members report being paid, on average, two to three weeks after they have delivered their coffee to their group.

tant because they provide evidence of revealed preferences. First, using a simple Dictator Game (DG), I find that elected leaders allocate almost twice as much as appointed leaders to anonymous association members, though they give slightly less than their appointed counterparts to anonymous strangers. Second, using a modified third-party punishment (TPP) experiment, I find that elected monitors are about *three times more likely* to sanction group members ("deciders") who transfer an unequal share of a joint endowment to other group members ("passive recipients"). These findings suggest that the selection process itself *changes the reciprocal expectations between leaders and members independently from reelection considerations* (Grossman, 2011*b*). Finally, I collect and analyze social network data in order to test for the existence of a "familiarity effect," or, an effect whereby growing familiarity between appointed officials and appointees creates a governance culture that is inimical to accountability. Indeed, I find evidence for such an effect, suggesting again that appointment-based systems are undesirable in this context.

These findings are consistent with Olken (2010), who shows that villagers who cast direct votes to select development projects experience dramatically higher satisfaction rates with the project, as well as greater perceived benefits. This is true even though the actual projects selected did not differ significantly across conditions. Findings presented here are also consistent with Huber and Gordon (2004), who finds that elected judges are more responsive to the public, but do not differ significantly from appointed judges in terms of their levels of education or experience. Similarly, I find evidence that direct elections yielded significantly more accountable and responsive leaders, and also find no evidence that appointed and elected leaders differ in terms of their observed attributes. These results may help explain the fact that, since their inception, membership in direct-vote farmer associations has increased at a much higher rate than appointed-based associations.

This paper is also related to recent experimental work on the impact of governance institutions in the context of community driven development (CDD) projects (Beath, Christia and Enikolopov, 2010, Olken, 2010). I aim to expand this nascent literature in three ways. First, the objective of the above studies has been to assess the relative effectiveness of methods for selecting development projects, whereas this paper focuses instead on the impact of the method for selecting local

leaders[10]. Also, whereas Olken (2010) and Beath, Christia and Enikolopov (2010) have focused on single-tier, village-level councils, this study considers membership-based, multi-tier organizations that span several neighboring villages. As I discuss in Section 4.3, the question of leadership selection is even more salient in multi-tier organizations than in single-tier organizations. Finally, this is the first leadership selection study to include behavioral experiments and social network analysis, which enables us to better identify and understand the causal mechanisms plausibly linking governance institutions to the behavior of leaders and constituents.

The paper proceeds as follows. Section 4.2 provides additional information on the specific context under consideration, including the organizational structure of APEP farmer associations. Section 4.3 presents a theoretical framework from which I derive testable hypotheses. Section 4.4 discusses the study's identification strategy in depth, and Section 4.5 provides information on the sampling scheme and data. Main results are reported in Section 4.6, and Section 4.7 explores possible causal mechanisms undergirding these findings. Section 4.8 concludes.

4.2 The Research Site: APEP

As mentioned above, the farmer associations considered in this paper were created as part of the Agriculture Productivity Enhancement Project (APEP). The project was funded by USAID and implemented by Chemonics International. The project, which began in 2004, covered 30 districts throughout Uganda and involved organizing 60,000 smallholder farmers into over 2,500 village-level producer organizations (POs). The goal of the project was to to support smallholder transition into commercial farming, and thereby promote economic development in Uganda.

[10]Olken (2010) and Beath, Christia and Enikolopov (2010) use randomization to assign villages with different methods for selecting development projects. In addition, Beath, Christia and Enikolopov (2010) vary the size of the selectorate: council members are either elected by a small subset of the village or by the village in large.

APEP Groups' Governance Structure

Two years into APEP, it became clear that farmer groups at the village level lacked the organizational capacity and bargaining power to become viable economic entities. In response, APEP decided to encourage neighboring POs to form larger associations. The resulting farmers' associations (or DCs) came to serve, on average, 200 farmers from ten neighboring villages. The DCs were designed specifically to exploit economies of scale, to reduce transaction costs for trading partners, and to increase farmers' bargaining power by increasing the volume of their transactions.

The organization of village-level groups into parish-level associations, however, did not come without drawbacks[11]. Villages are often information-rich environments. This allows village-level self-help groups to use the flow of information between members not only to coordinate activities, but also to minimize adverse selection and moral hazard problems. By contrast, the creation of associations (multi-tier organizations) likely reduces the efficiency of information flows, thereby hindering coordination and exacerbating agency problems. Regarding adverse selection, the creation of associations may reduce stakeholders' information concerning candidate qualifications[12]. As for moral hazard, associations increase the information asymmetry between leaders and the more widely dispersed members, while also plausibly reducing informal, norm-based constraints on leaders. The formation of associations also tends to introduce greater heterogeneity among members, thereby reducing their incentives and capacity for collective action.

These agency problems are reminiscent of the challenges that dispersed shareholders face in their relation with the management of traded companies[13] In fact, in their effort to overcome these challenges, APEP designed the structure of the farmer associations with a corporate governance template in mind[14]. Influenced by project field-facilitators who presided over the process of group

[11] In Uganda, districts are the central local administrative unit. Districts consist of 2-4 counties, each county has 3-6 sub-counties, each sub-county consists of 3-6 parishes, and each parish has about 5-15 villages. Between 200 and 800 households live in each village. Many services such as dispensaries, markets and trading centers, are organized on the basis of parish borders. DCs commonly serve one or two parishes, at most.

[12] Reducing the quality of the signal on candidates' quality should reduce the likelihood that high-ability members run for office, given positive candidacy costs and assuming that members are risk-averse.

[13] See Shleifer and Vishny (1997) for a comprehensive review of the literature that views corporate governance as a set of tools designed to mitigate against agency problems.

[14] This was confirmed in a series of interviews with APEP management and Chemonics consultants.

formation, all APEP associations share the following two-tier structure. At the village-level, producer organization (PO) members elect local leaders who are responsible for implementing decisions taken by the association. At the association level, registered PO members also elect two representatives to a board of directors, which functions as the association's main governance body. Thus, even in associations that I refer to as appointment-based associations, democratic participation is used to elect representatives to the board of directors. A graphical representation of the association governance structure is provided in Supplementary Appendix A, Fig. A.2.

Board directors have responsibilities as members of the board that include more than merely representing their POs at the association level. Apart from participating in plenary sessions, directors commonly serve on at least one of the association's committees. These include, among others, the influential executive committee as well as loans, auditing, marketing and procurement committees. In all APEP associations, the board is entrusted with the power to appoint directors to their respective committees[15]. In other words, though directors are elected by their respective POs to serve on the board as their representatives, association members do not have a say in the allocation of responsibilities at the board-level[16]. This is because the work of committees is bureaucratic in nature: committee members are expected to serve the entire association, not the narrower interests of their respective POs. Given the explicitly bureaucratic nature of these roles, there is no argument to have members directly elect board directors to their respective roles on the board.

What, then, is the case for allowing members to directly elect a board director to serve as the DC manager (leader)? First, DC leaders are responsible for the timely collection and quality of crops. They, therefore, interact frequently and directly with association members. The quality of this interaction crucially depends on the extent to which members trust their leader[17]. The interaction may be contentious, as leaders oversee the sanctioning of members whose actions undermine

[15]The most prestigious positions, such as chairman of the board, tend to be contested and are therefore decided through some voting procedure, usually a show of hands. Allocation to other positions tends to be based on seniority and voluntarism, and depends on the interests, expertise and time constraints of representatives.

[16]Put plainly, association members do not vote on which director will serve on what committee and at what role.

[17]Leaders can breach trust in several ways: when members deliver their crops to sell in bulk, managers can muddle with the weighting scales or exploit members' illiteracy when documenting the bulked volume. Some leaders have been caught reporting prices lower than that which was actually secured from buyers, pocketing the difference. Leaders can collude with buyers, selling the group's crop below market prices in exchange for a hefty commission, etc.

the association's marketing activities[18]. Third, because DC leaders travel frequently, representing the association in other contexts (e.g. larger towns and cities, or in higher level administrative settings), members are incentivized to become close to their leader. This is because leaders are believed to posses information that members value[19]. Finally, the effort that leaders put into their work can significantly impact the effectiveness of the association and thereby members' welfare (Grossman, 2011*a*). Members thus not only have a clear personal stake in the identity of the DC leader, but also tend to view him (always him) as their representative[20].

As mentioned earlier, we observe two selection methods for the DC managers across organizations. In over half of the associations, all registered members vote directly for their preferred candidate for the senior leadership position. In the remaining associations, board directors appoint the leader. Which method is more appropriate for selecting group leaders? What are the tradeoffs between direct vote and appointment in the context of self-help organizations? The importance of these questions transcends the boundaries of Ugandan farmer associations. Across the developing world it is becoming increasingly common to witness the amalgamation of several small-size groups into a single association. This process is especially prevalent in micro-insurance, micro-lending and farmer groups, where large gains can be made from increased bargaining power, reduction in transaction costs for trading partners, and from other scale considerations.

Before describing the study's strategy for identifying the impact of the different selection methods, the following section reviews the debate concerning elections and appointments in comparable contexts. From this discussion I derive a set of hypotheses, which I then test in Section 4.6.

[18]The bylaws commonly require members to apply "best agriculture practices." Managers have the power to reject crops due to low quality. In addition, members are required to refrain from side selling to middlemen. Violation may trigger some form of social sanctioning (e.g. public shaming), and in extreme situations, monetary fines.

[19]This information may or may not be related to the farmer association's core activities.

[20]Dozens of interviews with association members suggested that they perceive DC leaders to impact their livelihood at least as much, if not more than, local government officials.

4.3 Theoretical Framework

The theoretical and empirical work on the implication of differences in governance institutions across states (Besley and Case, 2003) and countries (Persson and Tabellini, 2003) is sizable and growing[21]. Within this broad topic, a political economy literature evolved around the pros and cons of methods for selecting public officials, where a plurality of selection methods co-exist in different jurisdictions. Examples include heads of regulatory agencies (Besley and Coate, 2003), school boards members (Hess, 2008), and judges (Hanssen, 1999), who are appointed by politicians in some jurisdictions, but are directly elected by the public in others[22].

As noted above, the literature on corporate governance is also relevant to this study. Scandals over manager compensation, fraudulent bookkeeping (e.g. the Enron scandal), and, more recently, the financial crisis of 2008, have prompted scholars to revisit basic tenants of corporate governance. This includes challenging the common practice of appointment of CEOs by a board of directors, rather then by shareholders (Frey and Benz, 2005). In what follows, I synthesize across these two literatures and formalize hypotheses.

Information, Independence and Accountability

Much of the debate between appointment and election-based systems focuses on the tradeoff between independence and information, on the one hand, and accountability and responsiveness, on the other. Most generally, appointment is preferred when concerns over independence and information asymmetries are strongest.

First, appointments arguably play an important role in facilitating the independence of officials, where independence is desired to mitigate time-inconsistency problems. The case for appointments is thus stronger when there is reason to believe that elected officials will fail to resist popular pressure to produce short-term results at the expense of longer-term gains (Alesina and

[21] This literature is especially interested in the effect of governance institutions on the size of government and on policies that have distributional consequences. For a recent review see Funk and Gathmann (2009).

[22] Unlike legislators, whose selection through popular direct-vote is a defining feature of modern democracies, the method for selecting other public officials is still a matter for debate.

Tabellini, 2008). This factor constitutes the basic justification for central bank independence (Rogoff, 1985). Similarly, an oft-cited defense of CEO appointment by board directors is that board directors are thought to have longer time-horizons than share-holders.

Second, appointments may also be preferred when the issue at hand is not salient, yet organized stakeholders have interests that do not coincide with the preference of the majority. In such cases, a dormant electorate may allow special interest groups to mobilize their members to influence the identity of public officials in ways that undermine the interest of the general public. This argument underlies the recent shift towards mayoral control of urban boards of education, for instance, which has replaced the election-based system put into effect during the Progressive Era (Hess, 2008)[23]. Similarly, the fear that large institutional investors will capture traded companies against the interest of small, dispersed or dormant shareholders is yet another justification for endowing board directors with the powers to appoint CEOs (Shleifer and Vishny, 1997).

Third, the relative benefits of appointment are often linked to information asymmetries, which undermine constituents' ability to meaningfully differentiate candidates, and to hold officials to account. Specifically, according to Maskin and Tirole (2004), constituents' ability to make informed choices and hold officials to an account diminishes as: (i) the technicality and/or unfamiliarity[24] of an issue increases, (ii) the cost of decision-relevant information increases, or (iii) the difficulty in obtaining feedback about the quality of decisions increases[25]. In such cases, the public — in the form of group members, shareholders, or citizen-voters — is poorly positioned to evaluate the performance of the official/CEO. Hence, one argument suggests that in order to minimize adverse selection problems and to reduce officials' incentive to pander in information-low environments (Gordon and Huber, 2002), it is best to keep the official accountable to a small number of knowledgeable appointees rather than to the public at large.

Indeed, in my interviews with APEP field-facilitators in Uganda, those who had encouraged the newly founded farmer associations to adopt an appointment rule justified their recommen-

[23] Hess (2008) finds that mayoral takeovers (i.e. appointment) improves student achievement and fiscal efficiency compared to directly-elected boards of education. His findings, however, are strongly contested.

[24] Familiarity here refers to how much exposure the electorate has with this sort of issue.

[25] The ability to map between an official's actions and outcomes is possible but only after a relatively long period.

dation by pointing to considerations similar to those described above. First, it was argued that board directors, who are significantly more educated and have larger plots of land than average members, are also more likely to have longer time horizons and see the "larger picture" when selecting a DC leader. In other words, directors are thought to be more likely to select candidates who can produce the highest possible value of public goods, rather than candidates who can provide private goods[26]. Secondly, APEP-facilitators pointed to the concern that if associations adopt a popular direct vote, a low turnout in the general assembly, the venue in which popular elections are commonly held, will enable nearby villages to elect "their" candidate, who may not be the most competent candidate. An appointment method, it was claimed, mitigates this concern, since each village is equally represented on the board of directors.

Turning to the information asymmetry argument, APEP facilitators who had encouraged an appointment rule also argued that, compared to the average member, board directors have more information about the relative competence and honesty of candidates due to their more prominent social and economic positions[27]. Put differently, ordinary members lack the information needed to make sound judgments regarding the relative qualities of various candidates, or so some of the APEP facilitators believed.

Additionally, APEP facilitators who had encouraged appointment often claimed that board directors are able to evaluate the contribution of incumbents to the value of the public good more accurately. It is difficult, in general, to track between a leader's actions and group-level outcomes in farmer associations because the crop price secured by a DC leader is typically affected by several exogenous factors, such as exchange rate and international coffee prices[28]. Compared to board directors, ordinary members are believed to be less informed about external "states of the world", which makes them more likely to vote out over-performers and retain under-performers[29]. This,

[26]Private goods refers here mostly to benefits that can be obtained through 'closeness' to a candidate, which may or may not be related to coffee production.

[27]Farmer associations serve, on average, ten villages; most farmers do not often travel outside their village. By contrast, prominent villagers are more likely to travel, engage with each other through commercial, political and social networks, and participate in activities at the parish and sub-county levels.

[28]Other factors include variability in level of competition between buyers in the region and the supply of coffee in Uganda, which is a function of rainfall conditions as well as pest outbreaks.

[29]Members are more likely to overestimate a leader's role in securing higher prices from trading partners. This

in turn, should reduce the quality of elected leaders. Finally, increasing reelection uncertainty of high-performers should also aggravate adverse selection in cases where high-ability members are risk-averse[30].

To summarize, if (i) board directors have longer time horizons, (ii) the election of DC leader is *not* politically salient, or (iii) members are less informed about candidate qualities and the "states of the world", then appointment-based associations should be better positioned than popular direct-vote associations to select better leaders, or leaders with "higher profiles" defined as having attributes that are conducive for producing higher valued public goods. Applying the above framework to APEP's Ugandan farmer associations, a testable hypothesis emerges:

Hypothesis 7 *Appointment-based associations are likely to select more able and/or honest leaders than associations that use popular direct-elections.*

It is possible, however, that accountability and responsiveness considerations may counterbalance the information and independence considerations. The key point is that appointment based systems can more easily be manipulated for personal gains while raising concerns about a loss of transparency[31]. For example, Shaub et al. (2005) demonstrate that growing familiarity between appointed officials and their appointees is conducive to a governance culture that takes shortcuts, cooks the books, or adopts other practices that do not effectively serve the interests of stakeholders. Appointing officials, they argue, makes it easier for them to control data, limit accountability, and reduce opportunities for stakeholder input. On a similar note, Shleifer and Vishny (1997) provide evidence that, in traded companies, although shareholders have the right to elect board directors, those representatives need not necessarily represent shareholders' interests and are in

is because trading partners will increase the price offered to farmer association – compared to a world in which the association does not exist – because the association reduces significantly their transaction costs. Interviews with leading coffee buying companies suggest that buying companies will return 25-40% of the transaction costs savings to the groups, to ensure their durability. This may lead members to require too little effort from leaders or set a reelection threshold too low. This too can encourage low ability members to enter the candidacy pool (Grossman, 2011*a*).

[30]For a more formal discussion of these issues, see Grossman (2011*a*).

[31]For example, a long-standing concern with regulated industries is that the regulators tend, over time, to become dominated by stakeholders that they are supposed to regulate. Similarly, elected judges may be beholden by criminal groups that control the electoral process.

fact likely to collude with, or be under the control of, the company's management[32]. Finally, in a recent review, Frey and Benz (2005) analyze the failure of board directors to check the powers of managers of traded companies. In response to what they see as a breakdown of accountability, they call for a shift towards direct democracy in corporate governance: allowing shareholders to vote directly on issues that were traditionally under the purview of board members. These include the selection of external auditors as well as the identity of senior management.

Empirical findings concerning outcomes of different selection methods for public officials suggest that elections tend to be consistent with higher levels of accountability and responsiveness. For example, in the literature on appointed versus elected regulators, consumer-prices data suggest that elected regulators produce more pro-consumer policies than appointed regulators, in areas as diverse as public utilities (Besley and Coate, 2003), telecommunications (Smart, 1994) and life-insurance companies (Fields, Klein and Sfiridis, 1997). Similar findings are reported in the literature on judicial selection. Huber and Gordon (2004), for example, find that elected judges become significantly more punitive the closer they are to standing for reelection. The authors interpret this as a sign that election increases *responsiveness* to the public, independent of the judges' personal and ideological attributes[33]. Hence theory and empirical evidence suggest that direct elections increase the responsiveness of representatives, broadly construed, and can mitigate the danger that managers and board directors collude against the larger group's interests.

Hypothesis 8 *Leaders of associations that use popular, direct elections are more likely to be accountable and responsive than their counterparts in appointment-based associations.*

With these hypotheses in mind, it is worth noting that although the literature discussed in this section contributes greatly to our understanding of the stakes of leadership selection rules, it is also plagued by endogeneity concerns. As a result, findings remain contested'[34]. The goal of

[32] Besley and Coate (2003) make a similar argument in the case of regulators. The fact that the public elects the appointee does not guarantee that the public's interest will be secured due to bundling effects.

[33] In a review of the literature on elected versus appointed judges, Champagne (2003, p. 416) concludes that appointed judges "do not have substantially different background characteristics than do elected judges". Huber and Gordon (2004) argue, therefore, that the selection institutions effect behavior independently of their effect on the profile of the selected judges.

[34] For example, against the backdrop of recent work (e.g., Besley and Coate (2003)) that finds that elected regulators

this paper is to contribute to the debate by exploiting plausibly exogenous variation in leadership selection rules among otherwise comparable institutions in order to identify their causal effect. The next section considers this empirical strategy in detail.

4.4 Plausible Exogenous Variation in Leader Selection Rules

As mentioned above, APEP's original mission was to organize dispersed smallholder farmers into producer organizations (POs), at the village-level. About two years into the program APEP decided that POs, which served approximately 20-25 members, lacked the capacity, bargaining power, capital and volume to become viable economic players. In response, APEP's management refocused the project: In addition to creating new farmer groups, they decided to encourage existing groups to join neighboring POs, thereby forming associations at the parish-level[35]. APEP hired and trained a set of local field-facilitators to visit the POs and oversee the establishment of larger-level associations (DCs).

This paper's identification strategy rests upon the fact that (i) APEP field-facilitators played a foundational role in establishing these groups, and (ii) the idiosyncratic preferences of individual field-facilitators are plausibly orthogonal to other characteristics of the groups they were hired to work with. To evaluate this strategy, it is necessary to reconsider processes of group formation and field-facilitator involvement.

The instrument

In their effort to establish larger associations, field-facilitators first secured the agreement of group members. They then worked closely with representatives from each PO on the association's design. Importantly, this process was guided by a 22-page long field manual, which outlined the

are more responsive to the public interests, Cavazos (2003) finds that consumer-price differences disappears after controlling for population density and percent of rural population. Cavazos argues that states that use direct elections are 'different' than those using appointments, pointing to the fact that the majority of states that elect public commissioners are low-population density states with a high proportion of rural residents.

[35] Local districts in Uganda are termed parishes.

steps field-facilitators should take to establish larger associations[36]. Through a series of workshops spread across three weekends, facilitators guided the PO representatives through discussion of issues, such as the DC organizational structure and the allocation of responsibilities between office holders, and between the POs and the DC. Facilitators made recommendations on a wide set of issues such as operating budgets, membership dues, annual registration fees, marketing commissions and other rules and bylaws, commonly enshrined in DC constitutions. These recommendations were derived from the manual, and were thus uniform across groups. The centrality and uniformity of the this process is attested to by the fact that all APEP associations — regardless of facilitator identity — ended up with similar organizational structures and division of power between the POs and the DC.

Yet, although it was elaborate, the manual did not specify every aspect of group formation. For this reason, although all associations share a basic organizational structure, DC constitutions differ in such aspects as the number of times the general assembly convenes and quorum requirements for passing resolutions and bylaws. Crucially, the manual was also silent on the question of leadership selection rules.

As a result, we observe that five of the eight facilitators operating in the study area encouraged associations to enact an appointment-based system, whereby the board of directors (where all POs are equally represented) would appoint a DC manager (leader), while the three other facilitators suggested that members directly elect their leader. In interviews, facilitators were clear on the fact that their recommendations were based on *their personal belief* concerning the most appropriate selection method. With this in mind, I propose to use facilitator recommendation as an instrumental variable (IV) for leadership selection rule. The counterfactual that forms the basis of the identification strategy is the outcome of interest (e.g. leader accountability), had APEP allocated to district S, a field-facilitator with different idiosyncratic priors with respect to the relative effectiveness of a given leader selection rule.

[36] A copy of the training kit/manual can be shared upon request.

Assessing the validity of the instrument

To formalize the identification strategy, let $Z \in \{0, 1\}$ be a binary variable indicating a facilitator's recommendation, where Z_{j0} indicates that the facilitator encouraged the jth DC to have the board directors appoint the DC manager, and Z_{j1} indicates that the trainer encouraged popular direct vote. Let $d \in \{0, 1\}$ indicate the *actual* selection rule a DC adopted, where d_{j0} indicates the adoption of an appointment method and d_{j1} indicates popular direct vote. Out of 50 farmer associations that were randomly sampled, 32 (64%) were encouraged to use appointments (Z_{j0}), and 18 DCs (36%) were encouraged to use direct elections (Z_{j1}). Take-up rates were high: 25 associations (78%) of the 32 that were advised to use representatives' vote followed (d_{j0}) this recommendations, and only 7 adopted instead a direct-vote rule (d_{j1}). Similarly, 17 of the 18 associations (94%) that were advised to use direct vote complied, and only a single DC did not. These high take-up rates further attest to the centrality of facilitator recommendations in the process of rule selection.

Following the literature on encouragement research design (Angrist and Krueger, 2001), I adopt APEP facilitator recommendations as an instrumental variable (IV). This allows me to identify an unbiased local average treatment effect (LATE) for "compliers": farmer associations that used direct vote, but would have used appointments had they been assigned a field-facilitator with different priors. This design requires satisfying several conditions (Angrist and Pischke, 2009, Ch. 4). First, the association between the instrument and the independent variable must be strong. Second, the instrument needs to be independent of potential outcomes. Third, the instrument must satisfy the exclusion restriction[37].

As for the association between Z and d, first-stage regression of leader selection rule on facilitator's recommendation produces a high and significant coefficient ($p\text{-}value = 0.00$). As for independence from potential outcomes, there are several good reasons to believe that the instrument (facilitator's recommendation) satisfies this condition. First, the allocation of facilitators to specific districts was "as good as random", since it was based on criteria that were *unrelated to spe-*

[37] There are no reasons to suspect that monotonocity and homogenous partial effects (Dunning, 2008*b*) – the fourth and fifth conditions – do not hold in our case.

cific group attributes[38]. Second, as mentioned above, interviews with field-facilitators claim to have based their recommendation on their prior beliefs regarding the appropriate selection method, not about the group they were interacting with.

What we require here are reasons to believe that facilitators did not tailor their recommendations to specific groups based on group attributes. Aside from their self-reports, how can we prove this is the case? First, had facilitators conditioned their advice on group characteristics, we should witness variability in their recommendations. The data shows, however, that field-facilitators made the *same* recommendation to *all* the DCs they helped create, including those that are out-of-sample (Appendix, Table 4.14). Second, farmer associations that were encouraged to use elections have similar pre-treatment covariates to associations encouraged to use appointments (Appendix, Table 4.13)[39]. The assignment to treatment Z_j, therefore, did not depend on factors such as the size of membership body or age of the association. Finally, it is worth noting that general member turnout to the DC general assembly in associations recommended to use direct vote is only slightly higher than turnout in associations that were recommended to use appointments (60% compared to 55%). This demonstrates that recommendations to use direct-vote (Z_{j1}) were not targeted only to those associations that have an engaged membership body, for instance.

There are still other concerns, however. For an unbiased IV estimation, the instrument Z_j must affect the outcomes of interest only via its impact on the instrumented variable, i.e., the choice of selection rule (d_j). There are two concerns with respect to possible violations of the exclusion restriction. First, it is possible that some facilitators are more competent than others and *that these differences are not orthogonal to* Z_j[40]. Second, facilitators may have taken additional actions that

[38]The first criterion guiding the allocation of facilitators to districts was local language proficiency. Since Uganda has almost 40 vernacular languages, and since English, the Ugandan official language, is not widely spoken in rural areas, the ability to communicate in a vernacular was a major concern for the APEP staff. The second criterion was the facilitator home district. APEP preferred to dispatch facilitators to areas other than their home districts. The rational was to reduce the likelihood that facilitators will engage in income generating activities outside the project.

[39]Unfortunately, APEP did not collect baseline data on the farmers, hence the small number of pre-treatment covariates at the association level. Given the large turnover in membership since the groups had formed, I was unable to use the sample of current members to assess balance.

[40]The key point is not whether facilitators differ in competence, but whether this difference is correlated with treatment assignment. To fix ideas, imagine how a controlled randomized trial would have unfolded. In an ITT design, groups would go through a similar facilitation process apart from the encouragement to adopt a certain randomly chosen leader selection rule. Even if assignment of groups to treatment is random, we cannot rule out the possibility that,

affected the outcomes of interest to this study, and *these actions may not have been orthogonal to* Z_j. For example, it is quite possible that facilitators who recommended the use of elections were more sensitive to accountability concerns, stressing governance issues more forcefully during the facilitation process.

The small number of facilitators (eight) as well as the inability to decouple facilitators' attributes from their recommendation or from district (fixed) effects are limitations of this identification strategy. Define γ as the direct effect of facilitator attributes on the study's outcome. Random assignment would have ensured that *in expectation* the impact of facilitation competence and sensitivity to governance issues can be ignored ($E[\gamma = 0]$), but not necessarily in a particular sample. This limitation is present even if we are willing to accept that the assignment of facilitators to districts is as good as random.

