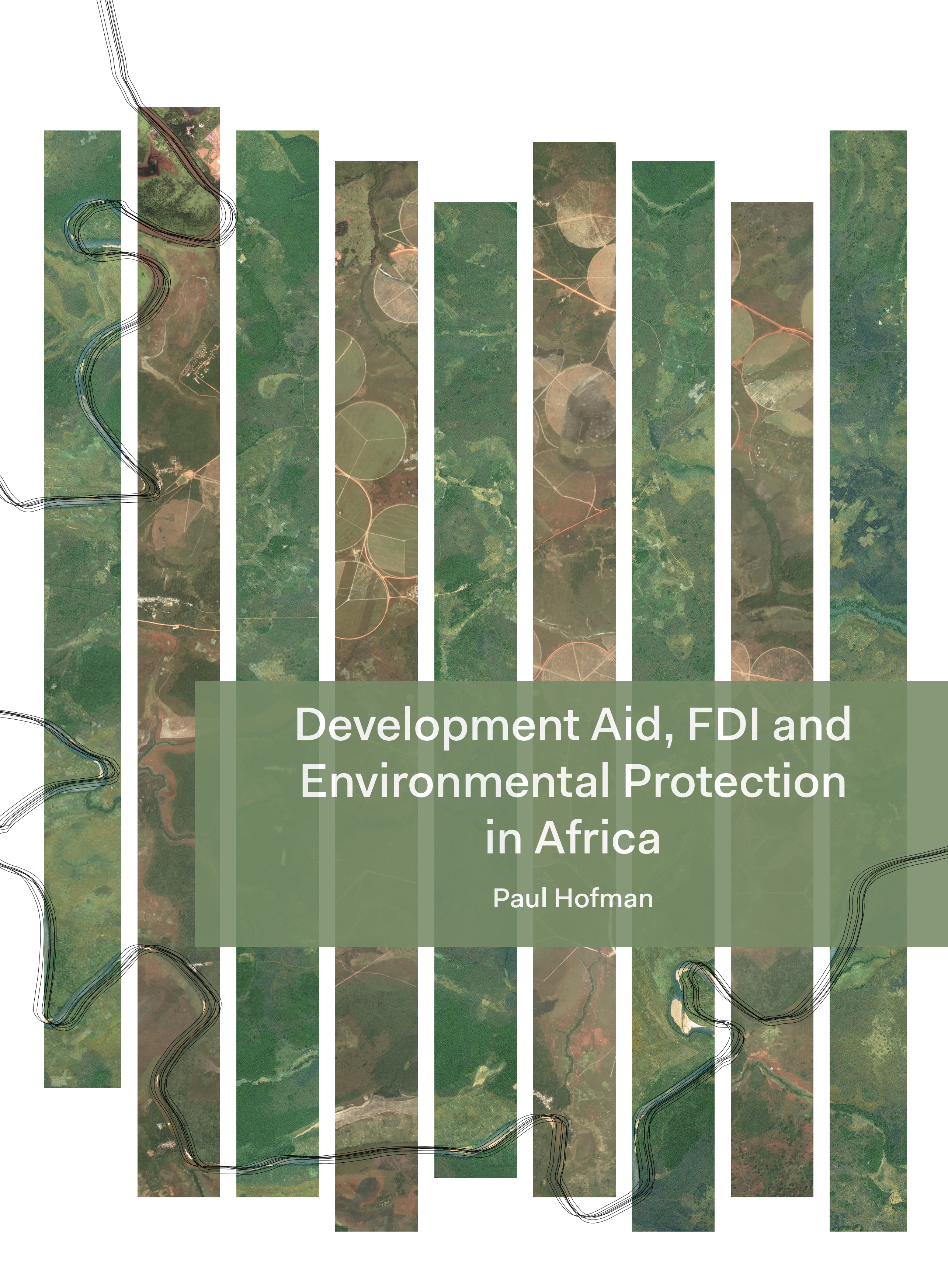


Development Aid, FDI and Environmental Protection in Africa

Paul Hofman

Propositions

1. Local factor markets should be the starting point for designing development interventions.
   (this thesis)
2. Using network position to optimize diffusion of resources increases inequality.
   (this thesis)
3. For more representative aggregation, hexagons are better than squares to tesselate an area.
4. Readability is more important than brevity when writing reproducible programming code.
5. Business for development is more for business than for development.
6. Research is a dance of independence with multiple partners.

Propositions belonging to the thesis entitled:

“Development Aid, FDI and Environmental Protection in Africa”.

Paul Hofman

Wageningen, 19 October 2020

Development Aid, FDI and Environmental Protection in Africa

Paul Hofman

Thesis committee

Promotor
Prof. Dr E.H. Bulte
Professor of Development Economics
Wageningen University & Research

Co-promotor
Dr M.J. Voors
Associate Professor, Development Economics Group
Wageningen University & Research

Other members
Prof. Dr Ewout Frankema, Wageningen University & Research
Dr Goylette F. Chami, University of Oxford, United Kingdom
Prof. Dr Jann Lay, GIGA Institute, Hamburg, Germany
Prof. Dr Jörg Peters, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, Germany

This research was conducted under the auspices of the Wageningen School of Social Science (WASS)

Development Aid, FDI and Environmental Protection in Africa

Paul Hofman

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 19 October 2020
at 4.00 p.m. in the Aula.

Paul Hofman
Development Aid, FDI and Environmental Protection in Africa
193 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summary in English

ISBN: 978-94-6395-541-6
DOI: https://doi.org/10.18174/531119

Dedicated to my grandmother Riet

Contents

Chapter 1

Introduction

1.1 Problem Statement

Sub-Saharan countries still suffer from very low incomes and high poverty rates, despite decades of development assistance. The traditional approach to development aid is Official Development Aid (ODA), which directly finances programs aimed to improve development outcomes. Another approach that is gaining popularity is to create development through Foreign Direct Investment (FDI). FDI can create local jobs, provide improved infrastructure and kickstart local productivity through technology spillovers. There is also an increase in protected areas: these aim to improve incomes around these protected areas through development programs or the development of alternative income sources, while contributing to environmental protection at the same time. These approaches have seen substantial shifts over the past three decades: Figure 1.1 plots the relative importance over time. ODA has halved from 6 to 3% of GNI.[1] FDI has increased from 0 to 2% of GDP. The percentage landmass of Sub-Saharan Africa that is environmentally protected has increased from 11 to 16% in 2014 (and is likely higher now). Can FDI and protected areas fill the gap in development aid that ODA has left? Can ODA maximize its impact by improving efficiency, and how does that affect local inequality? This thesis examines how ODA, FDI and protected areas can contribute to development.

If there are fewer resources available for development aid, this can be overcome by improving its efficiency. One way to do this is by using local horizontal institutions: social networks. Social networks can (a.o.) be used to optimize the diffusion of new technologies, a crucial method that can lead to development. This is done by using a network metric to select the 'optimal' spreader that causes maximum adoption and/or diffusion. Only a few papers have used this approach, with mixed success (Banerjee et al., 2013; Beaman et al., 2018; Emerick et al., 2016; Kim et al., 2015). However, these approaches could also exacerbate local inequalities when resources circulate mostly amongst these 'optimal' spreaders. Examining how these network approaches affect the distribution of resources is therefore an important question. Another application of social networks on development is to examine its relation

[1]While the GNI of African countries has also increased over the past 30 years, the drop in ODA has been relatively larger.

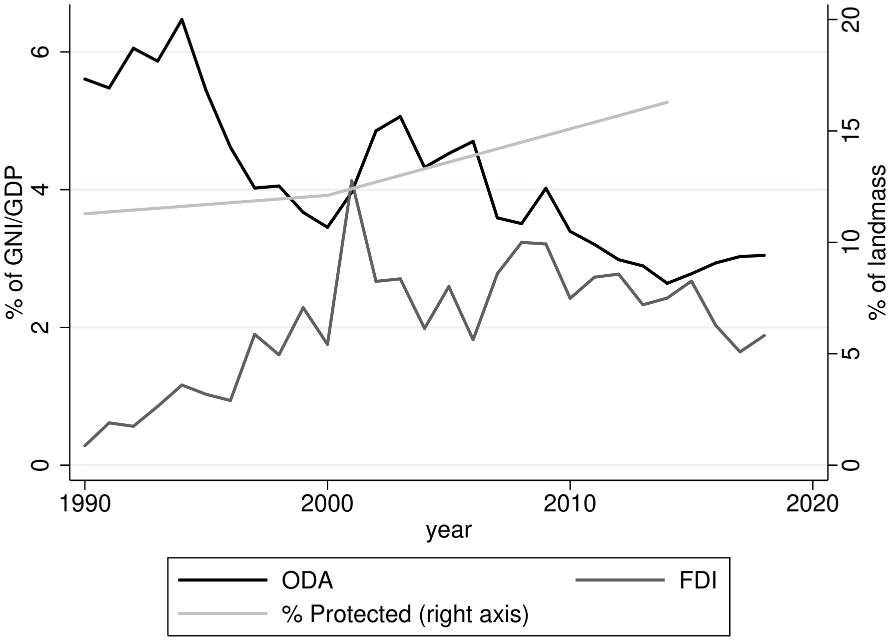


Figure 1.1 – Development trends in Sub-Saharan Africa

ODA is Official Development aid (% of GNI), FDI is Foreign Direct Investment (% of GDP). % protected is the total landmass that is in some form protected. Sources: World Bank (2020); Development Indicators Unit (2016)

with interpersonal trust. Trusting behavior is associated with improved market performance. Individuals might use their relations within the social network as collateral to complete trust-based interactions. Understanding the social network can then be used to predict trusting behavior. There is very little evidence on the network determinants of trust, especially in a developing country context. This thesis examines the distributional effects of network-based technology diffusion, and the relation between social networks and trust.

A more recent approach to development is through Foreign Direct Investment (FDI). There is an interest in donor countries in spurring development through private companies. For example, the Dutch government's goal for development is

'Aid & Trade' (Bitzer et al., 2017). Furthermore, Sub-Saharan Africa has seen improvements in the ease of doing business (World Bank, 2020). This has allowed more foreign companies to start doing business in Africa. The 2007-8 boom in food prices created an especially large interest in agricultural investments (sometimes financed from development budgets). This wave of (agricultural) FDI could improve development through the creation of well-paying jobs, improvements in local infrastructure and technology spillovers. Many action groups instead point to negative effects: land appropriation, higher inequality and social tensions. There is very little micro-level evidence on the impact of these kinds of investments. This thesis examines two cases of how local communities are affected through agricultural FDI, once with a focus on overall welfare and once examining the effect of technology spillovers.

Protected areas are another approach that also aims to improve development outcomes. The earth is currently undergoing the sixth mass extinction event, described as 'biological annihilation' of nearly half the world's species (Ceballos et al., 2017). Climate change is also affecting millions of lives through increased temperatures and extreme weather events. One method to jointly fix these issues is through the creation of protected areas. In many African countries these protected areas are coupled with management programs with the explicit goal to improve development outcomes. The goal of these management programs is twofold: increase local buy-in for the protection of the area, and reduce the need for the local population to engage in profitable but environmentally devastating mining, logging and slash-and-burn agriculture. The effectiveness of these programs on conservation and development has not been extensively researched (Ferraro and Pattanayak, 2006). This thesis will examine the environmental *and* developmental impacts of one such protected area.

While these three approaches to development are distinct and examined in separate self-contained research chapters, some insights transcend the chapters. Several chapters provide evidence that inequality is increased as a result of (development) interventions. This appears to be mainly caused by those with more power being able to steer resources to themselves. This means that (development) interventions do not reduce local inequality, but might instead increase it. One mechanism this thesis examines is through social networks: if resources flow following network lines those at the center will have easier access. How exactly the final distribution of resources is affected by the social network is explored through two chapters. Finally, in several chapters the main mechanism works through the local labor market. In these chapters the labor markets are characterized by severe shortages. The response of the labor market is crucial to understanding the transmission of the impacts. These insights are further worked out in the final chapter.

1.2 Literature

This section introduces the three kinds of literatures this thesis will be speaking to and identifies the research gaps this thesis will address.

1.2.1 Social Networks, Technology Adoption and Distribution

The adoption of new technologies can foster development. Widespread adoption of new technologies has been shown to improve political and economic development (Bailard, 2009; Carter et al., 2014; Cilliers et al., 2013; Grossman et al., 2014). However, willingness to adopt is low among generally risk-averse African farmers. A literature is emerging that uses existing social networks to improve diffusion and adoption (Banerjee et al., 2013; Beaman et al., 2018; Emerick et al., 2016; Kim et al., 2015; Chami et al., 2017). This works through calculating network-based characteristics that are then used to select the optimal farmer. One reason why this might work is that adoption is dependent on repeated exposure to the new technology. If this is the case, selecting farmers who have many social network connections and will therefore expose more people to the technology should increase adoption.

However, there is a 'dark side' to this approach. Social networks are a reflection of existing social and economic cleavages. Those on the fringes of the network are also likely to be excluded from profitable economic exchanges (for example because of lower interpersonal trust, see the next paragraph). Therefore, programs that exploit social networks should not only be concerned with efficiency, but also equity. As training/demonstrating a new technology is costly it will not cascade infinitely throughout a community, risking it reaching only the most connected. While there is an emerging literature on which approach helps optimize diffusion, these papers do not focus on the *distribution* of the new technology. It might be that these network approaches that rely on existing social relations perpetuate or increase inequality.

