NBER WORKING PAPER SERIES

RESHAPING INSTITUTIONS:
EVIDENCE ON AID IMPACTS USING A PRE-ANALYSIS PLAN

Katherine Casey
Rachel Glennerster
Edward Miguel

Working Paper 17012
http://www.nber.org/papers/w17012

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2011

We wish to thank the GoBifo Project staff—Minkahil Bangura, Kury Cobham, John Lebbie, Dan Owen and Sullay Sesay—and the Institutional Reform and Capacity Building Project (IRCBP) staff—Liz Foster, Emmanuel Gaima, Alhassan Kanu, S.A.T. Rogers and Yongmei Zhou—without whose cooperation this research would not have been possible. We are grateful for excellent research assistance from John Bellows, Mame Fatou Diagne, Mark Fiorello, Maryam Janani, Philip Kargbo, Angela Kilby, Gianmarco León, Tom Polley, Tristan Reed, Arman Rezaee, Alex Rothenberg and David Zimmer. Jim Fearon, Brian Knight, Ed Leamer, Kaivan Munshi, Biju Rao, Gerard Roland, Ann Swidler, Eric Werker and seminar audiences at Brown, CGD, the Econometric Society (North American Summer Meetings), IGC, MIT, NEUDC, Stanford, U.C. Berkeley, UCLA, UCSD and WGAPE have provided helpful comments. We gratefully acknowledge financial support from the GoBifo Project, the IRCBP, the World Bank Development Impact Evaluation (DIME) initiative, the Horace W. Goldsmith Foundation, the International Growth Centre, the International Initiative for Impact Evaluation, and the National Bureau of Economic Research African Successes Project (funded by the Gates Foundation). All errors remain our own. *Corresponding author: Edward Miguel (emiguel@econ.berkeley.edu) The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan
Katherine Casey, Rachel Glennerster, and Edward Miguel
NBER Working Paper No. 17012
May 2011, Revised October 2011
JEL No. F35,H41,O4

ABSTRACT

Although institutions are believed to be key determinants of economic performance, there is limited evidence on how they can be successfully reformed. Evaluating the effects of specific reforms is complicated by the lack of exogenous variation in the presence of institutions; the difficulty of empirically measuring institutional performance; and the temptation to "cherry pick" a few novel treatment effect estimates from amongst the large number of indicators required to capture the complex and multi-faceted subject. We evaluate one attempt to make local institutions more egalitarian by imposing minority participation requirements in Sierra Leone and test for longer term learning-by-doing effects. In so doing, we address these three pervasive challenges by: exploiting the random assignment of a participatory local governance intervention, developing innovative real-world outcomes measures, and using a pre-analysis plan to bind our hands against data mining. The specific program under study is a "community driven development" (CDD) project, which has become a popular strategy amongst donors to improve local institutions in developing countries. We find positive short-run effects on local public goods provision and economic outcomes, but no sustained impacts on collective action, decision-making processes, or the involvement of marginalized groups (like women) in local affairs, indicating that the intervention was ineffective at durably reshaping local institutions. We further show that in the absence of a pre-analysis plan, we could have instead generated two highly divergent, equally erroneous interpretations of the impacts—one positive, one negative—of external aid on institutions.

Katherine Casey
Stanford GSB (effective 1 Jan 2012)
Memorial Way
Stanford, CA 94305
katherine.e.casey@gmail.com

Rachel Glennerster
Abdul Latif Jameel Poverty Action Lab
MIT Department of Economics
E60-275
Cambridge MA 02139
rglenner@mit.edu

Edward Miguel
Department of Economics
University of California, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720
and NBER
emiguel@econ.berkeley.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w17012

1. Introduction

Many scholars have argued that the accountability and inclusiveness of government institutions are key determinants of economic performance. In particular, institutions that are egalitarian and protect individual rights have been tied to better economic outcomes in India (Banerjee and Iyer 2004), Brazil and the United States (Engerman and Sokoloff 1997), and former European colonies (Acemoglu, Johnson and Robinson 2001). However, there is no consensus on the specific policy reforms that will successfully engender better functioning institutions, or on whether it is possible (or desirable) for external actors like foreign aid donors to reshape local power dynamics in less developed countries. This debate has played out vigorously in discussions on aid policy reform: while some scholars argue that large infusions of foreign aid can themselves help build stronger institutions (Sachs 2005), others assert that historically rooted local institutions and social norms are difficult to understand, let alone transform (Easterly 2006), and that attempts by outsiders to create "better" institutions will be futile.

Progress toward resolving this debate is complicated by the difficulty of measuring institutional performance empirically, where change in institutional structure is rarely exogenous and the subject matter is difficult to pin down. The amorphous and contextually determined nature of institutions means that there are few standard indicators to draw from, and reliance on subjective measures risks bias from "halo effects" (see Olken 2009 on corruption measurement). Moreover, their multi-dimensionality leads to the collection of a large number of outcome variables, tempting the researcher to "cherry pick" the handful of results that will be statistically significant by random chance. In this paper we rigorously evaluate one attempt to transform local institutions in Sierra Leone. To address these measurement challenges, we exploit a randomly assigned governance intervention, develop objective real-world measures of institutions, and use a pre-analysis plan to bind our hands against data mining.

Among foreign aid donors, non-governmental organizations (NGOs) and governments in less developed countries today, imposing participation requirements to enhance the position of marginalized groups has become a popular strategy to promote the inclusiveness and accountability of local institutions. Giving greater representation to minority groups aims to foster learning-by-doing and demonstration effects that empower its members over the longer term. While skeptics might argue that the position of underrepresented groups will only change in the long-run if there are changes in underlying political power, demonstration effects have been found to be successful in enhancing the participation of women in India (Beamen et al. 2009). To test these questions, we evaluate a "community driven development" (CDD) project that combines block grants for local public goods with intensive training and requirements on minority inclusion designed to catalyze collective action and empower marginalized groups in local decision-making. Random assignment of the program presents a convenient opportunity to study exogenous variation in institutional reform, as well as evaluate the efficacy of the vast sums of aid currently channeled through these programs; Mansuri and Rao (forthcoming) estimate that the World Bank alone has spent 50 billion U.S. dollars on CDD initiatives over the past ten years. This type of governance reform attempts to bolster local *coordination* – for example, by setting up village development committees (VDC) and plans – and to enhance *participation* and inclusion, by requiring women and "youths" (adults aged 18 to 35) to attend project meetings and hold leadership positions. As arguments in favor of local ownership relate to those behind fiscal decentralization, CDD is often used to provide "bottom up" support for broader decentralization reforms in practice.

Yet while advocates of participatory local governance promise a long and varied list of benefits – ranging from more cost-effective construction of local infrastructure, to a closer match between project choice and village needs, to the weakening of authoritarian village institutions[1] – critics hold concomitant concerns that participation requirements serve as a regressive tax, widening political participation will clog up rather than expedite decision-making (Olson 1982), external resources may attract new leaders and crowd out the more disadvantaged (Gugerty and Kremer 2008), and that these additional resources will be captured by elites if the program is unable to change the nature of *de facto* political power (Bardhan 2002). While researchers have begun to explore these claims, few studies provide rigorous evidence on the real-world impacts of community driven development projects (Mansuri and Rao 2004).

This paper studies a large-scale randomized local governance project that was part of the Government of Sierra Leone's broader reform agenda to strengthen participatory democracy and decentralize public services. As background, scholars attribute the incompetence and elite

[1] For instance, Dongier et al. (2003) write that: "Experience demonstrates that by directly relying on poor people to drive development activities, CDD has the potential to make poverty reduction efforts more responsive to demands, more inclusive, more sustainable, and more cost-effective than traditional centrally led programs…achieving immediate and lasting results at the grassroots level."

domination of Sierra Leone's institutions – both the formal state and the traditional chieftaincy system – in the 1970's and 80's as key contributors to the brutal civil war that ensued from 1991 to 2002 (Richards 1996, Keen 2003). Emerging from war with widespread poverty and a dearth of public services, the country fell to the very bottom of the United Nations Development Program Human Development Index that measures standards of living, health and education (United Nations 2004). Housed within the government's Decentralization Secretariat and funded by the World Bank, the particular project we study, "GoBifo" (or "Move Forward" in Krio, Sierra Leone's *lingua franca*), provided both what we in this paper call "hardware" and "software" support to rural communities. The hardware included block grants of $4,667, or roughly $100 per household, for constructing local public goods and sponsoring trade skills training and small business start-up capital. Akin to community organizers in the U.S., GoBifo program facilitators also provided "software": technical assistance that promoted democratic decision-making, the participation of socially marginalized women and youth in local politics, and transparent budgeting practices. While the objective of making local government institutions more participatory aimed to address some of the perceived root causes of the civil war, GoBifo's design is similar to many other CDD projects in non-post-conflict societies.

This paper assesses the extent to which GoBifo achieved its goals of improving local governance in rural Sierra Leone communities, and in so doing makes two contributions of general interest, and two contributions specific to the development economics literature. The first broad contribution is the use of a pre-analysis plan. The research and project teams agreed to a set of hypotheses regarding the likely areas of program impacts in 2005 before the intervention began. As the project came to a close in 2009, we fleshed out this document with the exact outcome measures and econometric specifications that we would use to evaluate success, and archived this pre-analysis plan before analyzing the follow-up data (see supplementary Appendix A). Our decision to adhere rigorously to this plan eliminates the risk of data mining or other selective presentation of empirical results ("cherry-picking"), and generates correctly sized statistical tests, bolstering the scientific credibility of the findings. Moreover, such registration of experimental trials helps to round out the body of evidence available to researchers and practitioners, mitigating the publication bias that arises from underreporting of null or inconclusive results. Registration of medical trials, including a pre-analysis plan, is required by

U.S. law but, to our knowledge, this is among the first economics studies to adopt this approach.[2]

The second general contribution is the creation of new objective measures of local institutions and collective action. We combine rich household survey data with novel "structured community activities" (SCAs) that introduce three concrete, real-world scenarios that allow us to observe and measure how communities: (i) respond to a matching grant opportunity to purchase building materials at a subsidized price; (ii) make a communal decision between two comparable alternatives; and (iii) allocate a valuable asset (provided for free) among community members. We feel that these SCAs capture actual local collective action capacity, and uncover the decision-making processes that underlie it, more accurately than lab experiments, hypothetical vignettes or survey reports alone. The fact that these activities were carried out *after* the GoBifo program (and its financial resources) had ended allows us to isolate any persistent impacts on collective action and institutional performance generated by the program. We are unaware of other studies that have used these SCAs in practice, and believe they may be useful tools for other researchers interested in unobtrusively and objectively measuring local institutions.

A third contribution of particular interest to development economists and practitioners is a rigorous and comprehensive evaluation of a CDD project, which is important given the billions of dollars currently dedicated to this type of external assistance. We use a randomized experimental research design, which produces rigorous evidence on causal impacts in a relatively large study sample of 236 villages and 2,832 households. The extended timeframe of our study over four years (2005-2009) allows us to assess longer run impacts on institutional outcomes than is typically possible. While four years may be short in comparison to the lifetimes over which current institutions developed, it is not short in comparison to the time scales of most community development or other externally funded projects, or compared to learning-by-doing effects documented elsewhere. To guide our empirical work, we develop a theoretical framework for understanding the mechanisms through which financial assistance and participation requirements might impact public goods and collective action both during and after the program.