To address possible violations of the exclusion restriction, I enlist two approaches. First, to corroborate the study's findings I use results from the behavioral experiments I conducted as a measurement tool. In other words, while it is a plausible concern that facilitators influenced other aspects of group formation alongside leadership selection rule, it is less plausible that this influence should affect the experimental results other than through the institutions they helped put in place at the association's inception. Second, I re-estimate the IV regression using Conley, Hansen and Rossi (2010) sensitivity analysis. As I explain more formally in Appendix 4.C, this method allows me to conduct inference while relaxing the exclusion restriction assumption (i.e. assuming $\gamma > 0$).

Several features of APEP minimize, but do not eliminate, the extent to which violations of the exclusion restriction are of real concern. Primarily, the raison d'être of the project was group formation. Once a farmer group was formed, facilitators moved on to the next village to form another group. The large area of coverage as well as the large number of groups and associations under the purview of each facilitator necessarily limited the interaction between facilitators and groups once group formation was complete. This feature of APEP seriously constrained the ability

in practice, facilitators had different 'facilitation' abilities (see Humphreys, Masters and Sandbu (2006)).

of facilitators to influence outcomes after the initial setup of DCs. Also APEP maintained a very strict "no-handout" policy, discouraging field staff from going beyond their role as facilitators. This further reduced facilitators' ability to exert direct influence on associations outside of their influence during the brief facilitation period. A more formal treatment of the paper's identification strategy can be found in Appendix 4.A. In sum, the instrument of facilitator recommendation satisfies requirements for IV regression necessary to justify the analysis presented here.

4.5 Sampling and Data Sources

This section briefly describes the data used in this paper and how it was collected. To reduce crop-related variability, I first limited the study's target population to associations marketing coffee, the most common cash crop sold by the APEP groups. I then sampled 50 associations from 5 district-areas (strata) using a stratified, random, multistage cluster design[41]. Quantitative data for the empirical analysis was collected between July and September 2009 by a team of local interviewers.

Within each association, several different types of data were collected. At the association level, data was collected using a questionnaire completed in an interview with members of the DC executive committee. Data on the DCs' economic activities were also collected from association books and records. For each association, I sampled six producer organizations (POs), for a total of 287[42]. I also conducted interviews with the leaders of sampled producer organizations, from which I derive more group-level data. Additionally, I collected individual-level data: From each sampled PO, I further sampled approximately six members for a total of 36 members per association[43]. Sampled members were surveyed in person by trained interviewers in the respondents' local language for a total of 1,781 surveys. I refer to this data source as the "members' survey".

The set of PO representatives to the board of directors, including members of the executive

[41] Additional information on the sampling scheme, including a map locating DCs in the five strata, can be found in Supplementary Appendix A.

[42] In a few cases, when a farmer association had fewer than seven POs, I selected all of its village-level groups.

[43] The number of sampled members from each of the six *sampled* village-level groups was proportional to the size of the PO. This assured that the sample is self-weighted.

committee, form 'the complete network of DC representatives'. This network is also the complete pool of potential candidates for the DC manager's position. A significant effort was made to survey all members of the network of DC representatives. In each sampled farmer association I surveyed (i) the four DC executives, (ii) the chairmen of all village-level POs, whether or not their group was sampled, and (iii) one or two representatives from each village-level PO, irrespective of whether their PO was sampled. Thus, I have individual-level data on the complete set of potential candidates for the senior management position, for a total of 1,316 interviews. These "representatives' surveys" only partially overlap with the members' surveys, as they were tailored to capture the roles and responsibilities of representatives within an association's organizational structure.

In addition, each of the representatives completed a social network module based on a roster of the names of all DC representatives. This social network data allows me to analyze both the structure of the DC leadership network (e.g., its density), as well as the position of each member within the network (Wasserman and Faust, 1994). The survey team visited each association up to four times to reduce attrition, which was brought down to less than ten percent.

4.6 Results

In this section I report the study's main findings. The first set of results correspond to Hypothesis 8, namely whether elected leaders are more accountable and responsive than appointed leaders. Accountability refers here to the subjection of leaders to rewards and sanctions based on their behavior in office. Responsiveness refers to an incumbent's modification (or maintenance) of her behavior according to signals of group members' preferences (Przeworski, Stokes and Manin, 1999). Using a set of behavioral measures, I demonstrate that direct-elections result in leaders who are more accountable and responsive to association members than appointed leaders.

To corroborate this finding, I report a second set of results concerning perceptions of leader accountability both by members and by board directors. These perceptions are consistent with the finding that elected managers are more accountable and responsive than appointed managers.

The third set of results are tests of Hypothesis 7, which states that appointed leaders will have 'higher profiles' than elected leaders. Higher profile refers here to attributes that likely increase the ability of the manager to perform successfully her roles and responsibilities, thus increasing their likelihood of providing more highly valued public goods. I find no evidence to support Hypothesis 7 across any of the three dimensions of a leader's profile: ability, network centrality and other-regarding preferences.

Monitoring, Accountability and Responsiveness

I begin by testing whether DC leader accountability is a function of leader selection rules. Both the treatment variable (selection rule) and the variables used to measure monitoring, accountability and responsiveness are discrete. There are two common parametric approaches for estimating causal effects in such cases. One approach uses Linear Instrumental Variable estimation strategy that disregards the binary structure of the outcome and treatment variables (Altonji, Elder and Taber, 2005, Imbens and Angrist, 1994). The second computes maximum-likelihood estimates of a bivariate probit (BP) model, which assumes that the outcome and treatment are each determined by latent linear models with jointly normal error terms (Maddala, 1983, Neal, 1997).

For each of the dependent variables tested in this section I estimate both BP and linear IV models. The BP models allow me to report Average Treatment Effects (ATE) and Average Treatment Effects on the Treated (ATT), whereas the linear IV models allow me to report Local Average Treatment Effects (LATE). Following Chiburis, Das and Lokshin (2010) recommendation, I bootstrap standard errors in all parametric regression models. I adopt the most conservative approach, adjusting bootstrapped standard errors to account for clustering at the field-facilitator level. A more technical exposition of the linear IV and BP estimators can be found in Appendix 4.A.

Given the discrete nature of the dependent variables, the increase in probability attributed to a one-unit increase in a given predictor is dependent both on the values of the other predictors and the starting value of the given predictors. For this reason the interpretation of regression coefficients is not straightforward. To help appreciate the substantive significance of the models'

coefficients, I report and graph predicted probabilities in the main text and provide regression tables in Appendix 4.B[44].

Monitoring and Accountability

I use two measures of accountability. The first measure includes two parts, namely the extent to which an association employs external and/or internal auditors. This data is derived from interviews with the complete set of board directors. External auditors are licensed accountants who are *non-members* hired by the association specifically to conduct an audit of the group's books. Internal auditing refers to requests by board directors to personally view the association's books and records. I find that the predicted probability that board directors report hiring external auditors in appointment-based associations is 0.22 $(0.14 - 0.32)$. This figure more than doubles in the case of direct vote associations (0.53, ranging from $0.34 - 0.71$ at the 95% confidence interval). This difference is substantial and significant (*p-value*$= 0.00$). Regarding internal audits, I find that the predicted probability that a board director has ever asked to review the associations' books is almost 50% higher in direct vote associations (0.29, with $0.23 - 0.36$ at the 95% confidence interval), than in associations in which the DC manager is appointed (0.20, with $0.15 - 0.26$). This difference is both substantial and significant at conventional levels (*p-value*$= 0.09$)[45].

The second measure of accountability is the frequency of selection procedure. Interviews I conducted prior to the enumeration activity indicated that a non-negligible number of associations fail to hold a leader selection process — whether appointments or elections — in the frequency stipulated in their constitution. In other words, though leaders are selected for fixed terms, not all associations hold selection processes for the senior leadership position at the end of the fixed term. Importantly, this failure can be a source of contention, especially among aspiring board directors. Using data collected at the DC-level, I find that associations that use direct vote have held selection processes for the DC manager position more frequently than appointment-based

[44]Clustering standard errors at the field-facilitator level usually yields larger standard errors than clustering at the farmer association level. Throughout the section, predicted probabilities are calculated by using the largest standard errors, holding continuous variables at mean values and categorical variables at their median values.

[45]The full-set of regression results can be found in the Appendix, Table 4.7.

associations, controlling for pre-treatment covariates such as the association's age and original number of POs (*p-value*$= 0.00$). This result provides additional support to the finding that popular direct elections result in more accountable leaders.

Remarkably, the managers themselves attest that direct-vote associations are better at holding the DC leaders to account. Directly elected leaders are *more than two times* more likely than appointed leaders to report being warned by other board directors for not doing their job well or for not putting enough effort (*p-value*$= 0.01$)[46]. This finding, reported in Table 4.2 in Appendix 4.B, further increases our confidence in the reliability of the above accountability measures.

Responsiveness

I now turn to test whether higher levels of leader accountability are associated with greater responsiveness to ordinary members. I use two measures of leader responsiveness. The first measure is a binary indicator of whether the DC manager writes receipts to members who sell their crops through the group. Since members receive payments only about three weeks after they deliver their crops to the farmer association, written receipts serve as a guarantee against possible exploitation. Since written receipts constrain managers' ability to take advantage of ordinary members, the likelihood that members obtain receipts when selling in bulk can be interpreted as the extent to which managers acquiesce to a popular demand.

Using the sample of ordinary members, the predicted probability that a member in a direct-vote association will report obtaining receipts from her manager when selling in bulk is 0.94 $(0.92 - 0.95)$. This figure drops to 0.65 in the appointment-based associations ($0.44 - 0.82$ at 95% confidence intervals). This difference is significant (*p-value*$= 0.00$) and substantial. For robustness, I corroborate these results with interviews with the complete set of board directors. In direct-vote associations the predicted probability that a board director would claim that the manager provides receipts "at least sometimes" to members who sell in bulk is 0.92 $(0.88 - 0.95)$. This figure drops

[46]The question wording is: "where you ever warned by other PO or DC leaders because you were not your job well, or not putting enough effort"?

to 0.59 $(0.40 - 0.76)$ for appointed leaders. This difference is, again, significant and substantial. That the IV regression results, obtained from two different sources, produce very similar findings further increases our confidence in the study's conclusions.

Following Huber and Gordon (2004) who use the severity of sentences to measure judges' responsiveness, the second measure of leader responsiveness is the extent to which social sanctioning is used to enforce group norms of cooperation with respect to collective marketing. Elsewhere I have discussed at length the collective action problem associated with collective marketing, the single most important public good produced by the farmer group (Grossman, 2011*b*). It is sufficient here to recall that the price the DC leader can obtain from buyers depends crucially on the volume and quality of the group's coffee. To mitigate the social dilemma associated with quality enhancing investments, DCs commonly pass resolutions and bylaws requiring members to use best-agricultural practices[47]. To mitigate the social dilemma associated with marketing choice, DCs forbid members from side-selling to middlemen. Since the DC leader's prime responsibility is organizing and coordinating the group's marketing activity, he is also responsible for enforcing those rules and bylaws[48].

I find that compared to members in appointment-based associations, members in direct-vote associations are significantly more likely to report being sanctioned for producing low quality crops, or for side-selling. For example, the predicted probability that a member in direct vote association reports ever being sanctioned for side-selling is 0.27 $(0.23 - 0.32)$. This likelihood is about 2.5 times larger than the predicted probability of members in appointment-based associations (0.11, with $0.08 - 0.16$). Similarly, the predicted probability that a member in a direct-vote association reports ever having been sanctioned for not using "best agricultural practices" is 0.41 $(0.36-0.47)$, but only 0.27 $(0.2-0.35)$ when the DC leader is appointed. In both cases the difference in sanctioning probabilities is significant at the 95% confidence intervals.

[47] Best agriculture practices refer, for example, to the frequency of weeding, mulching and pruning, and to the methods used for picking and drying coffee beans. Sanctioning here commonly takes the form of rejecting member's coffee from participation in bulk marketing. Other punishments include public shaming during group meetings.

[48] That leaders are responsible for ensuring that rules and bylaws are enforced does not necessarily indicate that they execute punishments. Usually PO leaders will sanction defectors during a PO meeting through "naming and shaming". DC leaders, however, encourage and coordinate this type of decentralized social sanctioning.

Perceptions of accountability and responsiveness

The variables I use to measure leaders' accountability and responsiveness were chosen following dozens of semi-structured interviews with both members and board directors in the months leading up to full survey enumeration. However, in order to ensure that the interpretation I give to the above variables – e.g., writing receipts or hiring external auditors – aptly captures monitoring and accountability, I asked the survey respondents an additional set of questions designed to elicit their perceptions of their leader's level of accountability.

First, sampled members were asked whether they believe that board directors are monitoring the actions and activities of the DC manager. Consistent with the behavioral results, I find that members in direct-vote associations are significantly more likely to perceive their manager to be monitored. The predicted probability that an ordinary member reports that her elected leader is sufficiently monitored is 0.57 ($0.51 - 0.62$); this figure drops to 0.48 in the case of an appointed leader ($0.43 - 0.53$). To confirm these results, I asked the board directors if "there is anyone who is responsible for making sure the DC manager does their job diligently and transparently?" In direct-vote associations, the predicted probability that a board director responded positively is 0.71 ($0.61 - 0.79$), but only 0.57 when managers are appointed by the board ($0.48 - 0.65$).

Finally, members were asked how accountable or unaccountable they believe their DC leadership to be, in general. In direct-vote associations, the predicted probability that a member perceives the leadership to be *unaccountable* is 0.05 ($0.03 - 0.07$), which is significantly lower than in the case of appointed leaders (0.13, with $0.08 - 0.20$ at the 95% confidence intervals). Similarly, the predicted probability that the manager is perceived to be *very accountable* is only 0.21 in appointment-based associations ($0.17 - 0.25$), against 0.33 in direct-vote associations ($0.30 - 0.36$). Figure 4.1 and Fig. 4.2 provide graphical representations of this section's results for board directors and sampled members, respectively.

The validity of variables used to measure abstract concepts such as monitoring, accountability and responsiveness can naturally be questioned. However, the fact that results are consistent across a large set of behavioral and attitudinal measures, as well as across different data sources,

improves confidence in the study's main finding: leaders are more closely monitored and deemed more accountable where members participate directly in their selection.

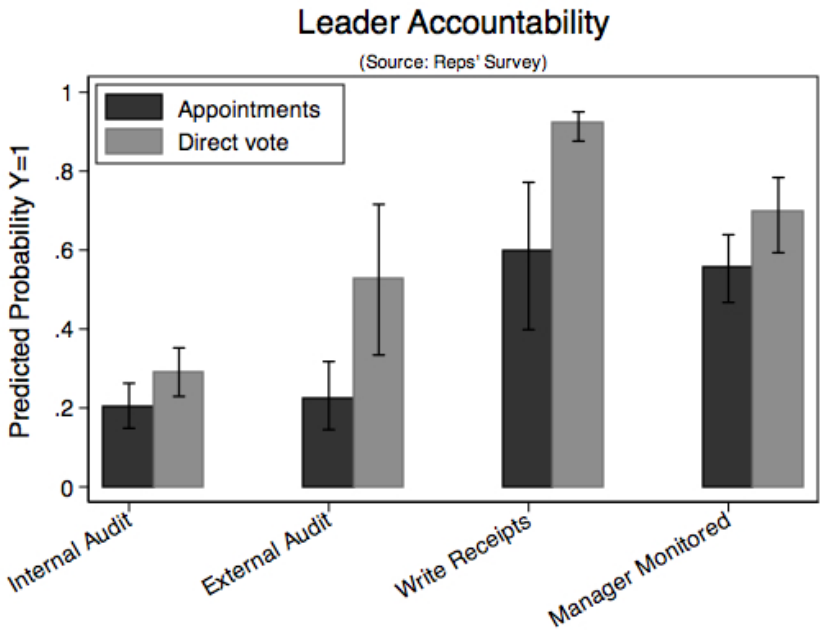

Figure 4.1: **Leader selection rule and accountability (representatives' survey).** Predicted values are estimated by setting continuous control variables to their mean values and categorical variables to their median values. Standard errors are bootstrapped using 1,000 iterations, adjusted for the clustering nature of the data (8 strata for field-facilitators).

Leader Profile: Ability, Socio-Demographics and Network Position

In this section I turn to test Hypothesis 7, which states that, compared to directly elected managers, appointed managers have a 'higher profile': attributes that likely increase the ability of managers to perform successfully their roles and responsibilities.

Since each DC has a single manager, the analysis in the section is based on a relatively small sample size ($n = 50$). The small sample size is especially problematic in IV parametric estimations, which can be biased when asymptotic property assumptions are not met (Chiburis, Das and Lokshin, 2010). This problem is compounded by the clustered nature of the data (i.e. associations nested within districts and attached to a facilitator), resulting in low statistical power. To mitigate

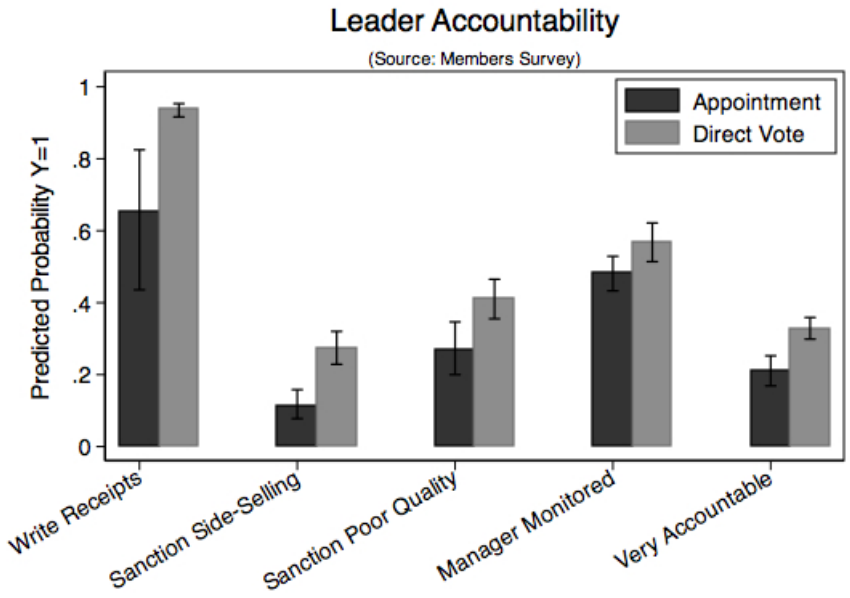

Figure 4.2: **Leader selection rule and accountability (members' survey).** Predicted values are estimated by setting continuous control variables to their mean values and categorical variables to their median values. Standard errors are bootstrapped using 1,000 iterations, adjusted for the clustering nature of the data (8 strata for field-facilitators).

concerns about these problems, I present two sets of results. First, I report Linear IV estimations, in which standard errors are bootstrapped using a large number of iterations[49].

Second, where there are reasons to fear possible type II errors, I report p-values based on non-parametric randomization inference tests. Recall that the counterfactual that informs the identification strategy is "what if the facilitator assigned to area S would have a different preference with respect to the appropriate selection rule". Since there are only eight facilitators in the study-area, with 3 recommending direct-vote (Z_{j1}) and 5 appointments (Z_{j0}), there are 56 possible combinations of "shuffling" facilitators between different S. By reshuffling Z_j, randomization inference allows to compute exact p-values by comparing observed statistics (here, mean difference $\bar{Y}_{j1} - \bar{Y}_{j0}$) to all possible mean differences *in the sample* (Keele, McConnaughy and White, 2008).

[49] Following a personal communication with Jeffery Woodbridge, when conducting analysis at the association-level, I use bootstrapped standard errors based on $10,000$ iterations that do *not* take into account possible clustering in the data generating process. Given that facilitators' recommendations (Z_j) are collinear with districts, and given the small number of clusters, additional clustering results in unreliable standard errors. Bootstrapped standard errors that ignore higher-order clustering turns to be more conservative, since bootstrapped standard errors that accommodate clustering are consistently lower than the ones reported in this section.

Leader ability

I use four proxy measures of manager's ability that are strongly and positively correlated with marketable skills in Uganda[50]: (i) manager's educational attainment on a six-category scale; (ii) English proficiency[51]; (iii) score on two cognitive tests[52]; and (iv) an ability composite index that is constructed from the above variables via principal component analysis[53]. In addition, I ranked the network of board directors according to their standardized ability composite index score, and tested whether appointment-based associations result in managers with higher *rankings*.

Using a Linear IV estimation strategy, I find little evidence in support of Hypothesis 7: appointed leaders do *not* score significantly higher than elected leaders on five out of six proxy measures of ability (Appendix, Table 4.8). First, differences between appointed and elected leaders with respect to English proficiency (*p-value*$= 0.41$), the composite ability index (*p-value*$= 0.84$) and managers' ability *ranking* (*p-value*$= 0.69$) are highly insignificant. Second, I find that appointed leaders score slightly *lower* on both cognitive tests. Lastly, though I find that appointed managers attained slightly higher education levels (at the 90% confidence level), exact p-value from a randomization inference test is not significant at conventional levels (0.16, see Appendix, Fig. 4.11). Taken together, the evidence suggests that the failure to reject the null hypothesis – that the impact of selection rules on ability is no different than zero – is likely *not* due to low statistical power.

In addition, I do not find evidence in the data that appointment-based associations result in DC leaders with other socio-demographic attributes that might enhance leader effectiveness, such as age, religiosity (measured as attendance in religious services), richness of associational life (mea-

[50]Fig. A.5 in the Supplementary Appendix A demonstrates the strong positive correlation between the composite ability measure and job status. Similarly, Fig. A.6 demonstrates the positive correlation between the composite ability measure and wealth. These correlations increase our confidence that the composite ability index captures relevant marketing skills.

[51]Because no local language is spoken by more than 20% of Ugandans, English is the lingua franca of the business and political class. English proficiency allows individuals to communicate with potential trading partners outside their small geographic areas.

[52]The cognitive tests included two assignments: solving a simple maze in less than two minutes and solving a raven test comprised of 12 questions in two minutes. For additional information see the Supplementary Appendix.

[53]All of the variables that are included in the ability index are positively correlated, and the first principal component was able to explain more than 61% of the variance. The ability index included also indicators for respondents' ability to read and write. I do not report these in Fig. 4.10 since all DC managers are literate.

sured by the number of social, economic and political groups and associations in which the leader participates actively), total land size in acres, and experience (measured as the number of years the leader has been growing coffee.) The full set of Linear IV regression results are reported in Table 4.9 in the Appendix.

Leaders' centrality of network position

Finally, I test whether appointed and elected leaders differ in their network centrality position. Centrality indices measure the importance of a node within a network. The intuition here is that the more central an actor is within a well-defined network, the better positioned she is to coordinate activities and transmit information between other network actors. Since much of the work of DC managers revolves around coordination and dissemination of information, the manager's node centrality is likely to be positively correlated with leader performance (Mehra et al., 2006). As briefly mentioned above, to test whether the DC leader's network position is a function of leader selection rule, I have collected network data using a complete roster of board directors in each of the 50 sampled associations[54]. In each DC, network data was collected the following three dimensions[55]: (i) Friendship[56], (ii) Prior Acquaintance[57], and (iii) Advice[58].

One natural way to measure the centrality of a node within a network is degrees – the number of direct connections a node has. Degree centrality builds on the basic intuition that a prominent network actor is one that has high involvement in many relations (Freeman, 1979). Since the three types of relation – friendship, prior acquaintance and advice – are directional, I measure the Indegree centrality or "Prestige" of all nodes: the extent to which a social actor in a network receives or serves as the object of relations sent by others in the network (normalized by the number of direct

[54]The term 'complete' refers here to the fact that the board of directors is a clearly defined organizational structure with clear and transparent membership rules. Using a simple roster, each board director was asked to provide information on his or her ties with all other directors.

[55]Network questions were asked only when respondents answered positively a basic sorting question "Do you know [NAME]?"

[56]Wording: "Is [NAME] a close friend or do you just know him or her? by close friend, I mean that you (a) eat together regularily; (b) you can leave your child with him or her if you need to travel for several days; and (c) he or she will help you in case of family death".

[57]Wording: "Did you speak to [NAME] on a regular basis before the creation of the DC?"

[58]Wording: "In the past 12 months, have you asked [NAME] for information or advice on matters related to the DC?"

ties). More formally, Prestige, or Indegree centrality, is defined as $C_D(n_i) = \frac{d_i(n_i)}{g-1}$ where $d_i(n_i)$ is the indegree of a node n_i and g is the number of all nodes in the network.

Another common centrality measure is 'Betweenness', which is arguably among the most important centrality indices (White and Borgatti, 1994). In most basic terms, Betweenness measures the extent to which a given node lies on non-redundant shortest paths (geodesics) between third parties, and is therefore concerned with how actors control or mediate the relations between dyads that are not directly connected. The more often that actor i is located on the shortest path between numerous dyads, the higher is the actor's potential to control interactions and information flows.

I find no evidence in the data that appointed managers interacted more frequently with other directors on a regular basis before the creation of the DC than elected leaders did. Similarly I do not find that appointed managers are nominated more frequently as friends or confidents of other DC board directors – attributes that are positively correlated with successful DCs. Thus I do not find support for hypothesis 7; appointed rulers are no more likely to have higher profiles than elected leaders. The full set of regression results is reported in Appendix 4.E, Table 4.10.

4.7 Causal Mechanisms

The paper's main finding is that, at least in the Ugandan farmer associations I study here, directly elected leaders of self-help associations are significantly more accountable and responsive than appointed leaders, even though they do not differ significantly in their observable attributes. In this final section of analysis, I turn to investigate the mechanisms that mediate between elections and accountability. Specifically, I test two possible mechanisms. The first mechanism is rooted in a rich social psychology literature that studies the way procedures condition the relation between leaders and followers in small-groups situations. Informed by this literature, I use a set of 'lab-in-the-field' behavioral experiments to test whether elections create higher expectations of elected leaders, compared to appointed leaders. Here, the impact of leader selection rule is thought to operate at the individual-level.

The second mechanism I test is derived from the corporate governance literature. Here I use social network analysis to test the hypothesis that a growing familiarity between appointed officials and appointees creates a governance culture inimical to accountability (Shaub et al., 2005). According to the "familiarity hypothesis", the impact of leader selection rules is thought to operate at the organizational level.

Reciprocal Expectations as expressed in behavioral games

Social psychologists have long argued that the method through which a leader obtains its authority creates different psychological environments for *both the leader and followers* (Read, 1974). Hollander and Julian (1970) argue that, compared to appointments, elections increase popular *demands* on the leadership role. This is because when group members elect a leader, the group is believed to support one specific candidate over others and thus in a way transfers hopes and expectations upon that individual. Ben-Yoav, Hollander and Carnevale (1983) also advance a demand-side explanation, demonstrating experimentally that subjects expect elected leaders to be more responsive to the needs of followers. More recent work has focused on supply-side explanations. Elgie, Hollander and Rice (1988) find that elected leaders express, on average, a greater sense of commitment and indebtedness towards group members. They stipulate that this is because members are the source of elected leaders' legitimacy. Finally, De Cremer and van Dijk (2008) use a laboratory experiment to test the hypothesis that leadership selection rules determine the share that leaders retain for themselves from a common resource. They not only find that appointed leaders take significantly more of the common than their elected counter-parts, but they also find that subjects are more accepting of norm violating behavior by appointed leaders.

In what follows, I employ a series of behavioral experiments to test the hypothesis that elections change the *reciprocal expectations* between leaders and member[59]. The use of behavioral experiments serves two purposes. First, the experiments' strong internal validity allows for clear

[59]In the above laboratory studies, the identity of the leader as well as the group affiliation were manipulated by the researcher. I improve on these studies here by using natural groups and leaders. See Grossman (2011*b*) for a discussion on external validity in the context of lab-in-the-field experiments.

identification of a well-defined causal relation (Berkowitz and Donnerstein, 1982). Second, recall the concerns regarding possible violations of the exclusion restriction. It seems quite unlikely that a facilitation process that took place 4-5 years ago will directly affect subjects' play in the behavioral experiments.