If training is costly, why do individuals engage in it in the first place? One explanation that moves beyond simple altruism is the theory of 'social collateral', developed by Karlan et al. (2009). It posits that connections in a social network have value because they can be used as social collateral. If A and B want to trade but either A or B can betray the other the trade will not take place. However, A and B can use their social network connections as collateral. If the value of their relationship, and the value of relationships that they share are higher than the costs of betraying the other the trade can take place. The assumption is that if A betrays B he will lose the relationship with B, and any relationships with shared relationships. If this is true, there is a value in developing and maintaining relationships, for example by training others in new technologies. There are very few papers that test the validity of this theory. This is mainly caused by social network data being very costly to collect, especially in a developing country context where it must be done fully through surveys.

1.2.2 Large-Scale Land Investments

Since the 2007-8 world food price crisis there has been a marked increase in Large-Scale Land Investments (Koning and van Ittersum, 2009; Arezki et al., 2013). These are investments (generally acquisitions) by foreign actors in agricultural land, to create large farms. This has been the subject of many media and academic articles (Chapter 4 presents 13 policy reports/academic papers surrounding one single investment, and this list is unlikely to be exhaustive). These investments (be it acquisitions or contract farming schemes or hybrid forms) mobilize land and labor in novel ways: through long-term labor contracts, exclusive supply contracts or others. These programs are often partially financed through development budgets, with the expectation that these will contribute to development (Engström and Hajdu, 2019; Kindornay and Reilly-King, 2013).

There is little micro-level evidence on the impacts of these investments. The papers that exist correct for selection bias using matching algorithms, generally on post-intervention data, requiring strong identifying assumptions (Herrmann and Grote, 2015; Herrmann, 2017; Bottazzi et al., 2018). Theoretical work by Dessy et al. (2012); Kleemann and Thiele (2015) have explored how the local economy is affected by these investments. Crucial in determining overall welfare is the response of the local land and labor markets. If they do not respond because labor and land are plentiful the local economy should not be affected. This is unlikely to be the case.

Another important mechanism is through productive spillovers. New techniques employed on the agricultural investment can spill over to local production (mainly if crops and production methods are similar) (Kleemann and Thiele, 2015; Crespo and Fontoura, 2007; Liu, 2008). The effect of FDI on development has been extensively researched at the macro level (Makiela and Ouattara, 2018; Li and Liu, 2005; Nwaogu and Ryan, 2015), but micro-level evidence is rare (Ali et al., 2019; Lay et al., 2018).

1.2.3 Conservation and Development

Conservation of forested areas plays an important role in combating climate change and reducing biodiversity loss. Deforestation is a serious problem in developing countries (Hansen et al., 2013). For example, 25% of Sierra Leone's landmass has seen deforestation in the last 19 years. One approach to overcome this are REDD+ programs (REDD stands for Reducing Emissions from Deforestation and forest Degradation). These programs focus on creating alternative, forest-friendly activities, or directly provide payments for not deforesting. In this way they try to protect the natural environment, while improving development outcomes at the same time.

It is an open question whether conservation and development can go hand in hand (Ferraro and Pattanayak, 2006). Existing evidence on the effectiveness of protected areas and REDD+ programs shows that these are effective in reducing deforestation (Jayachandran et al., 2017; Simonet et al., 2019; Roopsind et al., 2019). However, there is very little evidence that also examines how local livelihoods are affected. Additionally, there is also very little evidence on whether REDD+ affects buffer zones. These areas just around the protected area prevent the creation of ‘biodiversity islands’, which have reduced biodiversity potential. Examining all three of these aspects (deforestation in the protected area, its buffer zone and local development outcomes) is crucial to understand the full impact of these protected areas.

1.3 Objective and Research Questions

The overall objective of this thesis is to examine three approaches to development and to draw some overall lessons from these examinations. Based on the research gaps identified in the previous section, this thesis will try to answer the following five research questions:

1. *What are the distributional effects of network-based technology diffusion?*
   (Chapter 2)
2. *Can social network links explain interpersonal trust?*
   (Chapter 3)
3. *What is the impact of large-scale land investments on the local economy and welfare?*
   (Chapter 4)
4. *What are the productive spillovers of large-scale land investments?*
   (Chapter 5)
5. *How do protected areas affect deforestation, deforestation in buffer zones and development outcomes of the local population?*
   (Chapter 6)

1.4 Design

To answer these research questions this thesis employs a diverse set of methods, most of which have the explicit aim of establishing causality (under a set of weaker and stronger assumptions). Each of these methods is discussed in turn. See Abadie and Cattaneo (2018) for a more substantial review of these methods.

1.4.1 Randomized Control Trials

Randomized Control Trials (RCTs) are considered the 'gold standard' for causal inference. It has been used for decades in drug testing and has seen spectacular growth within development economics (Banerjee and Duflo, 2009; Baranov et al., 2020; Romero et al., 2020). 2019's Nobel prize was awarded to three development economists that popularized RCTs. RCTs work by randomly assigning some kind of treatment to a group (the treatment group), while the other group does not receive the treatment (control group). As the groups are randomly decided they are *by definition* similar in levels and trends of observable and unobservable characteristics. This allows any difference between the two groups after administering the treatment to be attributed to the treatment. This is subject to one main assumption: the Stable Unit Treatment Variation Assumption (SUTVA). This means, simply said, that the *only* difference between the two groups must be the treatment assignment. The most common SUTVA violation is through spillovers: the control group is affected by the treatment in the treatment group, for example through changes in the local markets. However, SUTVA can also be violated by differential attrition between the two groups, or through characteristics affecting take-up (if relevant). By carefully monitoring the treatment implementation a researcher can make an argument that SUTVA is not violated.

This thesis uses this method in Chapter 2 to asses the causal impact of two different methods of technology diffusion. The researchers had full control over the treatment, which allowed us to choose the strongest causal method. This control also allowed careful tracking of implementation to discover violations of the SUTVA.

1.4.2 Natural Experiments

Natural experiments are similar to RCTs, but in this case the treatment is not assigned by chance, but by 'Nature'. That is, through circumstances a group was assigned a treatment while another group was left out that might have received that treatment as well. Besides SUTVA, the biggest assumption is that this 'natural' division creates extremely similar groups, as if randomly selected. The strength of the method depends on the strength of this argument, which can be supported by data. Examples are Deschenes et al. (2020); Banerjee and Maharaj (2020).

This thesis uses this method for one paper, where a company had planned to work in a large set of villages but ended up dropping a portion of these for reasons not related to local farm-life. For these types of large-scale projects it is generally not possible to randomize, leaving us with weaker methods. Especially when combined with other methods, this can still lead to causal inference.

1.4.3 Difference-in-Differences

A very popular method in development economics is Difference-in-Differences (DiD, sometimes called double differences as well) (Abadie, 2005; Kearney and Levine, 2015; Berry et al., 2020). It uses repeated observations over time, for two groups, one of which is treated and one is not. It works by comparing the differences over time with the differences between the two groups. This approach corrects for all observed and *unobserved* time-invariant characteristics, which can very closely approach causal effects. The crucial assumption is that of parallel trends: that without the treatment both groups would have *trended* similarly. This cannot be proven, but a common approach is to examine pre-treatment trends. If trends were similar before the treatment, they would likely have been similar had there been no treatment. This requires the availability of pre-treatment data, which is uncommon. The approach's popularity stems from its relative ease of implementation: if it is known that a project will be implemented, all that is required is to find a suitable control group and then collect data before and after the implementation. This allows it to be used for projects where stronger causal approaches are not feasible. However, the Difference-in-Difference approach is especially powerful when combined with other methods. For example, combining it with an RCT ensures that the parallel trends assumption holds.

This thesis uses the DiD approach for all its analyses except for Chapter 3. The approach's versatility makes it a great fit for studies that do not lend themselves to stronger causal approaches. And when combined with these stronger methods the strength of the causal claim can be increased.

1.4.4 Control-variables approach

Finally, this thesis also uses the 'control-variables approach'. This approach examines the correlation between two variables of interest, controlling for as many other characteristics (variables) as possible. Recent examples are Michaelsen and Salardi (2020); Leonard et al. (2020). To make causal claims this approach assumes that all relevant covariates can be accurately measured and accounted for. In practice this is very difficult, and the approach is generally used to discover correlations. However, often these relations can still be of interest and can give input into the development of future research. This thesis uses this approach in Chapter 3, where it examines correlates of trust and social network characteristics. Since these are innate characteristics finding any kind of exogenous variation would be almost impossible, leaving us with this approach.

1.4.5 Data

Survey Data

The most important source of data in this thesis is survey data collected in face-to-face interviews. In all chapters (except Chapter 4, and parts of Chapter 6), this data collection was implemented by this thesis's author. Several measures were taken to improve the quality of the data that was collected. Data was collected using tablet computers using software from Open Data Kit (ODK). This gives several advantages over data collection using paper. First, it takes out the step of data *entry*, which is where the data is typed into a computer. This repetitive, uninteresting task is very susceptible to user error, for example typing a wrong value. There is also a risk of data on the paper survey not being legible due to poor handwriting or the paper being destroyed. Second, digital data collection allows the creation of elaborate skip logic without burdening this on the interviewer, as questions are automatically added and dropped based on earlier questions. Third, not having to carry stacks of paper substantially eases the burden for interviewers who won't have to carry these (sometimes for extended periods of time).[2]

Additionally, several measures were used to improve the quality of measurement. An important metric is welfare, but this can be hard to measure in subsistence farmers (Meyer and Sullivan, 2003). This thesis deals with this by examining several dimensions: incomes, expenditures and physical assets. In some cases these measures are be augmented with satellite data to improve measurement. Another important element was ensuring informed consent by walking through an extended information sheet with the participant. This ensured that participants were willing to participate and did not provide bogus answers.

Geographic Information Systems

This thesis also regularly uses data stemming from Geographic Information Systems (GIS), through satellite imagery. Satellites that orbit the earth take regular photographs of the earth's surface. These photographs can be used to measure many variables of interest such as deforestation, crop success and urbanization. One important advantage of this approach is that the data is (generally) available for the entire world, and has very long time series (this thesis uses GIS data dating back to 2000). Furthermore, there is very little risk of Hawthorne/Observer effects (where respondents change their behavior when they realize they are being observed). This thesis uses GIS data to examine forest loss and crop success in two chapters. In both cases the long time series is also used to examine the parallel trends assumption of the DiD approach.

1.5 Outline

Chapter 2 uses an RCT to examine whether social networks can help optimize technology diffusion processes. This chapter emulates the ODA approach to development by providing new technology at no cost. It uses farmer field schools to diffuse the new technology. It uses a social network metric to select the initial

[2]It is also easier to collect GPS location data and photographs.

farmers and thus examines how this affects the final distribution of the technology.

Chapter 3 digs into this social network element even further by examining how social network characteristics are related to behavior in a trust game. Specifically, it provides several tests of Karlan's 'social collateral' theory (Karlan et al., 2009). This can help shed further insight into the distribution processes observed in Chapter 2.

Chapter 4 examines the impact of a large-scale land investment (or agricultural FDI). This approach has seen a boom in the last decade and is often financed through development aid. This chapter examines whether it achieves its goal of kickstarting local development.

Chapter 5 examines another case of an agricultural foreign direct investment but instead focuses on the productive spillovers. This is another important channel through which local welfare is affected by large-scale agricultural investments.

Chapter 6 examines how nature conservation can improve development outcomes by providing unconditional aid surrounding a national park. By affecting the profitability of local activities and local social norms it aims to reduce pressure on a national park. By looking at both the effect on conservation and local livelihoods it examines if conservation and development can go together.

Chapter 7 looks across the chapters to develop some overall insights. It explores the role of a strained labor market in agriculture, how networks matter for the *distribution* of impacts and how local inequality can be affected by development programs.

Chapter 2

How Social Network-Targeted Interventions Perpetuate Inequality: Evidence from a Field Experiment in the Congo

From epidemics to marketing campaigns, networks spread things to more people if the most central are involved. This finding has inspired pro-development interventions that give technology to the most central to disseminate. However, political and economic development is not *solely* concerned with 'how many'. What are the distributional consequences of selecting the most central as entry points? Given that centrality tends to correlate with political and economic privilege, does selecting them concentrate an intervention's benefits? In 40 DRC villages, we randomize whether initial recipients of fertilizer are the most or least central in their village's network and hold constant the amount of fertilizer introduced. We find significant distributional differences: targeting the central reduces the likelihood that the least well-off receive fertilizer. Furthermore, we find no evidence that villages are nonetheless better off: the increase in use, knowledge, and valuation of fertilizer at the village level is the same.

Publication status: Hofman, P., Larson, J.M., Van der Windt, P. and Voors, M., 2020. How Social Network-Targeted Interventions Perpetuate Inequality: Evidence from a Field Experiment in the Congo. Working Paper.

2.1 Introduction

New technology fosters development. Countless studies show that if new technology becomes widely used in an area, political and economic development outcomes improve there (Bailard, 2009; Carter et al., 2014; Cilliers et al., 2013; Grossman et al., 2014). To become widely used, the technology must be adopted by enough people, which means enough people must know about it, know how to use it, believe it is useful, and acquire it. Social networks facilitate this process, allowing the technology itself or information about it to spread from person to person throughout an area.

A growing literature seeks to relate the precise arrangement of links in a social network to the ultimate reach of goods and information (Banerjee et al., 2013; Centola and Macy, 2007; Conley and Udry, 2010; Emerick et al., 2016; Foster and Rosenzweig, 2010; Larson and Lewis, 2017). Often borrowing insights from epidemiology, this line of research focuses on diffusion, where the outcome of interest is the number of people who ultimately receive whatever is spreading through a network. These studies yield insights about which network positions to choose as entry points for interventions designed to disseminate information or goods through communities, typically finding that the most central, well-connected network positions are ideal initial recipients to target to ultimately reach the most people (Banerjee et al., 2013; Beaman et al., 2018; Emerick et al., 2016; Kim et al., 2015).