[2]The Food and Drug Administration Modernization Act of 1997 launched this initiative and led to the creation of the NIH-sponsored web registry clinicaltrials.gov in 2000. The Food and Drug Administration Amendments Act of 2007 expands coverage to a broader group of trails, requires results reporting and imposes financial penalties for non-compliance. Moreover, since 2005, registration of clinical trials is a prerequisite for publication in any member journal of the International Committee of Medical Journal Editors. There are even further calls in medicine (Lancet 2010) and epidemiology (Bracken 2011) to register protocols for *observational* studies. The U.S. Department of Education Institute of Education Sciences launched a trail registry, although the information requirements and incentives are weaker than those found in medicine. See also Simes (1986) and Horton and Smith (1999).

As most existing evaluations of these kinds of projects have been relatively atheoretical, we believe this model could be useful in interpreting the findings of other studies as well.

Fourth and finally, we provide additional evidence to the growing literature that explores the potential demonstration effects of giving minorities control over decisions and resources. Research in India suggests that giving women and members of scheduled castes positions of political power shifts the composition of public spending toward goods preferred by these groups (Chattopadhyay and Duflo 2004, Pande 2003) and reduces statistical discrimination against female candidates (Beamen et al. 2009). By contrast, we find that requiring women and young adults to take on leadership positions, attend project meetings, and sign off on project finances does not have any lingering effect on their participation in local decision-making or on social norms regarding their leadership ability. One explanation for this difference may be that while Indian political reservations give representatives of historically excluded groups real power over resources as part of a formal state body (the *panchayat*), CDD takes a more indirect approach to *de jure* reforms—nudging communities to become more democratic and inclusive without explicitly attempting to weaken elites—and may in reality not change the identity of *de facto* local power holders (Acemoglu and Robinson 2008). Perhaps because sidelining the chiefs was not a program goal, male elders and chiefdom officials retained just as much control over village development committees in GoBifo communities as they held over comparable organizing bodies in control areas, despite requirements that women attend meetings and play a role in project management.

Our analysis explores an exceptionally wide range of outcome measures, which we divide into two broad groups: "hardware" or local public infrastructure outcomes (which we call family A), and "software" or institutional and collective action outcomes (family B). Under the first family, we find that the GoBifo project was well implemented: it successfully established the village-level organizations and tools to manage development projects in nearly all cases, and provided communities with the financing to implement them. The distribution of project benefits within communities was equitable for the most part, and the leakage of project resources was minimal. We further find immediate impacts on local public goods infrastructure. There is also more market activity in treatment communities, including the presence of more traders and items for sale, suggesting short-run economic gains.

However, such implementation and hardware gains appear insufficient to generate changes in for the second, arguably more important, domain of institutional reform. Here we find no evidence that the program led to fundamental changes in the "software" of collective action (family B) – namely, the ability to raise funds for local public goods, decision-making processes, or even social attitudes and norms. As an example, despite the new experiences many women in treatment villages gained by participating in GoBifo activities, they were no more likely to voice an opinion during observed community meetings after the project ended or to play a leading decision-making role (using a variety of metrics). Similarly, the establishment of a democratically elected village development committee that carried out multiple projects did not lead treatment villages to be any more successful at raising funds in response to the matching grant opportunity. These patterns, and the lack of significant effects across many other outcomes, indicate that the program did not reshape village institutions, empower minorities, or improve collective action beyond the activities directly stipulated by the project itself. The time horizon of the research over four years suggests that these results cannot be dismissed simply as the result of a short term study.

The finding that a well-implemented project with beneficial public goods and economic impacts did not trigger broader spillover effects on institutions and norms resonates with the mixed results seen in the emerging empirical literature on CDD programs. In the Philippines, Labonne and Chase (2008) find that CDD increased participation in village assemblies and interaction between residents and village leaders but did not initiate broader social change, and in fact, may have crowded out other avenues for collective action. Voss (2008) uncovers mixed impacts of the Kecamatan Development Program (KDP) in Indonesian household welfare and access to services. Focusing on roads constructed under the same KDP project, Olken (2007) finds that enhanced top down project monitoring—through guaranteed government audits—was more effective in reducing corruption than increased grassroots participation in village-level accountability meetings between residents and project officials. A related set of papers exploring the impacts of community mobilization on public service providers similarly finds mixed results with strong positive effects seen for healthcare in Uganda (Bjorkman and Svensson 2009) but no effect on education in India (Banerjee et al. 2010). Most similar to our study context and methods, Fearon, Humphreys and Weinstein (2009) concurrently conducted a randomized evaluation of a community driven post-war reconstruction project in Liberia in 83 communities. Their basic result of positive impacts on collective action and social cohesion – as measured by greater contributions to an experimental public goods game in the mixed-gender treatment arm (although there were no impacts in the women-only treatment arm) and reduced self-reports of inter-group tensions – accompanied by little effect on material welfare, appears quite the opposite of our findings. To explain this divergence, they note that the program they study explicitly forbade the income generating initiatives that many GoBifo communities chose to implement. Moreover, their public goods game may have more directly mimicked the NGO's intervention while our SCAs aimed to capture shifts in more general institutional practices and behaviors. Crucially, though, neither study finds compelling evidence of program spillovers on real-world, non-project collective activities including contributing to existing public goods (such as road maintenance, schools and wells), and attending or speaking up in community meetings.

The rest of the paper is structured as follows. Section 2 discusses the Sierra Leone context, the GoBifo intervention, and a theoretical framework of local collective action under a CDD program. Section 3 covers the research design, pre-analysis plan and econometric specifications. Section 4 discusses the empirical results and accompanying robustness checks, and the final section concludes.

2. Background

2.1. Institutions in Sierra Leone

Before describing the aims and content of the GoBifo program, let us first consider why participatory local government advocates might argue that existing institutions in Sierra Leone warrant reform. At a high level, the country has a dual system of governance (common in many African countries, Mamdani 1996) in which the national state apparatus based in the capital runs in parallel to the "traditional" local chieftaincy system, neither of which has historically been particularly democratic or inclusive. Regarding the former, authoritarian leaders in the 1970's and 1980's enriched themselves through illicit diamond deals while providing woefully inadequate public services (Reno 1995). President Siaka Stevens dismantled democratic institutions entirely, initially by abolishing elected district governments in 1972, and ultimately declaring the country a one-party state in 1978. One-party rule continued until the 1992 coup that roughly coincided with the start of the civil war (which ran from 1991 to 2002).

As background on the traditional system, the 149 paramount chiefs come from hereditary "ruling houses"; they serve for life once appointed or elected; and exert considerable control over resource allocation, including land and labor, as well as the local court system that reigns outside the capital. Many local public goods (such as road maintenance and community schools) are provided by local fundraising organized through traditional systems at the village level. Dominated by male elders, this system has continued to the present day to largely exclude women (who are not even eligible to serve as chiefs in much of the country) and young men from decision-making. Political exclusion, growing frustration with government incompetence and corruption, and grievances against heavy-handed chiefs are seen as destabilizing factors that contributed to war (Richards 1996, Keen 2003).

Emerging from the civil war with dismal standards of living, health and education (United Nations 2004), the government and its donor partners oversaw major reforms to restore multi-party democracy and stimulate economic growth. One of the most high-profile reforms was the reconstitution of district-level government in 2004 after over thirty years of dormancy. Housed within the government's Decentralization Secretariat, the GoBifo project was launched as a pilot initiative to extend this broader decentralization reform to even more local levels, providing funds and technical assistance to wards and villages.[3] To formally link project activities to higher tiers of government, Village Development Committees (VDC) were required to submit their village development plans to the appropriate Ward Development Committee (WDC) for review, endorsement and onward transmission to the district council for approval (GoBifo Project 2007). The government's broader reform agenda and joint sponsorship of the project (with the World Bank) gave weight to GoBifo's objectives of making local institutions more participatory and democratic.

2.2 The GoBifo Project

As mentioned in the introduction, the GoBifo project had two main components: i) financial assistance in the form of block grants of around five thousand dollars per village to sponsor local public goods provision and small enterprise development; and ii) intensive organizing to establish new structures to facilitate collective action (i.e. Village Development Committees) and institute participation requirements to elevate historically marginalized groups to positions of authority. As examples of the latter, GoBifo required that one of the three co-signatories on the

[3] The WDCs are the lowest formal government administrative unit, covering around 10,000 citizens on average, and the elected district councilor representing the ward serves as the WDC chair. While the project we study also operated at the ward level, only the village-level intervention was randomly assigned and is thus our focus.

community bank account be female; encouraged women and youths to manage their own projects (e.g., small business training for women); made evidence of inclusion in project implementation a prerequisite for the release of block grant funding tranches; and, as part of their internal review process, even required field staff to record how many women and youth attended and spoke up in meetings.

The process of establishing new village institutions, training community members, and promoting social mobilization of marginalized groups was intense and accounted for a large part of GoBifo human and financial resources. Specifically, all project facilitators were required to reside in one of the six villages assigned to them and spend approximately one day per week in each of the remaining villages. After the start of project work in January 2006 and through the completion of all village-level projects in July 2009, each village thus received roughly six months of direct "facilitation" over a three and a half year period (see the detailed timeline in Appendix B). Furthermore, while just under half of the total GoBifo budget was dedicated to village- and ward-level block grants (US$896,000 or 47%), the balance covered "capacity development" in village- and ward-level planning (US$589,732 or 30%), project management and contingencies (US$255,320 or 14%), and monitoring and evaluation (US$177,300 or 9%). Thus for every dollar spent directly on actual construction or small business projects, roughly one dollar was spent on capacity-building, facilitation and oversight.

GoBifo village projects were carried out in several areas. The largest share of projects, at 43%, was in the construction of local public goods, with 14% in community centers or sports fields, 12% in education (i.e., primary school repairs), 10% in water and sanitation (e.g., latrines), 5% in health (including traditional midwife posts), and 2% in roads. Another 26% was in agriculture, including seed multiplication and communal farming; 14% in livestock or fishing (i.e., goat herding); and 17% in skills training and small business development initiatives in blacksmithing, carpentry, and soap making. Leakage of GoBifo funds also appears minimal: when we asked villagers to verify the detailed financial reports that were given to the research team by project management, community members were able to confirm receipt for 86.5% of the 273 transactions that were cross-checked.[4]

[4]The discrepancies were of two types: i) the amounts in community records was markedly less than in project accounts; or ii) community members reported receiving building materials in kind and could not estimate their value. For each of the disputed transactions, the GoBifo accounting team produced hard copy payment vouchers signed by both a village representative (either the VDC Chair or Finance Officer) and a project field staff member.