The first experiment I conducted is a dictator game (DG), which allows me to obtain measures of *conditional altruism*. In DGs, two anonymous players are allotted a sum of money – in this case, 10 coins of 100 Ugandan Shillings. The first player ('the decider') offers a portion of this sum to a second player, who cannot reject the offer ('passive recipient'). The decider thus dictates the portions of the stake allocated to herself and to the passive recipient. In this one-shot, anonymous game, a purely self-interested individual would offer zero. It is common to interpret offers in the basic DG as a measure of altruism, as it is not directly linked to kinship, reciprocity, reputation, or the immediate threat of punishment (Camerer, Loewenstein and Rabin, 2004).

In my version of the DG, each of the DC managers was asked to make two separate allocation decisions: one to an anonymous resident of the sub-county (i.e., 'stranger' condition) and one to an anonymous member of the farmer association[60]. Figure 4.3 summarizes the results of this experiment (see also Appendix, Table 4.11). Note that the treatment in the DG is whether a leader was elected or appointed, which is instrumented by the facilitators' recommendation. In other words, the difference between allocations of appointed and elected leaders should be thought of as LATE.

First, I find that elected and appointed managers allocate a similar amount to 'strangers'. The predicted allocation of elected managers to strangers is 213 USH, whereas appointed managers allocate an average of 250 USH ($p\text{-}value = 0.53$)[61]. These allocations can be interpreted as a base measure of altruism, where social distance has been maximized. Importantly, the results of the 'stranger' variant corroborate our earlier finding that elected and appointment methods produce leaders with similar attributes – in this case other-regardness preferences.

[60] Members were told that they were receiving an allocation from one of the 'leaders of the DC'. Managers were informed prior to making their allocation decision exactly what members were being told about the identity of the sender.

[61] Predicted values are calculated by setting control variables to their mean values.

Though both elected and appointed leaders allocate, on average, to unknown group members more than to 'strangers', the effect of shortening the social distance on elected leaders is much more pronounced. Whereas appointed leaders allocate to anonymous members 339 USH, elected leaders allocate to members 70% more (574 USH, *p-value* = 0.09). Since DC managers were aware of the fact that the identity of the sender is unknown to the recipients, these results provide further evidence of a supply-based explanation, which cannot necessarily be reduced to reelection considerations. Namely, at least in the context of self-help associations, popular direct elections seem to trigger a stronger commitment from the part of local leaders, who reciprocate their election.

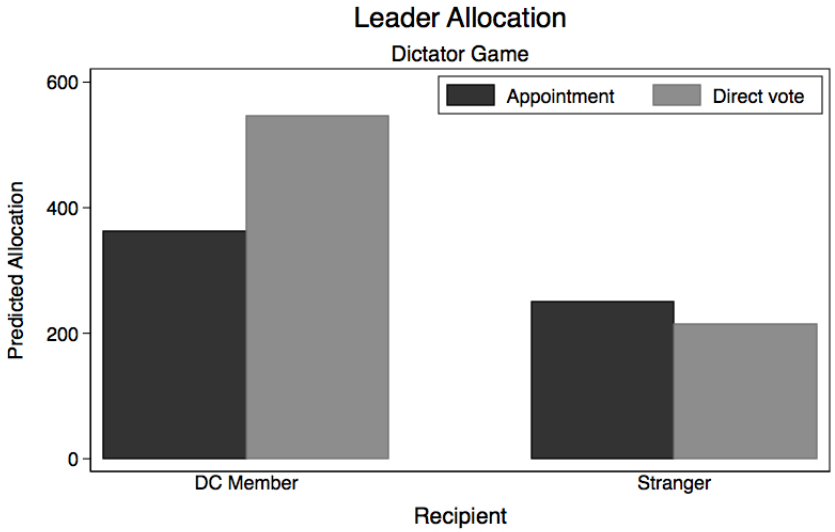

Figure 4.3: **Dictator allocation by Leader Selection Rule.** Predicted values are based on a Linear IV Estimation in which the DV is modeled as a function of an association's leader selection rule, instrumented by the DC facilitator's recommendation. Standard errors are bootstrapped using 10,000 iterations.

Results from a second set of experiments — a series of Third Party Punishment experiments (TPPs) — support this hypothesis. In TPPs, two players are allotted a sum of real money (10 coins of 100 Ugandan Shillings), and a third player gets one-half of this sum (5 coins). Player 1 (the 'decider') must decide how much of the stake to transfer to Player 2 (a 'passive recipient', who makes no decisions in the game). Then, before learning the actual amount Player 1 transferred to player 2, Player 3 must decide whether to pay 1 coin (which is 20% of his/her endowment) to

punish Player 1, causing Player 1 to suffer a deduction of 3 coins, for all possible transfers Player 1 might make. Thus I use the conventional strategy set protocol, (Henrich et al., 2010), in which punishment strategy of Player 3 is elicited *prior* to Player 1 actual allocation for all possible offers.

The game theoretical prediction of TPP is straightforward: Since the experiment is an anonymous one-shot interaction, a purely self-interested Player 3 would not pay to punish Player 1. Knowing this, a purely self-interested Player 1 should always offer zero to Player 2. Since the payoff of Player 3 depends only on her punishment decisions, an individual's willingness to punish provides a direct measure of the subject's willingness to suffer a personal cost in order to enforce a norm of fairness between *other* group members.

In my version of the TPP experiment, both the decider and the passive recipients were ordinary, randomly selected association members. The third player (the 'punisher') was one of the four DC executives[62]. Results are presented below in Fig. 4.7 and in Table 4.12 in the Appendix. The magnitude of these effects is striking. In all unequal allocations, directly-elected leaders are about two to three times more likely to punish deciders who kept more than 6 coins for themselves. Since punishment decisions are costly, and because DC leaders were aware of the fact that members do not know the identity of the monitor, these results further strengthen the hypothesis that the procedure of elections trigger a stronger obligation to reciprocate on the part of leaders. One of those manifestations is a stronger commitment to engage in costly punishment of members violating norms of cooperation.

The TPP findings can, therefore, help explain the observational data reported above: that compared to appointment-based associations, members in election-based associations are more likely to report being sanctioned by association leaders for producing low quality coffee, or for side-selling to middlemen.

[62]I report only results for the DC managers. Note that DC executives knew that players 1 and 2 are members, members knew that player 3 was 'one of the DC leaders', and players 3 knew what Players 1 and 2 knew.

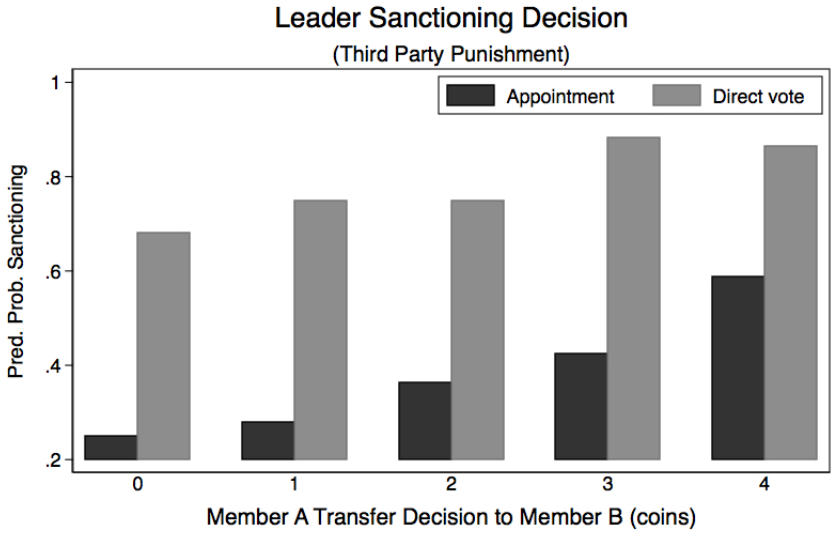

Figure 4.4: **Third Party Punishment by Leader Selection Rule.** Predicted values are based on a Linear IV Estimation in which the DV is modeled as a function of an association's leader selection rule, instrumented by the DC facilitator's recommendation. Standard errors are bootstrapped using 1,000 iterations. On the X-axis are the number of coins allocated by player 1.

Testing for a "familiarity effect"

I now turn to test for the presence of a "familiarity effect," wherein a growing familiarity between appointed officials and appointees creates a governance culture inimical to accountability. I use social network data collected for each of the 50 sampled boards of directors to test this mechanism. The analysis proceeds in two steps. First, I test whether the density of the friendship and the advice network differs between associations that use different leader selection rules. If a "familiarity effect" exists, we should see that a board of directors using appointment rules has a denser friendship network[63].

Using a linear IV regression, I find that the (Local Average Treatment) effect of moving from appointments to elections on the density of the friendship network is -0.49 (p-$value$= 0.00). Put differently, the friendship network density of boards of directors that use appointments is 53% larger

[63]Network density is one of the most widely used measures of social network structure (Wasserman and Faust, 1994). Density describes the general level of linkages among the points in a graph, and is defined as the number of actually occurring relations or ties as a proportion of the number of theoretically possible relations or ties.

than the friendship network density of boards of directors using direct-elections. Interestingly, I find that in associations in which leaders are directly elected, the *advice network* of board directors is denser, though the difference is not significant at conventional levels ($p\text{-}value = 0.31$). Secondly, I test for the existence of a correlation between the density of the friendship and advice networks and an association's level of monitoring and accountability. As Fig. 4.5 makes clear, the more dense the friendship network of board directors is, the *less likely* is the board is to use internal auditors (left panel), external auditors (middle panel), or to ensure that the manager writes receipts to members selling in bulk (right panel). By contrast, as demonstrated in Fig. 4.6 the more dense the advice network of board directors is, the *more likely* the board of directors is to perform its monitoring function. These findings are consistent with the idea that appointments create affinity between appointees and appointed officials that can be detrimental to monitoring.

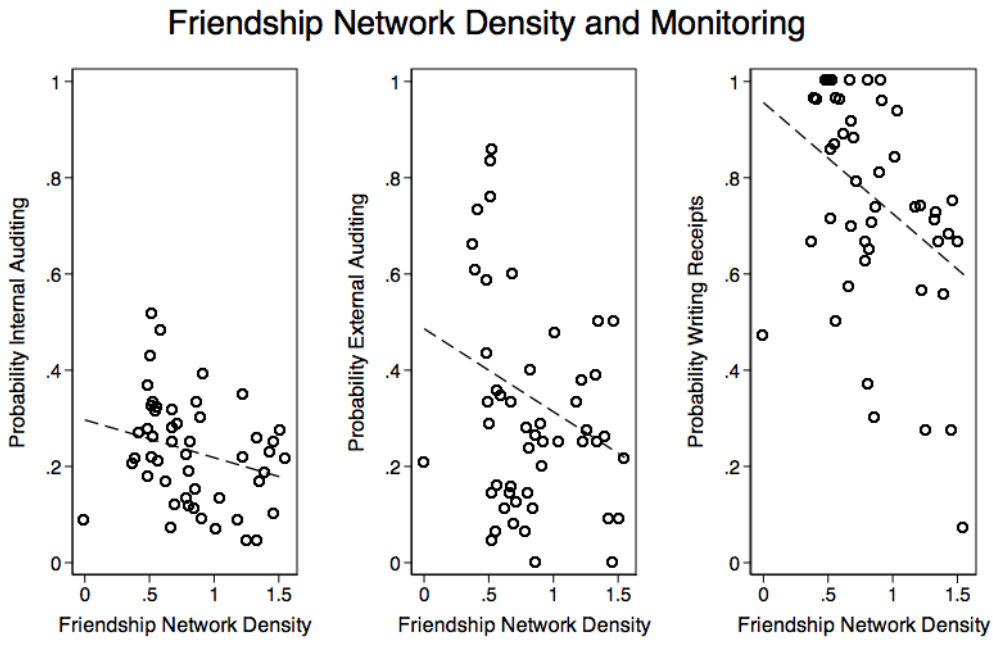

Figure 4.5: **Friendship Network Density and Monitoring.** Line represents the prediction for the DV (monitoring) from a linear regression of monitoring on network density.

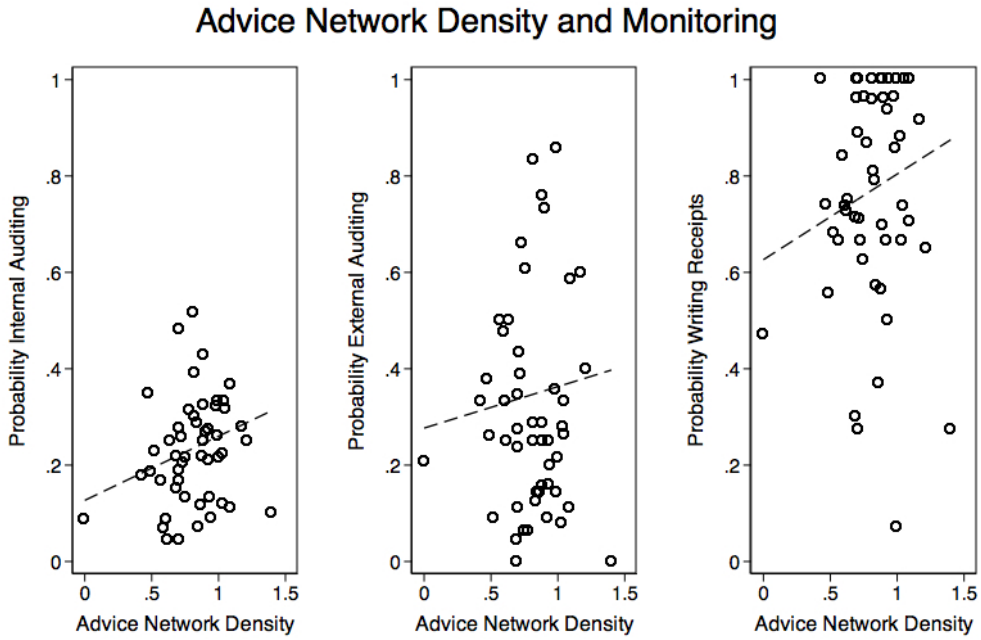

Figure 4.6: **Advice Network Density and Monitoring.** Line represents the prediction for the DV (monitoring) from a linear regression of monitoring on network density.

4.8 Conclusion

This paper studies the impact of varying leaderships selection rules—appointment or election—within the context of local, multi-tier self-help associations. Past work on leadership selection rules has been constrained by pervasive identification problems. This study addresses the identification challenge by exploiting unique conditions that resulted in plausibly exogenous variation in leader selection rules within Ugandan farmer associations.

Results presented here suggest that direct elections have a positive, substantial and significant effect on the accountability and responsiveness of local leaders, at least in this context. Put simply, governance institutions matter. This finding is consistent across a large number of behavioral and perceptional measures and holds across a variety of different data sources, including members, board directors and DC managers. This increases confidence in the study's main results.

Turning to causal mechanisms that might link elections to accountability, results from a set of

behavioral experiments provide relatively strong evidence that elections increase the commitment of local leaders towards ordinary group members, independently of reelection considerations. One mechanism that might explain this finding concerns reasons why appointed leaders *fail* to monitor as successfully as their democratically elected counterparts. Using original social network data, I present evidence of an 'affinity effect,' whereby appointment corresponds to affinity between appointees and appointed officials inimical to monitoring and accountability.

By contrast, I find little evidence that leadership selection rules impact the profile of selected leaders. Contrary to the expectations of the facilitators who recommended appointments, appointed leaders do not posses observed profiles or attributes that are more conducive to successful fulfillment of their roles and responsibilities. Recalling the study's theoretical framework, that appointed leaders do not have 'higher' profiles than elected leaders indicates that information asymmetry problems are not severely aggravated by moving from an appointment to an election-based rule. Two possible explanations might account for this. First, it is quite possible that parishes in Uganda are small enough to remain information-rich environments, and thus villagers have sufficient knowledge of potential candidates. Alternatively, it is possible that where popular direct elections take place, PO members consult their more informed representatives before casting their vote. In this scenario, election outcomes closely mirror appointment outcomes, since the opinions of local representatives carry a disproportionally large weight. Future research should investigate these dynamics in greater detail.

Findings presented here also indicate that *low saliency* is not a major concern, at least in the this context. This may be because the selection of an association leader has significant implications for the livelihood of members, as was described to me in innumerable interviews. Another possible reason for the relatively high salience of leader selection is that popular elections take place in the context of annual or biannual general assembly meetings. Members also vote on by-laws and resolutions, and receive useful information about their association's activities in these meetings. The fact that turnout in general assembly meetings exceeds 60% lends support to this interpretation.

There are several important caveats to the study's key findings, including external validity,

sample size, and length of study period. On the first point, there is some reassurance in the fact that main findings presented here are consistent with results from the study of Indonesian Village Councils (Olken, 2010) and of judicial decisions in the US (Huber and Gordon, 2004). The extent to which the results of this study apply to other types of associations, however, or to countries with different income levels is an open question for future research to address.

Concerning sample size, it should be repeated that this study is based on a relatively small number of associations (50), which were founded with the help of a small number of field facilitators (8). Though the results point to large and statistically significant effects of leader selection rules on accountability and responsiveness, some caution must be used in interpreting the lack of impact on leader profiles. That said, even if such effects exist (though they are not captured by randomization inference tests or by the Linear IV estimation), they are evidently much smaller in magnitude than the effects of direct-voting on accountability.

Finally, this paper considers only short-term effects of direct election on accountability and leader profiles. It is possible that in the long run, strategic adaptations and accumulative knowledge may impact the results. In addition, the long-term effectiveness of leaders with respect to the production of club goods and services — not addressed in this chapter — may further alter the evaluation of the different leadership selection rules over time. Each of these caveats points to fruitful avenues for future research. At present, this study's findings contribute to the broader debate over leadership selection rules. Whether by limiting "affinity effects" or by changing the reciprocal expectations between leaders and followers, this study finds that allowing group members to directly elect their own leaders has a strong, positive impact on leader accountability.

Appendix

Appendix 4.A Identification Strategy

There are two complications in estimating the effect of leader selection rules. First is the problem of self-selection: political and social units (be they groups, communities and countries) select rules for electing their leader in a complex political process that is endogenous to the unit's characteristics. One way for dealing with this problem is via an Instrumental Variable (IV) approach. If (a) the final stage is fully specified, and (b) equations for earlier stages include instruments to address endogeneity, then if (c) the dependent variable in the final stage is continuous and unbounded then instrumental variable techniques such as 2SLS are consistent (Roodman, 2009).

Throughout the analysis, when the dependent variable is continuous, I fit a linear 2SLS IV model[64]. The Linear IV model considers the effect of an endogenously chosen binary treatment d_j (leader selection rule) on a fully observed variable y_j, conditional on two sets of independent variables Z_j and X_j. The primary interest is in the regression function:

$$y_j = \alpha_y + \beta X_j' + \delta d_j + \varepsilon_j \tag{4.1}$$

where d_j is an endogenous dummy variable $E[\varepsilon|d \neq 0]$ indicating whether the association uses appointments (d_{j0}) or popular direct elections (d_{j1}) to select it's manager. X_j is a $n \times K$ matrix of exogenous control variables and ε_j is a group-level error term. The binary decision to obtain the treatment d_j is modeled as the outcome of a linear function of the exogenous covariates Z_j and a random component υ_j. Specifically,

$$d_j^* = \alpha_d + \Pi Z_j + \upsilon_j \tag{4.2}$$

where

$$d_j = \begin{cases} 1, & \text{if } d_j^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

It is also assumed that ε_j and υ_j are normally distributed with mean zero; $Cov(Z, \varepsilon) = 0$; $Cov(Z, \upsilon) = 0$, and $Cov(Z, d) \neq 0$. To be a valid instrument $Z \perp (Y_{j0}, Y_{j1}, d_{j0}, d_{j0})$. See Imbens and Angrist

[64] As mentioned in the main text, when using individual-level data, I estimate bootstrapped standard errors based on 10,000 iterations that take into account the nested structure of the data. By contrast, when using association-level data, I estimate bootstrapped standard errors based on 1,000 iterations that do not account for possible higher-level nesting.

(1994) for Local Average Treatment Effect derivation (LATE).

There are additional complications if, however, the dependent variable is binary. In such cases, I follow Chiburis, Das and Lokshin (2010) recommendation to use both Linear IV regressions (Altonji, Elder and Taber, 2005), as well as Maximum Likelihood (ML) Seemingly Unrelated Equations (SUR) Bivariate Probit (Neal, 1997, Roodman, 2009). In both estimation techniques, I use bootstrapping to estimate standard errors.

Equations in an SUR system seem unrelated in the sense that the dependent variables are generated in a process that is independent except for correlated errors, that share a multidimensional distribution. ML SUR can consistently estimate parameters in an important subclass of mixed-process simultaneous systems: ones that are both *recursive*, i.e., with clearly defined stages and that are *fully observed*, i.e., that endogenous variables appear on the right hand side not as latent variables, but only as observed variables. Therefore, such estimation procedure is ideal for encouragement study designs, such as our own. Define,

$$\begin{aligned} d_j^* &= \alpha_d + \Pi Z_j + \upsilon_j \\ y_j^* &= \alpha_y + \beta X_j' + \delta d_j + \varepsilon_j \end{aligned} \tag{4.4}$$

$$d_j = \begin{cases} 1, & \text{if } d_j^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

$$y_j = \begin{cases} 1, & \text{if } y_j^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

where y_j^* is a latent variable whose value determines the binary outcome y_j, d_j^* is a latent variable whose value determines the leader selection rule adopted by the association (i.e., appointment by directors d_{j0} or direct vote by ordinary members d_{j1}), Z_j is an exogenous variable affecting the value of d_j^*, $X = (x_1, \ldots, x_k)'$ is a column vector of covariates (individual and group-level controls), and β is a vector of coefficients.

We further assume that $(\upsilon_j; \varepsilon_j)$ are independent of Z and are jointly distributed as standard bivariate normal with correlation ρ, where ρ measures the endogeneity of d_j in the y_j^* equation. For the full derivation of the likelihood function of a two-stage probit, see Roodman (2009) and Chiburis, Das and Lokshin (2010).

Appendix 4.B Monitoring, Accountability and Responsiveness

Leader Accountability by Leader Selection Rule (Self-Report)

	Bivariate Probit	Linear IV	Bivariate Probit	Linear IV
Direct vote	1.03	0.37**	1.43**	0.45**
	(0.68)	(0.19)	(0.60)	(0.17)
Age of DC (centered)			0.57**	0.16***
			(0.23)	(0.06)
Original No. of POs (centered)				0.01
				(0.01)
Constant	-0.79**	0.21**	-1.02***	0.18*
	(0.38)	(0.11)	(0.34)	(0.10)
First-Stage Regression (DV: Direct vote)				
Z	2.54	0.76***	2.52	0.76***
	(685.42)	(0.11)	(35.08)	(0.11)
Original No. of POs (centered)	0.03	0.01	0.03	0.01
	(34.24)	(0.01)	(3.04)	(0.01)
Age of DC (centered)	0.10	0.01	0.04	0.01
	(58.24)	(0.05)	(14.49)	(0.05)
Constant	-0.98	0.17***	-0.98	0.17***
	(379.87)	(0.06)	(25.88)	(0.06)
athρ	-0.25		-0.41	
	(19.93)		(16.14)	
ATE	0.38*		0.45***	
	(0.23)		(0.16)	
ATT	0.37*		0.42***	
	(0.21)		(0.14)	
Observations	45	45	45	45
r^2		0.10		0.19

Bootstrapped standard errors (1,000 iterations) in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2: **Leader Accountability by leader selection rule (Self-Report).** The DV is modeled as a function of an association's leader selection rule, instrumented by the facilitator recommendation. In columns 1 and 3 I report regression results from maximum likelihood bivariate probit, whereas in columns 2 and 4 I report linear instrumental variable estimation. In all models standard errors are bootstrapped(1,000 iterations), ignoring possible dependencies in higher-order clusters. Note that the ATE in th third cloumn and the LATE estimate (forth column) are virtually identical.

Number of Selection Processes by Leader Selection Rule

	IV Poisson	Linear IV
‘ Direct vote	0.55***	0.89***
	(0.14)	(0.27)
Original No. of POs (centered)	0.01	0.01
	(0.01)	(0.02)
Age of DC (centered)	0.17***	0.29***
	(0.05)	(0.10)
Constant	0.02	0.95***
	(0.07)	(0.13)
First-Stage Regression (DV: Direct vote)		
Z	0.79***	0.79***
	(0.11)	(0.11)
Original No. of POs (centered)	0.01	0.01
	(0.01)	(0.01)
Age of DC (centered)	-0.02	-0.02
	(0.05)	(0.05)
Observations	43	43
r^2		0.28

Bootstrapped standard errors (10,000 iterations) in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: **Number of Selection Processes by leader selection rule.** The DV – Number of selection processes for the DC manager position since the creation of the farmer association – is modeled as a function of an association’s leader selection rule, which itself is instrumented by the DC facilitator’s recommendation.

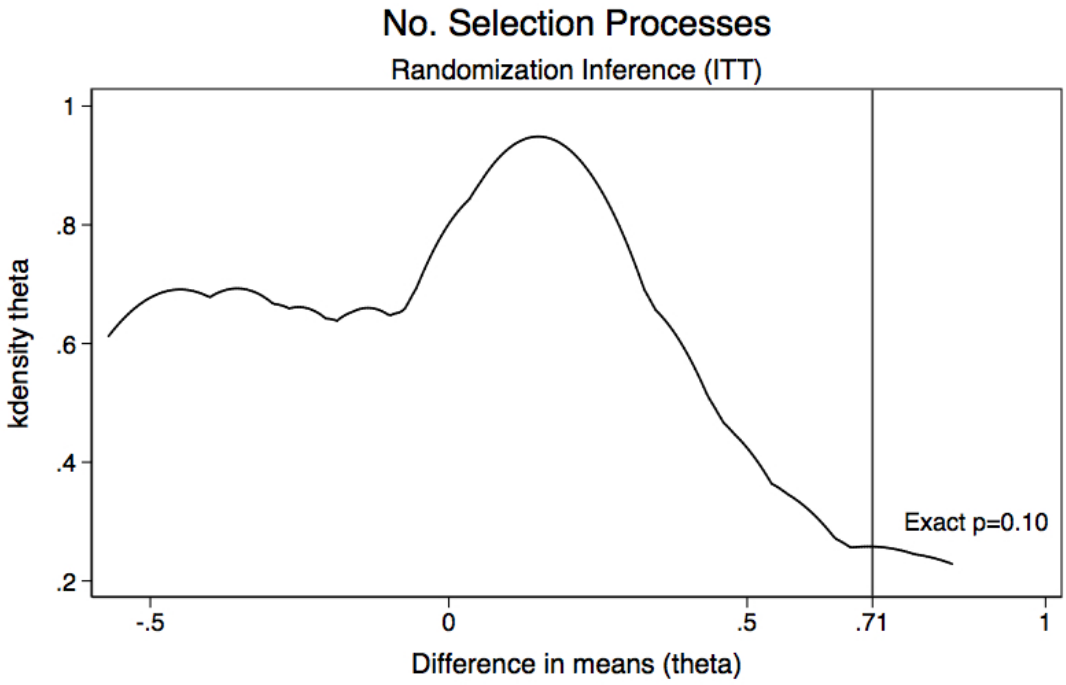

Figure 4.7: Figure describes the distribution of beta coefficients estimates from a reduced form regression of facilitator's recommendation on manager's ability ranking (ITT). The distribution of betas is estimated using randomization inference: all possible permutations of selecting three facilitators out of eight recommending direct-vote ($Z_j = 1$). Reported P-values are exact.