However, socially well-connected people tend also to be wealthier, more politically connected, and members of in-groups. Even if interventions targeting such people reach more people than those targeting the less well-connected, who is left out? And importantly, are the less well-connected ever reached? The question is especially important for interventions designed to spread technologies that drive development: if marginalized populations are never reached, then gains accrue to the already well-off. In short, the rich get richer.

We use a field experiment to directly test the distributional consequences of interventions that rely on social networks to disseminate goods and information. Specifically, we experimentally test whether the social network eigenvector centrality of the first receivers of new technology – individuals we refer to as ‘ambassadors’ – affects how that technology is ultimately spread, understood, valued, used, and, crucially, distributed within villages. We first precisely measure the social ties interconnecting households in 40 rural villages in the Democratic Republic of the Congo. These villages are all in the South Kivu region and are comprised of farmers. Subsequently, in 20 randomly selected villages, we identify the three most eigenvector central heads of household and select them to be ('central') ambassadors; in the other 20 villages, we select the three least eigenvector central heads of household to be ('isolate') ambassadors. In all 40 villages, the ambassadors are trained in the correct application of fertilizer and given a carefully packaged set of fertilizer to be distributed.

A core innovation of this study is in the way the fertilizer is packed for distribution. Each ambassador, both central and isolate, received a bundle containing a 1kg bag of fertilizer for him or herself to keep as well as three fertilizer 'kits'. Each kit contained three 1kg bags of fertilizer. Ambassadors were encouraged to keep the 1kg bag, distribute the three kits, and tell the kits' recipients that they could keep one 1kg bag and distribute the other two. This bundle made the spread of the new technology from an ambassador to villagers in one step, and from them to other villagers in a second step, convenient and traceable. Although our design included no incentives to comply with passing fertilizer along, 81% of ambassadors in fact complied (with equal compliance between central and isolate ambassadors). As the number of bags was limited per village we do not examine the effect on the number of people reached. Instead, we focus on distribution, adoption and attenuation.

We find, first, that by selecting the most socially central individuals as ambassadors, we were indeed selecting a privileged set of entry points. The central ambassadors were on average substantially wealthier, held more leadership positions, enjoyed greater political access, and were less likely to be a migrant than the isolate ambassadors. Second, we find that across all 40 villages the use and knowledge of fertilizer increased, but not villagers' willingness to pay for it. Furthermore, we find strong evidence that social networks are responsible for this increase, and document classic attenuation as the distance of spread increases from one to two to three degrees through the network.

Third, we find that inequality in entry positions does generate inequality in the recipients of the technology. In terms of reaching the most marginal, technology reached 40% fewer villagers in the bottom tercile of network centrality in villages with central ambassadors than in those with isolate ambassadors. Moreover, our unique data allow us to explore the transactions through the network that produced this result. We find that villagers with relatively high centrality and villagers with relatively low centrality give fertilizer to villagers with relatively high centrality. While we cannot definitively say why this behavior obtains, it may be based on a belief that central villagers are most likely to put the fertilizer and knowledge about it to better use, or because isolate receivers gift strategically to central villagers, perhaps to move ahead in the village (Sherry Jr, 1983;

Mauss, 2002; Adloff and Mau, 2006). Regardless, it shows that irrespective of who is initially targeted, the resources are likely to accrue to the already privileged.

In sum, our study suggests that the informal process of diffusion is not well suited to reach relatively isolated villagers. Interventions that use the most central as entry points reduce the chance that the more marginalized are reached. Of course, if the development-promoting technology has positive externalities, it is possible that having personal access to the technology is unimportant. It could be that so long as enough others in an area use the new technology, the entire area experiences increased prosperity. In fact, if the most central make more use of new technology, or better understand it, then reaching only them may be best for everyone.

Our study allows us to test this possibility since the bundled fertilizer design introduces the same quantity of fertilizer into the villages with central ambassadors and those with isolate ambassadors. We compare village-wide baseline and endline surveys to assess potential differences in village-level changes in outcomes including the extent of fertilizer use, knowledge, and willingness to pay. We find very small and statistically insignificant differences between villages based on the centrality of the initial recipient. This is true even though our design maximized the difference in the centrality of initial recipients. Selecting the most central as entry points (and having the technology disseminate to more central villagers) does not make the whole village better off than villages with the most isolate as entry points (and hence with more isolated villagers receiving the technology).

2.2 Research Context

This study is conducted in 40 rural villages in the Congolese province of South Kivu (Figure A.2.1 in the appendix).[1] Congo ranks at the bottom of the human development index (UNDP, 2016), the GDP per capita is low at 442US$, and over 85 percent of the population currently lives below the poverty line (World Bank, 2020). The region has been embroiled in violent conflict, spiking during the First and Second Congolese Wars (1996-1997 and 1998-2003). The latter was the deadliest war in modern African history (Coghlan et al., 2006). Hostilities remain up to this day.

[1] Villages were not randomly selected. Selection took place based on the following criteria: fewer than 100 inhabitants, road access, and proximity to larger villages where research assistants could spend the night.

The majority of the rural population are subsistence farmers. Farms are often small and fragmented. Due to conflict and underdevelopment agricultural yields remain stagnant, causing widespread malnutrition. Income from cash crops is low and most farmers struggle to improve their livelihoods. Most farmers have no access to important input markets for pesticides, fertilizer and (improved) seeds (Pypers et al., 2011). For example, just 3% of farmers report having used fertilizer previously in a recent survey conducted in South Kivu (Bulte et al., 2015). In addition to infrequent use, the average quantities applied are low. In 2014, on average, only 0.3 kilograms of fertilizer were applied per hectare in Congo, compared to 14kg for Sub-Saharan Africa as a whole and 166kg per hectare in Asia (FAO, 2015).

Villages in South Kivu, like in the rest of Congo, are small, typically comprising less than 200 households (Bulte et al., 2015). These villages are also isolated. Economic and social interactions take place within the village due to underdeveloped transport and ICT infrastructure.[2]

The absence of agricultural innovations and the localized nature in which economic and social life is organized make villages in Eastern Congo well-suited to learning about how social networks shape innovation adoption and diffusion.

2.3 Experimental Design

As part of our field experiment, each village was visited three times. The first visit entailed a baseline household survey, with all heads of households, to collect social network information as well as socioeconomic and agricultural production data. The intervention took place during the second visit, approximately one month later. The intervention consisted of training and the distribution of packets of fertilizer and information to three selected villagers, which we call 'ambassadors'. A third visit took place about two weeks later, during which all household heads were revisited and data about knowledge and use of fertilizer was collected. Table 2.1 provides an overview of the activities undertaken. We describe each visit in detail next.

[2] For example, among villagers in Eastern Congo, six percent do not know the location of the nearest public transport, and those that do know the location of the nearest transport claim that the average facility is on average 4.5 hours away on foot. Furthermore, 90 percent of villagers have not read a newspaper in the month before the survey, and 75 percent have not listened to the radio. The reach of government is limited, and customary leaders, like the village chief, are central in organizing the economic and social activities of raising taxes, settling disputes and allocating communal resources (van der Windt et al., 2019).

Table 2.1 – Experimental Design

Visit	Day	Activities	Date
1	1	Creation household list; household survey; mapping of household-level social networks; chief survey to obtain village level information	17 February to 13 March, 2015
2	31	Training ambassadors; distribution of fertilizer and fertilizer information; elicit willingness to pay for fertilizer	20 March to 29 March, 2015
3	45	Household survey; track fertilizer distribution	8 April to 29 April, 2015

Notes: Timeline and key activities of the field experiment.

2.3.1 Visit 1: Mapping the social network

During the first visit, research teams, with the aid of the village chief and other knowledgeable individuals, created a list of all households in the village, including the head of household's full name, age, and gender, and whether there were other adults present in the household.

After creating this list, each household head was visited for a survey. This baseline survey consisted of two parts. The first part of the survey collected information about basic socio-economic characteristics and agricultural practices, including fertilizer use and knowledge.

The second part of the survey collected social network data. We focus on three types of networks: the family network (whether the head of the household is biologically related with any member of the other household),[3] the field-neighbor network (whether the head of the household's field borders a field owned by any member of the other household) and the agriculture network (whether the head of one household discusses agricultural-related topics with any member of the other household). We chose these three networks for two reasons. First, for this same region, Kendzior et al. (2015) found that these three networks are the predominant channels via which agricultural resources are shared, which makes them the most appropriate networks for our intervention technology (discussed below). Second, three pilot studies found that the above three dimensions were the most distinct from one another and thus captured maximum variation while minimizing the number of network survey questions.[4] These pilots were conducted in the three months before

[3]Specifically, we use whether the other person is biologically related to a maximum of the third degree (this is a well-understood term in Congo).

[4]More specifically, we conducted three pilot studies that included two additional networks. We exclude the 'friends' network because everyone was everyone's friend. We also did not include whether individuals worked on each other's farms because it greatly overlapped with the other networks.

the start of the intervention and were designed to establish the appropriate networks to study.

To elicit network ties, the research teams first explained the network under study and then moved down the village household list asking for each household on the list if the network applies. A major benefit of this approach is that we are more likely to capture the full network. That is, in pilots we found that respondents offered up significantly fewer relations when simply asked who in the villages is in this network.[5] Our network data is thus at the household level for each of the three relationships covered.

2.3.2 Visit 2: Ambassador Training and Distribution of Fertilizer and Information

During visit 2, we implemented the intervention. In each village, three pre-selected individuals (the selection process is discussed below) took part in an extensive training session. The training of these 'ambassadors' was led by agronomists.[6] The training followed a set script (see Appendix Section 2.7.2) and the topics included information on types of fertilizer, benefits, application methods, expected market prices and access points in the nearby city of Bukavu. The training sessions focused on inorganic fertilizer containing NPK for its flexibility in being applied throughout the growing season and the positive effect on yields for a range of crops grown within the region.[7]

At the end of the training session, each ambassador received a single 1-kilogram bag of fertilizer that was theirs to keep. In addition, each ambassador also received three fertilizer 'kits'. Each kit consisted of three 1 kilogram bags of fertilizer. We asked the ambassadors to distribute each kit to different households in the village. We also asked the ambassadors to spread the information provided during the training. The recipients of the kits – 'second-stage ambassadors' – were asked to take one 1kg bag from the kit, and distribute the two remaining bags further. They were also asked to further distribute information about fertilizer. Figure 2.1 illustrates the suggested diffusion pattern for one ambassador. The structure allows for a maximum of 27 transfers per village. Note that there were no sanctions or incentives imposed to ensure this structure was followed in practice.[8]

We designed this intervention to mirror agricultural extension programs popular among NGOs in the developing world to promote technology adoption (Pypers et al., 2011). Such programs typically train and provide resources to a subset of intended recipients,

[5] One drawback of this approach is that it is time-consuming. To minimize survey fatigue, we

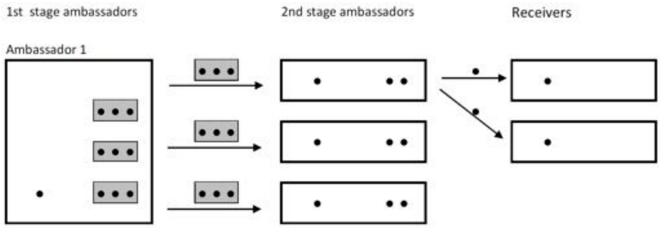


Figure 2.1 – Suggested Diffusion Pattern

Notes: Suggested diffusion pattern for one ambassador and one second-stage ambassador. Three ambassadors are selected in each village, adding up to 27 possible transactions per village.

frequently referred to as 'lead' farmers. These farmers are then asked to further disseminate resources and information within their village, aiming to reach wide distribution at low cost (Feder et al., 2003).

Finally, before the ambassador training, we measured the willingness to pay for fertilizer among a subset of villagers. To measure willingness to pay while avoiding ordering and anchoring effects, we follow Smith (2006) and implement a randomized card sorting game with eighteen selected respondents in each village. The respondents were selected (by the authors between visits 1 and 2) by sorting all household heads within the village by their centrality score and selecting the top six, bottom six and middle six. We showed each participant ten cards that displayed potential fertilizer prices from 100 Congolese francs (about 0.90 US$) to 5000 francs (about 4.50 US$).[9] Next, we asked participants whether they would be willing to pay that amount for a one kilogram bag of NPK fertilizer. The protocol can be found in Appendix Section 2.7.3.

decided to measure village social networks across only three dimensions.

[6] Each research team consisted out of at least one individual with a degree in agronomy.

[7] NPK fertilizers are three-component fertilizers providing nitrogen (N), phosphorus (P), and potassium (K).

[8] In theory, for example, a second stage ambassador could give the two fertilizer packages to one receiver. This receiver, in turn, can then act as a 'third-stage ambassador'. This happened only twice.

[9] The exact voucher amounts were: 100, 200, 500, 1000, 1300, 1500, 1700, 2000, 3500, and 5000.

2.3.3 Exogenous Variation in Initial Recipient

This study aims to understand the importance of the position of the initial receiver for the distribution of technology across a network. A major problem faced by researchers is that the initial receiver is unlikely to be chosen at random. NGOs often target specific individuals; for example, those that are literate or chosen by the village chief (Feder et al., 2010; Simpson et al., 2015). In response, we randomly assign villages to different initial entry points. We did so as follows.