The GoBifo project is quite representative of CDD initiatives in other less developed countries. The project implementation stages—establishing a local committee, providing facilitation that aims to shift social norms, and allocating block grants—are quite standard, as is the pervasive emphasis on inclusive, transparent and participatory processes. Compared to projects studied in other countries (Olken 2007, Labonne and Chase 2008), the most notable programmatic difference is that the village-level component of GoBifo did not involve any inter-community competition for funding. Regarding the scale of funding, GoBifo disbursed grants worth a bit under \$5,000 to communities with 50 households, or 300 residents, on average (so roughly \$100 per household, or \$16 per capita over three and a half years).[5]

2.3 A Framework of Collective Action and External Aid

We next lay out a stylized local collective action framework that clarifies how an external intervention that provides financing and participation requirements might change local decision making and institutions, and derive implications that then structure our empirical analysis (see Appendix C for the formal exposition). In the model, a social planner determines the optimal investment in local public goods and sets a corresponding tax schedule, which is implemented with perfect compliance. Individual residents then decide whether or not to voluntarily participate in the planning and implementation of the public goods projects, taking their individual tax burden as given. We feel this framework is a reasonable approximation to the context of rural Sierra Leone (and similar societies with strong village headmen), where the traditional chief has the authority to levy fines and collect taxes to provide basic public goods, but there is variation in how involved residents are in actual decision making and implementation. In this setting, the external intervention lowers the *marginal* cost of local public goods provision through financial subsidies, and affects the *fixed* costs of collective action by imposing participation requirements and instilling democratic norms. We allow for underrepresented groups (i.e., women) to have differential participation costs *ex ante*, which could be impacted by learning-by-doing or demonstration effects during project implementation.

We define three time periods that correspond to our data collection activities: the pre-

[5] The Fearon et al. (2009) Liberia project provided roughly \$20,000 to "communities" that comprised around four villages with two to three thousand residents, so \$8 per capita over two years; and villages received \$8,800 in Indonesia (Olken 2007). While the difference in total grant size may affect the maximum feasible project scale, the per capita funding differences are not substantial.

program period when the baseline survey was fielded; the program implementation phase, where the first follow-up survey captured activities that had been completed during the intervention (and launched the structured community activities); and the post-program period, where the second follow-up survey explored what happened with the SCAs after the project had finished. As the marginal cost reductions are tied directly to external financial assistance, while the fixed organizing cost reductions could be internalized and maintained, we can speculatively gain some leverage over which channels are at work by comparing impacts during and after project implementation. Moreover, studying the post-program period allows us to evaluate the persistence and "sustainability" of impacts.

First consider the individual's decision of whether to contribute time and voluntary labor to the planning and implementation of local public goods. While these decisions are taken in a decentralized fashion, they allow them to aggregate in a way that affects the costs of public goods provision facing the social planner. The fact that individuals ignore the aggregate effect of their voluntary labor captures the classic externality feature of collective action, and implies that even with perfect tax compliance, the planner will still not be able to achieve the first-best level of local public goods provision.

Individuals gain utility from consumption of the current stock of public goods, private consumption, and a psychic or social benefit of participating in collective action that captures the intrinsic value of civic involvement. Regarding the latter, Olken (2010) and Dal Bó et al. (2010) provide evidence that having a say in the decision-making process can have a large effect on satisfaction and cooperation even if the choice process has no impact on the final policy outcome *per se*. Given historical legacies of exclusion, we assume that while some women and youth may derive positive utility from participation, they face additional social costs of speaking up and thus, on average, their net benefits of civic participation are lower than for the traditional elder male elites. All residents face the same opportunity cost of participating, which reflects the cost of time spent engaging in public goods provision instead of wage-earning activities, and must pay the tax set by the social planner. The first order conditions imply that the individual thus chooses to participate in collective action if and only if the net benefits are nonnegative.

The social planner chooses the level of local public goods investment with the objective of maximizing the sum of individual utilities. The cost of public goods provision has a marginal component, capturing the price of construction materials, as well as a fixed coordination cost of collective action, which is a function of both the sum of individual participation decisions and the capacity of local institutions. Following the theory motivating participatory local development, we assume that the fixed costs of collective action are falling in both the capacity of local institutions and community participation; we assess the empirical validity of these assumptions below. The latter condition would be true if, for example, greater community involvement made public goods provision easier and if more involvement in decision making created greater support for the process. Importantly, even if participation has no effect on coordination costs at all, advocates argue that local civic engagement carries intrinsic benefits, and therefore project participation belongs in the individual utility function and its enhancement becomes an appropriate objective for intervention.

Standard first order conditions imply that the planner chooses the efficient level of local public goods investment if is affordable, or a smaller investment that exhausts the village budget (at a corner solution) if it is not. Given the extremely limited public services in rural Sierra Leone, it seems reasonable to assume the latter, where communities face a binding budget constraint that keeps public investment well below optimal levels. This means that there are plenty of public investments—in latrines, water wells, primary schools—whose village-wide marginal benefits exceed the marginal cost of construction, yet are simply unaffordable given the community's tax base and inability to borrow in light of pervasive financial market imperfections. Under these constraints, profitable investments become unaffordable because construction prices and/or coordination costs are prohibitively high.

Within this framework, participatory local governance interventions aim to have three separate impacts. First, by subsidizing the cost of construction materials, the financial grants reduce the marginal cost of public goods provision. Second, the leadership quotas and participation requirements for women and youth aim to increase the benefits of participation for these historically marginalized groups. Such requirements should automatically translate into greater participation in collective activities during project implementation for these groups. Moreover, if women and young men learn-by-doing, or if their participation exerts positive demonstration effects on others that shifts social norms, this experience could trigger a persistent increase in their benefits of participation, sustainably raising participation levels into the post-program period. Third and finally, this increase in community participation, accompanied by the establishment of village development committees, plans and bank accounts, aims to reduce the fixed coordination costs of collective action. The idea is that once an organizing body is in place and residents have reached consensus on local priorities, the next village project should be less costly to identify and execute. Note that this effect should be evident in both the implementation and post-program periods. As such, the original GoBifo project funding proposal emphasizes the sustainability, "durability" and broad mandate of these new structures, suggesting they will become "the focal point for development interventions" and other forms of local collective action in the future (World Bank 2004). This description is consistent with more general claims by CDD advocates that a temporary intervention can permanently enhance local public goods provision by reducing the costs of collective action.

This simple framework generates three empirical predictions to take to the data. First, the combination of financial subsidies and lower coordination costs should unambiguously increase public goods investment during the program implementation phase. To assess this, outcome family A includes project implementation indicators to first evaluate whether the grants were in fact delivered to villages and new institutions established on the ground, and then a set of "hardware" measures regarding the stock of local public goods to assess immediate impacts on public investment levels. Second, as we move from project implementation to the post-program period, the marginal investment costs return to baseline levels while the fixed costs (potentially) remain permanently reduced. To evaluate whether the establishment of durable village institutions lead to continued greater investment in public goods in the post-program period, family B includes take-up of the building materials vouchers (in SCA #1), as well as several other measures of collective action beyond the direct program sphere. Third, if participation requirements for women and youth trigger a permanent enhancement in their benefits from participation, we should see more women and youths attending community meetings and taking part in decision-making post-program. This is captured by the "software" outcomes in the gift choice component of SCA #2 and household survey responses concerning civic engagement in non-program spheres. Moreover, enhancing participation by marginalized groups could initiate broader changes in social norms and attitudes (for instance, regarding the desirability of female leadership), as captured in several additional hypotheses under outcome family B.

3. Research Design

3.1. Random Assignment

The 118 GoBifo treatment and 118 control villages were selected from a larger pool of eligible communities using a computerized random number generator. Two study districts were chosen to strike a balance in terms of regional diversity, political affiliation, and ethnic identity, while simultaneously targeting poor rural areas with limited NGO presence (see Appendix D for a map). Bombali district is located in the Northern region dominated by the Temne and Limba ethnic groups and traditionally allied with the All People's Congress (APC) political party, one of Sierra Leone's two largest parties. Bonthe district is in the South, where the Mende and Sherbro ethnic groups dominate and where the population is historically aligned with the other major party, the Sierra Leone People's Party (SLPP). Using the 2004 Population and Housing Census, the eligible pool of villages was restricted to communities considered of appropriate size for a CDD project, namely between 20 and 200 households in Bombali and 10 to 100 households in Bonthe (where villages are smaller), and once the final study sample was chosen, the villages were randomized into treatment and control groups, stratifying on ward.[6]

For each community in the study sample, government Statistics Sierra Leone staff randomly selected twelve households to be surveyed from the Census household lists. Given research interest in the dynamics of political exclusion and empowerment, the choice of respondent within each targeted household rotated among four different demographic groups in each subsequent household surveyed: non-youth male, youth male, non-youth female and youth female. All respondents are at least 18 years old, and note that the Government of Sierra Leone's definition of youth includes people up to 35 years of age (although in reality the definition of youth is a bit subjective, especially since some respondents do not know their exact age). This data collection strategy means that for each community, and for the overall sample,

[6]We ran 500 computer randomizations and saved all resulting assignments that generated no statistically significant differences (at 95% confidence) between treatment and control groups in terms of the total number of households per village and the distance to the nearest motorable road. Among these "balanced" assignments, one was then selected at random for the final allocation of GoBifo treatment and control villages. Bruhn and McKenzie (2009) argue correctly that this process of re-randomization to achieve balance on observables may lead standard errors to be either under- or over-estimated. They show that correct inference can be achieved by including the "balancing" observables in the regression analysis as control variables, and these variables are thus included in our standard set of regression controls in all results presented below. The treatment effect estimates are thus interpreted as impacts conditional on these observables. It is worth noting, however, that coefficient estimates and standard errors are nearly unchanged whether or not these controls are included in the analysis (not shown).There were two minor data issues in measuring community size and ward location that led to a partial re-sampling of a small number of villages, however these did not affect the integrity of the randomization (see web Appendix E).

responses are roughly balanced across the four demographic groups.[7]

The randomization procedure successfully generated two groups balanced along observable dimensions. Specifically, Table 1 lists the mean value in the control group and the treatment minus control pre-program difference for a variety of community characteristics (including total households, distance to nearest road, average respondent years of education, and indices for civil war exposure and local history of domestic slavery) as well as an illustrative selection of pre-program values for measures that fall under each of the two outcome "families" mentioned above. There are no statistically significant mean differences across the treatment and control groups in the 2005 values of any of these variables; Appendix F contains the same estimates for all 94 baseline measures and shows that the difference across treatment and control groups is significant at 90% confidence for only seven of these, roughly as expected by chance. Note that the analysis below controls for baseline values of the outcome under consideration where available, addressing any incidental imbalance across groups. One noteworthy pattern in the baseline data is the stark gender difference in local meeting involvement, with twice as many males (59%) than females (29%) speaking at village meetings.

3.2 Data Collection and Measurement

This analysis draws on three main data sources: household surveys from late 2005 (baseline) and mid-2009 (follow-up); village-level focus group discussions held in 2005 and 2009; and three novel structured community activities (SCAs) conducted in late 2009 shortly after GoBifo activities had ended. The SCAs were introduced with the initial post-program survey in May 2009 and then followed up in an unannounced visit five months later. The research team and enumerators were operationally separate from GoBifo staff at all stages of the project.

The 2005 household surveys collected data on baseline participation in many local collective activities, as well as detailed household demographic and socioeconomic information. To establish a panel, the field teams sought out the same respondents during the 2009 follow-up household surveys that they had previously interviewed, and the attrition rate was moderate: overall, 96% of the same households were located and 76% of the same individual respondents.

During the data collection visits in 2005 and 2009, the field team supervisor assembled

[7] These four demographic groups each comprise roughly a quarter of the adult population in these two districts in the 2004 Census (ranging from 21 to 31%), indicating that our sample is quite representative.

key opinion leaders—including VDC members, the village chief, as well as women and youth leaders, among others—to describe the condition of local infrastructure and answer questions about local collective processes and activities. Research supervisors also made their own physical assessments of the quality of construction as a cross-check on focus group responses.