Relationship between Leader Selection Rule and Accountability (Members Survey)

	Receipts	Side-Sell	Bad Quality	Monitored	Accountable
Direct vote	1.14***	0.61***	0.39***	0.22*	0.37***
	(0.27)	(0.13)	(0.14)	(0.12)	(0.11)
Sex	0.14	-0.05	-0.10**	0.11*	0.03
	(0.13)	(0.09)	(0.04)	(0.07)	(0.05)
Age (units of 10)	0.02	0.00	-0.09**	-0.01	0.06*
	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
Education (Std.)	0.07	0.06**	0.05**	0.15***	0.18***
	(0.06)	(0.03)	(0.02)	(0.04)	(0.03)
Years since Joining Group	0.05**	-0.01	0.05**	0.04**	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Church attendance	0.10	-0.05	0.05	0.13***	0.01
	(0.07)	(0.04)	(0.05)	(0.04)	(0.07)
Associational-life (std.)	0.12*	0.17***	0.06**	0.16***	0.09*
	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)
Manager's co-villager	0.02	0.09	0.17	-0.17	0.20**
	(0.17)	(0.12)	(0.12)	(0.13)	(0.08)
Middleman mistrust	0.23	-0.14	0.13*	0.08	0.35***
	(0.15)	(0.15)	(0.08)	(0.11)	(0.10)
Original No. of POs	0.00	0.00	0.00	-0.00	-0.02***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Age of DC	0.25***	0.17***	0.11**	0.05*	0.02
	(0.09)	(0.03)	(0.05)	(0.03)	(0.06)
Constant	-1.28***	-1.35***	-0.90***	-0.81**	-1.50***
	(0.48)	(0.24)	(0.26)	(0.33)	(0.26)
First-Stage Regression (DV: Direct vote)					
Z	2.50***	2.57***	2.58***	2.57***	2.53***
	(0.59)	(0.60)	(0.60)	(0.60)	(0.60)
Original No. of POs	0.03	0.03	0.03	0.03	0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Age of DC	0.02	0.01	0.01	0.01	0.01
	(0.11)	(0.11)	(0.11)	(0.10)	(0.11)
athρ	-0.69**	-0.39***	-0.13	-0.04	-0.14**
	(0.29)	(0.07)	(0.10)	(0.06)	(0.07)
ATE	0.30***	0.16***	0.14***	0.08**	0.12***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
ATT	0.37***	0.14***	0.14***	0.08**	0.11***
	(0.06)	(0.02)	(0.03)	(0.04)	(0.03)
LATE	0.28***	0.18***	0.14***	0.09**	0.13***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Observations	1359	1536	1536	1501	1465
ρ	-0.60	-0.38	-0.13	-0.04	-0.14
Log Likelihood	-1170.24	-1320.53	-1567.27	-1591.33	-1397.57

Standard errors clustered at the field-facilitator level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: In Model "Receipts", the dependent variable is binary, indicating whether the respondent reports receiving receipts from the DC manager when selling in bulk. In Model "Side-Sell" and "Bad Quality", the dependent variables are binary, indicating whether the respondent reports ever being punished for side-selling and for producing low quality, respectively. In Model "Monitored", the dependent variable indicates whether the respondent believes that the DC manager is monitored. In Model "Accountable", the dependent variable indicates whether the respondent perceives the DC manager to be very accountable. Standard errors for ATE, ATT and LATE are bootstrapped using 1000 iterations, adjusted for 8 strata (field-facilitators)

Relationship between Leader Selection Rule and Accountability (Board of Directors)

	External (N)	External (Y)	Internal	Receipts	Monitored
Direct vote	-0.77**	0.84**	0.28*	1.16***	0.37
	(0.34)	(0.33)	(0.17)	(0.28)	(0.34)
Sex	0.09	0.11	0.35***	0.12**	0.21***
	(0.09)	(0.12)	(0.10)	(0.05)	(0.05)
Age (units of 10)	-0.05	-0.06*	-0.16***	-0.11**	-0.02
	(0.04)	(0.03)	(0.05)	(0.05)	(0.02)
Years since Joining Group	0.06***	-0.03	0.10***	0.01	0.04
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
Rep ability (std.)	0.03	0.07	0.17***	-0.10*	0.19***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)
Associational-life (std.)	0.06**	0.00	0.10**	0.09	0.08*
	(0.02)	(0.02)	(0.05)	(0.06)	(0.04)
Wealth (std.)	0.02	0.02	0.14***	0.01	0.05
	(0.04)	(0.03)	(0.05)	(0.06)	(0.03)
Age of DC	-0.25***	0.25***	-0.08	0.25**	0.14*
	(0.08)	(0.09)	(0.06)	(0.11)	(0.07)
Original No. of POs	-0.02	0.02	-0.00	-0.00	0.03**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Constant	1.02*	-1.43***	-0.58	-0.13	-0.85**
	(0.56)	(0.54)	(0.38)	(0.62)	(0.42)
First-Stage Regression (DV: Direct vote)					
Z	2.43***	2.44***	2.53***	2.44***	2.48***
	(0.45)	(0.45)	(0.53)	(0.45)	(0.45)
Age of DC	-0.01	-0.01	-0.02	0.01	-0.01
	(0.09)	(0.09)	(0.20)	(0.09)	(0.09)
Original No. of POs	0.02	0.02	0.02	0.03	0.02
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
athρ	0.37	-0.34*	-0.15	-0.70***	-0.14
	(0.23)	(0.20)	(0.10)	(0.23)	(0.18)
ATE	-0.29***	0.30***	0.08**	0.33*	0.13***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
ATT	-0.29***	0.28***	0.08**	0.38*	0.13***
	(0.06)	(0.03)	(0.03)	(0.04)	(0.04)
LATE	-0.30***	0.32***	0.08**	0.32*	0.13***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
midrule Observations	1065	1065	1054	1058	1079
ρ	0.35	-0.33	-0.15	-0.61	-0.13
Log Likelihood	-1091.50	-1052.19	-929.76	-909.04	-1078.57

Standard errors clustered at the field-facilitator or farmer association level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: In Models "External (N)" and "External (Y)", the dependent variables are binary, indicating whether the respondent reports that the DC has ever used external auditors. In Model "Internal" the dependent variables are binary, indicating whether the respondent reports ever asking to review the group's books. In Model "Receipts", the dependent variable indicates whether the respondent believes that the DC manager provides receipts to members selling in bulk. In Model "Monitored", the dependent variable indicates whether the respondent perceives the DC manager to be monitored. Standard errors for ATE, ATT and LATE are bootstrapped using 1000 iterations, adjusted for 8 strata (field-facilitators)

Appendix 4.C Sensitivity Analysis

Using Dunning (2008*a*) framework, the instrumental variable applied in this paper is only 'plausibly exogenous'. In other words, the exclusion restriction that underlies the validity of the IV inference is suspect. To address this concern I build on recent work by Conley, Hansen and Rossi (2010) who develop practical methods for performing inference while relaxing the exclusion restriction. The key to Conley, Hansen and Rossi (2010) method is to define a parameter γ that reflects how close the exclusion restriction is to being satisfied in the following linear structural model:

$$Y = \alpha_y + \beta X + \delta d + \gamma Z + \epsilon \tag{4.7}$$

$$d = \alpha_d + \Pi Z + \upsilon \tag{4.8}$$

Where Y is the outcome variable, X is a matrix of endogenous treatment variables, $E[X\epsilon] \neq 0$, with β the treatment parameter. ϵ are unobservables; Z is a matrix of instruments that are assumed uncorrelated with ϵ; i.e., $E[Z'\epsilon] = 0$; and Π is matrix of first-stage coefficients. When X is endogenous, the parameters β and γ are not jointly identified, so prior information or assumptions about γ are needed in order to obtain estimates of β – the parameters of interest. The IV exclusion restriction is equivalent to the dogmatic prior belief that γ equals zero. Conley, Hansen and Rossi (2010) definition of plausible exogeneity considers the possibility that γ is near zero but perhaps not exactly zero. Allowing $\gamma \neq 0$ provides sufficient structure to allow for estimation and inference while relaxing the IV exclusion assumption.

Here I present results from one of the authors' proposed sensitivity tools. In this approach, the researcher merely specifies a set of possible values for γ, (i.e. the support of γ). Conditional on any potential value of γ, interval estimates for β, the treatment parameter of interest, can easily be obtained. Taking the union of these interval estimates across different values provides a conservative (in terms of coverage) interval estimate for β. The main advantage of this method is that it does not require complete specification of a prior distribution of γ, but only specification of a range of plausible values for γ. In the following analysis β should be interpreted as the LATE and γ as the average *direct effect* of Z on Y. For further information, see Conley, Hansen and Rossi (2010)[p. 6–8].

Applying this method to our study, I estimate the interval estimate for β for a range of plausible values for γ for two accountability measures: utilization of external auditors and obtaining

receipts from the DC manager when selling in bulk[65]. Starting with the use of external auditors, when $\gamma = 0$, the LATE estimate of moving from appointments to direct elections is 0.31 with $0.15 - 0.48$ at the 95% confidence intervals. Figure 4.8 demonstrates how the confidence intervals of β change as γ increase. When $\gamma = 0.12$ the lower bound of β intersects with zero. In other words, so long as the direct effect of Z on Y – i.e,. the direct effect of the field-facilitator on accountability that is not captured by the leader selection rule – is less than 38% of the direct impact of the leader selection rule, we can assume that elections have a positive impact on the likelihood that an association uses external auditors.

Turning to likelihood of writing receipts, when $\gamma = 0$, the LATE estimate of moving from appointments to direct elections is 0.32 with $0.18 - 0.47$ at the 95% confidence intervals. When $\gamma = 0.15$ the lower bound of β intersects with zero (Fig. 4.9). In other words, we cannot reject the null hypothesis that elections have no impact on the likelihood that managers write receipts, only when the direct impact of field-facilitators on writing receipts exceeds about 50% of the LATE estimate.

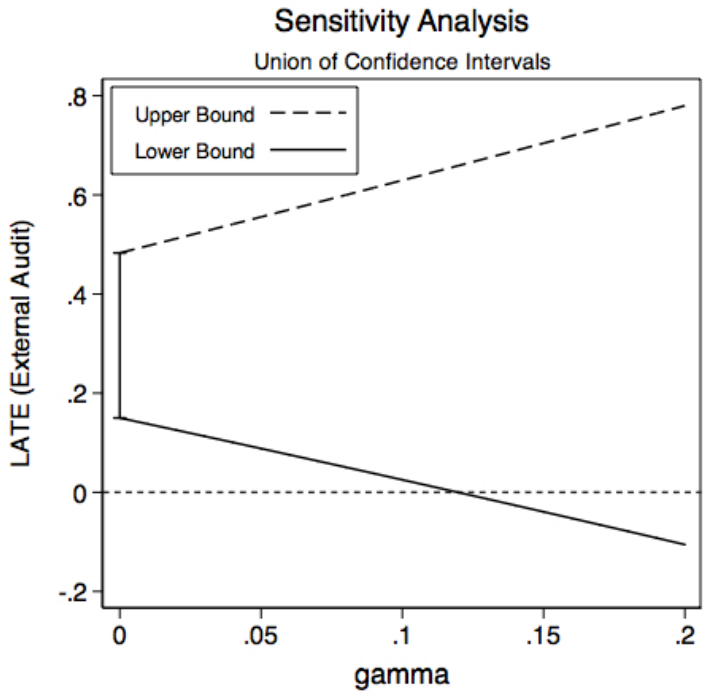

Figure 4.8: Union of Confidence Intervals with γ Support Assumption: External Auditing.

[65]The LATE estimates I report here use data from the representatives' survey, and are based on simple linear IV regression models (2SLS) with standard errors clustered at the association-level.

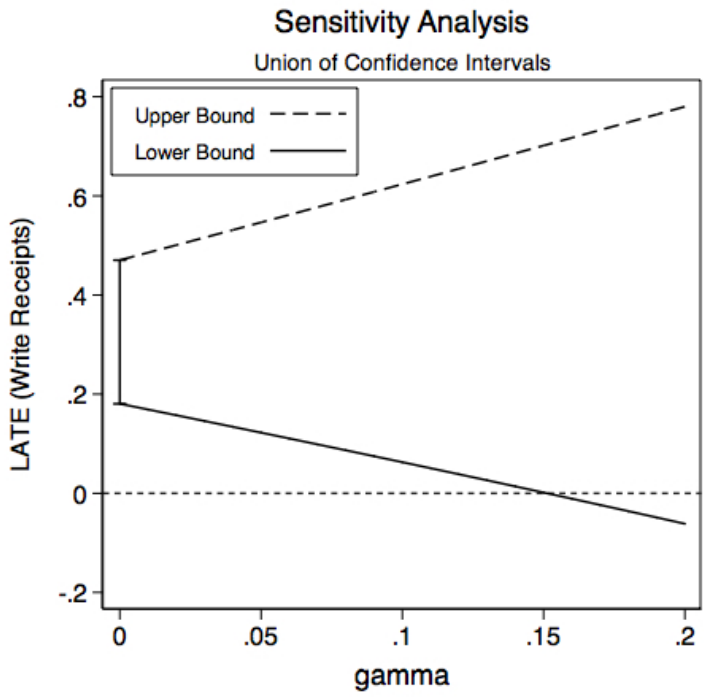

Figure 4.9: Union of Confidence Intervals with γ Support Assumption: obtaining receipts.

Appendix 4.D Leader Ability and Leader Profile

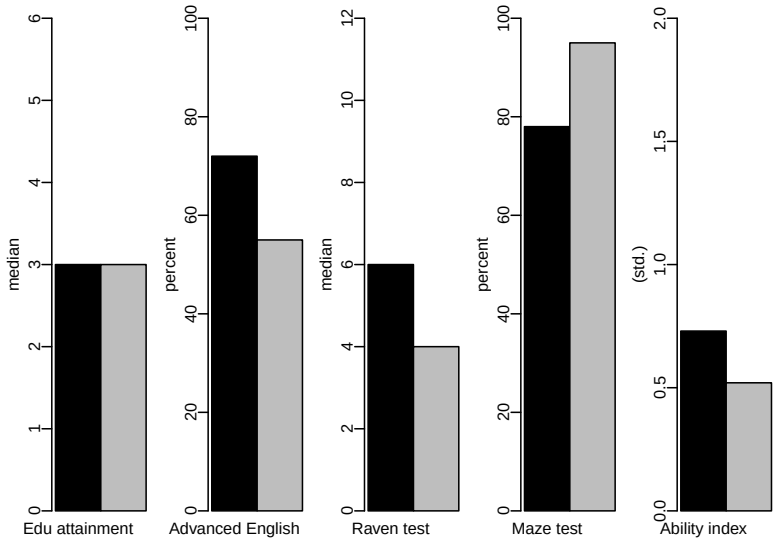

Figure 4.10: **Relation between leader selection rule and leader ability**, measured as (A) mean education-level on a six-point scale; (B) proportion of leaders that can communicate in advanced English; (C1) mean number of correct responses to a 12-question raven test; (C2) proportion of leaders who solved a maze; and (D) mean score of a composite ability index. Leaders appointed in black and directly elected in grey.

Leader Ability by Leader Selection Rule (LATE)

	Edu	English	Raven	Maze	Index	Rank
Direct vote	-0.88*	-0.18	0.13	0.31**	-0.05	1.03
	(0.46)	(0.21)	(1.31)	(0.13)	(0.25)	(2.59)
Original No. of POs (centered)	-0.00	0.00	0.02	0.00	0.01	-0.04
	(0.03)	(0.02)	(0.08)	(0.01)	(0.02)	(0.17)
Age of DC (centered)	-0.03	-0.03	-0.31	0.03	-0.01	0.75
	(0.19)	(0.08)	(0.46)	(0.06)	(0.10)	(0.93)
Constant	3.50***	0.72***	5.43***	0.72***	0.66***	7.16***
	(0.31)	(0.12)	(0.68)	(0.11)	(0.16)	(1.45)
First-Stage Regression (DV: Direct vote)						
Z	0.75***	0.75***	0.75***	0.75***	0.75***	0.75***
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Original No. of POs (centered)	0.01	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age of DC (centered)	0.01	0.01	0.01	0.01	0.01	0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
ATE		-0.16		0.31***		
		(0.22)		(0.11)		
ATT		-0.16		0.46***		
		(0.20)		(0.16)		
Observations	45	45	42	42	42	45

Bootstrapped standard errors (10,000 iterations) in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: Linear IV estimation in which the DV is a function of the leader selection rule, instrumented by the facilitator's recommendation: In model 'Edu' the DV is manager's education attainment on a 6-level scale; in Model 'English' the DV is binary indicating whether the manager can speak advanced English; in model 'Raven' the DV is the number of questions solved out of 12 in a simple raven test; in model 'Maze' the DV is binary indicating whether the manager solved a simple maze test in less than two minutes; in model 'Index' the DV is the standardized composite ability index, and in model 'Rank' the DV is the manager's ability rank from high to low. For the binary outcomes ('English' and 'Maze') I report, in addition, ATE and ATT from Bivariate Probit bootstrapped estimation (1,000 iterations).

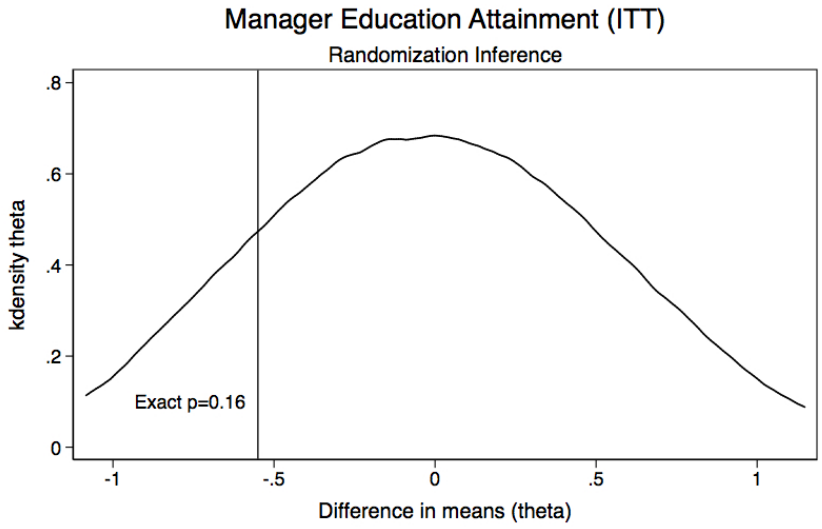

Figure 4.11: Figure describes the distribution of beta coefficients estimates from a reduced form regression of facilitator's recommendation on manager's education attainment (ITT). The distribution of betas is estimated using randomization inference: all possible permutations of selecting three facilitators out of eight recommending direct-vote ($Z_j = 1$). Reported P-values are exact.

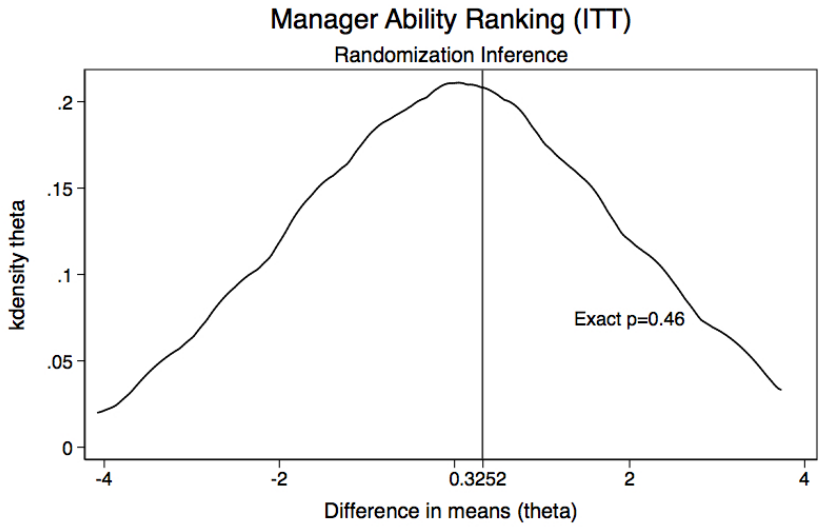

Figure 4.12: Figure describes the distribution of beta coefficients estimates from a reduced form regression of facilitator's recommendation on manager's ability ranking (ITT). The distribution of betas is estimated using randomization inference: all possible permutations of selecting three facilitators out of eight recommending direct-vote ($Z_j = 1$). Reported P-values are exact.

Leader Socio-demographic Profile by Leader Selection Rule (LATE)

	Age	Church	Associations	Expereince	Local
Direct vote	-0.12	-0.03	-0.61	-7.26	-0.66***
	(4.27)	(0.15)	(0.79)	(5.33)	(0.23)
Original No. of POs (centered)	-0.23	0.00	-0.07	-0.22	0.01
	(0.29)	(0.01)	(0.07)	(0.32)	(0.01)
Age of DC (centered)	-1.18	-0.10*	-0.01	-0.56	0.02
	(1.90)	(0.06)	(0.29)	(2.10)	(0.07)
Constant	46.41***	2.95***	5.85***	23.91***	0.94***
	(2.63)	(0.10)	(0.52)	(3.04)	(0.12)
First-Stage Regression (DV: Direct vote)					
Z	0.75***	0.75***	0.75***	0.75***	0.75***
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Original No. of POs (centered)	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age of DC (centered)	0.01	0.01	0.01	0.01	0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Observations	45	45	42	45	45
r^2	0.03	0.06	0.07	0.03	0.04

Bootstrapped standard errors (10,000 iterations)
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.9: Linear IV estimation in which the DV is a function of the leader selection rule, instrumented by the facilitator's recommendation: 'Age' is a continuous variable measuring the leader's age; 'Church' is a four category variable measuring the leader's frequency of attendance in religious services; 'Associations' is a continuous variable measuring the number of social, economical and political groups and organizations in which the leader participates actively; 'Experience' is a continuous variable measuring the number of years the leader has been growing coffee; and 'Local' is binary measuring whether the leader was born in his current village.

Appendix 4.E Social Network Analysis

Leader Network Centrality by Leader Selection Rule (LATE)

	Prestige			Betweenness		
	Friend	Prior	Advice	Friend	Prior	Advice
Direct vote	-1.86	0.42	5.13	15.94	5.47	-9.84
	(2.88)	(3.31)	(3.47)	(16.55)	(9.53)	(37.08)
Original No. of POs (centered)	0.62*	0.60*	0.66**	0.93	0.56	8.26
	(0.32)	(0.31)	(0.32)	(1.26)	(0.74)	(6.86)
Age of DC (centered)	0.70	1.25	2.12**	12.33**	4.25	30.38
	(1.06)	(1.02)	(1.08)	(5.85)	(2.94)	(22.16)
Constant	15.55***	16.93***	15.69***	29.75***	15.23***	74.19***
	(1.67)	(1.64)	(1.73)	(7.47)	(3.99)	(26.87)
Observations	44	44	44	44	44	44
r^2	0.23	0.19	0.21	0.10	0.07	0.17

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.10: **Leader Network Centrality.** Columns 2-4 report LATE of Leader selection rule on on the score of the manager's Indegree ('Prestige') network centrality for three types of network relations: (i) Friendship; (ii) Prior Acquaintance; and (iii) Advice. Columns 5-7 report LATE of Leader selection rule on the score of the manager's 'Betweenness' network centrality.

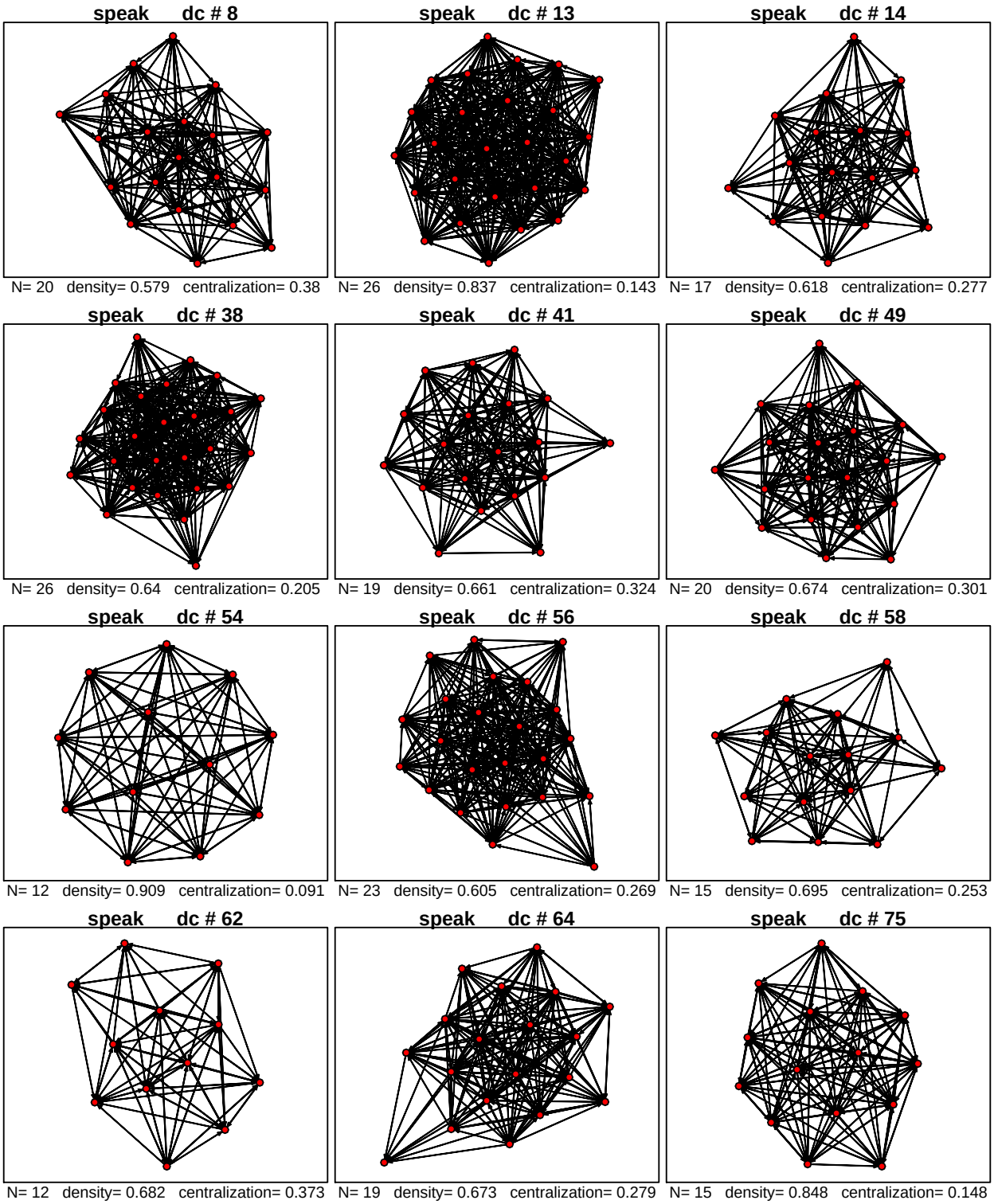

Figure 4.13: **Acquaintance Network (board directors) by farmer association.** Question wording: "Did you speak to [NAME] on a regular basis before the creation of the DC?"