First, based on the data collected during visit 1, we construct the family, field-neighbor and agriculture networks for each village, where a tie is present if at least one of the two households claims a tie to the other household. We then aggregate these three networks into one combined network.[10]

Next, we calculate for each individual her position within this combined network. The dimension of interest is eigenvector centrality (Bonacich, 1972). This centrality measure is based not only on the number of ties that a given node has (like degree centrality) but also weights each edge between nodes by the degree of the node that the edge leads to. Eigenvector centrality is therefore often used to proxy for an individual's level of influence within their social network. In the case of technology diffusion, a highly eigenvector central node may, by nature of its access and connectedness, be a more compelling source of novel information and for adoption. Alternatively, nodes with low eigenvector centrality may, by nature of homophily, be better positioned to reach nodes in the network that lay on the periphery of the network structure (Aral et al., 2009). Eigenvector centrality is thus particularly well suited for this study as a measure of centrality.

Next, we block randomize the 40 villages to one of two treatments.[11] In twenty randomly selected villages we choose the three individuals with the highest centrality score to be ambassador, whom we call 'central ambassadors'. In the other twenty villages, we choose the three individuals with the lowest centrality score as ambassador, whom we call 'isolate ambassadors'.[12] As discussed earlier, it is these selected ambassadors that receive training and fertilizer during visit 2. Figure A.2.2 and Figure A.2.3 in the appendix plot the 40 village networks and the selected central and isolate ambassadors.

Table A.2.2 in the appendix presents basic information on heads of household characteristics across the two treatments based on pre-treatment (visit 1) information. We find

[10] That is, if the tie exists in any of the three networks it also exists in the combined network.
[11] We block randomize on the five research teams to avoid enumerator effects.
[12] In only a few cases, an individual with the highest (or lowest) centrality score was not present in the village. When this happened, they were replaced by the next highest (lowest).

that the randomization was successful in creating two similar treatment groups.

2.3.4 Visit 3: Collecting Outcome Measures and Distribution Tracking

About two weeks after the intervention, we revisited all villages to collect outcome measures. We are interested in individuals' use, knowledge and willingness to pay for fertilizer. Furthermore, we are interested in exactly who received fertilizer and fertilizer information. To obtain this information our enumerators undertook three activities in each village. First, we revisited all households for another survey. Second, we conducted the random card sorting game for a second time with the same individuals. Third, enumerator teams undertook an extensive fertilizer tracking exercise to trace the transfer of packets of fertilizer throughout the village.[13]

Based on the first two data sources we create three outcome variables. The first is *fertilizer use*. To measure fertilizer use, we ask respondents whether they had applied chemical fertilizer on any of their fields during the agricultural season preceding the survey.

Second, we ask respondents about their *fertilizer knowledge*. In contrast to fertilizer use, knowledge about fertilizer is not a scarce commodity and may thus be governed by different distribution dynamics. To measure fertilizer knowledge, we asked each respondent questions related to the expected benefits of chemical fertilizer, method and timing of application, and market availability and pricing. These questions directly correspond to the information that we provided as part of the ambassadors' training during visit 2. We create a fertilizer knowledge score based on correct responses.

The third outcome of interest is an individuals' *willingness to pay* for fertilizer. Distributing fertilizer and information about it may influence how people value fertilizer. We obtain this information conducting the randomized card sorting game with the same eighteen individuals also visited during visit 2.

These same data were also collected before the intervention. As a result, we have panel data, which we use to increase statistical precision.

[13] I.e. research teams visited each first-stage ambassador and asked to whom they had distributed fertilizer, if at all. Next, the teams would visit these second-stage ambassadors to ask to whom they had distributed fertilizer, if at all. We only did a tracking exercise related to fertilizer packages, not knowledge about fertilizer, because transfer of the latter is difficult to verify.

2.3.5 Empirical Strategy

We first investigate what characteristics explain who receives fertilizer and knowledge about fertilizer. To do so, we use data collected during visit 3 and construct a dataset in which we list all dyads between each first-stage ambassador and all other household heads in the village; i.e. potential second-stage ambassadors. The resulting dataset thus contains $a = \{1, 2, 3\}$ times $i = \{1, ..., n-1\}$ dyads per village $j = \{1, ..., 40\}$.[14] We estimate:

$$\boldsymbol{Y}_{aij} = \beta_0 + \boldsymbol{\beta_1 X_{aj}} + \boldsymbol{\beta_2 X_{ij}} + \boldsymbol{\beta_3 X_{aij}} + \varepsilon_{aij} \tag{2.1}$$

where $\boldsymbol{Y}_{aij}$ is one of three dummy variables: 1) whether fertilizer was transmitted based on information from the tracking exercise, 2) whether fertilizer was transmitted based on survey data, and 3) whether information about fertilizer was transmitted based on survey data. $\boldsymbol{X_{aj}}$ is a vector with ambassador characteristics, $\boldsymbol{X_{ij}}$ is a vector with potential second-stage ambassador characteristics, and $\boldsymbol{X_{aij}}$ contains dyad-specific information. We cluster the error, ε_{aij}, at the village and the ambassador level. Then, we re-estimate Equation 2.1 and interact all individual and dyad level characteristics with the treatment assignment, T_j, to examine whether the distribution patterns are different in villages with central ambassadors compared to those villages with isolate ambassadors.[15]

To estimate the effect of initial recipient centrality on the use of, knowledge about and willingness to pay for fertilizer we estimate a difference-in-difference equation:

$$\boldsymbol{Y}_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 Endline_t + \beta_3 T_j * Endline_t + \boldsymbol{\beta \gamma_k} + \varepsilon_{ijt} \tag{2.2}$$

where $\boldsymbol{Y}_{ijt}$ represents an outcome for individual $i = \{1, \ldots, n\}$ in village $j = \{1, \ldots, 40\}$ at time $t = \{0, 1\}$, where t=0 for visit 1 (fertilizer use and knowledge) or visit 2 (willingness to pay) and t=1 for visit 3. T_j is our treatment variable at the village level, which takes the value $T_j = 1$ if the ambassador is central, and $T_j = 0$ if the ambassador is isolate. $Endline_t$ is a dummy that equals one if the data were collected during visit 3, and zero if

[14] Theoretically, each ambassador could give to the other two ambassadors.

[15] We also examine this result for second-stage ambassadors (shown in Appendix Table A.2.3.). In other words, we explore the transmission between second-stage ambassadors and all other household heads in the village; i.e. the potential "receivers" in Figure 2.1. The dataset used thus contains $a = \{1, \ldots, x\}$ times $i = \{1, ..., n-1\}$ dyads per village, where x are the number of second-stage ambassadors. Second-stage ambassadors could potentially give to first-stage ambassadors and other second-stage ambassadors.

the data were collected during visit 1. The coefficient of interest is β_3. Finally, $\boldsymbol{\beta\gamma_k}$ are research team fixed effects, our randomization blocking variable, and ε_{ijt} are Newey-West standard errors robust to heteroskedasticity and autocorrelation, clustered at the village level.

Finally, to learn about intervention attenuation, we examine the relationship between an individual's position within the distribution and that individual's use of, knowledge about and willingness to pay for fertilizer. We use data collected during visit 3 and estimate:

$$\boldsymbol{Y}_{ij} = \beta_0 + \beta_1 2nd\_amb_i + \beta_2 rec_i + \beta_3 not\_rec_i + \varepsilon_{ijt} \tag{2.3}$$

where $2nd\_amb_i$ equals one for second stage ambassadors, rec_i equals one for those that received from the second stage ambassadors, and not_rec_i equals one for those individuals that did not receive anything. Those individuals that were trained by our agronomists directly, the first-stage ambassadors, are the comparison group. Insofar that the strength of the intervention is strongest for those closest to the agronomist, we thus expect $\beta_1 > \beta_2 > \beta_3$. Subsequently, we re-estimate equation 2.3 and interact receiver type with the treatment assignment, T_j, to learn whether the entry point of the new technology affects attenuation. Note that these estimates are not experimentally identified.

2.4 Baseline Characteristics and Treatment Implementation

Our study population comprises 40 villages, 2,677 households and a potential 187,628 network ties. Figure A.2.4 in the appendix presents a CONSORT-style flow diagram with the targeted and collected data by treatment arm.

During visit 1, survey data was collected from 2,584 households. A total of 23,002 ties in the combined network exist across our study villages.[16] The rate of attrition (from sampling frame to baseline sample) was 3%.[17] This is low for these contexts as we undertook great efforts to trace all household heads.[18] During visit 2, agronomists

[16]The family network alone contains 15,071 ties, the field-neighbor network contains 6,950 ties and the agriculture network contains 5,906 ties.

[17]We find no evidence of selective attrition.

[18]Beaman et al. (2018), for example, were able to reach about 80% of targeted respondents. We undertook great efforts to trace all household heads. If a household head was not available, the research assistant returned a few days later. If the head of the household remained unreachable,

successfully trained three ambassadors in all villages, and a total of 717 households participated in the random card sorting game. Finally, during visit 3, we collected survey data from 2,305 households that were also visited during visit 1. Furthermore, 658 individuals participated in the random card sorting game that also participated during visit 2.

2.4.1 Villagers

Table A.2.2 presents descriptive statistics for our respondents based on baseline data. Variable definitions are available in Table A.2.1. Household heads are, on average, 46 years old and predominantly illiterate. On average, 68% of the household heads are male, 34% were not born in the village (we call them 'migrants') and almost all are members of the village's majority ethnic group. Many have also been exposed to violence, having experienced on average 3.3 out of 7 conflict events. A typical household owns 2.7 animals (chickens, goats and cows). Furthermore, 98% of households own at least one plot of land and those that do own on average 2.3 plots of land (not in table). The average farm size (adding up the different plots) is around five square km. To assess whether someone is viewed as influential among farmers, we follow Banerjee et al (2013) and ask households to whom we should speak if we want to spread information about a new agricultural technique; 22% of household heads were mentioned at least once. About 35% of household heads have a leadership role in the village, and the average household head interacts with the village chief around six times a month.

Finally, we turn to baseline values of our main outcome variables. Table A.2.2 shows that only 7% of households had ever applied fertilizer to their fields before the intervention. Fertilizer knowledge is also low, with an average sample score of 1.44 out of a possible score of 8.5. Finally, based on the randomized card sorting game, the typical household head is willing to pay approximately 1.45USD for 1 kg, somewhat below the market price of fertilizer in the nearby city of Bukavu (1.70USD during the study period).

---

we looked for an adult replacement within the household. Upon replacement, we asked the replacement about the characteristics (including the network characteristics) of the head of the household. In about 26% of households the head was replaced, generally by the spouse. We asked all replacements the reason for the absence of the head of the household. The most common reasons mentioned are: visit to the household's fields, visit to Bukavu, and temporary outmigration for work.

2.4.2 Initial recipients: Ambassadors

Next, we zoom in on the three village members selected in each village to receive fertilizer and fertilizer training, the ambassadors. Table 2.2 presents baseline information, where we separate the ambassadors by treatment condition. Central ambassadors (the three most central individuals in the village) differ significantly across a considerable number of characteristics from the isolate ambassadors (the three least central individuals). Central ambassadors are more likely to be male, literate and are on average more wealthy. Among isolate ambassadors, 47% are migrant, while this is only 10% among central ambassadors. About 52% of the central ambassadors were mentioned by at least one villager to be a lead farmer; this number is just 15% for isolate ambassadors. And while 50% of ambassadors have some sort of leadership position in the village, this is only 22% for isolate villagers. Finally, the average central ambassador interacts with the village chief about 11 times per month; this number drops to 5 for isolate ambassadors. Table 2.2 also shows that the ambassador selection process was successful. Central ambassadors have a significantly higher centrality score.

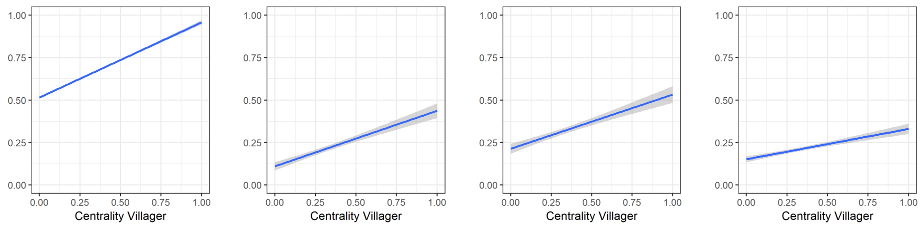
Table 2.2 – Ambassador Characteristics at Baseline

	Central Ambassadors			Isolate Ambassadors			
	N	Mean	SD	N	Mean	SD	Diff
Age (years)	60	49.10	14.26	60	45.23	17.45	3.867
Male (=1)	60	0.87	0.34	60	0.53	0.50	0.333***
Literate (=1)	60	0.55	0.50	60	0.32	0.47	0.233**
Migrant (=1)	60	0.10	0.30	60	0.47	0.50	-0.367***
Ethnic Majority (=1)	60	1.00	0.00	60	0.92	0.28	0.083
War exposure (7 events)	60	3.63	1.21	60	3.27	1.30	0.367
Wealth index	60	3.42	3.71	60	2.07	3.00	1.350**
Farm size	55	6.33	8.72	56	5.11	9.47	1.215
Likely Lead Farmer (=1)	60	0.52	0.50	60	0.15	0.36	0.367***
Leadership position	60	0.50	0.50	60	0.22	0.42	0.283***
Access to village chief	50	10.58	10.43	60	5.37	8.18	5.213**
Centrality score	60	0.88	0.11	60	0.09	0.14	0.785***
Used Fertilizer (=1)	59	0.08	0.28	60	0.02	0.13	0.068
Fertilizer knowledge (max 8.5 points)	60	1.89	1.68	60	0.82	1.28	1.075***
Willingness to pay for fertilizer (USD)	54	1.77	1.63	57	1.43	1.57	0.335

Note: Baseline data from 120 ambassadors. Reported p-values based on regressions with standard errors clustered at the village level. Variable definitions can be found in Table A.2.1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Related to the outcome measures, central ambassadors are more likely to have used fertilizer, know more about fertilizer and are willing to pay more for fertilizer. However, only the difference in knowledge about fertilizer is statistically significant (p<0.01).