Given the difficulties in gauging institutional dynamics and collective action through survey responses alone, the third main type of data was gathered through the SCAs. These were designed to measure how communities respond to concrete, real-world situations requiring collective action in three different dimensions: (i) raising funds in response to a matching grant opportunity; (ii) making a community decision between two comparable alternatives; and (iii) allocating and managing an asset that was provided for free. As opposed to hypothetical vignettes or laboratory experiments in the field, these exercises more directly, realistically, and less obtrusively capture institutional outcomes of interest. As we see the development of these measures as a key contribution of this paper, we discuss each SCA in detail below.

SCA #1 was designed to measure whether GoBifo produced persistent effects on villages' capacity for local collective action beyond the life of the project itself. Each community received six vouchers they could redeem at a nearby building materials store if they raised matching funds. Specifically, each voucher was worth 50,000 Leones (roughly US$17) only if accompanied by another 100,000 Leones (US$33) from the community. Matching all six vouchers generated a sizeable 900,000 Leones, or approximately US$300, for use in the store. As the materials could always be resold at value for a profit (given the subsidy), take-up of the vouchers was in the community's self-interest. Yet since individuals had negligible savings and faced credit constraints, take-up of the vouchers is a measure of local capacity for cooperation. Voucher redemption was recorded by clerks at the building materials stores. Enumerators returned to all villages five months after the initial distribution of the vouchers to assess the distribution of project contributions and benefits (i.e., did they buy zinc to build a new roof for the primary school or for the chief's private compound?), the quality of final construction, and how inclusive and transparent the management of the resulting project had been. In the context of the model, higher take up in treatment communities implies that the program reduced the *fixed* costs of collective action, as in this case the marginal component (i.e. the financial subsidies offered through the vouchers) was exactly the same for treatment and control villages. A treatment effect here would thereby suggest that GoBifo exerted a persistent effect on public investment by changing the nature of local institutions, norms and collective action capacity.

SCA #2 was designed to measure the extent to which community decision-making is democratic and inclusive, and to assess the level of community participation. The day before survey work, the enumerator teams met with the village head (the lowest level chiefly authority) and asked him/her to assemble the entire community for a meeting the next morning. At the subsequent meeting, the enumerators presented the community with a choice between two gifts each valued at roughly US$40—a carton of batteries (useful for radios and flashlights) versus many small bags of iodized salt—as a token of appreciation for participating in the research. The enumerators – who were Statistics Sierra Leone employees and not GoBifo staff – emphasized that the community itself should decide how to share the gift and then withdrew from the meeting to observe the decision-making process from the sidelines. The enumerators remained "outside" the community meeting circle and recorded how the deliberation evolved without making any comments of their own. Among other things, the enumerators recorded who participated in any side-meetings; the degree to which the chief, village head and elders dominated the discussion; the extent of debate in terms of time and the number of comments; and a subjective assessment of the apparent influence of different sub-groups (e.g., women) on the final outcome. This exercise thus provides concrete quantitative data on the relative frequency of female versus male speakers, and youth versus non-youth speakers in an actual community meeting.[8] Note that these are exactly the same metrics that the GoBifo facilitators were required to track during project meetings as part of their own internal performance assessment, to monitor strides in women's and youth participation, leadership and power in treatment communities (GoBifo Project 2008). Any effects on women and youth's participation would imply that the program's minority representation requirements exerted a durable impact on behavior through learning-by-doing or demonstration effects.

SCA #3 was designed to gauge the extent of elite capture of resources—a common concern for decentralization reforms—as well as the broader nature of collective action. During the first follow-up visit in 2009, the enumerators gave each village a large plastic tarpaulin sheet as a gift. Tarpaulins are frequently used in Sierra Leone as makeshift building materials for roofing (40% of households have potentially leaky thatched roofs), and in agriculture as a

[8] Of the four enumerators, one focused their data collection on the participation of youths, one on women, one on all adults and the fourth kept careful track of each person who spoke publicly.

surface for drying grains (fewer than a quarter of villages have a functional drying floor). During the second 2009 follow-up visit five months later, enumerators recorded which households had had access to the tarpaulin in the intervening period. This activity also captures an element of collective action, as enumerators assessed whether villages had been able to decide on a use for the tarp at all, and whether it had been put mainly towards a public (e.g., a communal grain drying floor) or private end (patching the roof of an individual's home).

3.3 The Pre-Analysis Plan and Econometric Specifications

In what follows, we present results for the specific hypotheses described in our pre-analysis plan, a document that was finalized before we analyzed any follow-up data. The genesis of the plan was a pre-program 2005 agreement between the research and project teams that set out the areas GoBifo was likely to impact and how success in these areas would be assessed. Building on this early document, we drafted a formal analysis plan that specified the exact outcomes under each of eleven hypothesized areas of impact and the econometric specifications to be used, which we archived with the Abdul Latif Jameel Poverty Action Lab randomized evaluation archive on August 21, 2009 (through hypotheses@povertyactionlab.org). We believe we are among the first economics studies to use this approach.[9] Pre-analysis plans limit data mining, specification searches, or an *ex post* rationalization that selectively highlights only positive (or negative) effects—or those on specific subgroups—discovered during analysis. The establishment of a public registry could further help mitigate publication bias by expanding the body of evidence readily available to researchers and practitioners (Rasmussen et al. 2011).[10]

Towards these ends, the plan has several components. First, it defined both the sets of explanatory and dependent variables (Leamer 1983) and econometric models (Leamer 1974) before data analysis began. While the randomized framework naturally imposes much of this narrowing (i.e., the treatment indicator is the leading explanatory variable), the plan also details the set of interaction terms and population subgroups we would use to explore heterogeneous treatment effects. Second, the large number of outcome variables we consider means that several individual treatment effects will be statistically significant due simply to random chance. To

[9] See also Olken et al. (2010), Schaner (2011) and Alatas et al. (forthcoming). Finkelstein et al. 2011 archived a pre-analysis plan on December 3, 2010.
[10] For an illustration of the problem of specification searching and publication bias in the empirical labor economics literature regarding minimum wage impacts, see Card and Krueger (1995).

account for this, the plan commits us to a mean effects approach that reduces the effective number of tests we conduct by identifying in advance which outcome variables would be grouped together to jointly identify the different hypotheses laid out in the 2005 document (see O'Brien 1984; Kling and Liebman 2004; Anderson 2008). While the mean effect index is the primary metric by which we evaluate hypotheses, we also provide results for the outcome measures individually to provide a sense of their magnitude and economic significance. Third, for full transparency, we disclose the complete results for all 318 unique outcome variables, including the exact wording of the survey question, in supplementary web Appendix G.

There are two minor deviations from the original pre-analysis plan in what we present below. We added a twelfth hypothesis (called hypothesis 1 below) by pulling together outcomes that had already been explicitly included within the original eleven hypotheses. Thus no new outcome measures were added or excluded in what we present below. Those who wish to consider only the results as exactly laid out *ex ante* can ignore hypothesis 1. However, we feel it was an oversight to exclude a project implementation hypothesis beforehand and thus still find the results of hypothesis 1 useful to consider. Perhaps more important is that we group the hypotheses into two "families" for ease of comprehension and to facilitate links to the theory. While we did not specify these families beforehand, we believe that the groupings—the development "hardware" of project implementation, public goods and economic activity (family A), and the "software" of local collective action (family B)—are intuitive.

Under each hypothesis, we evaluate specific treatment effects using the following model:

$$Y_c = \beta_0 + \beta_1 T_c + X'_c \Gamma + W'_c \Pi + \varepsilon_c \tag{1}$$

where Y_c is an outcome (i.e., local school construction) in community c; T_c is the GoBifo treatment indicator; X_c is a vector of the community level controls, including those used to assess treatment versus control group balance in the original computer randomizations; W_c is a fixed effect for geographic ward, the administrative level on which the randomization was stratified; and ε_c is the usual idiosyncratic error term. Elements of X_c include distance from road, total number of households, an index of violence experienced during the recent civil war and a measure capturing the historical extent of domestic slavery. The parameter of interest is β_1, the average treatment effect. Note that while some outcomes are measured at the household (e.g., radio ownership) or individual level (e.g., political attitudes), the natural unit of analysis is the village and we thus measure all variables at this level, taking village averages as necessary (analysis at the household level yields nearly identical results, not shown).[11]

For the subset of outcome variables that were collected in both the baseline 2005 survey and in the 2009 follow-up surveys, the analysis exploits the panel data structure:

$$Y_{ct} = \beta_0 + \beta_1(T_c * POST_t) + \beta_2 T_c + \beta_3 POST_t + X'_c\Gamma + W'_c\Pi + \varepsilon_{ct} \tag{12}$$

where Y_{ct} is a particular outcome for community c at time t, where $t = 0$ in the 2005 baseline survey and $t = 1$ in the 2009 follow-up. The additional indicator variable *POST* denotes the follow-up period. The parameter of interest is again β_1, the average treatment effect, and here the disturbance terms are clustered at the village level. Results are robust to the exclusion of the vector of community controls and to limiting our analysis to only the post-program data (as shown in the sparse specifications in supplementary web Appendix G). As set out in our analysis plan, we further assess the degree of heterogeneous treatment effects by including interaction terms of treatment with respondent gender, age, village remoteness, community size, war exposure, the local history of domestic slavery, and location in each of the two study districts. As we do not find any evidence for heterogeneous effects along any of these dimensions, we have excluded this discussion from the main text (see supplementary appendix H for summary results by outcome family).

The mean effects index for a hypothesis captures the average relationship between the GoBifo treatment and the K different outcome measures grouped in that hypothesis. Following Kling and Liebman (2004), estimation of the index first standardizes outcome variables into comparable units by translating each one into standard deviation units (by subtracting the mean and dividing by the standard error of the control group) before regressing each outcome on the vector of independent variables. The index coefficient is the mean of these K standardized treatment effects. The estimation method calculates the standard error of the index itself, which depends on both the variances of each individual $\beta_{1,k}$ as well as any covariances between $\beta_{1,k}$ and $\beta_{1,\neg k}$, requiring a seemingly unrelated regressions (SUR) system approach to test the cross-equation hypothesis that the average index of K coefficients equals zero.

[11] Our pre-analysis plan states that we would carry out regressions at each of these three levels of aggregation, and since these different levels accommodate different control variables, this leaves an extra degree of freedom that we could have eliminated. Fortunately, the results are consistent across these three different specifications. If we were to create the pre-analysis plan again, however, we would pre-specify analysis at the village level only with the accompanying set of village-level controls.

4. Empirical Results

Table 2 presents a concise summary of the mean effect results for all twelve hypotheses, grouped into the two outcome families. The positive and significant (at 99% confidence) mean effect estimate of 0.352 standard deviation units for family A (hypotheses 1, 2 and 3) indicates that GoBifo achieved its most immediate objective of providing the organizational and financial means to encourage local public goods construction and small enterprise development. Specifically, the coefficient on hypothesis 1 indicates that the program was well executed, perhaps more so than many other real-world development projects: GoBifo increased measures of local organization and linkages to facilitate collective action by 0.687 standard deviations on average. This strong implementation performance in turn led to immediate impacts on local infrastructure and other hardware. The estimated mean effect of 0.164 for hypothesis 2 reflects positive effects on the stock and quality of local public goods; while the 0.399 coefficient for hypothesis 3 reflects gains in general economic welfare. Reflecting back on the theoretical framework, these increases provide strong support for the prediction that the combination of lowering the marginal cost of public goods through grants, as well as reducing coordination costs through the establishment of new institutions, led to greater public investment during project implementation. The next question is how much of this effect was driven by changes in institutions, norms and collective action capacity.