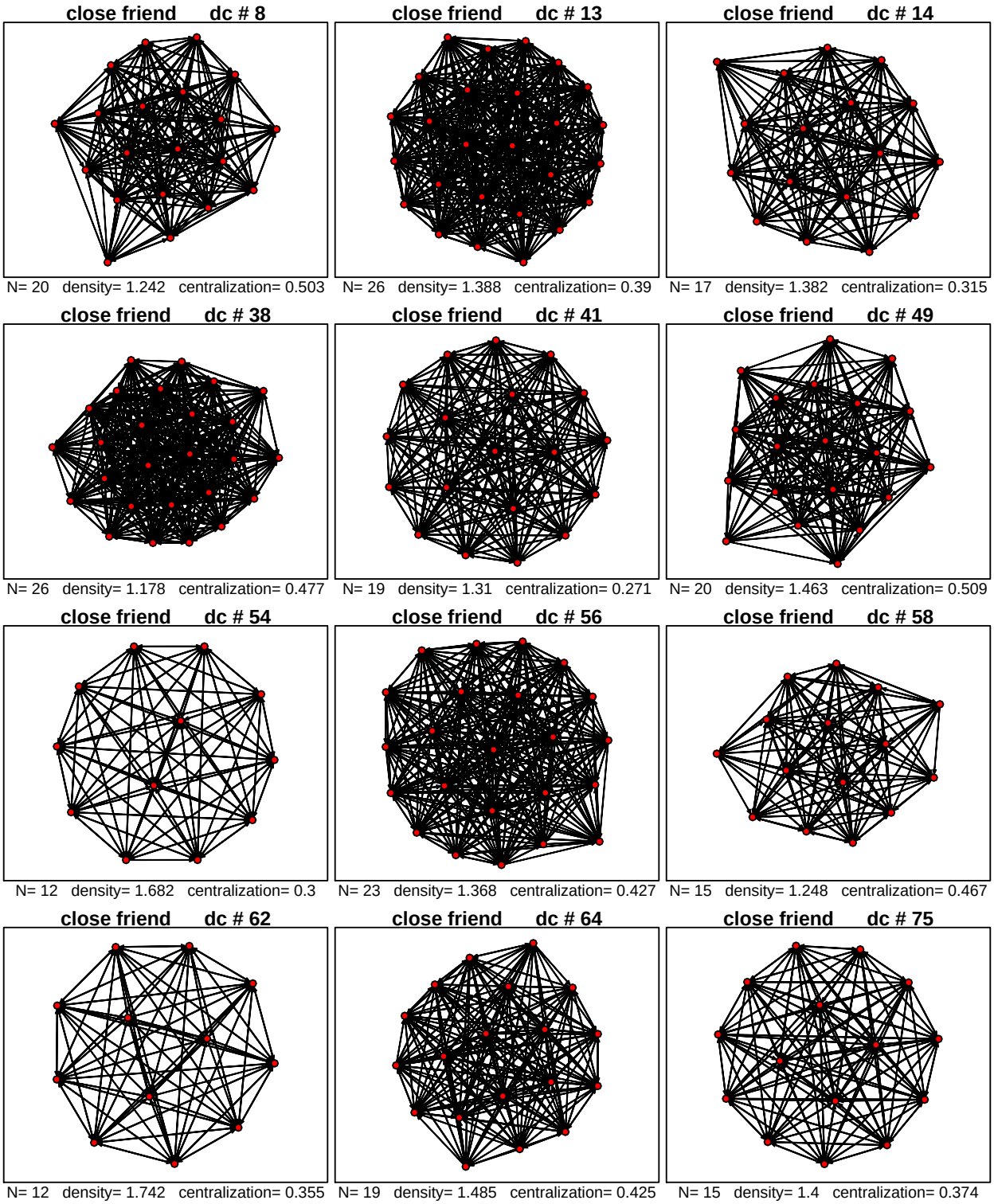

Figure 4.14: **Acquaintance Network (board directors) by farmer association.** Question wording: "Is [NAME] a Close Friend?"

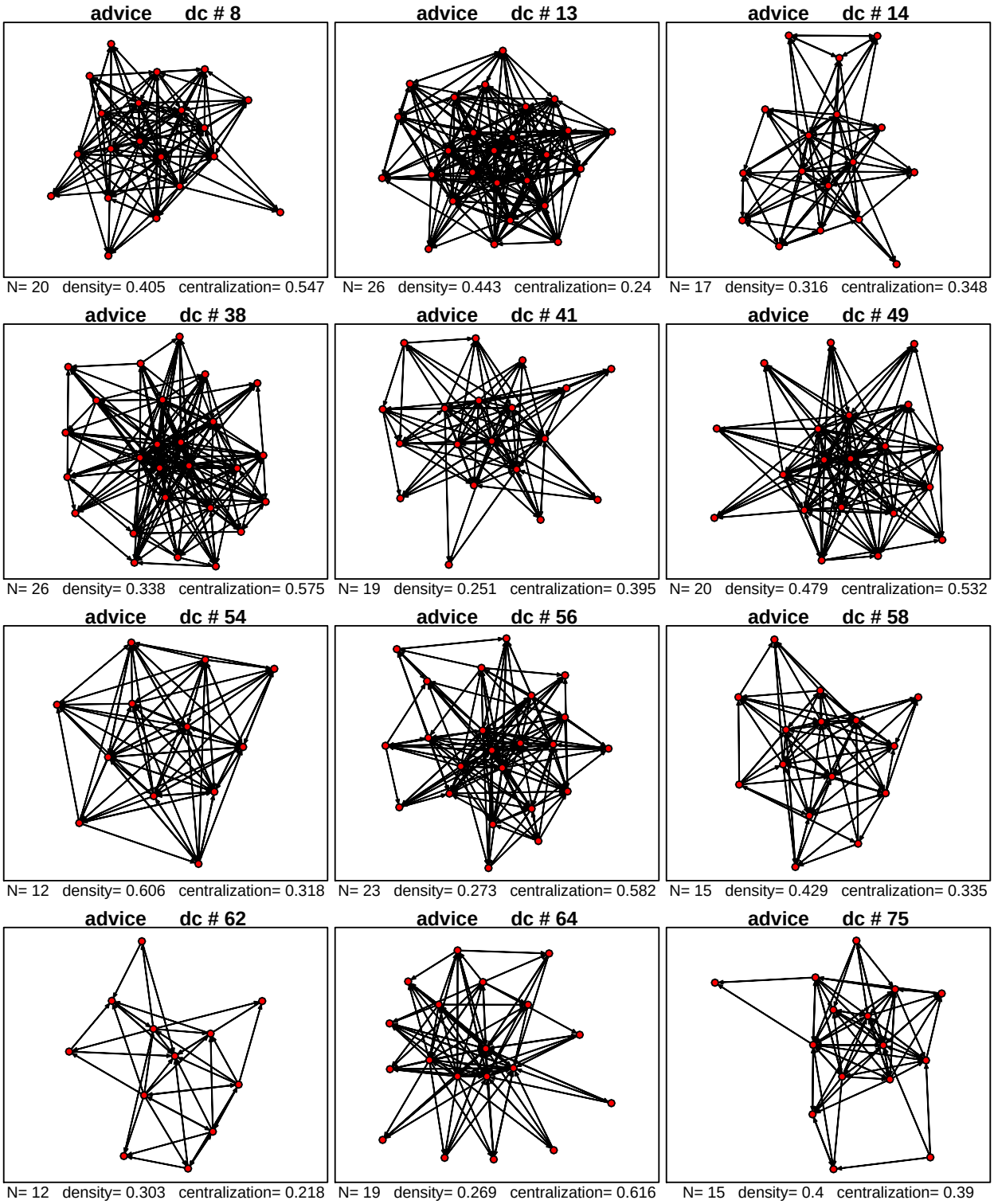

Figure 4.15: **Acquaintance Network (board directors) by farmer association.** Question wording: "In the past 12 months, have you asked [NAME] for information or advice on matters related to the DC?"

Appendix 4.F Behavioral Games

Leader Allocation Decisions: Dictator Game

	Stranger		DC Member	
	A	B	A	B
Direct vote	-46.95	-36.91	167.24	183.53*
	(60.11)	(59.46)	(152.13)	(108.61)
Original No. of POs (centered)		-4.21		10.23
		(3.63)		(7.13)
Age of DC (centered)		19.14		72.25*
		(20.95)		(40.38)
Constant	254.36***	250.10***	370.97***	361.54***
	(33.39)	(33.46)	(70.11)	(46.59)
First-Stage Regression (DV: Direct vote)				
Z	0.76***	0.76***	0.76***	0.76***
	(0.11)	(0.11)	(0.11)	(0.11)
Original No. of POs	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Age of DC	0.01	0.01	0.01	0.01
	(0.05)	(0.05)	(0.05)	(0.05)
Observations	44	44	44	44
r^2	0.05	0.09	0.03	0.17

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.11: **Linear IV Estimation of Leader Allocation Decision in the DG.** The DV is modeled as a function of an association's leader selection rule, instrumented by the DC facilitator's recommendation Z. Since The control variables have been centered the constant is the predicted allocation for appointed managers, and the coefficient on direct vote is the increase in allocation for elected manager. Standard errors are bootstrapped using 10,000 iterations.

Leader Third-Party Punishment Experiment

	10	9	8	7	6
Direct vote	0.43*	0.47**	0.38*	0.46**	0.28
	(0.23)	(0.22)	(0.23)	(0.20)	(0.19)
Original No. of POs (centered)	-0.01	-0.01	-0.01	-0.01	-0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Age of DC (centered)	-0.08	-0.09	-0.06	-0.09	-0.13
	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)
Constant	0.25*	0.28**	0.36***	0.42***	0.59***
	(0.13)	(0.14)	(0.14)	(0.14)	(0.13)
First-Stage Regression (DV: Direct vote)					
Z	0.76***	0.76***	0.76***	0.76***	0.76***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Age of DC	0.01	0.01	0.01	0.01	0.01
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Original No. of POs	0.01	0.01	0.01	0.01	0.01
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Observations	43	43	43	43	44

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12: **Linear IV Estimation of Leader Punishment in TPP experiment.** The DV is modeled as a function of an association's leader selection rule, instrumented by the DC facilitator's recommendation. Since The control variables have been centered the constant is the predicted probability of sanctioning for appointed managers, and the coefficient on direct vote is the increase in predicted probability of sanctioning for elected manager. Standard errors are bootstrapped using 1000 iterations.

Appendix 4.G Balance

Covariate	Reps Mean	GA Mean	OLS P-value	K-S P-value	Reps N	GA N
DC Age	3.00	2.56	0.268	0.226	32	18
	(1.02)	(0.92)				
Years under APEP	3.84	3.33	0.043	0.054	32	18
	(1.11)	(1.08)				
No. Original POs	9.56	10.17	0.808	0.930	32	18
	(4.89)	(8.18)				
No. Original Members	161	178	0.781	0.496	22	16
	(106)	(179)				
PIO	-0.15	0.28	0.344	0.107	32	18
	(1.08)	(0.78)				

Table 4.13: **Assessing balance across selection rules (DC Characteristics).** In column 5 I report $p - values$ from OLS regressions of each covariate on treatment assignment. Column 6 reports exact $p - values$ from two-sample Kolmogorov-Smirnov test for equality of distribution functions. PIO stands for Private Income Opportunities.

APEP Trainer	**Region**	**District**	$Z_j = 0$	$Z_j = 1$	$d_j = 0$	$d_j = 1$	**Sampled DCs**
Elisa Tegyeza	West	Ibanda	6	0	5	1	6
Joseph Katto	West	Mbarara	0	1	0	1	1
Joseph Katto	West	Kiruhura	0	7	1	6	7
Wilburforce Tibairira	East A	Iganga	5	0	4	1	5
David Baligindwire	East B	Kamuli	11	0	9	2	11
Vincent Okoth	Central A	Masaka	7	0	5	2	7
Daniel Kambale	Central A	Rakai	3	0	2	1	3
Edison Kawalya	Central B	Mityana	0	1	0	1	1
Edison Kawalya	Central B	Mubende	0	3	0	3	3
Noa Kuluse	Central B	Mubende	0	6	0	6	6
Total			32	18	26	24	50

Table 4.14: **Assessing exogenous variation in leader selection rule.** Table provides information on the leader selection rule recommendation Z_j and leader selection rule adoption d_j by APEP field-facilitator.

Chapter 5

Concluding Remarks

The chapters that make up this dissertation have shown that how leaders are selected and who decides to run for office have a strong impact on the effectiveness of small self-help groups. The effectiveness of self-help groups in public goods production, in turn, has a positive and significant impact on their members' welfare. In the concluding remarks I wish to point to some conflicting and puzzling findings, to reflect on the external validity of the study's result and to suggest future avenues for research.

Leader Selection Rule, Monitoring and the Self-Selection Effect

The empirical findings of chapter 4 are in tension with the results of chapter 3. This tension calls for a closer examination of the role monitoring institutions play in public goods production. I briefly point towards some possibilities that will hopefully be taken up in future work. In chapter 4, I demonstrate that popular direct vote results in leaders who are more tightly monitored, and hence more accountable and responsive. In chapter 2, I demonstrate that accountability has a positive effect on cooperation. By contrast, in chapter 3, I show that more and better monitoring institutions do not necessarily lead to better outcomes because they may result in higher-ability group members opting out of candidacy through the self-selection effect. Beyond a certain point, monitoring has negative consequences for the value of the public good (recall Fig. 3.4).

The model I present in chapter 3 suggests that increase in leader rewards may offset the negative effects from increased monitoring. In a recent study, Ferraz and Finan (2010) reach similar conclusions. By contrast, Mattozzi and Merlo (2008) argue that an increase in politicians' salary *decreases* the average quality of individuals who become politicians, because it exacerbates the problem of adverse selection. The intuition is that politics becomes relatively a more attractive option for all levels of political skills. The extent to which higher rewards may affect the quality of the candidacy pool depends on the strength of the competence signals that *potential* candidates emit. In other words, whether higher rewards can offset the negative impact of monitoring depends on the richness of information that group members have on each other.

My model assumes that self-help groups have perfect information on members' abilities. In such a case, an increase in leader rewards should have a positive effect on the value of the public good produced. Whereas, this may be a relatively reasonable assumption when studying parish-level associations in rural Uganda (as findings from chapter 4 seem to confirm), it may not be a reasonable assumption in other contexts. Relaxing the assumption of perfect information and addressing possible interactions between monitoring institutions, information levels and reward schemes, both theoretically and empirically, offer interesting avenues for future research.

Public Goods Games

Results from the public goods games have also raised a set of new puzzles and questions that future work should address. On one hand, in the public goods experiment (chapter 2) I did not find much evidence suggesting that elected and random monitors punish differently. On the other hand, in chapter 4, I find that elected leaders punish significantly more frequently than appointed leaders using both observational and experimental data (e.g., Third Party Punishment game). Evidently more work is needed to reconcile those conflicting findings.

Secondly, one of the central finding of the public goods games is that in the presence of a centralized sanctioning authority, groups can reach higher levels of cooperation and that monitors are willing to bear the cost of punishing in order to increase cooperation. As I noted above, these results are qualitatively similar to those obtained using peer-punishment institutions, with the *possible* advantage that a centralized system of monitoring will be more efficient than a decentralized one. Future studies should investigate this claim directly: *under what conditions* — type of players, group-size, size of social-return, cost of punishment, etc. — *are centralized-sanctioning regimes more efficient than decentralized peer-punishment*?

Leaders' Socio-demographic Attributes

One of the relatively surprising findings of this dissertation is that, in both the observational and the experimental data, I can only uncover a small and insignificant impact of leaders' socio-demographic attributes on cooperation. Similarly, though I find that leader ability has a positive effect on members' cooperation, it's impact becomes insignificant once I account for other factors, such as leader effort (chapter 3) or leader selection rules (chapter 4). How should we interpret these counter-intuitive findings?

First, that I do not find that leaders' socio-demographic attributes play a large role in inducing cooperation may simply be due to the characteristics of the APEP groups. As noted, the APEP associations cover a fairly narrow demography, wealth disparities between members are small, and the membership body is relatively homogenous religiously and ethnically. Alternatively, it may be that leaders' attributes matter even in our case, but the inability to reject the null hypothesis results from low statistical power. In chapter 4 I provide, however, some reasons to believe that low statistical power is not driving these results.

Finally, it might be that some leader attributes matter, but such impact is not related to easily observable variables such as age or wealth, but to less directly measured variables such as status in one's community. For example, it may be interesting to examine the effectiveness of self-help groups in which the group leaders are also community leaders, against the effectiveness of groups where leadership in the groups and the community do not overlap. In this dissertation I did not examine directly the impact of the centrality of leaders' network position nor their standing in the community. I plan to address those topics in subsequent studies.

Network Effects

The use of network data in chapter 4 demonstrates the utility of focusing on the *type* of social ties, in addition to the common focus on the *strength* of social ties. When examining the relationship

between the structure of the board of directors network and a group's level of monitoring, I find that dense 'friendship' networks are negatively correlated with monitoring and accountability, whereas dense 'advice' networks are positively correlated. Unfortunately, since I do not have baseline network data from the founding days of the APEP associations, I am reluctant to make here any causal claims. Designing studies that can identify the casual effect of different types of social ties is a promising and exciting future avenue of research.

More so, in the dissertation I have focused only on the relation between the features of a network and group-level monitoring. However, there are some goods reasons to believe that the *position or role of group members within a social network* — including their ties with the group leader — may affect their cooperative behavior. I plan to address this question directly in future work.

External and Scope Validity

Finally, the above discussion naturally raises the following question: *to what extent are the study's empirical findings conditional on the specific characteristic of the APEP groups*; for example, the fact that they operate in information rich environments? Since I have embedded some discussions about external and scope validity — as defined in Shadish, Cook and Campbell (2001) — in the different chapters, I will address this issue here only in passing by reflecting on the external and scope validity of two of the study's key findings.

Consider the key findings of the public good game regarding the independent impact of leader selection rules and the mediating role that legitimacy plays in tying between elections and cooperation. I argue that these results have a strong external validity and scope validity. First, I demonstrate in chapter 2, that these findings do *not* depend on the richness of information that characterizes the APEP groups[1]. Second, Uganda is a 'hard case' for testing the legitimating role of elections, given its long history of fraudulent elections. Third, Ugandan farmer associations

[1]This is because I designed the PGG experiments (i) to include a large number of subjects (ten), (ii) from a pool of ordinary members that did not hold any leadership roles, (iii) using a random sample of subjects from six different villages.

are also a 'hard case' for testing whether the design of intuitions has an independent impact on cooperative behavior. The farmer groups are relatively small, group members know each other well, and ties between members extend to spheres other than their joint membership in the farmer groups. In such conditions one might expect that local norms and informal rules would prevail. Demonstrating the impact of institutional design in a hard case such as this, increases the confidence that leader selection processes matter in larger political units where formal rules have 'a stronger bite'.

Consider, by contrast, the external validity of one of the key results of chapter 4, that appointed leaders are not more able or altruistic than elected leaders. In that chapter I hypothesized that this finding likely depends on the fact that the APEP groups serve a small geographic area, in which communities tend to be relatively homogenous in terms of wealth, religion and ethnicity. Since these conditions support information flows, it remains to be tested whether appointments result in more able leaders where the quality of information is lower. Similarly I stipulated that this result might be due to the fact that electing leaders of farmer groups is a salient issue, given the leaders' ability to significantly impact the members' welfare. In other types of groups, where the ability of leaders to affect members' welfare is smaller, we may find different results to the ones reported in chapter 4. The extent to which this finding depends on (i) the quality of information flows and/or (ii) the saliency of leadership selection, is an open question that will hopefully be addressed in future work.

Part II

Appendices

Appendix A

Empirical setting and data

A.1 Selection of Development Project

In this section I summarize some of the reasons for studying the APEP groups to inform the empirical parts of the dissertation. Briefly, the APEP project presents many advantages. First, the project's scope and size allow me to conduct a large-scale quantitative study within the boundaries of a single nation, thus securing the homogeneity of the political and legal environments, as well as many project-related factors.

Second, as noted above, the process of group formation occurred under the lead of a few project field-trainers. As a consequence, APEP groups, which were formed around the same time and for the same purpose, have similar governance structures and leadership positions whose roles and functions are comparable across sites.

Third, while all the APEP associations have similar governance structures, there exists a significant amount of variation across associations in variables of interest, such as the value of the public good produced [chapter 2], availability of private income opportunities [chapter 3] and leader selection rules [chapter 4].

Fourth, the Ugandan farmer associations that I study fit the features of small self-help groups, described in the theoretical model [chapter 3], well. The associations are made up of farmers from several nearby villages. All group members share a common main goal, obtaining higher prices for their outputs, with secondary goals including obtaining lower input prices and learning about better farming practices. The group leaders spend only part of their time working for the group, with the rest devoted to farming their land or working at other off-farm jobs. When leaders negotiate higher prices for their crops, they benefit directly and significantly from the group public good through the higher price that they receive for their own crops.

Finally, that the study's subjects are members of pre-existing groups contributes to the external validity of the study's findings. This is especially true given the study's reliance on behavioral 'lab-in-the-field' experiments for identification of causal relations. It has been argued that laboratories are a rather poor setting for testing how subjects balance the tension inherent in social dilemmas (Burnham and Johnson, 2005). Laboratories strip individuals from context, and are limited in their ability to replicate the mutual trust, past experience, shared norms and group identity which are central for balancing tension between private and public interests (de Rooij, Green and Gerber, 2009, Henrich et al., 2004, Levitt and List, 2007). This consideration has led researchers to argue that *pre-existing groups* are the most natural setting to test theories about the emergence of cooperative behavior (Herrmann, Thóni and Gächter, 2008, Nowak, 2006, Rand et al., 2009).

A.2 Sampling Design

This section briefly describes the sampling scheme used in this dissertation. To reduce crop-related variability, I limited the target population to only those associations that marketed the same crop. Coffee was selected since it was the most common cash crop marketed by the APEP groups. Limiting the sample to coffee reduced the universe of farmer association in about half: from 204 to 113. An additional 8 DCs were excluded due to the following: I excluded 2 associations from Bugiri because coffee turned to be very peripheral in that district. I further excluded 5 associations from Busheni district because those groups were formed many years before APEP, and were not comparable in terms of their organizational capacity. Finally I excluded from the sample an association from Kamwenge because it was the single DC in that district and surveying it would have been logistically complicated and prohibitively expensive. The final universe of cases comprises of 105 farmer associations, all created by APEP facilitators after 2005. Once the target population was chosen, I used a stratified, random, multistage cluster design to select our sample.

Step 1: Define Strata Though the universe of farmer associations is spread over 9 districts, I grouped associations into 5 strata. Strata were defined by meaningful district-areas: neighboring districts that were covered by the same project field trainers and trading partners, and that share a dominant ethnicity and/or were historically part of the same district. Figure A.1 presents a map showing the location of the sampled districts.

Step 2: Farmer Associations (DCs) Based on a power calculation performed on simulated data, I sampled 50 farmer associations. I used unequal probability sampling without replacement to sample associations within strata (proportional to their size). The number of associations that were sampled from each stratum was proportional to the number of associations in each strata. According to this scheme, each sampled associations is representative of his stratum, without a need to further use weights.

Step 3: Village-level Producer Organizations (POs) Prior to sampling group members, I used an independent random sample to select six producer organizations (or POs) from each association, for a total of 287 POs. In few cases, when a farmer association had fewer than seven POs, I selected all of its village-level groups.

Step 4: Group Members I sampled 36 respondents (36) from each farmer association. The number of sampled members from each of the six *sampled* village-level producer organizations was proportional to the size of the groups. This assured that the sample is self-weighted. Total sample size is thus 50 DCs X 6 POs X 6 members per PO = 1,800. I succeeded in surveying 1,781 out of the 1,800 sampled group members. I refer to this data source as the "members' survey".

Step 5: Relevant Pool of Potential Candidates The DC board of directors, composed of all representatives from the PO groups, including those that become the leader, is the relevant pool of candidates for the leader's position. A significant effort was made to survey each of the DC board members. To keep the size of the relevant pool of candidates manageable, I adopted the following rule: when the number of groups that make the DC was up to ten, I surveyed the PO chairperson, the two most active representatives from each farmer group, and any members holding DC executive positions (this includes the DC manager as well as a Chairperson, Secratary, and Treasurer). When the number of groups was larger, I invited only the PO chairperson plus the most active other PO representative, plus those members holding DC executive positions. For example, if a farmer association is comprised of eight village-level producer organizations, I surveyed 28 prominent members based on their position: 4 DC executives + 2*8 PO representatives + 8 PO chairmen). The survey team visited each association up to four times to reduce attrition-rate, which was brought down to less than ten percent. A summary of the sample scheme is provided in table A.1 below.

Table A.1: **Sample Design**

Step	Sampling Unit (SU)	Number of SUs	Sampling Method
1	Target Population	105 DCs	Coffee growers
2	District-area	5	Stratified – proportional to # of DCs in strata
3	Farmer Associations (DCs)	50	unequal probability without replacement
4	Produce Organizations (POs)	6 per DC	Clustered – simple random sample.
5	Group members	36 per DC	Clustered – probab proportional to group size.
6	Potential Pool of Candidates	$\sim$ 28 per DC	No sample: Complete Network.

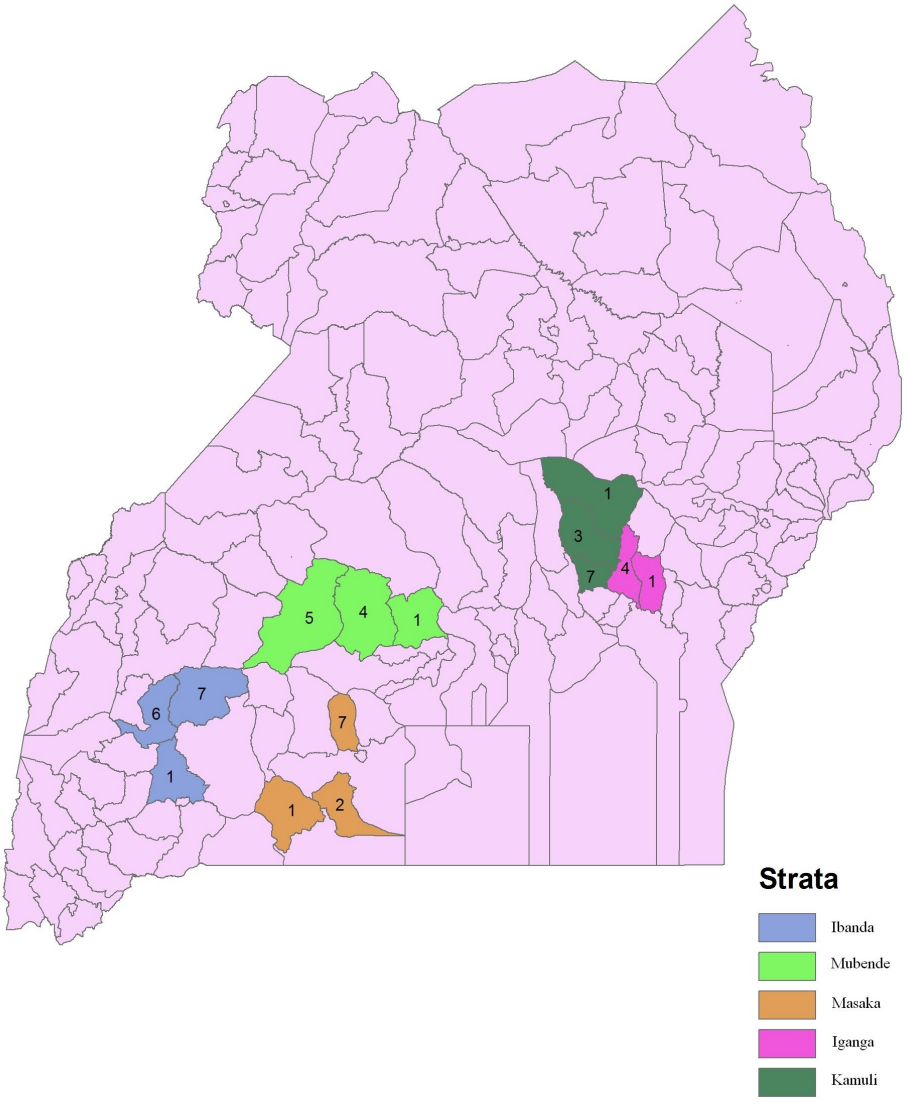

Figure A.1: **Location of sampled farmer associations by strata.** Colors indicate sample strata. Numbers indicate the number of farmer associations sampled within each strata and district.