The differences between central and isolate ambassadors suggest that centrality is correlated with attributes associated with, among others, political marginalization. That is, the less central the individual is within the village network, the more politically marginalized the villager. This observation corresponds with other studies (e.g. Larson et al. (2019)). We explore this dynamic further by focusing on four characteristics that largely qualitative scholars of Congo have suggested as indicators for political marginalization in this research setting: migration status, likely lead farmer, leadership position, and access to village chief.[19] Figure 2.2 plots estimation lines from a simple OLS regression of the political marginalization indicator on villagers' centrality scores, using data on all household heads. The estimations include village fixed effects, effectively controlling for all village level characteristics. Gray bounds indicate 95% confidence intervals. Moving from the least to the most central villager increases the chances of being a native (the inverse of a migrant) by 48 percentage points ($p<0.01$), being a lead farmer by 35 percentage points ($p<0.01$), having a leadership role by 34 percentage points ($p<0.01$), and having had interaction with the village chief by 20 percentage points ($p<0.01$). We thus find that villagers' centrality score is very much correlated with indicators of political marginalization.



(a) Native **(b)** Lead Farmer **(c)** Leadership **(d)** Access Chief

Figure 2.2 – Receiver Centrality and Political Marginalization

Notes: Figures plot villagers' centrality scores (x-axis) on indicators of political marginalization: (a) whether the villager is village native (1 if yes), (b) a lead farmer (1 if yes), (c) a village leader (1 if yes), and (d) the number of times the individual has interacted with the village chief in the previous month divided by 30. Based on visit 1 data.

2.4.3 Distribution of Fertilizer and Fertilizer Knowledge

All 120 ambassadors received ten packages of fertilizer and information. Was this new technology distributed further? In the household survey, we asked all respondents whether

[19] Another indicator would be ethnic group membership, but this variable has too little variation.

they had received fertilizer recently, and if so how much and from whom. In total, our dataset records 549 transfers of fertilizer. Of the 120 first-stage ambassadors, 97 (or 81%) distributed fertilizer.[20] Those first-stage ambassadors who distributed fertilizer gave away, on average, 4.5 kilograms, about half of the suggested 9 kilograms (see Figure 2.1).[21] This fertilizer was distributed to an average of 3.7 individuals; thus more than the suggested three second-stage ambassadors. Distribution beyond the first-stage ambassadors' receivers is much lower. Of the 354 villagers that received something from first-stage ambassadors, only 123 distributed fertilizer further. Those second-stage ambassadors who distributed fertilizer gave on average 1.6 kilograms (instead of the suggested 2 kilograms), and to an average of 1.58 receivers (instead of the suggested two villagers).

The household survey also asked respondents whether and from whom they received fertilizer information. Given that information is not a scarce good, it is not surprising that information about fertilizer was shared more often than fertilizer. In total, 646 transfers of information were made. Of the 120 first-stage ambassadors, 102 (or 85%) distributed knowledge. In total, they gave information to 392 village members (or 3.8 per ambassador). Of those villagers that received information from first-stage ambassadors, 168 distributed the information to another 254 village members.

Information reported by households may be biased because individuals may respond strategically or forget who gave information. In response, during visit 3, we also undertook a fertilizer tracking exercise (discussed in Section 2.3.4), which recorded who gave fertilizer to whom. The exercise records 507 transfers of fertilizer. Among the first-stage ambassadors, 97 distributed fertilizer. When they did, they shared with an average of 3.39 households. Distribution beyond the first-stage ambassador is lower. Of the 329 second-stage ambassadors, only about a third (102) distributed fertilizer further, sharing with a total of 178 other households.

In sum, despite the absence of sanctions or incentives for distribution we find that the fertilizer and information about fertilizer was distributed through the villages. In the next section, we explore whether the centrality of ambassadors had an impact on these distribution patterns.

[20] We find no differences in compliance by ambassador type.

[21] Debriefing interviews with first-stage ambassadors during visit 3 reveal that those ambassadors who did not distribute their fertilizer kept it for their own use, or said that they were waiting until the next planting period to distribute the fertilizer.

2.5 Results

2.5.1 Distributional Consequences of Technology Seeding

What are the distributional consequences of the type of ambassador selection? Which recipients do ambassadors seek out to disseminate the packs of fertilizer and information about its use? And does this distribution depend on the centrality of the ambassador? Table 2.3 presents results assessing receiver and dyadic characteristics of recipients. To readily compare across coefficients, we standardize all independent variables. Columns 1 and 2 are based on information reported by households during the survey in visit 3. As a robustness check, column 3 reports results based on data from the fertilizer tracking exercise.

We first focus on columns 1 to 3. We find that the centrality of the receiver correlates positively with receiving fertilizer. On average, an increase in centrality of one standard deviation makes an individual 1.3 percentage points more likely to receive fertilizer ($p<0.01$). We find a similar result based on data from the tracking exercise (1.1 percentage points, $p<0.05$). More central villagers are also more likely to receive information about fertilizer, although this effect is not statistically significant. Previous war exposure also correlates positively with both types of outcome, perhaps as victims are singled out due to recent losses. Two dyad characteristics stand out. Ambassadors are more likely to give to those to whom they are connected in the agricultural network. Furthermore, villagers that live closer to the ambassador are also significantly more likely to receive fertilizer.

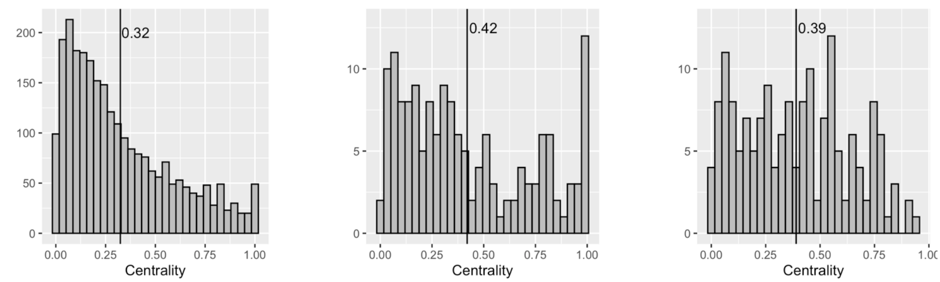
Columns 4 to 6 in Table 2.3 interact the receiver and dyad characteristics with the treatment indicator to assess whether the distribution of technology depends on ambassador centrality. Focusing on the interactions, receiver centrality stands out both in magnitude and statistical significance. Villagers with higher centrality scores are more likely to receive fertilizer in villages with isolate ambassadors, compared to villages with central ambassadors. We find the same dynamics for the distribution of information (column 5) and the distribution of fertilizer based on data from the tracking exercise (column 6). We now further explore the magnitude and statistical significance of this conditional effect.

Table 2.3 – Distribution Behavior by First-Stage Ambassadors

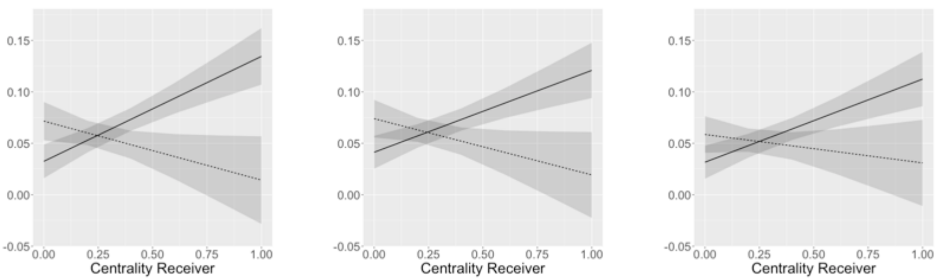
	(1) Fertilizer (Endline Survey)	(2) Information (Endline Survey)	(3) Fertilizer (Tracking Exercise)	(4) Fertilizer (Endline Survey)	(5) Information (Endline Survey)	(6) Fertilizer (Tracking Exercise)
Centrality (R)	0.013*** (0.004)	0.008* (0.004)	0.011** (0.005)	0.027*** (0.005)	0.021*** (0.005)	0.020*** (0.006)
Age (R)	0.001 (0.004)	0.001 (0.004)	0.009** (0.004)	0.005 (0.005)	0.006 (0.004)	0.011** (0.004)
Male (R)	0.002 (0.004)	−0.001 (0.004)	0.013*** (0.003)	−0.003 (0.005)	−0.004 (0.004)	0.009*** (0.003)
Literate (R)	0.005 (0.003)	0.004 (0.003)	0.010** (0.004)	0.014*** (0.003)	0.010*** (0.003)	0.014*** (0.005)
Migrant (R)	0.000 (0.004)	−0.002 (0.003)	0.001 (0.004)	−0.001 (0.004)	−0.006 (0.004)	0.000 (0.005)
Ethnic Majority (R)	0.003*** (0.001)	0.001 (0.002)	−0.004 (0.003)	0.003 (0.002)	0.002 (0.002)	−0.007** (0.003)
War Exposure (R)	0.009** (0.004)	0.008** (0.004)	0.006 (0.004)	0.001 (0.005)	0.000 (0.004)	0.003 (0.004)
Wealth (R)	0.001 (0.003)	0.000 (0.003)	0.003 (0.004)	−0.005 (0.004)	−0.005 (0.005)	0.005 (0.006)
Farm Size (R)	0.000 (0.003)	0.004 (0.003)	0.000 (0.003)	0.001 (0.004)	0.002 (0.004)	−0.002 (0.003)
Family	−0.001 (0.005)	0.001 (0.004)	0.001 (0.005)	0.015 (0.013)	0.014 (0.013)	0.008 (0.015)
Field Neighbors	0.003 (0.005)	0.008 (0.005)	0.005 (0.005)	0.004 (0.007)	0.022** (0.009)	0.013 (0.010)
Agriculture	0.010** (0.005)	0.009** (0.005)	0.012*** (0.004)	0.009 (0.010)	0.008 (0.010)	0.011 (0.008)
Physical Distance	−0.027*** (0.005)	−0.022*** (0.006)	−0.020*** (0.004)	−0.024*** (0.006)	−0.022*** (0.007)	−0.015*** (0.005)
Social Distance	0.002 (0.004)	0.000 (0.004)	−0.001 (0.004)	0.004 (0.005)	0.003 (0.004)	0.000 (0.003)
Treatment = village has central entrypoint				−0.011 (0.008)	−0.010 (0.007)	−0.006 (0.009)
Treat * Centrality (R)				−0.041*** (0.007)	−0.035*** (0.007)	−0.027*** (0.009)
Treat * Age (R)				−0.008 (0.009)	−0.011 (0.008)	−0.006 (0.009)
Treat * Male (R)				0.009 (0.008)	0.006 (0.007)	0.008 (0.007)
Treat * Literate (R)				−0.016** (0.006)	−0.010** (0.005)	−0.006 (0.008)
Treat * Migrant (R)				0.001 (0.007)	0.007 (0.007)	0.002 (0.008)
Treat * Ethnic Majority (R)				0.002 (0.002)	−0.003 (0.004)	0.012*** (0.004)
Treat * War Exposure (R)				0.015** (0.007)	0.016** (0.006)	0.008 (0.008)
Treat * Wealth (R)				0.010 (0.007)	0.009 (0.007)	−0.003 (0.008)
Treat * Farm Size (R)				−0.002 (0.005)	0.003 (0.006)	0.003 (0.005)
Treat * Family				−0.010 (0.014)	−0.007 (0.014)	−0.004 (0.016)
Treat * Field Neighbors				0.000 (0.010)	−0.019* (0.010)	−0.011 (0.011)
Treat * Agriculture				0.003 (0.012)	0.003 (0.011)	0.001 (0.009)
Treat * Physical Distance				−0.010 (0.008)	−0.002 (0.011)	−0.016** (0.007)
Treat * Social Distance				−0.009 (0.007)	−0.006 (0.007)	−0.008 (0.007)
Observations	4614	4818	4717	4614	4818	4717

Notes: Standard errors clustered at the village and sender level. Randomization fixed effects included. Based on data from visit 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variable definitions can be found in Table A.2.1.

Panel (a) in Figure 2.3 shows the histogram of the centrality scores of all villagers. The average centrality score is 0.32 (standard deviation equals 0.26). Do isolate ambassadors give only to other isolate villagers? Do central ambassadors prefer to gift new technology to other central villagers? Panels (b) and (c), which show the histograms for those individuals that received fertilizer from the first-stage ambassador in villages assigned to isolate and central ambassadors, respectively, show that this is not the case. Isolate ambassadors, those villagers with centrality scores of around 0, give to villagers with an average centrality score of 0.42 (standard deviation of 0.25). Central ambassadors, those villagers with a centrality score of around 1, give to villagers with a lower average centrality score of 0.39 (standard deviation 0.31).