The small and not statistically significant mean effect estimate for family B (hypotheses 4 through 12) suggests that the experience of working together in GoBifo, and the introduction of new institutions and processes, did not durably change the nature of local collective action. The program's democratic decision-making and "help yourself" approach did not appear to spill over into other realms of village life nor to persist into the post-program period. We find no evidence that GoBifo led to fundamental changes in local capacity to raise funds and act collectively outside of the project, the nature of decision-making, the influence of women or youths, or a range of social capital outcomes. In the context of the model, these null results suggest that GoBifo did not permanently increase the benefits of civic engagement for marginalized groups and that the organizing institutions established did not persistently reduce the fixed costs of collective action. In the subsections that follow, we flesh out these results with an illustrative sample of outcomes under each family. Those interested in any particular outcome omitted from the discussion below (due to space constraints) should refer to Appendix G for the entire inventory of results for all twelve hypotheses.

4.1 Family A: Development Infrastructure or "Hardware" Effects

The first hypothesis focuses on project implementation and measures the extent to which GoBifo successfully established Village Development Committees (VDCs); helped communities draw up development plans and open bank accounts; and created links between the villages and their local government representatives. The first two panels of Table 3 present results for several outcomes under this hypothesis, where the first seven "full sample" outcomes in Panel A apply to all communities within the sample; while the remaining six "conditional" outcomes in Panel B are conditioned on t he existence of public infrastructure and thus only apply to those communities that have the particular good. All of these treatment effects are greater than zero and nine are statistically significant at 95% confidence.

Regarding interpretation, the treatment effect estimate from the panel specification in the first row of Table3 indicates an increase of 34.1 percentage points in the existence of a VDC. VDCs already existed in many Sierra Leonean villages when GoBifo was launched, having been introduced by humanitarian assistance groups during the war-torn 1990's (Richards et al. 2004). By the post-program period, 86.3% of GoBifo communities had a VDC compared to 45.8% of controls, a large effect. The corresponding coefficient in the second row indicates that GoBifo increased the likelihood that a community was visited by a member of its Ward Development Committee in the past year by 15.6 percentage points. Row 3 shows a positive treatment effect on the existence of village development plans by 29.6 percentage points, nearly a 50% increase on the base of 61.7% in the controls. Row 4 reveals an increase in having a village bank account of 70.6 percentage points, capturing a tenfold increase. In Panel B, the household survey asked respondents whether a member of the Ward Development Committee or district council was "directly involved in the planning, construction, maintenance or oversight" of several local public goods. Note that the treatment effect is positive and significant for nearly all outcomes. This suggests that GoBifo successfully led local politicians to increase their involvement in village projects, consistent with its objective of supporting the broader decentralization process.

Moving from project implementation to impacts on development hardware, hypothesis 2 explores treatment effects on the quantity and quality of local public goods. While combining measures within a single hypothesis into sub-indices was not specified in our pre-analysis plan, the outcomes under hypothesis 2 naturally form three sub-groups: the stock of local public goods, the quality of such goods, and community financial contributions to their construction and upkeep. Regarding the stock, the first five rows of Panel C in Table 3 present impacts for an illustrative sample of goods. Note that four of these treatment effects are positive and three are statistically significant. Specifically, there are marked increases in the proportion of villages with a functional traditional midwife post by 17.5 percentage points, community center by 24.1, and latrine by 21.0. Calculating a mean index on the entire sub-group reveals a highly significant increase of 0.258 standard deviations (s.e. 0.049, not shown).[12]

Turning to the next sub-group, the first three rows of Panel D show positive GoBifo impacts on the quality of construction of three of the most common public goods—primary schools, latrines and grain drying floors—where quality was determined through direct physical assessment. The effects are all positive as is the quality sub-group index overall, which shows an increase of 0.296 standard deviation units (s.e. 0.077).[13] These measures combine impacts from the GoBifo funded infrastructure projects, as well as any potential effects from better local collective action in maintaining existing infrastructure. However, as there is no evidence that management practices did in fact change in treatment villages, the leading interpretation is that the positive impacts are being driven by the grants.

The last three rows of Panel D present illustrative results for the final sub-group of outcomes that concern community financial contributions to existing infrastructure. Two of these are negative and one is statistically significant. Looking across nine different local public goods, the sub-group index is negative (at -0.113 standard deviations) but not statistically significant (s.e. 0.104). Combined with the negative and marginally significant effect on whether the community approached another NGO or donor for financial support (in row 15 of Table 3), these provide suggestive evidence that GoBifo funds served as a substitute, rather than a complement, for the community's own resources. At a minimum, they indicate that the GoBifo grants did not serve as a catalyst for additional fund-raising nor did project experiences encourage participants to seek out further development assistance. The SCA findings (discussed in Section 4.2 below) reinforce this finding.

Hypothesis 3 relates to at general economic welfare, since roughly one sixth of the grants

[12] The mean effect index for the sub-group includes impacts on six additional goods not presented due to space constraints: water wells, peripheral health unit, market, grain store, sports field and sports uniforms.

[13] The quality of construction sub-group index uses two measures for each good and includes effects on water wells.

were used to launch projects dedicated to job skills training or small business development—such as carpentry, soap-making and seed multiplication initiatives—that, if well implemented, could translate into higher small business profits, and perhaps lead to sustainably higher future earnings. Moreover, GoBifo injected cash grants into very poor communities, and as with any assistance, a portion of the funds are surely fungible. Via potentially all of these mechanisms, this third hypothesis considers project impacts on measures of community-wide economic activity and household welfare.

The first two outcomes in Panel E of Table 3 refer to village-level outcomes, where we see a 30% increase in the number of petty traders (0.7 more traders on a base of 2.4 traders in the control group) and a 13% increase in goods locally available for sale. The last four outcomes are aggregated from household survey reports. We observe improvements in an asset ownership score (derived using principal components analysis), where the underlying assets include common household durables (e.g., radios, mobile phones), amenities like drinking water source and sanitation, and the materials used in the roof, walls and floor of the dwelling. The project tripled the proportion of respondents who had recently participated in skills training: an 11.9 percentage point increase on a base of 6.1% in control communities. We find no impact on total household income in 2009, however, this is difficult to measure among households engaged in subsistence agriculture and the treatment effect estimate is relatively imprecise.

4.2 Family B: Impacts on "Software": Local Institutions and Norms for Collective Action

The positive treatment effects for outcome family A suggest that investment in local public goods did increase substantially during the project as predicted by the theoretical model. To determine the role played by more effective local institutions (versus the block grants), we next examine post-program outcomes after the block grants had been spent. The first hypothesis under the software family (hypothesis 4) covers outcomes relating to collective action and contributions to local public goods. The mean effect for this hypothesis is not statistically distinguishable from zero (0.041 standard deviations with a standard error of 0.042); and of the 59 full sample and conditional outcomes evaluated, only seven treatment effects are significant at 95% confidence, with five positive in sign and two negative. The subset of outcomes relating to the matching grant opportunity (SCA #1) provides the most succinct and concrete illustration of the lack of program impacts in this area. The top panel of Table 4 shows that there was no differential take-up of the subsidized building vouchers: 62 treatment (52.5%) and 64 control villages (54.2%) redeemed vouchers at local supply stores; nor is there any difference in the number of vouchers redeemed, as most of the villages that cashed in any vouchers used all six. The ability to mobilize around a new opportunity and raise funds for it is close to the essence of local collective action. This finding implies that the program did not have durable effects on collective action capacity.

Other outcomes under this hypothesis consider household contributions to existing local public goods, where we expand the set of contributions to include labor, local materials, or food for project workers, yet continue to find no treatment effect. There are also no differences in contributions to several local self-help groups (i.e., rotating savings groups and labor gangs) nor in financial support for community teachers. Lastly, while treatment villages were more likely to have a communal farm, by 23 percentage points (significant at 99% confidence), the total number of respondents in treatment areas who had worked on a communal farm in the past year was no higher. This presents a telling example of how project-funded activities—for example, the subsidized provision of seeds and tools for a community farm—exerted a proximate effect on the establishment of a local organization established to capture that funding, but did not have any lasting impacts on actual communal cultivation in subsequent years.

These findings raise troubling questions about GoBifo's long term impacts. Clearly, community members gained experience in working together to successfully implement local development projects over the nearly four years of the project. Yet their GoBifo-specific experiences did not lead to greater capacity to take advantage of new opportunities that arose after the program ended. Most strikingly, while GoBifo often created new structures designed to facilitate local development by reducing organizational costs—the VDC, a development plan, a bank account, and a communal farm—these structures left them no better able to take advantage of the realistic matching grant opportunity in SCA #1.

The second "software" hypothesis includes outcomes relating to the civic involvement of socially marginalized groups. Since the inclusion of women and youth held great prominence in GoBifo's objectives and facilitator operating manuals, it also received special attention in the data collection. Covering an exhaustive battery of measures, the mean effect is a precisely estimated zero (see hypothesis 5 in Table 2) indicating no overall impact on the role of women or youth in local decision-making, or on the transparency and accountability of decision-making more generally. Of the 72 distinct outcomes considered, only six were statistically significant at 95% confidence, dividing equally between positive and negative treatment effects.

Enumerator observations during SCA #2, when villages met to decide between salt versus batteries, provide a clear illustration of this zero result. In Panel B of Table 4 there are no treatment effects on the total number of adults, women and youths who attended the meeting or spoke publicly during the deliberation. To illustrate: on average, 25 women attended these meetings but just two of them made a public statement during the discussion about which item to choose. The difference between the number of women who spoke in treatment versus control communities is only -0.19 (s.e. 0.22), and the proportion of males who spoke during the meeting remained twice as high as the proportion of females in the treatment villages, the same as at baseline. We similarly find no impact on whether any smaller "elite" groups broke off from the general meeting to make the gift choice without broader consultation; the duration of the deliberation; or how democratic the decision-process appeared to the enumerators (e.g., by holding a direct vote). These results are further substantiated by respondent reports recorded immediately after the meeting of how the tarpaulin allocation choice in SCA #3 was made, including which individuals had the final "say" and to what extent the decision was dominated by local elites (i.e., village headmen and male elders). Moreover, respondent opinions collected during the second 2009 follow-up survey (five months later) also find no treatment effects on reports about how decisions were made to distribute the salt or batteries (SCA #2); how to use the tarp (SCA #3); whether to raise funds for the building materials vouchers, and if so, how to mobilize funds, which items to purchase, and how to manage any construction (SCA #1).

Despite all of the effort in GoBifo to elevate the position of women and youth, we thus do not observe any improvement in their role relative to older men in community decision making. Even for relatively low cost actions like speaking up in meetings, the nearly four years of GoBifo project activities did not translate into greater apparent voice for marginalized groups. In the context of the theory, this suggests no persistent gains in the individual benefits of participation for these groups, and provides additional evidence that the increase in public investment observed during project implementation was likely driven by the financial subsidy rather than fundamental changes in local institutions or *de facto* power.[14]

[14] However, we cannot rule out that the subsidy was particularly effective (i.e., led to such notable increases in public goods) in part because of the project's facilitation and emphasis on participation and transparency.