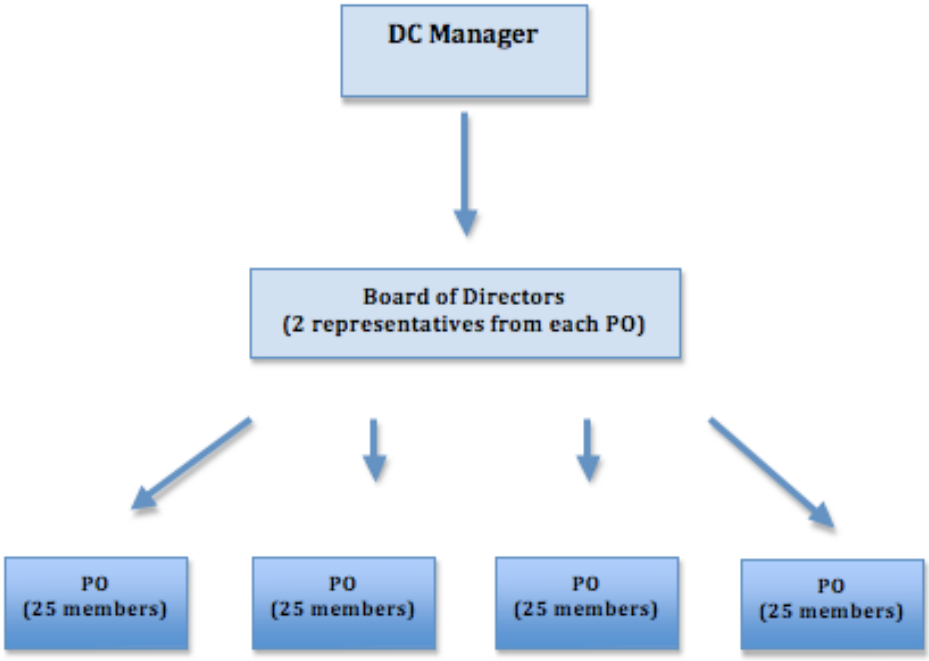

Figure A.2: **APEP associations organizational structure.** In this Figure I present the organization structure of the APEP associations. Each farmer association (known as DC) is comprised of about 10 village-level producer organizations (POs), themselves comprised of about 20-25 members. Each of the POs that make up the association elects two representatives to serve on the DC board of directors. The DC is led by a DC manager, who may be thought as the CEO. Responsibilities of the board of directors include, monitoring the work of the DC executives (including the manager), representing the opinions of PO members at the associational level, and aiding in the transmission of information. The manager's most important responsibilities include negotiating input and output prices, organizing the collection of crops (including hiring and supervising employees), and searching for buyers. Additionally, leaders help coordinate group activities and facilitate the flow of information throughout the association.

A.3 Data Sources

I use five main data sources to construct the variables used in the empirical parts of the dissertation: (i) Members' Survey; (ii) Representatives' Survey; (iii) Network Data (iv) PO questionnaire; and (v) DC questionnaire.

Members' Survey: An individual-level survey executed to a random sample of group members who do not hold any leadership position in the association. Trained enumerators administered the members' surveys in a face-to-face interview, for a total of 1,781 surveys.

Representatives' Survey: An individual level survey of directors, who represent the village-level POs at the farmer association board. These include the four members of the executive committee, PO representatives who serve as directors at the DC-level, and the chairpersons of the village-level POs. Trained enumerators administered the members' surveys in a face-to-face interview, for a total of 1,313 surveys.

Network Data: Each of the representatives (i.e., board directors) that completed a 'Representatives' Survey' was also asked to complete a network module based on a roster of the names of all DC representatives. The social network data allows me to analyze both the structure of the DC leadership network (e.g., its density), as well as the position of each member within the network.

PO questionnaire: This questionnaire was administered to 3-4 leaders from each of the six sampled village producer organizations (POs), for a total of 287 group-questionnaires. These questionnaire had two parts. In the first part, PO leaders were asked to provide information at the group-level (e.g., the group's year of foundation, its number of members, etc.) At the second part, group leaders were asked to provide additional information on the group members using a complete list of members. This part included information such as the leader's assessment of the crop quality of each of the members, and whether members sold their coffee to middlemen.

DC questionnaire: In each of the 50 sampled farmer associations, a trained enumerator interviewed the four DC executives together (DC manager, DC chairman of the board, DC Treasurer, and DC Secretary). This questionnaire was designed to capture information at the association-level (e.g. DC year of creation, number of POs, identity of APEP trainer, rules and procedures, buyers information, etc.). In addition, the executives were asked to provide information on the marketing activities of the association using the DC's books and records.

A.4 Implementation

The survey instruments were piloted during the first two weeks in July 2009, and were translated to one of three local languages. Data was collected between late July 2009 and September 2009 by a group of 60 experienced local interviewers (enumerators), who administered all instruments in the native language of the respondents. Enumerators, who were hired directly by the PIs, were divided into three "language" teams. The eastern team covered 16 farmer associations in Iganga and Kamuli districts, where Basoga is the primary local language. The central team covered 20 DCs from Mubende, Mityana, Masaka and Rakai districts, where locals speak Luganda. Finally, the western team covered 14 DC from Kiruhura, Mbarara and Ibanda districts, where Ranyankole is the lingua franca. Enumerators went through a lengthly training in class (4 days) and in field setting (4 days), which included also training on human subjects issues as well as survey techniques. Enumerators were supervised by team leaders by ratio of 1:5.

In each sampled association, data was collected in four rounds. First, an interviewer scheduled a meeting with the DC executives. In that meeting the interviewer introduced the study and asked for the DC leaderss' cooperation. In addition, in that meeting s/he administered the DC-level questionnaire, and obtained a list of all DC board directors. In the second day of enumeration, the research team conducted interviews with group representatives to the DC board and with the chairmen of all village-level groups, who were mobilized by the DC executives to a central location. In addition to individual-level interviews, leaders from each sampled village-group were

asked to respond to a PO questionnaire, and to provide a complete list of all group members. Between the second and third round, I sampled 36 members from each sampled DC (including 8 replacements). Immediately after the sampling procedure, an interviewer travelled back to meet with the associations' leadership. In that meeting, the interviewer gave the DC leaders the list of sampled members and coordinated with them the next round of interviews. Once again, I relied on the DC leaders to mobilize the sampled members to a centralized location. In the third day of enumeration, individual-level interviews were conducted with the sampled members and with representatives who were not present in the previous day. Finally, the survey team traveled to each association for an additional day in order to reach sampled members or representatives who, for any reason, were not present in the main enumeration days.

A.5 Missing Data

Great care was taken to reduce missingness. The research team administered association-level questionnaires in all 50 DCs PO questionnaires in 287 out of 289 sampled village-level farmer groups. Out of a sample of 1,800 "ordinary" members (i.e. farmers with no leadership role) the research team managed to conduct individual-level surveys with 1,781 farmers. As for data in surveys that were missing at random (MAR), I used Patrick Royston's ICE multiple imputation package in Stata, which applies a chained equations approach. I imputed missing data only for some asset variables and demographic characteristics, but chose not to impute data for farmers' activities as group members, such as marketing decisions and agricultural practices.

A.6 Sample Descriptive Statistics

Tables A.2, A.3, and A.4 provide some descriptive statistics, respectively, for the sample of 'ordinary' members and group, DC representatives, and the associations overall.

Table A.2: **Members: Summary Statistics**

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Dependent Variables					
Bulked at least once with DC (self-report)	0.61	(0.49)	0	1	1746
Bulked at least once with DC (leaders-report)	0.61	(0.49)	0	1	1712
Welfare Increase (std.)	0	(1)	-4.52	3.04	1759
Independent Variables					
Sex (male)	0.68	(0.47)	0	1	1781
Age	45.58	(14.42)	14	95	1781
Education (Std.)	0	(1)	-1.53	1.88	1781
Seasonal Coffee Yield (units of 100)	3.80	(5.45)	0	100	1756
Years since Joining Group	3.71	(1.76)	1	8	1769
Middleman honesty	0.91	(0.28)	0	1	1698
Leader's co-villager	0.11	(0.31)	0	1	1782
Associational-life (std.)	0	(1)	-3.14	2.85	1633
Community leadership experience	0.46	(0.49)	0	1	1583
Farming Primary Occupation	0.56	(0.5)	0	1	1766
Village distance to District Capital	29.18	(18.01)	0	96.60	1705
Village Distance to Trading Center	1.09	(2.72)	0	28.98	1719
Village Distance to DC Crop Collection	2.49	(2.99)	0	19.32	1715

Table A.3: **Representatives: Summary Statistics**

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Sex (male)	0.8	(0.4)	0	1	1316
Age	46.83	(11.88)	20	86	1316
Ability with cognitive tests (std.)	0	(1)	-3.69	1.9	1058
Ability without cognitive tests (std.)	0	(1)	-2.66	1.69	1316
Seasonal Coffee Yield (units of 100)	5.38	(7.65)	0	100	1307
Years since Joining Group	4.1	(1.65)	1	8	1314
Associational-life (std.)	0	(1)	-2.93	2.37	1140
Community leadership experience	0.74	(0.43)	0	1	1105
Wealth (std.)	0	(1)	-2.39	9.15	1310
Good health	0.84	(0.37)	0	1	1305
Born in village	0.51	(0.5)	0	1	1316
Household Head	0.89	(0.32)	0	1	1316

Table A.4: **Farmer Associations: Summary Statistics**

Variable	Mean	(Std. Dev.)	Min.	Max.	N
Age of DC	2.84	(1)	1	5	50
No. of group members	211.92	(111.59)	61	536	50
Leader's effort (std.)	0	(1)	-2.66	2.5	50
Monitoring level (std.)	0	(1)	-2.35	2.23	50
Leader's ability (std.)	0	(1)	-1.69	1.49	46
Leader's ability (std.)	0	(1)	-1.86	1.65	42
Mean seasonal yield (units of 100)	3.79	(1.8)	0.89	7.76	50
Ethnic fractionaliztion	0.23	(0.25)	0	0.8	50
Mean Associational-life	0	(1)	-2.08	1.74	50

A.7 Measurement of Key Variables

This section provides some supplementary information on the key variables used in the analysis.

Value of the Public Good

In the analysis, our primary measure of member's marketing decisions was derived from PO leader's reports of member's marketing decisions. This measure was used because the PO leader has little incentive to provide misleading information about how actively members sell through the associations. In contrast, members may feel uncomfortable mentioning that they generally do not sell through the association, since this is generally a violation of association rules. The two basic measures of the member's marketing decisions are, 1) an indicator variable capturing whether a member sold his crops via the association at least once in the past season, and 2) the share of a member's total seasonal coffee yield that was sold via the farmer group in the past season. Figures A.3 and A.4 show that these marketing decisions vary significantly across groups.

Measures of member's marketing decisions based on member's self-reports showed similar results. We also check our results against alternative measures of collective marketing, gathered from interviews with group leaders and the associations' books and records. These measures also deliver similar results.

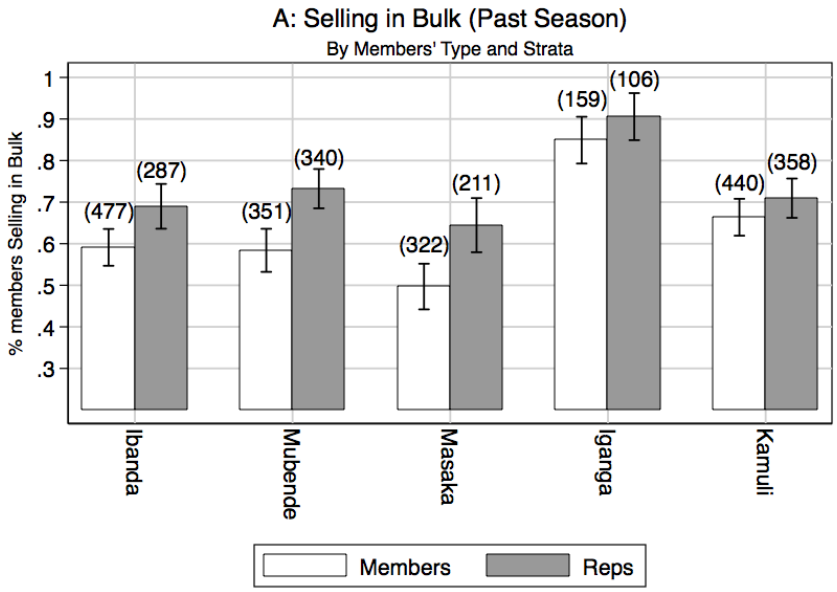

Figure A.3: Proportion of respondents reporting to sell coffee via their farmer group, at least once, in the past season, by type of member and region. No. of observations in parenthesis. Caps represent 95% CI.

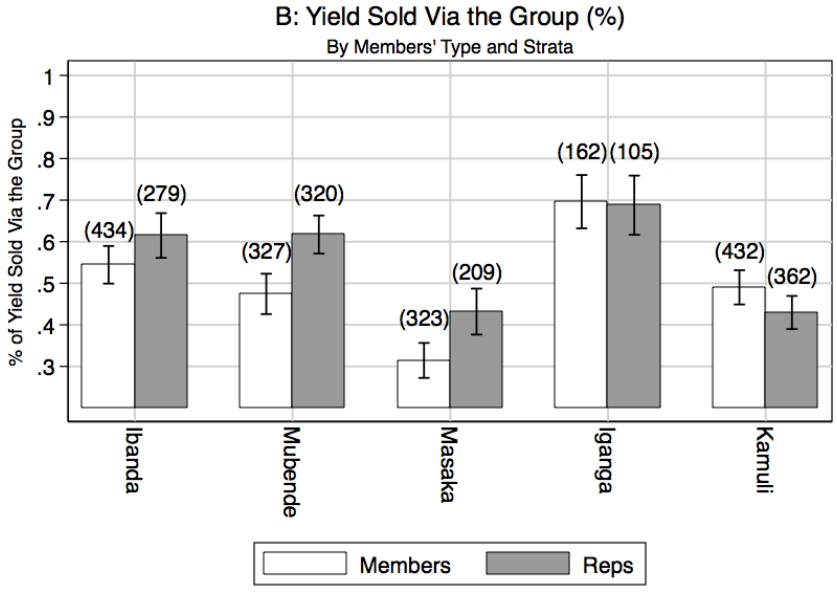

Figure A.4: Mean of the proportion of a member's total seasonal yield sold via her group, by type of member and region. No. of observations in parenthesis. Caps represent 95% CI.

Leader's Ability

To construct individual's ability measure we used information on respondent's ability to read and write, a 7-scale variable that measures the respondent's educational attainment, and English proficiency[1]. Respondents also completed two types of cognitive tests: solving a simple maze in less than two minutes and solving a raven test comprised of 12 questions in two minutes[2]. Several checks give us confidence in the ability measure. In Figure A.5 Panels A-C, we see that group members who hold high-skilled off-farm jobs have significantly higher ability than those that do not . In Figure A.5 Panel D, we observe that the representatives elected by the village-level groups have higher ability on average than ordinary members . Finally, we find that ability is highly correlated with wealth, as shown in Figure A.6.

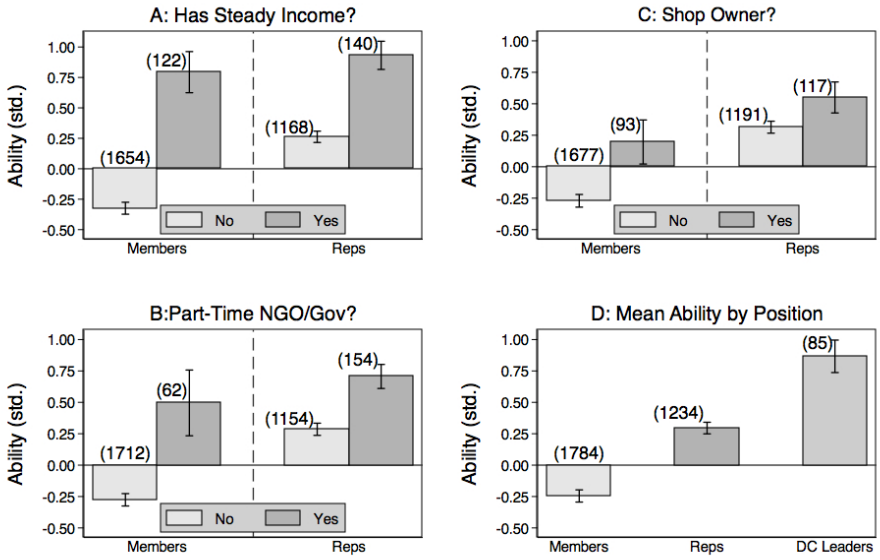

Figure A.5: Relationship between the composite measure of ability and a binary indicator of employment status, by position in the association (ordinary members and board directors). In each panel, point estimates represent the mean ability score for survey respondents who report having any steady source of off-farm income (panel A), work part-time in an NGO or a local government (panel B) or own a store (panel C), against the mean ability score for those who do not have such jobs. Caps represent 95% confidence intervals.

[1]Because no local language is spoken by more than 20% of Ugandans, English is the lingua franca of the business and political class. English proficiency allows individuals to communicate with potential trading partners outside their small geographic areas.

[2]The cognitive tests can be shared with readers upon request.

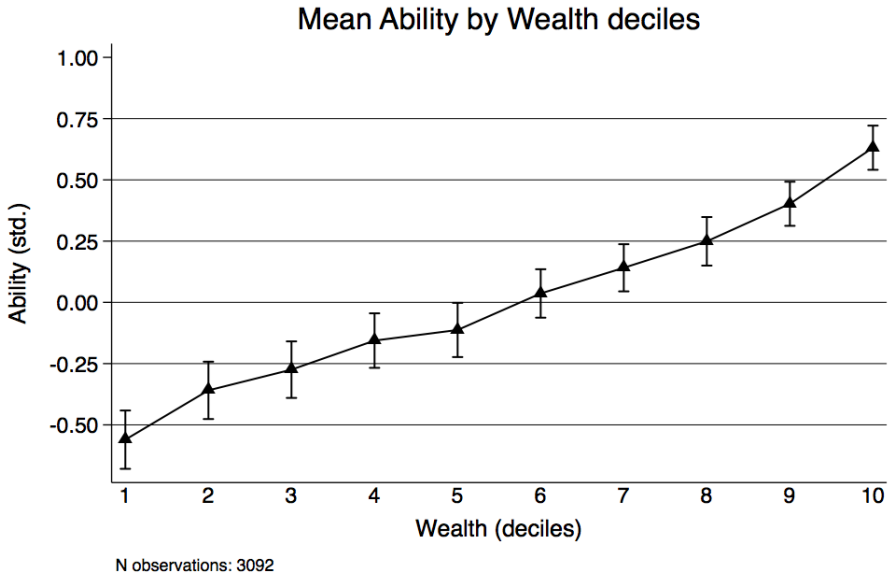

Figure A.6: Relationship between wealth (deciles) and ability (std), for the entire sample (N=3092).

Monitoring

The tables below provide some additional information on the variables used to measure group monitoring institutions. Table A.5 identifies the variables used to construct the monitoring measures. Table A.6 shows that these veriables tend to be positivively correlated, often significantly. This suggests that groups that implement one type of monitoring institution are also likely to have implemented others.

Table A.5: **Variables used to measure group monitoring**

Variable	Behavioral Measures	Type	From
irsf15	Whether there is a rule regarding the leader's time commitment	dummy	Reps
irsf20b	The extent to which external auditors are used	categorical	Reps
irsf21	Whether respondents have asked to review the DC's books	dummy	Reps
sf222	Whether someone is responsible for monitoring the DC manager	dummy	Reps
	Attitudinal Measures		
irsf24b	Whether the manager is transparent	dummy	Reps
imsf28c	The extent to which the manager is accountable	categorical	Members
imsf32	Whether the manager is monitored	dummy	Members
imsf33	The extent to which the manager is transparent	categorical	Members

Table A.6: **Correlation matrix of the variables used to measure group monitoring**

	(1)							
	irsf15	irsf20b	irsf21	sf222	irsf24b	imsf28c	imsf33	imsf32
irsf15	1							
irsf20b	0.577***	1						
irsf21	0.318*	0.449**	1					
sf222	0.672***	0.657***	0.306*	1				
irsf24b	-0.0679	-0.0762	0.110	-0.141	1			
imsf28c	0.0334	0.0100	0.250	-0.0409	0.468***	1		
imsf33	0.281*	0.410**	0.545***	0.187	0.169	0.629***	1	
imsf32	0.199	0.487***	0.325*	0.252	0.0679	0.186	0.495***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Welfare effect

Figure A.7 provides information on the distribution of welfare changes, for the entire sample.

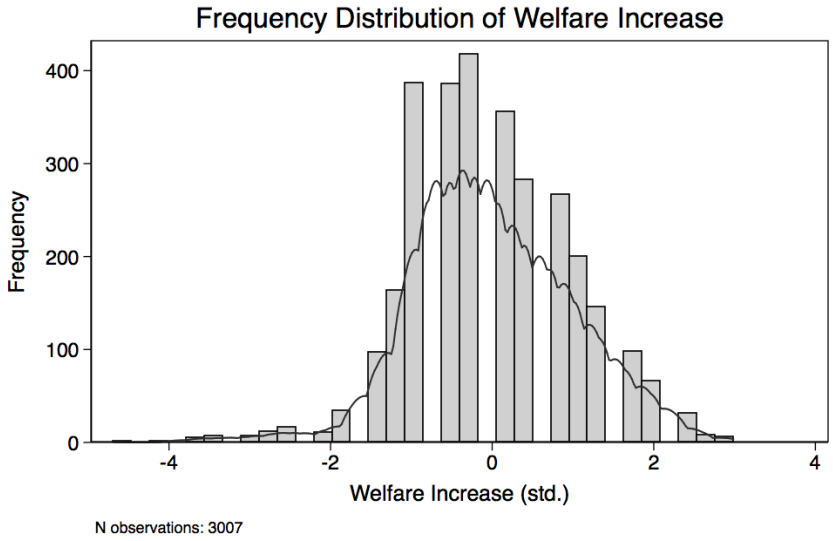

Figure A.7: **Distribution of Self-Reported Welfare Effect Since Joining the Farmer Group**

Appendix B

Chapter 2: Public Goods Games Manual and Scripts

B.1 Interviewer Manual

[These are instructions given to the interviewers during training in addition to the script of the behavioral game.]

Framework:

This game will be played with about 36 Producer Organization (PO) members per Farmer Association. The 36 PO members will be split up into three even groups, with 8-12 PO members in each (Note that the instructions are for situations with 12 players). We will mix PO members to maximize the heterogeneity of the group. Each group will go to a different location, and two interviewers will be responsible for each group. Each group will play one of three variants: The first group will play a baseline public goods game (with no monitor), the second will play a public goods game with a randomly selected monitor, and the third group will play a public goods game with an elected monitor. The baseline public goods game will have six rounds. The variants will

each have two "traditional" preliminary rounds with no monitors, followed by four rounds with monitors. Given our focus on the effects of monitor selection, players are assigned to the same group for all six rounds.

Games Summary:

In Public Good Games, players must decide how much of their endowment to keep in their private pocket and how much to put in the group pot. Each player, for each round of the game, receives an endowment of ten 100USH coins totaling to 1000USH. The total amount donated to the group pot is then doubled and redistributed evenly among the players. Payoffs will be rounded up to the nearest 50USH. The contributions will be made public to all the players, but the identity of the contributor will not. The total donation, as well as the average donation, and some exemplary individual payoffs will also be made known to the players.

In the monitor variant there will be a monitor who will be able to spend 100USH to reduce another players' payout by 300USH. The monitor will choose which contributions he wants to reduce. The players will remain anonymous to everyone but the interviewer. In the randomly-selected monitor variant, the monitor will be chosen at random. In the elected variant, the monitor will be elected by the group.

Player i's payoff in round t is calculated as $\pi_{it} = (10 - x_{it}) + \frac{2\sum_{i=1}^{N} x_{it}}{N} - 3P_{it}$, where x_{it} is i's contribution to the public account in round t, and $P_i \in \{0, 1\}$ indicates whether i was punished in round t, and N is the number of players.

Logic:

On an individual level, it is more profitable for a person to "free-ride", and have everyone else contribute to the collective fund. In contrast, the most profitable outcome for the group is if all players contribute all their tokens to the public fund. This game examines how players will balance their own self-interest and the well-being of the collective.

We will be looking at two factors: the effect of monitoring, and the effect of election. We expect

that the potential reduction of payoff by the monitor will increase the total contributions to the collective fund by the group. We expect that the election of the monitor will have an even larger effect on players, since players may be more responsive to monitoring by someone they have elected. We are also interesting in observing monitors' willingness or sense of duty to sacrifice money to sanction someone who donated little or nothing to the collective.

Materials (Per Group):

– 2 large sheets of paper and tape

– 12 3-sided cardboard screens

– 12 sets of ten 100USH

– 12 "Personal Pocket/Collective Pot" boards

– 2 markers

– cards (1-12) for Group 2 only

– calculators

– voting materials for Group 3 only

Setup:

For each Group, there will be two interviewers. The two interviewers will be assigned to be "interviewer 1" and "interviewer 2", as the script will have the interviewers doing different things. Interviewer 1 will talk more, while interviewer 2 will be more focused on data collection and presentation during the game. Interviewer 2 should be prepared to do more math.

Each room/area should have 12 cardboard privacy screens set up with a "Personal Pocket/ Collective Pot" board and ten 100USH coins on the "Personal Pocket" side. There also should be a demonstration set at the front of the room and two large sheets of paper (to serve as a chalkboard). The ideal set up is a circle (Figure C.1);

if this is not possible, have the participants sit in rows, as shown in Figure B.2.

Figure B.1: Setting of the Game.

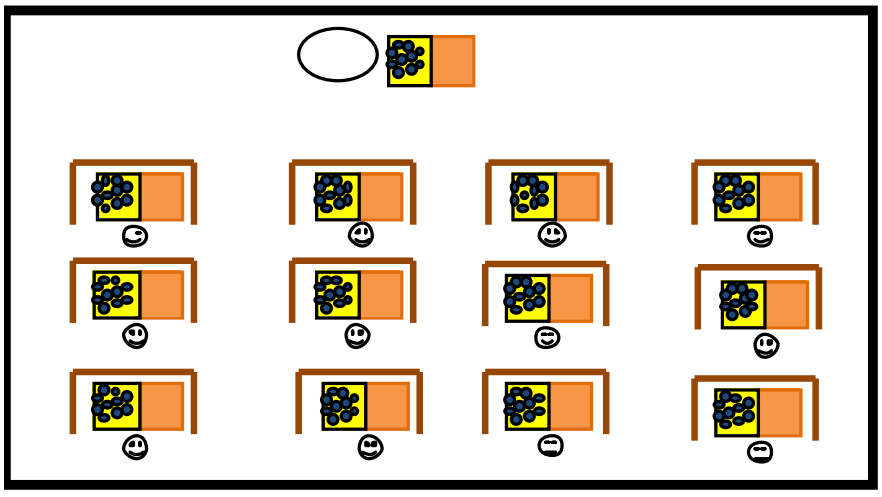

Figure B.2: Alternative setting of the game.

Basic Explanation of the Procedures of the Game:

There are some examples to run through before the start of the first round. These are thorough and each demonstrate one full round. The format of data presentation on the large sheet of paper at the front of the room will be the same for each example and each round.

Here is Example 1: The first series of digits represent how much each individual in the group has donated to the group pot. This has to be reported in increasing order. This data will be recorded by interviewer 2 walking around the room to each player's screen. This list of numbers is very important because it allows players to all see how much everyone in the room donated to the collective, but also allows everyone to remain anonymous.

	Average	Total	Double	Payback
0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2	100	1200	2400	**200 USH**

It will be important to go through the math of all of them. The "Average" is the sum of the list of numbers and divided by the number of players (12 in rounds without a monitor, 11 in rounds with a monitor). This allows players to see how much each player in the room contributed on average.

The "Total" is the sum of the list of numbers and the "Double" is that sum doubled. "Payback" is the "Double" divided by the number of players in the room (12 in rounds without a monitor, 11 in rounds with a monitor). This ensures that the collective good is now redistributed evenly to each player in the room.

The examples will make clear that: 1) people that donate more to the group pot will make less than people who do not donate or donate fewer coins; 2) the group as a whole gets the most money if everyone puts in all ten coins; 3) if a player puts a lot and everyone else puts in little or nothing, the player can end up with less than the 1000 shillings s/he started out with.