(a) All villagers **(b)** Receivers (From isolate ambassadors) **(c)** Receivers (From central ambassadors)



(d) Fertilizer (survey) **(e)** Information (survey) **(f)** Fertilizer (tracking)

Figure 2.3 – Distribution and Receiver Centrality by Treatment

Notes: Panel (a) displays a histogram of receiver centrality for all villagers, panel (b) for the receivers in villages with isolate ambassadors, and panel (c) for receivers in villages with central ambassadors. In panels (d) to (f), the y-axis is the probability of a transfer, the x-axis is the centrality of the receiver, isolate (central) ambassadors are indicated by the solid (dashed) line, and bounds are 95% confidence intervals

Panels (b) and (c) also show that these averages mask considerable differences in the distribution of receiver centrality scores. To quantify the conditional effect of ambassador centrality we explore the marginal effect at every observed value of receiver centrality. Panels (d) to (f) show the result from simulating the estimated models in Table 2.3 's columns 4 to 6, respectively, for different levels of receiver centrality and by treatment status, keeping all other variables at their mean value. We find that isolate ambassadors (solid line) contribute more fertilizer and fertilizer information to more central villagers. Focusing on panel (d), a villager with the lowest centrality score has a 3.31% chance of receiving fertilizer from an isolate ambassador, while a villager with the highest centrality score has a 13.51% probability of receiving fertilizer from the same ambassador. The slope is statistically significant ($p<0.01$). In contrast, we find that central ambassadors are less likely to contribute fertilizer and fertilizer information the more central the villagers. Focusing on panel (d), the most isolate villager has a 6.84% chance of receiving from a central ambassador, which decreases to 1.18% for the most central villager. The slope is again statistically significant ($p<0.05$).

Figure 2.3's panels (d) to (f) show that isolate ambassadors, compared to central ambassadors, are considerably more likely to contribute to the very central villagers ($p<0.01$). In addition, central ambassadors are more likely to contribute to the most isolate individuals compared to isolate ambassadors, although this difference is smaller and not statistically significant.

One potential criticism of this result is that it is mechanical: because we stipulated (though did not enforce) that ambassadors could not give to others within their household, central ambassadors were not able to give to the most central, and isolate ambassadors could not give to the most isolated. We explore this result in Appendix Section 2.7.9, where we replace transfers to counterfactual ambassadors (those who would have been ambassadors had the opposite treatment been assigned) with transfers to those just across the cutoff. The results hold, though with lower magnitude.

In sum, we find that the type of ambassador has a strong impact on who receives fertilizer and knowledge about fertilizer. Isolate ambassadors, compared to central ambassadors, are significantly more likely to gift new technology to those with high centrality scores. We find the same dynamics for knowledge about fertilizer.[22]

[22] Table A.2.3 in the appendix explores the distribution of fertilizer and fertilizer information by second-stage ambassadors. We find similar results.

2.5.2 Fertilizer Use, Knowledge and Valuation

Does the type of initial recipient affect villagers' use of fertilizer, knowledge about fertilizer, and their willingness to pay for fertilizer? Table 2.4 presents results. The intervention has a strong impact on the use of fertilizer in the village. Column 1 shows that, at the onset of the intervention, an estimated 7% of villagers in control areas had ever used fertilizer. This number more than doubles after the intervention ($p<0.01$). However, we find no evidence that this increase is stronger in villages where we trained central ambassadors.

Table 2.4 – Network Entry Point and Fertilizer Use, Knowledge and Valuation

	(1) Fertilizer Use	(2) Fertilizer Knowledge	(3) Willingness to Pay for Fertilizer
Treatment: Central entry point	−0.016	0.026	−0.099
	(0.019)	(0.139)	(0.192)
Endline	0.089***	0.905***	−0.111
	(0.029)	(0.112)	(0.160)
Treatment * Endline	−0.010	−0.018	0.048
	(0.036)	(0.171)	(0.218)
Outcome in control at baseline	0.074	1.44	1.60
SD in control at baseline	0.26	1.69	1.47
Observations	4673	4718	1216
# Clusters	40	40	40

Notes: Standard errors clustered at the village level. Randomization fixed effects included. Based on data from visits 1, 2 and 3. Column (3) based on data from participants in random card sorting game only. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variable definitions can be found in Table A.2.1.

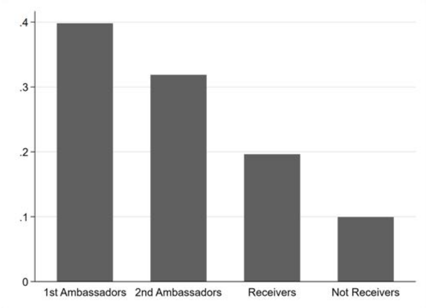
Column 2 shows that villagers also know much more about fertilizer after the intervention. Before the intervention, the average villager in a control village scores just 1.4 out of 8.5 points. This score increases to 2.3 points after the intervention ($p<0.01$). We again find no evidence, however, that this increase is different by type of entry point.

Finally, in column 3 we find that individuals' valuation of the fertilizer – as measured by their willingness to pay using the random card sorting game – is the same before and after the intervention. Two opposing effects can be at play. On the one hand, improved knowledge about the advantages of fertilizer may increase an individual's willingness to pay for fertilizer. On the other hand, the intervention increased local supply, which may push valuations down. We are unable to decompose these effects. We find no difference between villages with central and those with isolate ambassadors.

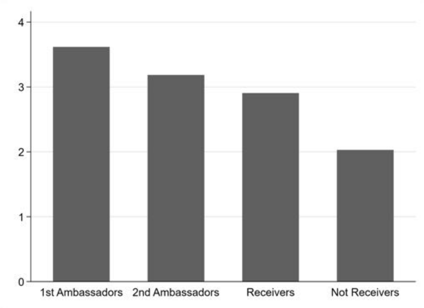
2.5.3 Intervention Attenuation

Next, we look at the attenuation of the intervention, and whether this differs by ambassador type. We expect the impacts on fertilizer use, fertilizer knowledge and willingness to pay for fertilizer to diminish as it spreads through the village. In other words, an individual that receives information about fertilizer use directly from our agronomist is more informed than an individual that received this information indirectly through network diffusion. Also, we expect the centrality of the ambassador to play a role as centrality is closely correlated with socioeconomic indicators of education, and villagers may be more willing to accept information from central ambassadors.

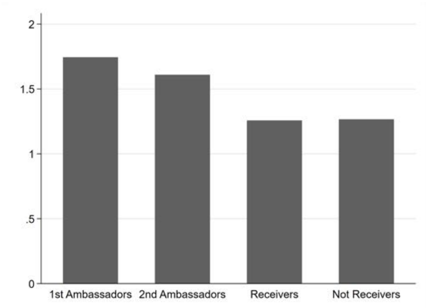
We divide each village into four groups following Figure 2.1: 1) first-stage ambassadors, 2) second stage ambassadors, 3) receivers, and 4) never receivers. Figure 2.4 plots fertilizer use, fertilizer knowledge, and willingness to pay for fertilizer across these four groups. We find that the intervention weakens when moving away from the entry point.



(a) Fertilizer use



(b) Fertilizer Knowledge



(c) Fertilizer WTP

Figure 2.4 – Intervention Attenuation

Notes: Figures present the average fertilizer use, fertilizer knowledge, and willingness to pay for fertilizer by type of villager. Based on data from visit 3. Variable definitions can be found in Table A.2.1

Columns 1 to 3 of Table 2.5 quantifies these differences, where the comparison group is the first-stage ambassadors; the village entry points. Across all three outcomes we find that the coefficients are negative and increasing in size when moving further away from these ambassadors. About 40% of first-stage ambassadors have used fertilizer. This number equals 32% for second-stage ambassadors, 19% for receivers and only 10% for not receivers. Related to knowledge about fertilizer, first-stage ambassadors score 3.6 out of the 8.5 points on our knowledge test, which decreases to 3.3 for second-stage ambassadors, 3.0 for receivers and just 2.1 for non-receivers. The bottom two rows show that these differences are statistically significant. Finally, willingness to pay for fertilizer follows a similar pattern. First-stage ambassadors are willing to pay 1.72US$ for fertilizer, which is slightly more than second-stage ambassadors (1.60US$). Receivers and non-receivers are willing to pay just 1.24US$ for fertilizer.

Table 2.5 – Ambassador Centrality and Intervention Attenuation

	(1) Fertilizer Use	(2) Fertilizer Knowledge	(3) WTP	(4) Fertilizer Use	(5) Fertilizer Knowledge	(6) WTP
2nd-stage ambassador	−0.080	−0.366**	−0.120	−0.044	−0.301	0.117
	(0.059)	(0.157)	(0.166)	(0.086)	(0.244)	(0.208)
Receiver	−0.209***	−0.713***	−0.489***	−0.186**	−0.673**	−0.479**
	(0.054)	(0.203)	(0.142)	(0.079)	(0.277)	(0.189)
Not Receiver	−0.303***	−1.541***	−0.484**	−0.266***	−1.592***	−0.445
	(0.053)	(0.170)	(0.206)	(0.070)	(0.258)	(0.265)
Treatment: Central entry point				0.036	−0.071	0.107
				(0.112)	(0.329)	(0.291)
Received from 1st stage * Central				−0.067	−0.118	−0.481
				(0.118)	(0.327)	(0.323)
Received from 2nd stage * Central				−0.043	−0.097	0.031
				(0.109)	(0.423)	(0.276)
Never received * Central				−0.075	0.106	−0.070
				(0.105)	(0.345)	(0.395)
Outcome first-stage ambassador	0.40	3.64	1.72	0.40	3.64	1.72
SD first-stage ambassador	0.49	1.55	1.47	0.49	1.55	1.47
Observations	2228	2241	524	2228	2241	524
# Clusters	40	40	40	40	40	40
P-value Received from 1st stage=Received from 2nd stage	0.0017	0.022	0.0042			
P-value Received from 2nd stage=Non-Receiver	0.019	0.000	0.97			

Notes: Standard errors clustered at the village level. Randomization fixed effects included. Based on data from visits 3. Columns (3) and (6) based on data from participants random card sorting game only. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variable definitions can be found in Table A.2.1.

Finally, we investigate whether this attenuation varies with the network position of the first-stage ambassadors. This might be because their social status makes their words carry more weight, or they might be more effective teachers. Columns 4 to 6 of Table 2.5

show no evidence that this is the case.

2.5.4 Discussion

Why do isolate ambassadors donate new technology to people more centrally located within the network, instead of fellow isolate villagers? We explore several explanations.

First, if the distribution of centrality is skewed towards central positions in the network, then even if isolate ambassadors distribute fertilizer randomly, we would observe upward giving. Panel (a) of Figure 2.3 , however, shows that this distribution is in fact skewed towards isolate villagers instead of centrals.

A second explanation relates to efficiency. Differences in agricultural productivity across and within households have been well documented (e.g. Udry (1996)). Central villagers may be more likely to make optimal use of fertilizer. As a result, an isolate ambassador contributing to central villagers would make sense from an economic efficiency point of view. To investigate this claim we regress several farm-related characteristics at baseline on villagers' centrality scores. We include village fixed effects to control for any village-level characteristics. Columns 1 to 5 in Table 2.6 show that more central villagers are more likely to be mentioned as lead farmers, have larger farm plots, are more likely to have used fertilizer and know more about fertilizer use at the onset of the intervention. These results are consistent with an efficiency argument.

Table 2.6 – Villager's Network Position and Farm Characteristics

	(1) Lead Farmer	(2) Farm size	(3) Fertilizer Use	(4) Fertilizer Knowledge	(5) Fertilizer WTP
Villager's centrality score	0.350***	2.893***	0.102***	1.319***	0.235
	(0.031)	(0.844)	(0.019)	(0.127)	(0.151)
Constant	0.092*	0.963	0.072**	1.340***	1.420***
	(0.054)	(1.519)	(0.036)	(0.233)	(0.366)
Observations	2584	2366	2447	2479	693

Notes: Village fixed effects included. Based on data from visits 1 and 2. Column (5) based on data from participants random card sorting game only. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, isolate ambassadors may have gifted their fertilizer to central villagers. In the household survey, we asked those respondents who received fertilizer about how they had received the fertilizer: for free, purchased, exchanged or other. We find that all fertilizer was given for free. Isolate ambassadors may have used the introduction of a new and scarce resource as a way to build their social networks in the village. There is ample evidence from the developing world to suggest that strategic gifting enables those at the edges of society to move ahead (Mauss, 2002; Sherry Jr, 1983). Gifting creates expectations for future interaction and may open up opportunities for social interactions, build networks and bridge social capital (Adloff and Mau, 2006; Putnam and Others, 2000). One remaining worry is that central villagers were able to pressure isolate ambassadors to 'gift' fertilizer to them, or that village leadership forced isolate ambassadors to do so. Careful qualitative work after the intervention, including debriefings with ambassadors in each village, did not find any instance in which isolate ambassadors were coerced to give fertilizer.

2.6 Conclusion

The diffusion of new technologies is a key component of political and economic development. While it is generally believed that the type of initial recipient impacts distribution patterns, few studies have put this to the test. This study investigates whether the network position of initial recipients affects the use, knowledge about and willingness to pay for a new technology. We also explore whether attenuation of the technology depends on the centrality of the initial recipient. Finally, we consider whether the network position of the initial recipients affects *who* receives the new technology and related information.

We implement a field experiment in 40 villages in Eastern Congo. As part of the experiment we select three initial recipients in each village, provide them with fertilizer (and training on its correct application), and ask them to distribute fertilizer and information about fertilizer. In half of the villages ambassadors are those most (eigenvector) central in the village; in the other half of villages we train the three most isolate individuals.

We find that centrality measures based on family, farming, and agricultural discussion relationships are strongly correlated with observable characteristics. Most prominently, an individual's centrality score is closely related to their level of political marginalization: the least central villagers are also those that are least likely to be native or mentioned as a lead farmer by others. They are also much less likely to have a leadership position in the village or interaction with the village chief.

Experimentally, we find no evidence that the position of the initial recipient has an impact on average village-level fertilizer use, knowledge or valuation. However, we do find that the network centrality of the initial recipient affects which village members gain access to new technologies. Both central and isolate farmers prefer sharing along existing social ties but the sharing behavior of isolate versus central ambassadors differ depending on the centrality of the receiver. Central ambassadors are more likely to give to individuals with a low centrality score, while isolate ambassadors are significantly more likely to gift new technology to those with high centrality scores. We suggest that the latter result may obtain because isolate villagers believe that centrals may put the fertilizer and knowledge about fertilizer to better use, or because isolate ambassadors gift strategically to central villagers to move ahead in the village.