The third software hypothesis (which was included in the pre-analysis plan but was not an official aim of GoBifo project management) asks whether by espousing more democratic ways of managing local development, the project led to changes in the role of the traditional chiefly authorities. Taking all outcomes together, the mean effect for hypothesis 6 is also zero (Table 2). Many outcomes under this hypothesis estimate the extent to which the village head and elders dominated the SCA decisions. While we find variation in how these decisions are made—at one extreme, in two villages the Chief decided between the salt and batteries in less than one minute without anyone else's input, while at the other an open discussion lasted nearly an hour and was followed by a formal vote—as mentioned above, we find no systematic differences across treatment and control villages.

A leading explanation for the lack of institutional change, with some support in the data, is that elites exerted substantial control over the new organizations GoBifo created. As an example, traditional elites retained their leadership of the VDC: in both treatment and control villages (for the roughly half of control communities with a VDC in 2009), 88% of VDC chairs are men, 87% are older than 35, and 52% are traditional chiefdom authorities and elders. While participation requirements translated into some gains for women (a 6.6 percentage point increase in the proportion female members and a near doubling of the proportion of female Treasurers (57% versus 31%)), the representation of youths remained at the same low level as in control areas (at 26%). These patterns highlight a tension inherent in the CDD approach: leveraging the capacity of existing institutions may be expedient for immediate project implementation while simultaneously limiting the likelihood of fundamental institutional transformation or changes in *de facto* power for marginalized groups.

We therefore tested the related hypothesis that CDD may enable local elites to capture a disproportionate share of economic benefits. We explored this issue directly by distributing a new public asset—the tarp (SCA #3)—in villages during the first 2009 follow-up visit, and then observing how it was being used in the unannounced second visit five months later. While the analysis finds no treatment effects on the extent of elite capture, it also reveals that the level of elite capture is, perhaps surprisingly, relatively low in the study communities. Panel C of Table 4 shows that for the 90% of communities that had used the tarp by the time of the second visit, 86% had put the tarp towards a public purpose, such as a communal rice drying floor or local ceremony. The most obvious example of elite capture would be use of the tarp to patch the roof of a single individual's house, which happened in fewer than 3% of all villages. That said, several communities had not yet used the tarp and were storing it at a private residence, which either suggests a failure to agree on the appropriate way forward, or signals the risk of future elite capture, or both.

The next three hypotheses explore proxies for "social capital"—self-expressed trust of others (hypothesis 7), involvement in local groups and networks (hypothesis 8), and access to information (hypothesis 9)—emphasized alongside collective action and inclusion in the official GoBifo project objectives (World Bank 2004, GoBifo 2007). Despite exploring a wealth of measures, the analysis reveals no treatment effects on social capital and the three mean effects indices are all indistinguishable from zero. Beginning with trust, the only significant effect is an increase in reported trust of NGOs and donor projects: residents in treatment communities were 5.4 percentage points more likely to agree that NGOs or donors "can be believed" (close to the Krio translation for trust) as opposed to you "have to be careful" in dealing with them. There are no effects on the remaining eleven indicators, which combine respondent self-reports regarding how much they trust various groups with concrete examples of trusting behavior, such as entrusting money to a neighbor to purchase market goods on your behalf.

Second, enumerators asked respondents whether they were a member of a local self-help group (such as a credit/savings group, communal labor gang, school committee, funeral savings group, fishing cooperative, women's group or youth group, among others) and if so, whether they had attended a meeting and contributed financially or in labor in the past month (hypothesis 8). We find no significant treatment effects on these indicators nor on other measures of local cooperation, such as whether the respondent had helped a neighbor re-thatch the roof of their house, a time-intensive activity that one cannot easily do alone.

There is also no evidence of treatment effects on households' access to information about local government or governance (hypothesis 9). Among 21 outcomes, only one—the proportion of villages visited by a WDC member, discussed above—shows statistically significant effects. The collection of zero effects includes measures of how much respondents know about what the community is doing with the building vouchers (SCA #1) and tarp (SCA #3); whether they can name their district council and chiefdom leaders; and their ability to answer objective questions about how local taxes are collected and used.

While the mean effect index for participation in local governance in Table 2 (hypothesis

10) is positive and statistically significant, it is largely driven by the outcomes already discussed under family A. Specifically, we find large impacts on the existence of VDCs and village plans, and increases in the oversight of local public goods by chiefdom authorities that mirror earlier results on the involvement of local government representatives. There is no systematic evidence, however, that these stronger links with either set of local officials translated into more active individual political engagement, such as self-reported voting or running for local office. Similarly, treatment communities were no more likely to use the building materials and tarp in the SCAs for goals specified in their village development plan. Reinforcing earlier results, this disconnect between the articulation of a development plan and its real-world application suggests that few communities applied GoBifo project tools to initiatives beyond the program.

There are no impacts on crime and conflict in treatment villages or in the mechanisms through which they are resolved, leading to a zero mean effect for hypothesis 11 (Table 2). Of the ten indicators considered, only one—the 2 percentage point reduction in household reports of physical fighting over the past year—is significant at 95% confidence. While the nine null results imply that project efforts to enhance conflict management capacity may not have created lingering benefits, on the positive side it provides some reassurance that the infusion of block grants into the treatment communities at least did not spark increased conflict.

The twelfth and final hypothesis concerns the nature of individual political and social attitudes. The GoBifo program's emphasis on the empowerment of women and youth, and the transparency of local institutions, may have engendered a more equitable or "progressive" outlook toward politics and society more generally. Even if there are no changes in actual decision-making processes or local collective outcomes (as above), a marked change in expressed attitudes might still mean that the seeds for future social change had been planted. Enumerators gauged attitudes using pairs of opposing statements, such as "As citizens, we should be more active in questioning the actions of leaders" versus "In our country these days, we should have more respect for authority," and asking respondents which they agreed with more. These paired statements capture respondent views on a diverse range of topics including the acceptability of the use of violence in politics (a particularly salient issue in post-war Sierra Leone), domestic violence, youth and women in leadership roles, paying bribes, and coerced labor. Once again, there are no significant program effects, despite the concern that social desirability bias might lead some respondents to express views promoted by the program. The only significant impact is a positive 3.8 percentage point increase in agreement with the statement that young people can be good leaders. However, recall that this change in opinions did not translate into more youths holding actual leadership positions on the VDC, or to more youth participation in the SCA meetings. Attitudinal change may be a necessary step toward changing future behavior, but almost four years of an intensive community driven development program did not lead to detectable changes in a wide array of expressed attitudes.

4.3 Robustness and Validity Checks

This section evaluates the robustness of the results. To start, we consider typical threats to randomized experiments. Fortunately, there were no problems with treatment non-compliance: all communities assigned to the treatment group received the program and none of those in the control group participated; and respondent attrition rates are no different in treatment and control areas. The baseline statistics presented in Table 1 and supplementary Appendix F also suggest that the randomization process successfully created two groups of villages that were similar along a wide range of observables. Note further that the analyses use the baseline value of the outcome of interest as a control variable wherever such panel data is available. Thus in order for spurious differences between the two groups to explain the positive impacts, the treatment group would on average have had to be on a different trajectory than the controls, but there is no reason to believe this should systematically be the case given the randomized research design.

We next consider reasons why the treatment effect estimates might be underestimates. First, significant program spillovers from treatment to control communities could lead us to underestimate program impacts, since the control communities would also be receiving program benefits, albeit indirectly. For this to be true, we might expect the coefficient on the $POST_t$ indicator (in equation 12) to often be positive and significant, but this is not the case: across all the outcomes in Appendix G where panel data is available, there are exactly as many (21) positive as negative coefficient estimates on $POST_t$ that are statistically significant, and thus it seems unlikely that the results are biased by spillovers across communities.

A further concern is that the projects GoBifo simultaneously implemented at the ward level systematically benefited the control group at the expense of the treatment group. There was a separate pot of funding for each ward that was allocated by the Ward Development Committee (see footnote 2). Bias could result if WDC members took into account the placement of GoBifo village-level projects in deciding where to locate the ward projects and targeted those areas that had not already benefited, perhaps as a way of compensating them for losing out on village-level assistance. However, there are no meaningful differences in the targeting of ward-level projects across treatment and control villages, and, if anything, treatment villages are slightly more likely to benefit: while 15.2% of respondents in treatment areas reported that a household member benefited from a ward-level project, only 6.1% of respondents in control areas reported benefits.

A final concern is that the outcome measures were simply insufficiently refined to detect subtle decision-making, institutional, political or social differences between treatment and control communities. While some of our measures are certainly better than others, our main strength lies in the diversity and multiplicity of measures we use and the fact that they all produce similar results. We combine different data collection approaches, for example, employing both survey self-reports on the percentage of female and male respondents who spoke during the SCA meetings with direct enumerator observation of how many men and women they saw speaking during the meeting. The research teams also gathered information from a variety of sources: they interviewed men and women in their own homes, held focus group discussions with key opinion leaders, observed a community decision as it unfolded, and recorded their own independent assessment of the construction quality of local infrastructure. Lastly, we examine a large number of outcomes. Taking all these data together, the "zero" GoBifo program effects are quite precisely estimated. To illustrate, the maximum true positive treatment effect on the proportion of women speaking (in the salt versus battery SCA #2 deliberation) that we may have incorrectly ruled out at 95% confidence is one additional female speaker per every 4.3 villages we visited, which is quite small. In the mean effects analysis, which combines many outcome measures, confidence intervals are considerably tighter.

4.4 Alternative Interpretations and the Perils of Data Mining

Section 4.2 shows that evaluating the institutional change outcomes jointly under their pre-specified hypotheses generates no evidence for program impacts under family B. Yet without the discipline of the pre-analysis plan and mean effects approach, we could have instead selected an assortment of individual treatment effects from amongst our 146 variables to tell basically any story we liked about the impact of external aid on institutions. While such data mining poses a risk for any analysis, it is particularly problematic for measuring institutions. The multidimensionality of institutions—governing political, economic and social behaviors—implies a large number of outcomes, some of which will be statistically significant by pure chance. Moreover, because institutions are amorphous and contextually determined, there is no commonly agreed set of standard measures defining the core of each domain. (By contrast, any study regarding the returns to education would by necessity focus on individual wages.) The combination of these two characteristics allows great flexibility in determining which institutional outcomes to measure and report, possibly tempting the researcher to "cherry-pick" a novel set of treatment effects whose selectively is difficult to detect from the outside. To underscore just how misleading such data mining can be, Table 5 uses our data to construct two alternative interpretations—one negative, one positive—about GoBifo's impacts on institutions.

The selective collection of individual treatment effects in the first panel of Table 5 suggests that the heavy emphasis placed on participation during GoBifo implementation activities created "meeting fatigue" within treatment villages, which eventually translated into poor management of local development projects and political apathy. Specifically, respondents were less likely to report that they had attended a meeting to decide what to do with the tarp after the research teams had left the village. Tracing this initial backlash against participation through the course of the tarp SCA, we see that villagers were less likely to: report that "everyone had equal say" in deciding how to use the tarp; know what the tarp was being used for; actually put the tarp to use; or be able to produce the tarp for inspection by the survey team. This deterioration in community participation appears to have further manifested in declining civic engagement more broadly, as evidence by decreased interest in holding local office (as a VDC member) and lower turnout in the recent local government elections.