Rounds:

Group 1, playing the Baseline (no monitor) version, will simply repeat the rounds six times. Groups 2 and 3 playing with monitors, will complete two rounds as Baseline, and then 4 rounds according to the monitor variant. The players should not be told how many rounds they are going to play, because there is a tendency to contribute very little when people know they are playing the "final" round.

For Groups 2 and 3, after having played two traditional rounds, the interviewer will inform the players that they will play the game again, but a little differently than in the previous rounds. A monitor will be selected after the first round for the rest of the game. For Group 2, this will be random. For Group 3, the monitor will be elected by the room through written ballots. In both cases, one of the players will become a "monitor". The monitor will not be participating as a regular player. He would therefore not benefit from the decisions of the players and his decision will be simply affected by whether he approves or not the behavior of the players.

The monitor will receive 1000 USH and given the option to give-up 100USH to take away 300USH from contributors whose donation s/he is dissatisfied with. The monitor may do this to as many contributors as he would like, but every single contributor can be penalized only once. The interviewer must also emphasize that the choice is up to the monitor; s/he does not have to reduce anyone's payoff. The interviewer will demonstrate two examples and will ask the monitor a question to ensure comprehension.

After the round is played and all the contributions have been reported on the board, the monitor will have to give the interviewer 100USH from his own endowment for each contributor he wants to penalize and then walk up to the board and mark the numbers corresponding to the contributions s/he wants to "reduce" by drawing an X below the numbers. The identity of the contributors will remain anonymous, but all the players will know the amount of donations that the monitor considers not satisfactory.

Payoff:

At the end of the last round, the interviewer will ask the player sitting at the front right corner to pick a card from a deck numbered 1-6. This determines which round the group will be paid for at the end of the day.

B.2 Script of the Public Goods Game: Elected Monitor Condition

[Each treatment condition had a slightly different script. The baseline and random monitor conditions are available upon request.]

Game introduction and instructions:

(Interviewer 1 begins instructions, interviewer 2 records everyone's ID numbers on the record.)

Hello my name is [interviewer 1] and my colleague's name is [interviewer 2]. Please take a seat at one of the stations. We would like to thank you all for being cooperative and for participating in the various activities.

For this activity, there must be absolute silence. You are not allowed to talk to each other. While in the group, you cannot ask questions or talk. This is very important. Please be sure that you obey this rule, because it is possible for one person to spoil the activity for everyone. If one person talks about the activity while sitting in the group or with other people later, we will not be able to continue the activity. Do not worry if you do not completely understand everything as we go through the examples here in the group. We will take questions when we are finished explaining.

This activity will have a few rounds of decisions. However, only one round will count for payment, which will be chosen randomly once the activity is completed. For the round that is chosen, the money will be yours to take home and use as you please. Since we do not know which round will count, you should decide in each round as if you were deciding on real money.

In front of you there is a board. One side represents your Personal Pocket, the other side

represents a Group Pot. There are now 1,000USH in your personal pocket. In this activity, you will have to decide how many Shillings you would like to keep for yourself in your personal pocket, and how many you would like to contribute to the group pot.

You must understand something very important about the group pot. The group pot will include only the contributions from all the people participating in this activity. Once everyone has decided how much to give to the group pot, I will add up the total amount and the research team will double the amount. The group pot will then have twice the amount of money people contributed to it. I will then redistribute the total earnings equally among all twelve of the people participating in the activity. Each person's payoff will be rounded up to the nearest 50USH.

Everyone's donations and decisions will be anonymous, and the screens are here to ensure that. No one will know another's contributions. Do not look at other people's boards. Only the interviewer will know how much each person has donated and the interviewer will never tell anyone else. The number of coins contributed by each person will be reported once all the contributions have been made, but I will not say who donated each number of coins. Names or ID numbers of the people here will not be used throughout the activity.

Let's go over a few examples.

Interviewer 1 explains and demonstrates the coins. Interviewer 2 writes results on the large sheet of paper. Participants should be able to see all three examples at one time by the end of the explanations.

Interviewer 1: One way the activity might turn out is that six participants decide to give nothing and the other six decide to give 200 USH to the group pot. Demonstrate by moving two coins from Personal Pocket to Group Pot.

Interviewer 2 writes on large sheet of paper and explain each number:
"0,0,0,0,0,0,200,200,200,200,200,200", total (1200), average (100), doubling (2400), and Payback (200).

0, 0, 0, 0, 0, 0, 200, 200, 200, 200, 200, 200	Average 100	Total 1200	Double 2400	Payback **200 USH**

Interviewer 1: In this case, the results will look like this. We will have a total 1,200 USH in the

group pot. The average donation is 100 USH. The research team double the total contribution, so the group pot now has 2,400 USH which will be divided equally amongst all the participants. In this case, everyone will get 200 USH.

The participants that donated 0 to the group pot will get 1000 shillings from their personal pocket and 200 from the group pot, making 1,200 USH. The participants that donated 200 USH will get 800 USH from their personal pocket and 200 USH from the group pot making 1,000 USH.

Interviewer 2: As interviewer 1 explains the individual payoffs, demonstrate on board like the following:

0, 0, 0, 0, 0, 0, 200, 200, 200, 200, 200, 200 1000 + 800 + 200 = 200 = 1200 total 1000 total	Average 100	Total 1200	Double 2400	Payback **200 USH**

Figure B.3: Interviewers at work.

Let's try another example.

In this case, everyone gives all 1000 shillings to the group pot.

Interviewer 2 write and explain each number: "1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000," total (12000), average (1000), double (24000), payoff (2000)

Interviewer 1 explain: The total contribution was 12,000 USH. In this case, the average contribution was 1,000 USH. Now the research team will double this amount. The group pot now has 24,000 USH. Everyone will get 2,000 USH. Since everyone donated everything, everyone has 0 in their private pocket, and everyone gets 2,000 USH from the group pot. Everyone has made 2,000 USH.

Let's try a final example, in this case 7 people don't give anything to the group pot, 4 people give 100 USH, and 1 person gives 800 USH. The results will look like this:

Interviewer 2 write: "0, 0, 0, 0, 0, 0, 0, 100, 100, 100, 100, 800," total (1200), average (100), double (2400), payoff (200). Also write out the three different contributions and their total pay off during explanation.

Interviewer 1: The total donation is 1200 USH. The average donation is 100 USH, The research team will then double this number so that the group pot has twice the money, or 2400 USH. The payoff is 200 shillings for each person.

The participants that donated nothing will get 1000 USH from their personal pocket, as well as 200 USH from the group pot. This makes 1,200 USH. While the participants that donated 100 will get 900 from their personal pocket, as well as 200 from the group pot. This makes 1100 USH. The participants that donated 800 USH will have 200 USH from their personal pocket, as well as 200 USH from the group. This makes 400 USH.

Here are some things to remember:
People that donate coins will make less than people who do not donate or donate fewer coins. In the first case, 6 people donated 200 USH and 6 people donated 0. The people that donated 200 made 1000 USH, but the people that donated 0 made 1,200 USH.

The group as a whole gets the most money if everyone puts in all 1000 USH. Remember the second example. Everyone donated everything, and everyone made 2,000 USH.

Also keep in mind that if you put a lot and everyone else puts in little or nothing, you can end

up with less than the 1000 USH you started out with. Remember the third example. Some made 1,200 USH, while the person who donated the most made only 400 USH.

Before we begin the activity, does anyone have any questions?

[Interviewers should answer questions, but they should stick to the script as much as possible. Also, they do not have to invite too many questions. They have to keep it short!]

Rounds 1 and 2

Interviewer 1: Now please use the screen we have provided and decide how much you would like to contribute to the group pot. We will go around and record your decisions. Place the amount of coins you would like to contribute to the group pot on the right side of the board, with the picture of the group of people. Remember you do not have to contribute if you do not want to. The coins in the group pot will benefit everyone in the room.

Interviewer 2 goes around and records decision on data sheet. Individual contributions are recorded next to the ID of the contributor. Calculate average, total, double, payoff.

Interviewer 2 writes the results of the donations on the large sheet of paper at the front of the room by writing all contribution amounts in increasing order.

Interviewer 1: Thank you. Please move all your coins back to your personal pocket. Here are the results of all the donations in the room.
Explain each number. Just to clarify, let's go over how much two participants are going to get.
Pick the third and eight donated amount and go over their payments:
The person that donated X will have Y in his personal pocket, and will get Z from the group pot, totaling Y+Z. The person that donated A to the pot will have B in his personal pocket, and get C from the group pot, making B+C.

[Repeat for round 2]

Interviewer 1: Now we will repeat the activity again.

Interviewer 2 walks around to make sure the setup has been followed.

(. . .)

[After round 2]

Monitor Election:

Interviewer 1: We will now repeat the activity again, but a little differently than the previous round. In this round one of the participants will become a "monitor". The monitor will be elected by all of the participants in this activity.

The monitor will not be participating in this activity like everyone else. He will not donate anything to the group pot, and he will not receive any payment from the group pot. He will be given 1000 USH and he can spend 100USH to take away 300USH from the private pocket of contributors he is dissatisfied with. He may do this to as many participants as he would like but he may only use 100USH for each participant. The monitor does not have to reduce from anyone and may keep all 1000 USH.

The monitor will not know the names of the contributors, only how much they donated. He will see the same results as everyone else. The monitor will only choose who to reduce money from by writing an "X" under the number indicating the size of the contribution. Now, we will elect the monitor.

Each of you has a piece of paper and a pen at your stations. Please write the ID number of the person that you would like to elect to be your monitor. An interviewer will walk around to collect your ballots and will tally the results. Please take a moment to look around at the ID tags on the other participants, so you know who you can vote for.

[Interviewer 2 collects ballots, and tallies results, and tells interviewer 1]

Interviewer 1 to elected monitor: You have been elected by the group as the monitor. Please step to the front of the room, and stand beside the large paper. Starting from the next round of

activity, you will be able to reduce 300 USH from any contributor, by giving up 100 USH. If you would like to reduce, you will give 100 USH to [interviewer 2], walk up to the board, and mark an X beneath the contributions you would like to reduce shillings from. So if you would like to reduce 300 USH from a person who contributed (point to the second from the left), place an X below it. If you would like to reduce 300 USH from a person who contributed (point to second from the right), place an X beneath it. Give one coin for each X to [interviewer 2].

Rounds 3-6

Interviewer 1: Now please use the screen we have provided and decide how much you would like to contribute to the group pot. We will go around and record your decisions. Place the amount of coins you would like to contribute to the group pot on the symbol. Remember you do not have to contribute if you do not want to, and the coins in the group pot will benefit everyone in the room.

Interviewer 2 goes around and record decision on data sheet. Individual contributions are recorded next to the ID of the contributor. Calculate average, total, double, payoff.

Interviewer 1 writes the results of the donations on the large sheet of paper at the front of the room by writing all contribution amounts in increasing order.

Interviewer 1: Please move all your coins back to your personal pocket. This is the result of all the donations in the room.
Let's go over how much two participants are going to get.

Pick the third and eight donated amount and go over their payments:
The person that donated X will have Y in his personal pocket, and will get Z from the group pot, totaling Y+Z. The person that donated A to the pot will have B in his personal pocket, and get C from the group pot, making B+C. Now, the monitor can decide if he would like to reduce 300 USH from a contributor.

Interviewer 1 to monitor: Would you like to reduce the 300 USH from any contributor? If yes, please

mark an X by the ones you want to reduce from and give 100 USH to [interviewer 2] for each X you mark. If not, let us know now.

Explain results to the room The Xs mean that 300 USH have been reduced from the person that donated the amount indicated to the group pot. So for example, one person that donated X has just had 300 USH reduced from their payoff. Instead of earning X, they will now earn Y.

Announce repetition Now, we will repeat the activity again. After you've made a decision the elected monitor will again be given the possibility to take some money away from you. Please move your coins back to their original position so that there are ten coins in your private pot.

(. . .)

[This process should be repeated for a total of 4 rounds.]

Payoff (end of 6 rounds):

The cards are numbered from one to six. We will ask one of you to pick a card. The number that will be picked will determine which round you will be paid for at the end of the day. For example, if you pick card number one you will be paid for round 1 (point to round 1 results). If you pick card number three you will be paid for round 3 (point to round 3 results). Remember, you will get both what you kept in your private pocket, and the payoff from the group pot.

Interviewer 1 asks the front right participant: Please pick a card from this shuffled deck.

Interviewer 2 records accordingly.

Thank you for your time. You may now leave.

Appendix C

Chapter 3: Dictator and Third-Party Punishment Games' Scripts

C.1 Dictator Game (Executives): Interviewer Manual

Materials (Per Group):
– 1 large cardboard privacy screen
– 30 × 100USH coins (20 for enumerator, 10 for demonstration set)
– 1 Generic Demonstration set
– 1 Demonstration set (large Stranger, DC member boards and 20 coins)
– 1 board symbolizing Stranger/Personal Pocket
– 1 board symbolizing DC member/Personal Pocket – 1 set of Payment cards (each deck 2 cards: Stranger, DC member) for choosing payment
– Activity 1 Activity 1 Executive Record sheet

Setup:

The group will sit together while instructions for the activity and examples are introduced. Enumerator will designate a private area/booth close by, which should be set up as follows:

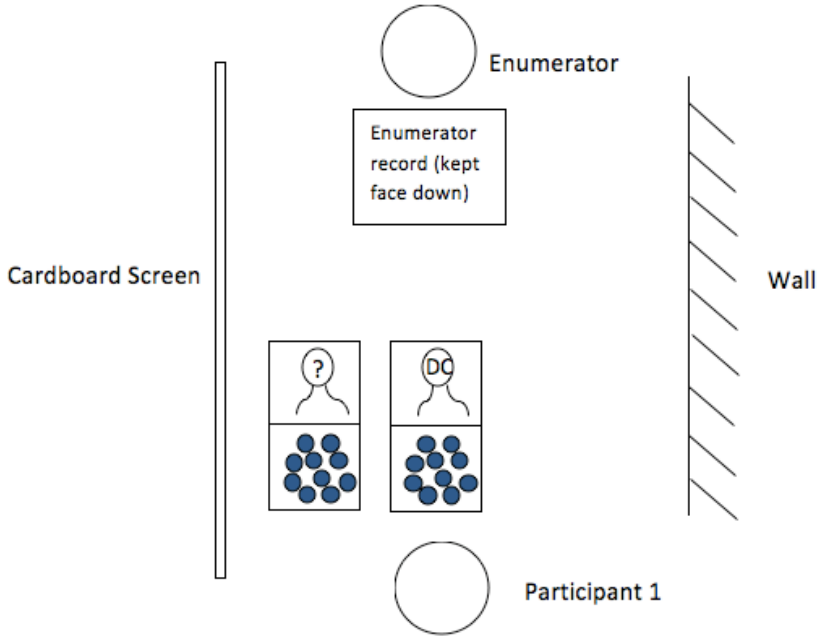

Figure C.1: Game Setting: Dictator Game (Executives)

Introduction to activities

Thank you all for participating to this activity. My name is *[enumerator 1]*.

In this activity we will be playing with real money. You should know that whatever money you win in this and the other activities will be yours to keep and take home. This money is for your individual use, goes to your private pocket, and you can spend it as you wish. You will receive the money from this activity, as well as your compensation, at the end of the day. The money comes from two American Universities, Princeton and Columbia. You should understand that this money does not come from our own (private) pocket. It is money given to the research team by the University to use for research.

Before we begin, I want to tell you the rules that we must follow. I am about to explain the first activity and it is important that you listen as carefully as possible, because only people who understand the activity will actually be able to participate. We will run through some examples here while we are all together. While in the group, you cannot ask questions or talk. This is very important. Please be sure that you obey this rule, because it is possible for one person to spoil the activity for everyone. If one person talks about the activity while sitting in the group or with other people later, this person will not be allowed to participate in todays activities and will be sent home. Do not worry if you do not completely understand the activity as we go through the examples here in the group. Each of you will have a chance to ask questions in private to be sure that you understand.

Finally, you will be playing in groups of two individuals, but you would not be told the name

of the person and you would not be able to see the person you are playing with. Similarly, we would not reveal your name to the person you are playing with or to anybody else. Only us, the research team, will know about your decisions, and well register them using an identification code, and not your name.

This first activity involves pairs of individuals. Each pair will be given a total of ten 100USH coins, totaling 1,000USH. The First participant, the Decider, has to decide how to allocate the money between her/himself and the Second participant, the Receiver. The Decider will take home whatever he has decided to allocate to himself, and the Receiver will take home whatever he has been allocated by the Decider.

All of you have been chosen to participate as Deciders. As a Decider, you have to decide how to split the 1,000 USH between you and a Receiver. Here are 4 examples of how you could split the money:

Demonstrate on the "Generic" Board:

- if you give 1,000 USH (10 coins) to the Receiver, the Receiver will receive 1,000 USH and you will walk away with nothing at the end of the day.

- if you give the Receiver 200USH (2 coins), s/he will receive 200 USH and you will walk away with 800USH (8 coins).

- if you give the Receiver 700 USH (7 coins), s/he will receive 700 USH and you will walk away with 300 USH, (3 coins).

- if you give the Receiver 0 coins, s/he will receive nothing and you will walk away with 1,000 USH (10 coins).

We will ask you to make a decision of how to allocate 1,000 USH two times for two different receivers. In the first case youll be asked to be the Decider and allocate money between you and a Stranger. You will use this board [pull out stranger board and set up 10 coins.]After you make your decision, a stranger will receive a box with the money. The Receiver will not be told your name or given any information about you. This stranger is a person who you do not know. We have chosen this person in the following way: from the list of all the people in the sub-county, we have randomly selected a few people. The person that will be receiving is one of them.

The second time, you will be asked, again, to allocate 1000USH between yourself and a second person. But this time the Receiver will be a regular member of the DC. Will use this board *[pull out a PO leader board and set up 10 coins.]* After you make your decision, a regular member of the DC will receive a box with the money. The Receiver will not be told your name or given any information about you. He will only be told only that the money comes from a DC executive. This

person might be here today or not. The only thing you know is that he is a regular member of the DC.

You will meet with the enumerators one-on-one. After you have decided how to allocate the money in each situation, we will determine which activity you will be paid for. Youll have to pick between two cards. One card has the symbol of the Stranger on it, while the other has the symbol of the DC member on it *[show cards]*. You will pick one of the cards, and you will take home the money you have made in the activity that you have randomly select. Be careful, since you do not know in advance for which activity you are going to be paid for, you have always to decide as if you were dividing real money.

Before making your decision, the enumerator will explain the activity again, and let you ask any question you might have. While waiting during the one-on-one meetings with enumerators, no one should talk about the activity or their decisions. If you have a question, please ask the enumerator, not another participant. Also, you are not allowed to speak about the activity with anybody during the day. You can talk about it only on your way home, after all the activities are over.

Now we are going to call each of you one-by-one to meet with an enumerator.

One-on-one meeting

Each booth should be set up before the participant comes for their one-on-one interview. The boards with the two symbols should be placed in front of the participant with ten 100USH coins on each board on the Personal Pocket side. Make sure to record the participants ID number in the "ID" Column.

You have been selected to be a Decider. You have been given a total of ten 100USH coins, totaling 1,000USH, and have to decide how to allocate the money between you and a Receiver. You will make both decisions when I have finished explaining the activity. Here you are dividing the money between yourself and a Stranger *[Point to the corresponding board]*. The Stranger is someone from your sub-county who has been selected at random from a complete list of all the members of the sub-county. This person will NOT be told your name or given any information about you.

Here you are dividing the money between yourself and a regular member of the DC *[Point to the corresponding board]*. This person will NOT be told your name or given any information about you. He will be told only that the money comes from one of the DC executives. This person might be here today or not. The only thing you know is that he is a regular member of your DC.

In both cases, you will take home whatever you have decided to allocate to yourself, while the Receiver will take home whatever you have decided to give him/her at the end of the day. You do not have to give any of these people any money if you do not want to. Remember that you are dividing real money, and you will be paid for one of these decisions.

Lets start with an example. If you decide to give a person 300USH (3 coins), then you will have 700USH (7 coins) for yourself.
Do you have any question for me?

Let me now ask you a question: If you give 600USH (6 coins) to a person, how many do you keep to yourself?
If the participant responds correctly, go ahead with the Game.

If not, give the answer, and ask another question:
If you give 200 USH (2 coins), how much are you keeping for yourself?

If the participant responds correctly, go ahead with the Game.
If not, provide the answer and explain, but put an $\star$ next to the participants decision.

You now have a few minutes to make their decisions for both situations. When you are finished, please let me know.

Record the number of coins the participant has given to the Stranger in the Offer - Stranger Column, and the number of coins the participant has given to the DC member in the "Offer DC member" Column. Show the participant the 2 payment option cards.

C.2 Third Party Punishment Game: Interviewer Manual

Materials (Per Group):
– 2 large private screens (1 per enumerator)
– 45 $\times$100 USH coins (15 for each enumerator, 15 for demonstration)
– 1 Demonstration TPP board
– 2 boards: 1 Decider and 1 Monitor board
– Record/pens

Each enumerator should have – A large privacy screen
– 10 $\times$ 100USH coins
– 1 board (Enum 1: Decider board; Enum 2: Monitor board)
– Record/pens

Setup:
Enumerators should decide who is Enumerator 1 and who is Enumerator 2. Enumerator 1 will be speaking more. Group 1 should have all odd-numbered PO members. First, the group will sit together to listen to instructions. Each enumerator should designate an area set up.

Group instructions:

Enumerator 1: This activity involves three individuals. All of them are members of a PO. Two players will be given a total of ten 100USH coins, totaling 1,000USH. The First player, the Decider, has to decide how to allocate the money between her/himself and the Second player, the Receiver. The Receiver takes home whatever he has been allocated by the Decider. The Decider might take home whatever he has decided to keep for himself. But the Decider has to wait for the decision of the third player, the Monitor. The Monitor is allocated 500 USH and, in case he is dissatisfied with the Decider decision, he may spend one 100USH coin to take away three coins (300USH) from the Decider. In such case the monitor will remain with 400USH. The monitor may also leave the decision as is. In such case the monitor gets to keep all the 500USH for himself.

The identity of each player will remain unknown to everyone except for us. Nobody will be told the name or any other information about the other players. The only think you know is that all the players are PO members. Half of you will be randomly assigned to be Deciders, the other half is randomly assigned to be Monitors. After all the participants have made their decision, we will pair up Deciders and Monitors at random.

Here are two examples: *[Use demonstration board and 15 coins]*. If the Decider gives the Receiver 300 USH and keeps 700 USH for him/herself, the Receiver will go home with 300 USH, while the Decider is left with 700 USH. However, how much the Decider will keep in the end depends on the decision of the Monitor. If the Monitor is dissatisfied with the allocation of money between the Decider and Receiver, he can decide to spend 100 USH to take away 300 USH from the Decider private pocket. In this case the Decider will remain with 400 USH in his pocket and the Monitor will have 400 USH. If instead the Monitor is satisfied with the allocation of money, the Decider will have 700 USH and the Monitor will keep 500 USH.

Lets take another example. If the Decider gives the Receiver zero USH and keeps 1,000 USH for him/herself, the Receiver will go home with nothing, while the Decider is left with 1,000 USH. If the Monitor is dissatisfied with the allocation of money between the Decider and Receiver, he can decide to spend 100 USH to take away 300 USH from the Decider private pocket. In this case the Decider will remain with 700 USH in his pocket and the Monitor will have 400 USH. If instead the Monitor is satisfied with the allocation of money, the Decider will keep all 1,000 USH and the Monitor will keep 500 USH.

Again, the transaction will be completely anonymous, and no one will know the names of who is giving, receiving, or monitoring. Now, you will each meet with the enumerators one-on-one to make your decision. Some of you have been randomly assigned to play as Deciders, others as Monitors. The enumerator will tell you your role. While waiting, no one should talk about the game and their decisions.

One-on-one interview

Enumerator 1 will take the 9 players with numbers from 1 to 17. They will be Deciders.
Enumerator 2 will take the 9 players with number 19 and higher. They will be Monitors.

Enumerator 1 (Instructions for 'Deciders'):
Record respondent's ID number
Make sure you have the appropriate board

You have been selected to be a Decider. A PO member will play as a Receiver, and another PO member will be the Monitor. You now have 1,000 USH (10x100USH coins) in front of you, and have to decide how to allocate the money between you and the Receiver. After you decide, the Monitor will be informed about your decision. If s/he is not satisfied with your decision, s/he will be able to take away 300 USH, three coins, from your private pocket.

Do you have any question? *[Answer questions if any, but keep it short]* Now you can make your decision. Please let me know when you have finished.

Record number of coins they have given to the other PO member. Rearrange coins back to original set up arrangement.

We will determine how much you have made in this activity once the Monitor has decided whether to take away 300 USH from your private pocket or not. We will give you the money at the end of the day. Thank you. You may now sit down.

Enumerator 2 (Instructions for 'Monitors'):
Record respondent's ID number
Make sure you have the appropriate board

You have been selected to be a Monitor, and have to oversee a transaction between two PO members. One PO member, the Decider, will be asked to divide 1,000 USH between him/herself and another PO member. You are given 500 USH. If you are unsatisfied with the Deciders choice, you are allowed to spend 100 USH to take away 300 USH form the Decider. If you spend a coin, you will get 400 USH. If not, you will get 500USH. I am now going to ask you what youd decide to do in different situations.

Do you have any question? *[Answer questions if any, but keep it short]*

Enumerator, place the coins to demonstrate each question and record after each answer: "1" if Player decides to spend 1 coin to monitor; "" if player decides not to spend any coins to monitor

1. The first PO member keeps 10 coins and gives the second PO member 0 coins, will you spend 1 coin so that the first PO member will receive 7 coins instead of 10 and the second still gets

0 coins? Or will you leave things as they are?

- *If they decide to reduce the Deciders private pocket, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

2. The first PO member keeps 9 coins and gives the second PO member 1 coins, will you spend 1 coin so that the first PO member will receive 6 coins instead of 9 and the second still gets 1 coins? Or will you leave things as they are?

   - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

3. The first PO member keeps 8 coins and gives the second PO member 2 coins, will you spend 1 coin so that the first PO member will receive 5 coins instead of 8 and the second still gets 2 coins? Or will you leave things as they are?

   - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

4. The first PO member keeps 7 coins and gives the second PO member 3 coins, will you spend 1 coin so that the first PO member will receive 4 coins instead of 7 and the second still gets 3 coins? Or will you leave things as they are?

   - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

5. The first PO member keeps 6 coins and gives the second PO member 4 coins, will you spend 1 coin so that the first PO member will receive 3 coins instead of 4 and the second still gets 4 coins? Or will you leave things as they are?

   - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

6. The first PO member keeps 5 coins and gives the second PO member 5 coins, will you spend 1 coin so that the first PO member will receive 2 coins instead of 5 and the second still gets 5 coins? Or will you leave things as they are?

   - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

7. The first PO member keeps 4 coins and gives the second PO member 6 coins, will you spend 1 coin so that the first PO member will receive 1 coin instead of 4 and the second still gets 6 coins? Or will you leave things as they are?

- *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

8. The first PO member keeps 3 coins and gives the second PO member 7 coins, will you spend 1 coin so that the first PO member will receive 0 coins instead of 3 and the second still gets 7 coins? Or will you leave things as they are?

    - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

9. The first PO member keeps 2 coins and gives the second PO member 8 coins, will you spend 1 coin so that the first PO member will receive 0 coins instead of 2 and the second still gets 8 coins? Or will you leave things as they are?

    - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

10. The first PO member keeps 1 coins and gives the second PO member 9 coins, will you spend 1 coin so that the first PO member will receive 0 coins instead of 1 and the second still gets 0 coins? Or will you leave things as they are?

    - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take three away from the first PO member. After they confirm this, replace coins back and set up the scenario in the next question.*

11. The first PO member keeps 0 coins and gives the second PO member 10 coins, will you spend 1 coin so that the first PO member will receive 0 coins instead of 0 and the second still gets 0 coins? Or will you leave things as they are?

    - *If they decide to punish, visually demonstrate the effect: take one coin from their personal pocket and take the rest away from the first PO member. After they confirm this, replace coins back to original set up arrangement.*

We will pair up your decisions with the allocation that two PO members have made. We will determine the amount that you will make after everyone in your group has played and will give you the money at the end of the day. Thank you.

Part III

Bibliography

Bibliography

Acemoglu, Daron and Simon Johnson. 2005. "Unbundling Institutions." *The Journal of Political Economy* 113(5):949–995.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91(5):1369–1401.

Agrawal, Arun and Sanjeev Goyal. 2001. "Group Size and Collective Action." *Comparative Political Studies* 34(1):63–93.

Ahlquist, John and Margaret Levi. 2011. "Leadership: What It Means, What It Does, and What We Want to Know About It." *Annual Review of Political Science* 14(1).

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

Alesina, Alberto and Guido Tabellini. 2008. "Bureaucrats or politicians? Part II: Multiple policy tasks." *Journal of Public Economics* 92(3-4):426–447.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* XL(4):791–821.

Angrist, Joshua D. and Alan B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *The Journal of Economic Perspectives* 15(4):69–85.

Angrist, Joshua David and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Ashraf, Nava, Xavier Gine and Dean Karlan. 2009. "Finding Missing Markets (and a Disturbing Epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya." *American Journal of Agricultural Economics* 91(4):973–990.

Atleson, James B. 1969. "Union Fines and Picket Lines: The NLRA and Union Disciplinary Power." *UCLA Law Review* 17:681–757.

Baker, George. 2000. "The Use of Performance Measures in Incentive Contracting." *The American Economic Review* 90(2):415–420.

Barr, Abigail. 2001. "Social dilemmas and shame-based sanctions: Experimental results from rural Zimbabwe." *Center for the Study of African Economies Working Paper* .

Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14:19–42.

Bates, Robert H. 1981. *Markets and states in tropical Africa: The political basis of agricultural policies*. Berkeley: University of California Press.

Bates, Robert H and Michael F Lofchie. 1980. *Agricultural development in Africa: issues of public policy*. New York: Praeger.

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2010. "Elite Capture of Local Institutions: Evidence from a Field Experiment in Afghanistan.".

Beetham, David. 1991. *The legitimation of power*. Atlantic Highlands, NJ: Humanities Press International.

Ben-Yoav, Orly, Edwin P. Hollander and Peter J. D. Carnevale. 1983. "Leader Legitimacy, Leader-Follower Interaction, and Followers' Ratings of the Leader." *Journal of Social Psychology* 121(1).

Benz, Matthias and Stephan Meier. 2008. "Do people behave in experiments as in the field?—evidence from donations." *Experimental Economics* 11(3):268–281–281.

Berdegué, Julio S. Antonio. 2001. Cooperating to Compete: Associative Peasant Business Firms in Chile PhD thesis Wageningen Universiteit en Researchcentrum Netherlands: .

Berkowitz, Leonard and Edward Donnerstein. 1982. "External validity is more than skin deep: Some answers to criticisms of laboratory experiments." *American Psychologist* 37(3):245–257.

Bernard, Tanguy, Marie-Héléne Collion, Alain de Janvry, Pierre Rondot and Elisabeth Sadoulet. 2008. "Do Village Organizations Make a Difference in African Rural Development? A Study for Senegal and Burkina Faso." *World Development* 36(11):2188–2204.

Bernheim, Douglas B. and Navin Kartik. 2010. "Of Candidates and Character." *Unpublished* .

Besley, Timothy. 2004. "Paying Politicians." *Journal of the European Economic Association* 2(2-3):193–215.

Besley, Timothy. 2005. "Political Selection." *Journal of Economic Perspectives* 19(3):43–60.

Besley, Timothy. 2006. *Principled Agents?* Oxford University Press.

Besley, Timothy and Anne Case. 2003. "Political Institutions and Policy Choices: Evidence from the United States." *Journal of Economic Literature* 41(1):7–73.

Besley, Timothy and Stephen Coate. 1997. "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics* 112(1):85–114.

Besley, Timothy and Stephen Coate. 2003. "Elected Versus Appointed Regulators: Theory and Evidence." *Journal of the European Economic Association* 1(5):1176–1206.

Bianco, William T. and Robert H. Bates. 1990. "Cooperation by Design: Leadership, Structure, and Collective Dilemmas." *The American Political Science Review* 84(1):133–147.

Bingen, Jim, Alex Serrano and Julie Howard. 2003. "Linking farmers to markets: different approaches to human capital development." *Food Policy* 28(4):405–419.

Birchall, Johnston. 2003. *Rediscovering the cooperative advantage Poverty reduction through self-help*. Cooperative Branch International Labour Office, Geneva.

Bjorkman, Martina and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Ugnada." *Quarterly Journal of Economics* 124(2):735–769.

Björkman, Martina and Jakob Svensson. 2010. "When is Community-based Monitoring Effective? Evidence from a randomized experiment in Primary Health in Uganda." *Journal of the European Economic Association* 8(2-3):571–581.

Bosc, Pierre-Marie, Didier Eychenne, Karim Hussein, Bruno Losch, Marie-Rose Mercoiret, Pierre Rondot and Mackintosh-Walker Sadie. 2002. "The role of rural producers organizations in the World Bank rural development strategy." Rural development strategy background paper : World Bank, 8.

Boyd, Robert, Herbert Gintis and Samuel Bowles. 2010. "Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare." *Science* 328:617–620.

Boyd, Robert, Herbert Gintis, Samuel Bowles and Peter J. Richerson. 2003. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences of the United States of America* 100(6):3531–3535.

Burnham, Terence C. and Dominic Johnson. 2005. "The biological and evolutionary logic of human cooperation." *Analyse & Kritik* 27(2):113–35.

Camerer, Colin. 2003. *Behavioral game theory: experiments in strategic interaction*. New York, NY: Russell Sage Foundation.

Camerer, Colin, George Loewenstein and Matthew Rabin. 2004. *Advances in behavioral economics*. New York: Russell Sage Foundation.

Cardenas, Juan Camilo and Jeffrey Carpenter. 2008. "Behavioural Development Economics: Lessons from Field Labs in the Developing World." *Journal of Development Studies* 44(3):311–338.

Carpenter, Jeffrey P. 2004. "Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research: Elinor Ostrom, James Walker (Eds.); Russell Sage Foundation, New York, NY, 2003." *Journal of Economic Behavior & Organization* 55(1):93–99.

Casari, Marco and Luigi Luini. 2009. "Cooperation under alternative punishment institutions: An experiment." *Journal of Economic Behavior & Organization* 71(2):273–282.

Caselli, Francesco and Massimo Morelli. 2004. "Bad Politicians." *Journal of Public Economics* 88(3-4):759–782.

Cassar, Alessandra, Luke Crowley and Bruce Wydick. 2007. "The effect of social capital on group loan repayment: evidence from field experiments." *The Economic Journal* 117(517):F85–F106.

Cavazos, Roberto J. 2003. "The Political Economy of Electricity Deregulation: Appointed vs. Elected Utility Commissioners." *Review of Policy Research* 20(2):255–262.

Champagne, Anthony. 2003. "The Politics of Judicial Selection." *Policy Studies Journal* 31(3):413–419.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72(5):1409–1443.

Chiburis, Richard, Jishnu Das and Michael Lokshin. 2010. "A Practical Comparison of the Bivariate Probit and Linear IV Estimators." *Working Paper* .

Conley, Timothy G, Christian B Hansen and Peter E Rossi. 2010. "Plausibly Exogenous." *Review of Economics and Statistics* .

Dal Bo, Pedro, Andrew Foster and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100(5):2205–2229.

De Cremer, David and Eric van Dijk. 2008. "Leader-Follower Effects in Resource Dilemmas: The Roles of Leadership Selection and Social Responsibility." *Group Processes & Intergroup Relations* 11(3):355–369.

de Rooij, Eline A., Donald P. Green and Alan S. Gerber. 2009. "Field Experiments on Political Behavior and Collective Action." *Annual Review of Political Science* 12(1):389–395.

Dewan, Torun and David P. Myatt. 2008. "The Qualities of Leadership: Direction, Communication, and Obfuscation." *American Political Science Review* 102(03):351–368.

Dickson, Eric S., Sanford C. Gordon and Gregory A. Huber. 2009. "Enforcement and Compliance in an Uncertain World: An Experimental Investigation." *The Journal of Politics* 71(04):1357–1378.

Dunning, Thad. 2008*a*. "Improving Causal Inference." *Political Research Quarterly* 61(2):282–293.

Dunning, Thad. 2008*b*. "Model Specification in Instrumental-Variables Regression." *Political Analysis* 16(3):290–302.

Dweck, Carol S., Therese E. Goetz and Nan L. Strauss. 1980. "Sex differences in learned helplessness: IV. An experimental and naturalistic study of failure generalization and its mediators." *Journal of Personality and Social Psychology* 38(3):441–452.

Easterly, William. 2001. *The elusive quest for growth: economists' adventures and misadventures in the tropics*. Cambridge, Mass.: MIT Press.

Eckel, Catherine C., Enrique Fatas and Rick Wilson. 2010. "Cooperation and Status in Organizations." *Journal of Public Economic Theory* 12(4):737–762.

Eguíluz, Víctor M., Martín G. Zimmermann, Camilo J. Cela-Conde and Maxi San Miguel. 2005. "Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks." *American Journal of Sociology* 110(4):977–1008.

Elgie, Donna M., Edwin P. Hollander and Robert W. Rice. 1988. "Appointed and Elected Leader Responses to Favorableness of Feedback and Level of Task Activity from Followers." *Journal of Applied Social Psychology* 18(16):1361–1370.

Fafchamps, Marcel and Bart Minten. 2001. "Social Capital and Agricultural Trade." *American Journal of Agricultural Economics* 83(3):680–685.

Fearon, James D., Macartan Humphreys and Jeremy M. Weinstein. 2009. "Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-conflict Liberia." *American Economic Review* 99(2):287–291.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic punishment in humans." *Nature* 415(6868):137–140.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1/3):5–25.

Ferraz, Claudio and Frederico Finan. 2010. "Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance." *Unpublished* .

Fields, Josepha., Linda Klein and James Sfiridis. 1997. "A market based evaluation of the election versus appointment of regulatory commissioners." *Public Choice* 92(3):337–351.

Filmer, Deon and Lant H. Pritchett. 2001. "Estimating Wealth Effects Without Expenditure Data – Or Tears: An Application to Educational Enrollments in States of India." *Demography* 38:115–132.

Fischbacher, Urs, Simon Gächter and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71:397–404.

Freeman, Linton C. 1979. "Centrality in social networks conceptual clarification." *Social Networks* 1(3):215–239.

Frey, Bruno S. and Matthias Benz. 2005. "Can Private Learn from Public Governance?" *The Economic Journal* 115(507):F377–F396.

Frye, Timothy. 2000. *Brokers and bureaucrats: building market institutions in Russia*. Ann Arbor: University of Michigan Press.

Funk, Patricia and Christina Gathmann. 2009. "Does Direct Democracy Reduce the Size of Government? New Evidence from Historical Data, 1890-2000." *Unpublished* .

Gagliarducci, Stefano and Tommaso Nannicini. 2010. "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection." *CEIS Tor Vergata, Research Paper Series* .

Gagliarducci, Stefano, Tommaso Nannicini and Paolo Naticchioni. 2010. "Moonlighting politicians." *Journal of Public Economics* 94(9-10):688–699.

Gehlbach, Scott, Konstantin Sonin and Ekaterina Zhuravskaya. 2010. "Businessman Candidates." *American Journal of Political Science* 54(3):718–736.

Gibson, Clark C. and Stuart A. Marks. 1995. "Transforming rural hunters into conservationists: An assessment of community-based wildlife management programs in Africa." *World Development* 23(6):941–957.

Gibson, James L. 1989. "Understandings of Justice: Institutional Legitimacy, Procedural Justice, and Political Tolerance." *Law & Society Review* 23(3):469–496.

Gibson, James L., Gregory A. Caldeira and Lester Kenyatta Spence. 2005. "Why Do People Accept Public Policies They Oppose? Testing Legitimacy Theory with a Survey-Based Experiment." *Political Research Quarterly* 58(2):187–201.

Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr. 2005. *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. MIT.

Gittell, Ross J. and Avis Vidal. 1998. *Community organizing: building social capital as a development strategy*. Thousand Oaks, CA: Sage Publications.

Goette, Lorenz, David Huffman and Stephan Meier. 2006. "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups." *The American Economic Review* 96(2):212–216.

Gordon, Sanford C. and Gregory A. Huber. 2002. "Citizen Oversight and the Electoral Incentives of Criminal Prosecutors." *American Journal of Political Science* 46(2):334–351.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *The American Economic Review* 83(3):525–548.

Greif, Avner. 2006. *Institutions and the path to the modern economy: lessons from medieval trade*. Cambridge: Cambridge University Press.

Greif, Avner, Paul Milgrom and Barry R. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *The Journal of Political Economy* 102(4):745–776.

Grossman, Guy. 2011*a*. A Theory of Candidacy Selection in Small Groups - With Evidence from Ugandan Farmer Associations PhD thesis Columbia University.

Grossman, Guy. 2011*b*. Centralized Sanctioning and Legitimate Authority in the Provision of Public Goods: Evidence from a "Lab in Field" Experiment In Uganda. PhD thesis Columbia University.

Grossman, Guy. 2011*c*. Do Leader Selection Rules Impact Accountability? Exploiting Exogenous Variation in the Formation of Ugandan Farmer Associations PhD thesis Columbia University.

Guth, Werner, M. Vittoria Levati, Matthias Sutter and Eline van der Heijden. 2007. "Leading by example with and without exclusion power in voluntary contribution experiments." *Journal of Public Economics* 91(5-6):1023–1042.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *The American Political Science Review* 101(4).

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2009. *Coethnicity: diversity and the dilemmas of collective action*. New York: Russell Sage Foundation.

Hanssen, F. Andrew. 1999. "The Effect of Judicial Institutions on Uncertainty and the Rate of Litigation: The Election versus Appointment of State Judges." *The Journal of Legal Studies* 28(1):205–232.

Hechter, Michael. 2009. "Legitimacy in the Modern World." *American Behavioral Scientist* 53(3):279–288.

Heckathorn, Douglas D. 1996. "The Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61(2):250–277.

Hellin, Jon, Mark Lundy and Madelon Meijer. 2009. "Farmer organization, collective action and market access in Meso-America." *Food Policy* 34(1):16–22.

Henrich, Joseph and Francisco J. Gil-White. 2001. "The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission." *Evolution and Human Behavior* 22(3):165–196.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan C. Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer and John Ziker. 2010. "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment." *Science* 327(5972):1480–1484.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press.

Herrmann, Benedikt, Christian Thóni and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science* 319(5868):1362–1367.

Hess, Frederick M. 2008. "Looking for Leadership: Assessing the Case for Mayoral Control of Urban School Systems." *American Journal of Education* 114(3):219–245.

Hibbing, John R. and John R. Alford. 2004. "Accepting Authoritative Decisions: Humans as Wary Cooperators." *American Journal of Political Science* 48(1):62–76.

Hoffmann, Elizabeth A. 2005. "Dispute Resolution in a Worker Cooperative: Formal Procedures and Procedural Justice." *Law & Society Review* 39(1):51–82.

Hollander, Edwin P. and James W. Julian. 1970. *Studies in Leader Legitimacy, Influence, and Innovation*. Vol. Volume 5 New York: Academic Press.

Huber, Gregory A. and Sanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind when It Runs for Office?" *American Journal of Political Science* 48(2):247–263.

Humphreys, Macartan, William A. Masters and Martin E. Sandbu. 2006. "The Role of Leaders in Democratic Deliberations: Results from a Field Experiment in São Toméand Príncipe." *World Politics* 58(4):583–622.

Hussi, P., J. Murphy, O. Linberg and L. Brenneman. 1993. *The Development of Cooperatives and Other Rural Organisations*. Washington DC: World Bank.: World Bank Technical Paper No.199.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2):467–475.

Isham, Jonathan. 2002. "The Effect of Social Capital on Fertiliser Adoption: Evidence from Rural Tanzania." *J Afr Econ* 11(1):39–60.

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):835–864.

Jones, Dan. 2010. "A WEIRD View of Human Nature Skews Psychologists' Studies." *Science* 328:1627.

Karlan, Dean S. 2007. "Social Connections and Group Banking." *Economic Journal* 117(517):F52–84.

Kasara, Kimuli. 2007. "Tax Me If You Can: Ethnic Geography, Democracy, and the Taxation of Agriculture in Africa." *American Political Science Review* 101(01):159–172.

Keele, Luke J., Corrine McConnaughy and Ismail White. 2008. "Statistical Inference For Experiments." *Working Paper* .

King, Andrew J., Dominic D. P. Johnson and Mark Van Vugt. 2009. "The Origins and Evolution of Leadership." *Current Biology* 19(19):R911–R916.

Kotakorpi, Kaisa and Panu Poutvaara. 2010. "Pay for Politicians and Candidate Selection: An Empirical Analysis." *Journal of Public Economics* .

Krishna, Anirudh. 2002. *Active social capital: tracing the roots of development and democracy*. New York: Columbia University Press.

Lake, David A. 2009. "Hobbesian Hierarchy: The Political Economy of Political Organization." *Annual Review of Political Science* 12(1):263–283.

Lattin, James M., J. Douglas Carroll and Paul E. Green. 2003. *Analyzing multivariate data*. Pacific Grove, CA: Thomson Brooks/Cole.

Laury, Susan K. and Laura O. Taylor. 2008. "Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good?" *Journal of Economic Behavior & Organization* 65(1):9–29.

Lazear, Edward P. 1986. "Salaries and Piece Rates." *The Journal of Business* 59(3):405–431.

Lele, Uma. 1981. "Co-operatives and the poor: A comparative perspective." *World Development* 9(1):55–72.

Levati, M. Vittoria, Matthias Sutter and Eline van der Heijden. 2007. "Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information." *Journal of Conflict Resolution* 51(5):793–818.

Levi, M. and A. Sacks. 2009. "Legitimating beliefs: Sources and indicators." *Regulation & Governance* 3(4):311–333.

Levi, Margaret. 1989. *Of Rule and Revenue*. Berkeley: University of California Press.

Levi, Margaret. 1997. *Consent, dissent, and patriotism*. Cambridge: Cambridge University Press.

Levi, Margaret, Audrey Sacks and Tom Tyler. 2009. "Conceptualizing Legitimacy, Measuring Legitimating Beliefs." *American Behavioral Scientist* 53(3):354–375.

Levitsky, Steven and M. Victoria Murillo. 2009. "Variation in Institutional Strength: Causes and Implications." *Annual Review of Political Science* 12:115–133.

Levitt, Steven D. and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *The Journal of Economic Perspectives* 21(2):153–174.

Lieberman, Evan S. 2007. "Ethnic Politics, Risk, and Policy-Making." *Comparative Political Studies* 40(12):1407–1432.

Linz, Juan J and Alfred C Stepan. 1996. *Problems of democratic transition and consolidation: southern Europe, South America, and post-communist Europe*. Baltimore: Johns Hopkins University Press.

List, John A. 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *The Journal of Political Economy* 114(1):1–37.

Locke, John. 1990. *Second Treatise on Civil Government [1690]* . Indianapolis, IN: Hackett.

Lubell, Mark and John T. Scholz. 2001. "Cooperation, Reciprocity, and the Collective-Action Heuristic." *American Journal of Political Science* 45(1):160–178.

Maddala, G. S. 1983. *Limited-dependent and qualitative variables in econometrics*. Vol. 3 Cambridge: Cambridge University Press.

Martinez-Bravo, Monica, Gerard Padro-i Miquel, Nancy Qian and Yang Yao. 2010. Accountability in an Authoritarian Regime: The Impact of Local Electoral Reforms in Rural China. Presented at the Annual Meeting of the Northeast Universities Development Consortium (NEUDC), Boston.

Maskin, Eric and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *The American Economic Review* 94(4):1034–1054.

Mattozzi, Andrea and Antonio Merlo. 2008. "Political careers or career politicians?" *Journal of Public Economics* 92(3-4):597–608.

McEwen, Craig A. and Richard J. Maiman. 1986. "In Search of Legitimacy: Toward An Empirical Analysis." *Law & Policy* 8(3):257–273.

McKinzie, David J. 2005. "Measuring Inequality with Assett Indicators." *Journal of Population Economics* 18(2):229–260.

Mehra, Ajay, Andrea L. Dixon, Daniel J. Bras and Robertson Bruce. 2006. "The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation." *Organization Science* 17(1):64–79.

Messner, Matthias and Mattias K. Polborn. 2004. "Paying politicians." *Journal of Pubic Economics* 88(12):2423–2445.

Miguel, Edward and Mary Kay Gugerty. 2005. "Ethnic diversity, social sanctions, and public goods in Kenya." *Journal of Public Economics* 89(11-12):2325–2368.

Murphy, Kristina. 2005. "Regulating More Effectively: The Relationship between Procedural Justice, Legitimacy, and Tax Non-compliance." *Journal of Law and Society* 32(4):562–89.

Narayan-Parker, Deepa. 2002*a*. *Empowerment and poverty reduction: a sourcebook*. Washington, DC: World Bank.

Narayan-Parker, Deepa. 2002*b*. *Empowerment and poverty reduction: a sourcebook*. Washington, DC: World Bank.

Neal, Derek. 1997. "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics* 15(1):98–123.

North, Douglass C. 1990. *Institutions, institutional change, and economic performance*. Cambridge: Cambridge University Press.

Nowak, Martin A. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314:1560–63.

O'Gorman, Rick, Joseph Henrich and Mark Van Vugt. 2009. "Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation." *Proceedings of the Royal Society B: Biological Sciences* 276(1655):323–329.

Olken, Benjamin A. 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104(02):243–267.

Olson, Mancur. 1965. *The logic of collective action: public goods and the theory of groups*. Vol. 124 Cambridge, Mass.: Harvard University Press.

Osborne, Martin J. and Al Slivinski. 1996. "A Model of Political Competition with Citizen-Candidates." *The Quarterly Journal of Economics* 111(1):65–96.

Ostrom, Elinor. 1990. *Governing the commons: the evolution of institutions for collective action*. Cambridge: Cambridge University Press.

Ostrom, Elinor. 2000. "Collective Action and the Evolution of Social Norms." *The Journal of Economic Perspectives* 14(3):137–158.

Ostrom, Elinor, James Walker and Roy Gardner. 1992. "Covenants With and Without a Sword: Self-Governanceis Possible." *American Political Science Review* 86(2):404–17.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11(6):619–632.

Persson, Torsten, Gerard Roland and Guido Tabellini. 1997. "Separation of Powers and Political Accountability." *The Quarterly Journal of Economics* 112(4):1163–1202.

Persson, Torsten, Gérard Roland and Guido Tabellini. 2000. "Comparative Politics and Public Finance." *The Journal of Political Economy* 108(6):1121–1161.

Persson, Torsten and Guido E. Tabellini. 2003. *The Economic Effects of Constitutions: What Do the Data Say?* Cambridge, Mass.: MIT Press.

Ponte, Stefano. 2002. *Farmers and Markets in Tanzania: how policy reforms affect rural livelihoods in Africa*. Oxford: Heinemann.

Przeworski, Adam, Susan C. Stokes and Bernard Manin. 1999. *Democracy, accountability, and representation*. Cambridge, U.K.: Cambridge University Press.

Rand, David, Anna Dreber, Tore Ellingsen, Drew Fudenberg and Martin Nowak. 2009. "Positive Interactions Promote Public Cooperation." *Science* 325(5945):1272–1275.

Rawls, John. 1971. *A theory of justice*. Cambridge, MA: Belknap Press of Harvard University Press.

Read, Peter B. 1974. "Source of Authority and the Legitimation of Leadership in Small Groups." *Sociometry* 37(2):189–204.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *The Quarterly Journal of Economics* 100(4):1169–1189.

Roodman, David. 2009. "Estimating Fully Observed Recursive Mixed-Process Models with cmp." Center for Global Development, Working Paper Number 168.

Rothstein, Bo. 2009. "Creating Political Legitimacy." *American Behavioral Scientist* 53(3):311–330.

Scholz, John T. and Wayne B. Gray. 1997. "Can Government Facilitate Cooperation? An Informational Model of OSHA Enforcement." *American Journal of Political Science* 41(3):693–717.

Sexton, Richard J. 1990. "Imperfect Competition in Agricultural Markets and the Role of Cooperatives: A Spatial Analysis." *American Journal of Agricultural Economics* 72(3):709–720.

Shadish, William R., Thomas D. Cook and Donald Thomas Campbell. 2001. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shaub, Michael K, Frank Collins, Oscar Holzmann and Suzanne H Lowensohn. 2005. "Self-Interest vs. Concern for Others." *Strategic Finance* 86(9).

Shepherd, Andrew. 2007. *Approaches to linking producers to markets. A review of experiences to date*. Rome: Food and Agriculture Organization (FAO).

Shleifer, Andrei and Robert W. Vishny. 1997. "A Survey of Corporate Governance." *The Journal of Finance* 52(2):737–783.

Sigmund, Karl. 2007. "Punish or perish? Retaliation and collaboration among humans." *Trends in Ecology & Evolution* 22(11):593–600.

Simmons, Richard and Johnston Birchall. 2008. "The role of co-operatives in poverty reduction: Network perspectives." *Journal of Socio-Economics* 37(6):2131–2140.

Smart, Susan R. 1994. "The Consequences of Appointment Methods and Party Control for Telecommunications Pricing." *Journal of Economics & Management Strategy* 3(2):301–323.

Staatz, John M. 1987. "Recent Developments in the Theory of Agricultural Cooperation." *Journal of Agricultural Cooperation* 2:74–95.

Stringfellow, Rachel, Jonathan Coulter, Trevor Lucey, Colin McKone and Ambereene Hussain. 1997. "Improving the Access of Smallholders to Agricultural Services in Sub-Saharan Africa: Farmer Cooperation and the Role of the Donor Community." *Natural Resource Perspectives* 20.

Sykuta, Michael E. and Michael L. Cook. 2001. "A New Institutional Economics Approach to Contracts and Cooperatives." *American Journal of Agricultural Economics* 83(5):1273–1279.

Taylor, Michael. 1982. *Community, anarchy, and liberty*. Cambridge, MA: Cambridge University Press.

Tyler, Tom R. 1997. "The Psychology of Legitimacy: A Relational Perspective on Voluntary Deference to Authorities." *Personality and Social Psychology Review* 1(4):323–345.

Tyler, Tom R. 2005. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57(1):375–400.

UBOS. 2002. *Uganda Population and Housing Census*. Kampala: Uganda Bureau of Statistics.

Urquiola, Miguel and Eric Verhoogen. 2009. "Class-Size Caps, Sorting, and the Regression-Discontinuity Design." *American Economic Review* 99(1):179 – 215.

Wasserman, Stanley and Katherine Faust. 1994. *Social network analysis: methods and applications*. Vol. 8 Cambridge: Cambridge University Press.

Weber, Max. 1922. *Economy and Society*. University of California Press [1978].

White, Douglas R. and Stephen P. Borgatti. 1994. "Betweenness centrality measures for directed graphs." *Social Networks* 16(4):335–346.

Wilson, Rick K. and Carl M. Rhodes. 1997. "Leadership and Credibility in N-Person Coordination Games." *Journal of Conflict Resolution* 41(6):767–791.

Woolcock, Michael and Deepa Narayan. 2000. "Social Capital: Implications for Development Theory, Research, and Policy." *The World Bank Research Observer* 15(2):225–249.

Zelditch, Morris. 2001. "Processes of Legitimation: Recent Developments and New Directions." *Social Psychology Quarterly* 64(1):4–17.