We highlight two implications of our results for program design. First, practitioners with strong priorities on *who* should benefit from an intervention (such as politically marginal households) should target those households directly. Practitioners should not rely on households at the edges of social networks to diffuse technologies to households with similar characteristics. Regardless of initial targeting, the resources are likely to end up with the most influential. Second, our study shows that the quality of information and the rate of adoption decreases with each step away from the initial injection point. This attenuation implies that practitioners who aim to leverage diffusion dynamics as opposed to direct targeting to reach a large number of village members should consider periodic reinforcement of resources and knowledge. Designing programs that incorporate multiple practitioner-village interactions may help alleviate some of these inherent attenuation effects.

Although previous studies tell us how many people receive something through their networks and how to maximize this number, they rarely tell us who the recipients are, or specifically who gets left out. But development can easily go hand in hand with inequality, especially when the driver of development– access to new technology– is itself distributed unequally.

2.7 Appendix

2.7.1 Map of Villages

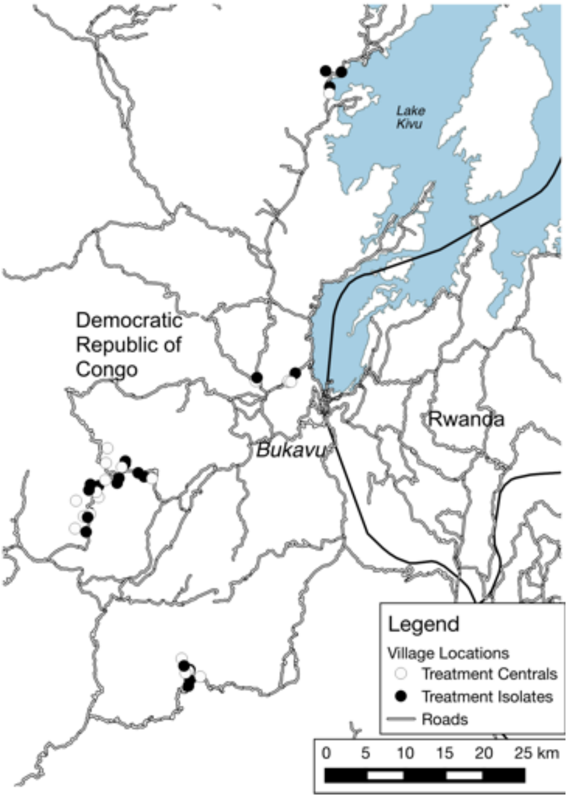


Figure A.2.1 – Map of Villages

Notes: Map displays location of the 40 study villages: black circles are villages assigned to isolate ambassadors, white circles are villages assigned to central ambassadors.

2.7.2 Ambassador Training Script

This section presents the script used during ambassador training, in French.

Aujourd'hui nous voulons parler à vous sur l'agriculture et le sol. Nous avons vous choisi pour être ambassadeurs pour ce village avec l'objectif à partager la connaissance sur l'engrais chimique et la gestion du sol. Nous voulons vous expliquer l'importance de l'engrais chimique puisque vous pouvez diffuser cette information à les autres villageois. Nous allons vous donner une petite formation et après nous allons vous donner quelque engrais chimique. Une partie de l'engrais est pour vous et la reste vous devez partager avec certaines autres personnes. Vous devez partager aussi l'information dans cette formation.

Pouvez-vous nous dire qu'est-ce que va arriver si on cultive la même culture dans le même champ pour plusieurs saisons consécutives?

- *Réponse cherché: si on cultive la même culture dans le même champ pour plusieurs saisons consécutives, les nutriments dans le sol diminuèrent. Les plants sont plus petits et les rendements sont moins.*

Selon vous, il y a un problème dans ce village avec la fertilité du sol?

- *Laissez-ils discuter si la qualité du sol est un problème – il n'y a pas une réponse correct ou incorrect a cette question. C'est simplement une opportunité pour ils donner leur opinion sur le sol. Assurez qu'on discute pourquoi la fertilité est ou n'est pas un problème.*

Selon vous, quelles sont les façons pour améliorer la fertilité du sol?

- *La meilleure façon est de suivre la Gestion Intégrée de la Fertilité du Sol (GIFS). GIFS inclus:*
  - *Jachère et la rotation des cultures, surtout avec les légumineuses. Si on laisse un champ pour une saison ou plus, il y a du temps pour les nutriments à revenir. Les insectes et les autres animaux vont donner quelque nutriment. Le plus temps que on attend, le plus le sol peut récupérer. Les légumineuses donnent azote au sol. Le défis: on ne peut pas récolter pendent la jachère.*

- *Engrais organique. Il y a beaucoup de nutriments dans les déchets des animaux, les déchets des champs (e.g. les feuilles, les racines, et l'herbe), et les déchets de la cuisine. Si on mélange les déchets et on laisse à décomposer, il devienne une substance riche en nutriment (le compost). Puis on peut appliquer ça aux champs pour améliorer la fertilité du sol. Le défis: ça prend du temps à fabriquer le compost et c'est difficile produire assez pour tous les champs.*
- *Engrais chimique. Similaire à l'engrais organique, mais on fabrique dans les usines avec un processus industriel. Les nutriments dans l'engrais chimique sont très concentrés. Donc une petite quantité d'engrais chimique peut amener la même quantité de nutriments qu'une grande quantité d'engrais organique.*
- *Les terrasses, la gestion des maladies, les canaux pour la pluie, etc.*

Maintenant nous allons expliquer les types d'engrais chimique.

- *NPK: C'est un engrais d'utilisation générale, particulièrement bon pour presque toutes les plantes et l'utilisation avant la plantation. C'est le type d'engrais vous recevrez aujourd'hui.*
- *Urée: Bon à utiliser plus tard dans la saison, et est alors utile pour presque toutes les plantes, mais c'est nécessaire qu'il pluie au maximum deux jours après.*
- *DAP et KCl : ils sont bon pour les cultures que forment les fruits.*

Maintenant nous voulons discuter comment et quand appliquer l'engrais chimique.

- *Mettez l'engrais dans un petit trou, couvrez-le avec du sol, puis mettez le grain au-dessus (épandage localisé avant de semi). Comme ça c'est assurez que le grain est proche aux nutriments de l'engrais, mais en évitent que le grain est en contact avec l'engrais.*
- *Mettre à côté d'une plante individu (épandage localisé après que les plants apparaissent). Il est également possible de mettre l'engrais chimique à côté des plantes que vous cultivez. Cela permet de garantir que les éléments nutritifs sont donnés directement et uniquement à la plante. Cependant, il faut avoir assez du temps ou des travailleurs passer par chaque plante. Et il peut être difficile de savoir à quelle plante vous avez déjà donné l'engrais. C'est important que l'engrais chimique ne touche pas les racines.*
- *Jeter l'engrais sur le champ (épandage à la volée). Le plus simple et le plus rapide pour appliquer des engrais chimiques est en le jetant sur le champ. Les engrais*

chimiques fonctionnent toujours là, mais ils sont moins efficaces. On ne peut pas contrôler la quantité d'engrais qu'on mette proche à chaque semence ou plante. Certaines plantes peuvent facilement reçoivent trop d'engrais, autres très peu.

Quantité approprié pour les cultures importantes:

Culture	Kg/Ha (100m x 100m)
Maïs	300
Manioc	300
Haricots	200

Que pensez-vous est la meilleure méthode pour appliquer des engrais chimiques?

- *Laissez-les discuter quelle est la meilleure méthode. Toutes les méthodes sont efficaces, il y a des avantages et des inconvénients à chacun.*
  - *Jeter sur le terrain est rapide mais moins efficace*
  - *Avant de semi, épandage localisé dans les trous est efficace (engrais ciblé à chaque plante), mais pas possible après le semi*
  - *Après de semi, l'épandage localisé à côté de chaque plante est très efficace (engrais ciblé à chaque plante) et peut être utilisé lorsque les plantes poussent ainsi. Mais il faut avoir plus de temps ou de travailleurs.*

Maintenant nous allons discuter certains problèmes avec l'engrais chimique.

- *Si l'engrais chimique est en contact avec une semence ou une racine, il peut bruler la plante. La plante ne va pas pousser.*
- *Montrez combien de l'espace il faut laisser entre la plante et l'engrais (e.g. une main).*
- *Si on utilise trop d'engrais chimique il peut détruire la plante. Mais si on utiliser très peu d'engrais, le récolte ne sera pas grande.*
- *L'engrais chimique est un poison verser les êtres humains. Il faut ne pas de manger l'engrais chimique, ou même le toucher et puis toucher la bouche. C'est important de porter les gants ou utiliser un sachet sur les mains quand on manipule l'engrais chimique.*
- *L'engrais dissoudre dans l'eau. Il faut garder l'engrais dans un endroit sèche, pas un endroit humide. Si vous avez une palette, vous pouvez mettre l'engrais là.*

- *Certains gens croient que l'engrais chimique changer le goût des cultures. Ce n'est pas vrai. Le goût reste le même.*

Comment on peut obtenir l'engrais chimique?

- *À Bukavu il y a plusieurs pharmacies vétérinaires que vendent l'engrais chimique (e.g. ADVS, proche à Regideso ; ou Lobiko, à Nyawera). Un kg de NPK cout environs 1,7 US$. Les commerçants peuvent donner aussi des conseils sur quel type d'engrais est approprié pour vos cultures.*
- *On peut aussi acheter l'engrais chimique dans certains grands centres. Le prix peut être plus haut qu'à Bukavu, mais le cout de transport pour vous peut être moins.*

Sommaire de la formation.

- *Il y a plusieurs façons pour restaurer les nutriments dans le sol: jachère, la rotation des cultures (surtout avec les légumineuses), l'engrais organique, et l'engrais chimique, et les combinassions de chacune.*
- *Il y a plusieurs façons à appliquer l'engrais chimique: mélanger avec le sol; mettre à cote de la plante; jeter.*
- *Il y a plusieurs moments dans la saison culturelle quand l'engrais chimique est avantageux.*
- *Vous pouvez augmenter les rendements si vous utiliser l'engrais chimique correctement!*

Maintenant nous allons discuter vos responsabilités comme Ambassadeurs.

- *Nous voudrions que vous alliez partager cette formation avec les autres habitants de ce village.*
- *Nous voudrions aussi que vous alliez donner une certain quantité d'engrais chimique à autres chefs de ménage de ce village. En fait, nous allons vous donner 1kg de NPK pour vous, et trois paquets pour donner à trois autres personnes –aussi seulement aux chefs de ménage. Chaque paquet a trois petite sachet de 1kg chacune. C'est essentiel que chaque bénéficiaire prenne 1kg et donne l'autre 2kg à deux autres personnes (1kg a chacun). Comme ça une grande partie du village peut expérimenter avec l'engrais chimique.*
- *Nous allons revenir dans 1-2 semaines pour assurer que l'engrais a était partagé. Pour chaque paquet il y a trois autocollants et trois stylos. Quand vous donnez le*

paquet à le chef de ménage, le bénéficiaire va mettre un autocollant sur le sachet que il va garder. Il va écrire votre nom (le donneur), le date, et l'heure. Pour l'autres deux sachets, quand il les donne a les autres, il va écrire son nom, le date, et l'heure. Cette a dire, chaque fois que un sachet est donné à quelqu'un, il faut afficher un autocollant et écrire le nom du donneur, le date, et l'heure (le nom doit être d'un chef du ménage). Quand nous revenons, nous allons noter qui a donné à qui.

- *Merci pour votre assistance avec ce projet! Avez-vous des questions?*

2.7.3 Protocol Random Card Sorting Game

This section presents the protocol used for the random card sorting game, in French.

Utilisez le Fiche Cout Engrais Chimique. Chaque enquêteur va utiliser une fiche. Cherchez une ID dans votre liste et amenez le participant à un lieu privé. Dites au participant : 'On va commencer avec quelque questions sûr le cout d'engrais chimique que vous pourriez accepter. Les questions sont hypothétiques; nous n'allons pas vendre d'engrais chimique'.

Montrez au participant les dix cartes avec prix de 100FC jusqu'à 5000FC. Demandez au participant d'imaginer qu'il y a un kg de NPK disponible à chaque prix. Demandez au participant à arranger les cartes dans trois piles.

1. *Les prix que le participant ne paierait pas. Cette à dire, quels prix sont trop élevé?*
2. *Les prix que le participant pourrait payer ou pas. Cette à dire, les prix pour lesquelles le participant n'est pas très certain.*
3. *Les prix que le participant paierait. Cette à dire, quels prix sont certainement acceptable?*

Dans le Fiche Cout Engrais Chimique, écrivez :

- *Le prix plus bas parmi les prix que le participant ne paierait pas (Prix Minimum Inacceptable).*
- *Le prix plus élevé dans les prix que le participant paierait (Prix Maximum Acceptable).*

2.7.4 Village Networks

Figure A.2.2 and Figure A.2.3 present the combined networks of the twenty villages in which the ambassadors are central, and those in which they are the isolated villagers, respectively. Selected ambassadors are indicated with black dots.