The second panel of Table 5 presents quite the opposite story: these treatment effects suggest that the positive experiences communities gained implementing GoBifo projects catalyzed other collective activities and encouraged villagers to incorporate the new democratic practices into other realms of decision making. These shifts in collective norms and behaviors in turn created space for new leaders in the community and incited greater interest in politics more generally. More specifically, individual outcomes in Panel B reveal gains in non-project specific collective action, like increased training for community teachers and a greater prevalence of women's groups. Broad adoption of the democratic norms introduced by the project is evidenced by increased minute taking at community meetings, a greater likelihood of storing building materials in a public place, and local chiefs playing a less dominant role in managing the tarp. Finally, the CDD experience instilled a more accepting attitude towards youths taking on leadership roles, and increased citizen awareness of national politics, as seen by the greater ability to correctly name the Section Chief and the date of the next general election.

These two plausible, completely opposite, and equally erroneous interpretations highlight the fallacy of choosing the set of outcomes to present *ex post* based on statistical significance alone. Our pre-analysis plan eliminated this possibility by specifying in advance the complete set of outcomes and how the individual treatment effects would be aggregated to evaluate the overall success of the intervention. Thus the statistically significant impacts scattered throughout family B do not contradict the estimation of a zero overall treatment effect on institutions.

5. Conclusion

This paper evaluates the effectiveness of a well-implemented program that sought to provide public goods and change institutions by establishing organizational structures to streamline collective action and imposing participation requirements to enhance the influence of marginalized groups in Sierra Leone. Our evidence suggests that the intervention was successful in setting up new village structures, improving the stock of local public goods and enhancing economic welfare, but did not lead to any lasting changes in village institutions, local collective action capacity, social norms and attitudes, or the nature of *de facto* political power.

The results contradict the currently popular notion in foreign aid circles that community driven development (CDD) is an effective method to sustainably catalyze collective action or fundamentally alter local decision-making processes. The establishment of local committees, development plans and bank accounts did not lead to permanent reductions in the fixed organizing costs of collective action, likely because communities did not adopt and apply the new structures and tools to communal endeavors beyond the immediate project sphere. Exposure to democratic project processes similarly did not make traditional elites more willing to seek out the views of others in making community decisions, nor were villages any better able to raise funds in response to a matching grant opportunity. While "good" institutions may be critical for successful economic development, our findings provide another piece of evidence that institutions and social norms are difficult to change. At the same time our results challenge the aid pessimist's view that external assistance cannot improve the lives of the poor in countries with weak institutions. While we should not be so naïve as to think that structural factors like social organization and institutions are easily transformed (Easterly 2006, Kremer and Miguel 2007), we find that well allocated external aid can have a positive impact on welfare. Indeed our results suggest that the comparative advantage of the World Bank and other donors may lie more in providing development "hardware," and less in instigating large-scale institutional and social change, at least not using current tools such as CDD.

Our results further suggest that participation requirements did not foster learning-by-doing or demonstration effects large enough to change attitudes, norms or behaviors towards marginalized groups taking on leadership roles in this context. Despite project requirements on the inclusion of women and youth in project related decision making, intensive facilitation designed to build these groups capacity to engage, and the experience of being in decision making roles throughout the project, nearly four years later we see that women and youths are no more likely to voice opinions about how the community should manage new public assets. Returning to the comparison between informal interventions focused on reshaping norms, like the program studied here, and changes to the rules of formal institutions, like female leadership quotas, the limited existing evidence suggests that the latter may be a more effective way to alter *de facto* power dynamics and social perceptions in a modest timeframe (Chattopadhyay and Duflo 2004; Beaman et al. 2009). Importantly, however, we cannot rule out that part of GoBifo's success in using grants to deliver hardware impacts was due to its emphasis on transparency and the inclusion of marginalized groups.

Turning to empirical methods, our research underscores the importance of registering a pre-analysis plan to prevent data mining and to generate appropriately sized statistical tests. In the absence of a pre-analysis plan, we show how misleading an undisciplined interpretation of the individual treatment effects can be by presenting two opposing and equally erroneous narratives based on our actual data. In addition, our structured community activities (SCAs) provide an example of participatory and contextually specific tools to measure institutional behaviors that are at the same time concrete and standardized. As opposed to relying on subjective survey data, lab experiments or hypothetical vignettes alone, we feel that using this kind of objective and realistic measures lends confidence to our results.

As our results concern one program in one country, our more general policy implications are clearly speculative. However, we can conclude with certainty that far more research is needed to identify the precise reforms and external interventions that can successfully reshape institutions to enhance collective action capacity while promoting accountability and inclusion. This study further emphasizes the importance of registering a pre-analysis plan and using objective measures of institutions to enhance the scientific credibility of such pursuits.

REFERENCES

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369-1401.

Acemoglu, Daron and James A. Robinson. 2008. "Persistence of Power, Elites, and Institutions." *American Economic Review,* 98(1): 267-293.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken and Julia Tobias. Forthcoming 2011. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association,* 103(484):1481-1495.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy*, 2(1): 1-30.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review,* 95(4): 1190-1213.

Bardhan, Pranab. 2002. "Decentralization of Governance and Development." *Journal of Economic Perspectives*, 16(4): 185-205.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics*, 124(4): 1497-1540.

Bjorkman, Martina and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics,* 124 (2): 735-769.

Bracken, Michael B. 2011. "Preregistration of Epidemiology Protocols: A Commentary in Support." *Epidemiology*, 22: 135-137.

Bruhn, Miriam, and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development field experiments."*American Economic Journal: Applied Economics*, 1(4): 200-232.

Card, David and Alan B. Krueger. 1995. "Time-Series Minimum-Wage Studies: A Meta-analysis." *American Economic Review*, 85(2):238-243.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica,* 72(5): 1409-1443.

Dal Bó, Pedro, Andrew Foster and Louis Putterman. 2010. "Institutions and Behavior:

Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100(5): 2205-2229.

Dongier, Philippe, Julie Van Domelen, Elinor Ostrom, Andrea Rizvi, Wendy Wakeman, Anthony Bebbington, Sabina Alkire, Talib Esmail and Margaret Polski. 2003. "Chapter 9: Community-Driven Development." In *The Poverty Reduction Strategy Sourcebook Volume 1*, 301-331. Washington DC: The World Bank.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Little.* New York: Penguin.

Engerman, Stanley L. and Kenneth L. Sokolof. 1997. "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States." In *How Latin America Fell Behind*. Stanford: Stanford University Press.

Fearon, James, Macartan Humphreys and Jeremy M. Weinstein. 2009. "Development Assistance, Institution Building, and Social Cohesion after Civil War: Evidence from a Field Experiment in Liberia." Center for Global Development Working Paper 194.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. 2011. "The Oregon Health Insurance Experiment: Evidence from the First Year." MIT Working Paper, http://econ-www.mit.edu/files/6796.

GoBifo Project. 2006. "GoBifo Overall Budget." Project mimeograph.

GoBifo Project. 2007. "Operations Manual: Version 2 June 2007." Project mimeograph.

GoBifo Project. 2008. "Results towards Goals and Objectives." Project mimeograph.

Gugerty, Mary Kay and Michael Kremer. 2008. "Outside Funding and the Dynamics of Participation in Community Associations." *American Journal of Political Science,* 52(3):585-602.

Horton, R. **and R. Smith.** 1999. "Time to register randomized trials." *British Medical Journal*, 319:865.

Kling, Jeffrey R. and Jeffrey B. Liebman. 2004. "Experimental Analysis of Neighborhood Effects on Youth." Princeton University Manuscript.

Keen, David. 2003. "Greedy Elites, Dwindling Resources, Alienated Youths: The Anatomy of Protracted Violence in Sierra Leone." *International Politics and Society,*2: 67-94.

Kremer, Michael and Edward Miguel. 2007."The Illusion of Sustainability." *Quarterly Journal of Economics*, 122(3): 1007-1065.

Labonne, Julien and Robert Chase. 2008. "Do Community-Driven Development Projects Enhance Social Capital? Evidence from the Philippines." The World Bank Policy Research Working Paper 4678.

Lancet Editors. 2010. "Should protocols for observational studies be registered?" *Lancet:* 357–348.

Leamer, Edward E. 1974. "False Models and Post-Data Model Construction." *Journal of the American Statistical Association*, 69(345):122-131.

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review,* 73(1): 31-43.

Loder E, Groves T, MacCauley D. 2010. "Registration of observational studies: the next step toward research transparency."*British Medical Journal,* 340:375–376.

Mamdani, Mahmood. 1996. *Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism*. Princeton, NJ: Princeton University Press.

Mansuri, Ghazala and Vijaynedra Rao. 2004. “Community-Based and –Driven Development: A Critical Review.” *The World Bank Research Observer,* 19(1): 1-39.

Mansuri, Ghazala and Vijayendra Rao. Forthcoming. *Localizing Development: Does Participation Work?* Washington, DC: The World Bank.

Oates, Wallace E. 1999. “An Essay on Fiscal Federalism.”*Journal of Economic Literature*, 37(3):1120-1149.

O’Brien, Peter C. 1984. “Procedures for Comparing Samples with Multiple Endpoints.” *Biometrics,* 40(4):1079-1087.

Olken, Benjamin A. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy,* 115(2): 200-249.

Olken, Benjamin A. 2009. “Corruption Perceptions vs. Corruption Reality.” *Journal of Public Economics,* 93(7-8): 950-964.

Olken, Benjamin A. 2010. “Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia.” *American Political Science Review*, 104(2):243-267.

Olken, Benjamin A., Junko Onishi, and Susan Wong. 2010. “Indonesia's PNPM Generasi Program: Interim Impact Evaluation Report.” Jakarta: The World Bank.

Olson, Mancur. 1982. *The Rise and Decline of Nations.* New Haven: Yale University Press.

Pande, Rohini. 2003. “Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India.” *American Economic Review*, 93(4):1132-1151.

Rasmussen, Ole Dahl, Nikolaj Malchow-Møller and Thomas Barnebeck Andersen. 2011. “Walking the talk: the need for a trial registry for development interventions.” Manuscript.

Reno, William. 1995., *Corruption and State Politics in Sierra Leone.* Cambridge and New York: Cambridge University Press.

Richards, Paul. 1996. *Fighting for the Rainforest: War, Youth and Resources in Sierra Leone*. London: James Currey & Portsmouth, NH: Heinemann for the International African Institute.

Richards, Paul, Khadija Bah and James Vincent. 2004. “The Social Assessment Study: Community-driven Development and Social Capital in Post-war Sierra Leone.” Unpublished manuscript.

Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin Press.

Schaner, Simone. 2010. “Intrahousehold Preference Heterogeneity, Commitment and Strategic Savings: Theory and Evidence from Kenya.” MIT Working Paper, http://econ-www.mit.edu/files/6221.

Simes, R.J. 1986. “Publication bias: The case for an international registry of clinical trials”, *Journal of Clinical Oncology*, 4:1529-1541.

Voss, John. 2008. “Impact Evaluation of the Second Phase of the Kecamatan Development Program in Indonesia.” The World Bank, Jakarta Research Paper.

United Nations. 2004. *Human Development Report 2004*. New York: United Nations Development Program, Oxford University Press.

World Bank. 2004. “Japan Social Development Fund Grant Proposal: Capacity Development to Strengthen Social Capital in Sierra Leone.” Project mimeograph.

World Bank. 2007. *World Development Report 2008: Agriculture for Development* Washington, DC: The World Bank.

Table 1: Baseline (2005) Comparison between Treatment and Control Communities

	Baseline mean for controls	T-C difference at baseline	N
	(1)	(2)	(3)
Panel A: Community Characteristics			
Total households per community	46.76	0.30	236
		(3.67)	
Distance to nearest motorable road in miles	2.99	-0.32	236
		(0.36)	
Index of war exposure (range 0 to 1)	0.68	-0.01	236
		(0.02)	
Historical legacy of domestic slavery (range 0 to 1)	0.36	0.03	236
		(0.06)	
Average respondent years of education	1.65	0.11	235
		(0.13)	
Panel B: Selected Outcomes from "Hardware" Family A			
Proportion of communities with a Village development committee (VDC)	0.55	0.06	232
		(0.06)	
Proportion visited by Ward Development Committee (WDC) member in past ye	0.15	-0.01	228
		(0.05)	
Proportion of communities with a functional grain drying floor	0.23	0.05	231
		(0.05)	
Proportion of communities with a functional primary school	0.41	0.08	230
		(0.06)	
Average household asset score	-0.06	0.11	235
		(0.08)	
Proportion of communities with any petty traders	0.54	-0.01	226
		(0.06)	
Panel C: Selected Outcomes from "Software" Family B			
Respondent agrees that chiefdom officials can be trusted	0.66	-0.01	235
		(0.02)	
Respondent agrees that Local Councillors can be trusted	0.61	0.00	235
		(0.02)	
Respondent is a member of credit / savings group	0.25	-0.03	235
		(0.02)	
Among males who attended a community meeting, respondent spoke publicly	0.59	-0.02	235
		(0.04)	
Among females who attended a community meeting, respondent spoke publicly	0.29	0.03	229
		(0.04)	
Respondent claimed to have voted in last local elections	0.85	-0.01	235
		(0.02)	

Notes on table: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors; iii) the T-C difference is the pre-program "treatment effect" run on the baseline data aggregated to the village-level mean, using a minimal specification that includes only fixed effects for the district council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road); iv) regressions for the two balancing variables in rows 1 and 2 exclude the outcome from the set of controls; and v) see Appendix F for the T-C difference for all 94 outcomes collected in the baseline survey.

Table 2: GoBifo Treatment Effects by Research Hypothesis

Hypotheses by family	GoBifo Mean Effect (std. error)
Family A: Development Infrastructure or "Hardware" Effects	
Mean Effect for Family A (Hypotheses 1, 2 and 3; 37 total outcomes)	**0.352****
	(0.030)
H1: GoBifo creates functional development committees (7 outcomes)	0.687**
	(0.062)
H2: GoBifo increases the quality and quantity of local public services infrastructure (16 outcomes)	0.164**
	(0.040)
H3: GoBifo improves general economic welfare (14 outcomes)	0.399**
	(0.047)
Family B: Institutional and Social Change or "Software" Effects	
Mean Effect for Family B (Hypotheses 4, 5, 6, 7, 8, 9, 10, 11 and 12; 146 total outcomes)	**0.029**
	(0.019)
H4: GoBifo increases collective action and contributions to local public goods (15 outcomes)	0.041
	(0.042)
H5: GoBifo enhances inclusion and participation in community decisions, especially for vulnerable groups (43 outcomes)	0.001
	(0.031)
H6: GoBifo changes local systems of authority (25 outcomes)	0.048
	(0.036)
H7: GoBifo enhances trust (11 outcomes)	0.042
	(0.064)
H8: GoBifo builds groups and networks (12 outcomes)	0.033
	(0.044)
H9: GoBifo increases access to information about local governance (19 outcomes)	0.003
	(0.039)
H10: GoBifo increases participation in local governance (15 outcomes)	0.114**
	(0.047)
H11: GoBifo reduces crime and conflict (8 outcomes)	0.028
	(0.054)
H12: GoBifo fosters more liberal political and social attitudes (9 outcomes)	0.034
	(0.041)

Notes on table: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors clustered by village for panel data; iii) includes fixed effects for the district council wards (the unit of stratification) and the following control variables: total households per community, distance to nearest motorable road, index of war exposure, and index of history of domestic slavery; iv) these mean effect estimates are limited to the full sample set of outcomes that excludes all conditional outcomes (i.e. those that depend on the state of another variable--for example, quality of infrastructure depends on the existence of the infrastructure); and v) for the complete list of all full sample and conditional variables under each hypothesis--including the exact wording of survey questions and treatment effect estimates for each distinct outcome measure--see Appendix G.

Table 3: Family A: Illustrative Treatment Effects

Outcome variable	Mean in Controls	Treatment Effect	Standard Error	N
	(1)	(2)	(3)	(4)
Panel A: Hypothesis 1 - Project Implementation				
Village development committee	0.458	0.341**	(0.077)	467
Visit by WDC member	0.212	0.156*	(0.070)	462
Village development plan	0.617	0.296**	(0.048)	221
Community bank account	0.081	0.706**	(0.045)	226
A local politician was involved in managing the infrastructure:				
Primary School	0.415	0.181**	(0.055)	138
Grain drying floor	0.243	0.140*	(0.061)	115
Latrine	0.219	0.155**	(0.040)	169
Traditional midwife post	0.399	0.002	(0.106)	70
Community center	0.251	0.244**	(0.053)	95
Panel B: Hypothesis 2 - Local Public Goods				
Functional primary school in the community	0.462	-0.007	(0.050)	464
Functional grain drying floor in the community	0.237	0.104	(0.066)	459
Functional traditional midwife post in the community	0.079	0.175**	(0.035)	235
Functional latrine in the community	0.462	0.210**	(0.059)	234
Functional community center in the community	0.212	0.241**	(0.063)	469
Community took a proposal to an NGO or donor for funding	0.292	-0.156+	(0.081)	460
Supervisor's physical assessment of construction quality (index from 0 to 1):				
Primary School	0.583	0.116*	(0.055)	123
Grain drying floor	0.375	0.142+	(0.076)	101
Latrine	0.270	0.177**	(0.055)	154
Infrastructure funding/supplies provided at least in part by the community:				
Primary School	0.554	-0.007	(0.112)	242
Grain drying floor	0.105	0.086	(0.124)	184
Latrine	0.761	-0.197*	(0.093)	126
Panel C: Hypothesis 3 - Economic Welfare				
Total petty traders in village	2.432	0.719*	(0.344)	225
Total goods on sale of 10	4.449	0.560*	(0.240)	236
Household asset score	-0.170	0.212*	(0.090)	471
Attended trade skills training	0.061	0.119**	(0.018)	235
Income from top 3 cash earning sources (in 1,000 Leones)	746.94	-21.773	(73.069)	236

Notes on table: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) treatment effects are estimated on panel data where available; iii) robust standard errors in parentheses, clustered by village for panel specifications; iv) includes fixed effects for the district council wards (the unit of stratification) and the following control variables: total households per community, distance to nearest motorable road, index of war exposure, and index of history of domestic slavery; and v) where indicated, outcomes are conditional on the existence of functional infrastructure in the community.

Table 4: Structured Community Activities (SCAs): Illustrative Treatment Effects

Structured Community Activity (SCA) Outcome:	Mean for Controls	Treatment Effect	Standard Error
	(1)	(2)	(3)
Panel A. Collective Action and the Building Materials Vouchers			
GoBifo Mean Effect for SCA #1 (13 outcomes in total)	**0.00**	**-0.06**	**(0.05)**
Proportion of communities that redeemed vouchers at building materials store	0.54	-0.01	(0.06)
Average number of vouchers redeemed at the store (out of six)	2.95	0.11	(0.35)
Proportion of communities that held a meeting to discuss the vouchers	0.98	-0.05*	(0.02)
Panel B. Participation in the Gift Choice Deliberation			
GoBifo Mean Effect for SCA #2 (32 outcomes in total)	**0.00**	**0.01**	**(0.04)**
Duration of gift choice deliberation (in minutes)	9.36	1.60	(1.13)
Total adults in attendance at gift choice meeting	54.51	3.50	(3.20)
Total women in attendance at gift choice meeting	24.99	1.99	(1.68)
Total youths (approximately 18-35 years) in attendance at gift choice meeting	23.57	2.10	(1.38)
Total number of public speakers during the deliberation	6.04	0.24	(0.40)
Total number of women who spoke publicly during the deliberation	1.88	-0.19	(0.22)
Total number of youths (approximately 18-35 years) who spoke publicly	2.14	0.23	(0.24)
Proportion of communities that held a vote during the deliberation	0.10	0.07	(0.04)
Panel C. Community Use of the Tarpaulin			
GoBifo Mean Effect for SCA #3 (18 outcomes in total)	**0.00**	**-0.03**	**(0.05)**
Proportion of communities that held a meeting to discuss use of the tarp	0.98	-0.03	(0.02)
Proportion of communities that stored the tarp in a public place	0.06	0.06	(0.04)
Proportion of communities that had used the tarp (5 months after receipt)	0.90	-0.08+	(0.04)
Given tarp used, proportion of communities using the tarp in a public way	0.86	0.02	(0.05)

Notes on table: i) significance levels denoted by + $p < 0.10$, * $p < 0.05$ and ** $p < 0.01$; ii) robust standard errors; iii) treatment effects estimated on follow-up data; (iv) includes fixed effects for the district council wards (the unit of stratification) and the following control variables: total households per community, distance to nearest motorable road, index of war exposure, and index of history of domestic slavery; and v) sample size varies between 225-236 for all outcomes save the last, which is conditional on having used the tarp and has N = 161.

Table 5: Erroneous Interpretations under "Cherry Picking"

Survey question	Mean for controls	Treatment effect	Standard error	N	Hypo
	(1)	(2)	(3)	(4)	(5)
Panel A: Institutions "Deteriorated"					
Attended meeting to decide what to do with the tarp	0.812	-0.037+	(0.021)	236	H5
Everybody had equal say in deciding how to use the tarp	0.509	-0.106+	(0.058)	232	H5
Correctly able to name what the tarp was used for	0.589	-0.08+	(0.048)	236	H9
Community used the tarp (verified by physical assessment)	0.897	-0.079+	(0.044)	233	H4
Community can show research team the tarp	0.836	-0.116*	(0.051)	232	H5
Respondent would like to be a member of the VDC	0.361	-0.043*	(0.021)	236	H10
Current (or acting) village chief/Headman is younger than 35	0.044	-0.038+	(0.023)	229	H12
Respondent voted in the local government election (2008)	0.851	-0.036*	(0.016)	236	H10
Panel B: Institutions "Improved"					
Community teachers have been trained	0.471	0.122+	(0.066)	173	H4
Respondent is a member of a women's group	0.235	0.060**	(0.021)	236	H8
Someone took minutes at the most recent community meeting	0.295	0.140*	(0.063)	227	H5
Building materials stored in a public place when not in use	0.128	0.246*	(0.098)	84	H5
Chiefdom official did not have the most influence over tarpaulin use	0.543	0.058*	(0.029)	236	H6
Respondent agrees with "Responsible young people can be good leaders" and not "Only older people are mature enough to be leaders"	0.762	0.038*	(0.017)	236	H6, H12
Correctly able to name the Section Chief for this section	0.533	0.053+	(0.032)	234	H9
Correctly able to name the year of the next general elections	0.192	0.038*	(0.018)	236	H9

Notes on table: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors; iii) treatment effects estimated on follow-up data; and iv) includes fixed effects for the disctrict council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road) as controls.