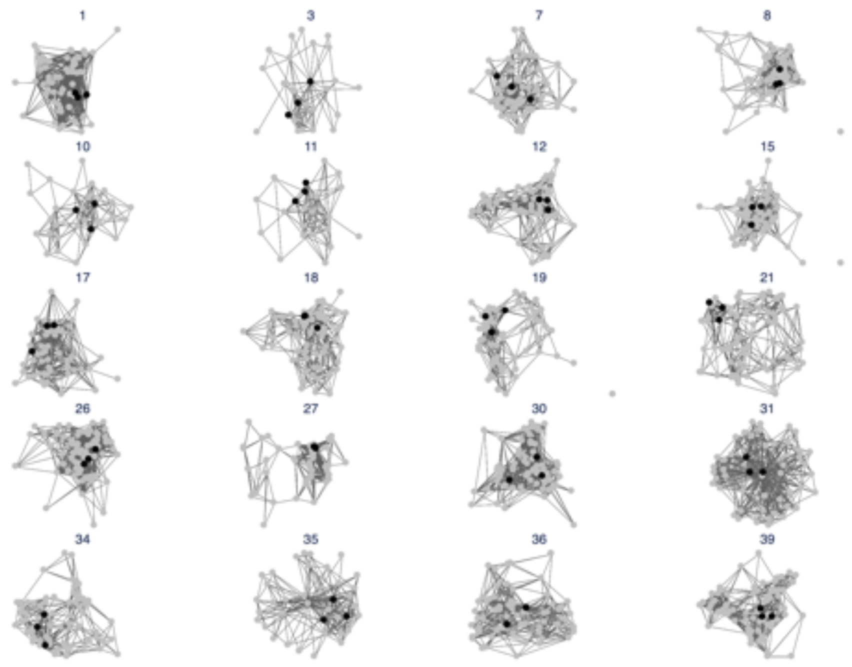


Figure A.2.2 – Central Ambassador Villages

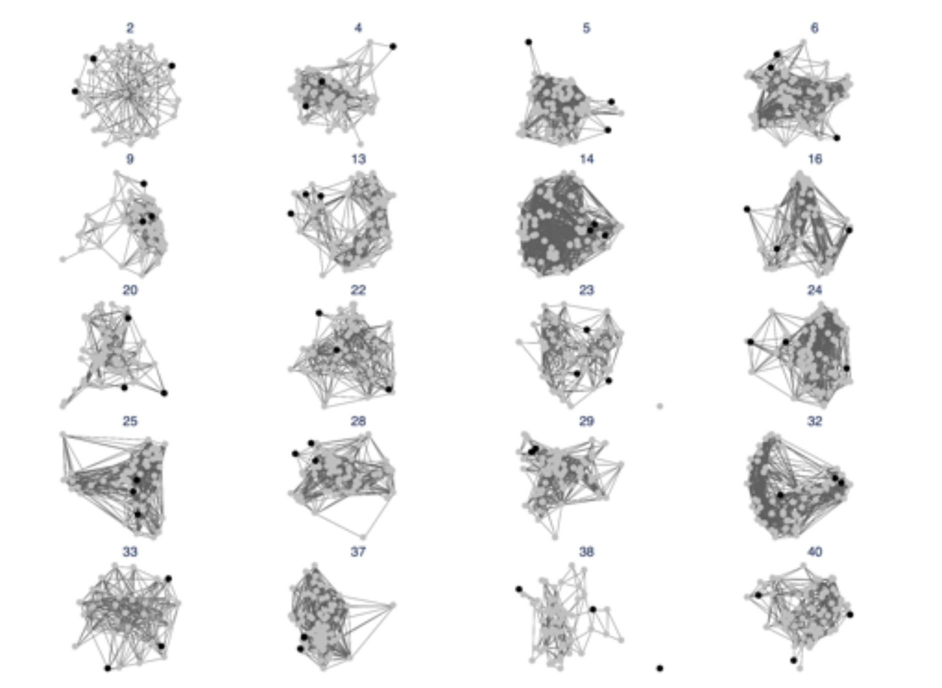


Figure A.2.3 – Isolate Ambassador Villages

2.7.5 Variable Definitions

Table A.2.1 – Variable Definitions

Variable	Visit	Description
Age	1	Continuous. Age in years.
Male	1	Binary. Respondent's gender.
Literate	1	Binary. Whether the respondent can read and write.
Migrant	1	Binary. Whether the person is born in the village (native) or not (migrant).
Ethnic majority	1	Binary. Whether the respondent belongs to the village's largest ethnic group.
War exposure	1	Continuous. Score from 0 to 7. Sum of events experienced by the respondent: Saw fighting; property damage/loss; injured in the war; family member injured in the war; family member killed in the war; migrated because of the war.
Wealth	1	Continuous. Adding up the number of chickens, goats and cows owned.
Farm size	1	Continuous. Size of all respondents' plots added up. Truncated at <100km2.
Likely lead farmer	1	Binary. Whether the respondent is mentioned by at least one other villager to be a potential lead farmer.
Leadership position	1	Binary. Respondent holds any of the following positions: village chief, elder, neighborhood chief, religious leader, teacher, school director, doctor, women leader, youth leader, president agriculture group, other.
Access to village chief	1	Continuous (<30). Number of times the respondent meets with the village chief.
Centrality score	NA	Continuous (0-1). Eigenvector centrality score created by the authors between visit 1 and visit 2.
Family	NA	Binary. Individual indicates to be family with the other individual, or vice versa.
Field neighbor	NA	Binary. Individual indicates to be a field neighbor of the other individual, or vice versa.
Agriculture	NA	Binary. Individual indicates to discuss agriculture with the other individual, or vice versa.
Combined network	NA	Binary. Individual indicates to be family, a field neighbor or discusses agriculture with the other individual.
Physical distance	NA	Continuous (kilometers). Distance between two individuals based on individuals' GPS coordinates.
Social distance	NA	Continuous. Shortest number of ties in the combined network to reach the other individual. Those dyads that are not in the same network receive the (arbitrary) score of 10.
Used fertilizer	1 & 3	Binary. Respondent has ever used fertilizer.
Fertilizer knowledge	1 & 3	Continuous. Score from 0 to 8.5, based on responses to the following five questions: 1) what is the effect of fertilizer on yields? ('Increases yields' is 1 point); 2) what are other effects of fertilizer? ('Earlier harvest' and 'Kills seeds' is each 0.5 points); 3) when is it effective to apply fertilizer? ('Before planting', 'During planting' and 'After planting' is each 1 point); 4) what is the best method to apply fertilizer? ('mix in soil before planting' and 'Put next to the seed/plant' are each one point, and 'Throw on the field' is 0.5 points); and 5) what is the price of fertilizer in Bukavu? (any amount between 1.5 and 1.7 dollars is 1 point, between 1.2 and 1.5 dollars or between 1.7 and 2 dollars is each 0.5 points).
Willingness to pay for fertilizer	2 & 3	Continuous (US$). Based on random card sorting game.
Fertilizer transmission	3	Binary. Whether fertilizer was transferred between a dyad. Based on two sources. The household survey where we asked receiver from whom they received fertilizer. And a tracking exercise.
Information transmission	3	Binary. Whether fertilizer information was transferred between a dyad. Based on the household survey where we asked respondents whether and from whom they received fertilizer information.

Notes: Centrality score, family, field neighbor, agriculture, combined network, physical distance and social distance are constructed by the authors between visits 1 and 2.

2.7.6 Balance

Table A.2.2 shows characteristics of heads of households, collected during visit 1 before the intervention, by treatment status. We have balance across treatment conditions.

Table A.2.2 – Household Head Characteristics at Baseline

	Central			Isolate			
	N	Mean	SD	N	Mean	SD	Diff
Age (years)	1306	45.26	17.14	1173	46.75	17.83	1.491
Male (=1)	1306	0.68	0.47	1173	0.68	0.47	0.003
Literate (=1)	1305	0.43	0.50	1173	0.49	0.50	0.058
Migrant (=1)	1305	0.31	0.46	1173	0.36	0.48	0.048
Ethnic Majority (=1)	1306	0.97	0.17	1173	0.99	0.10	0.020
War exposure (7 events)	1303	3.20	1.29	1171	3.41	1.35	0.215
Wealth index	1306	2.63	3.60	1173	2.73	4.27	0.097
Farm size	1240	4.90	9.99	1126	5.21	11.25	0.311
Likely Lead Farmer (=1)	1352	0.21	0.41	1232	0.22	0.41	0.009
Leadership position	1306	0.31	0.46	1173	0.33	0.47	0.015
Access to village chief	1281	6.48	8.50	1152	6.08	8.11	-0.399
Centrality score	1352	0.32	0.26	1232	0.33	0.27	0.008
Used Fertilizer (=1)	1287	0.07	0.26	1160	0.06	0.24	-0.015
Fertilizer knowledge (max 8.5 points)	1306	1.44	1.69	1173	1.45	1.72	0.011
Willingness to pay for fertilizer (USD)	349	1.60	1.47	344	1.50	1.34	-0.105

Note: Visit 1 data from all household heads, except for willingness to pay that is based on visit 2 data from a subset of households. P-values for significance stars based on regressions with standard errors clustered at the village level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.7.7 CONSORT Diagram and Attrition

Figure A.2.4 presents a CONSORT-style flow diagram with the details on the number of villages targeted and visited, and surveys targeted and collected.

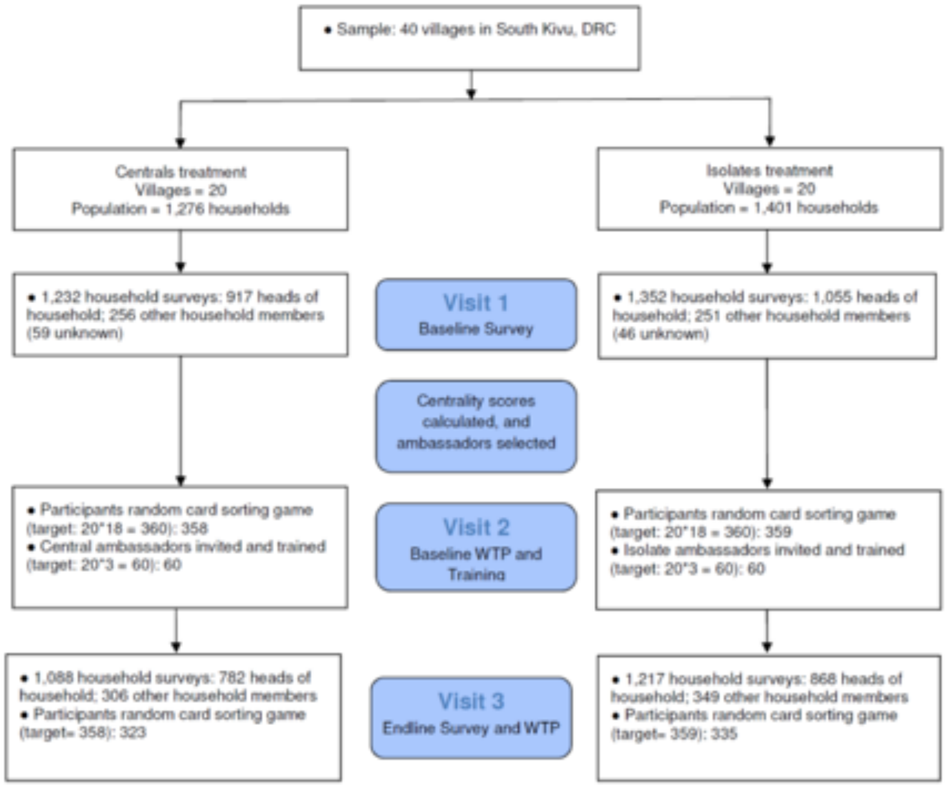


Figure A.2.4 – CONSORT Diagram

Notes: CONSORT diagram summarizing the organization of units, assignment and measurement strategies.

2.7.8 Distribution Behavior by Second-Stage Ambassadors

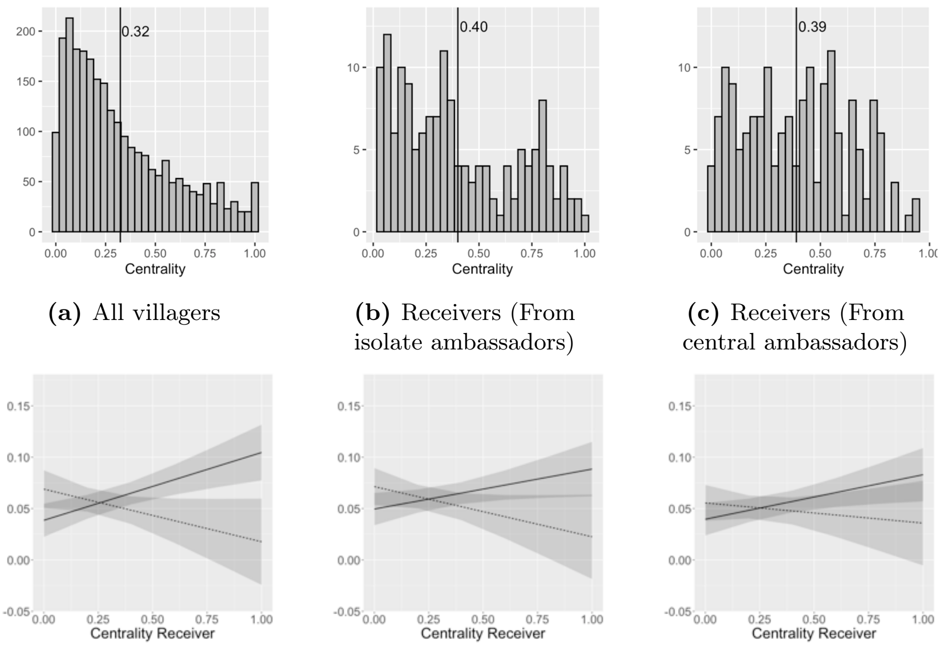
Table A.2.3 – Distribution Behavior by Second-Stage Ambassadors

	(1) Fertilizer (Endline Survey)	(2) Information (Endline Survey)	(3) Fertilizer (Tracking Exercise)	(4) Fertilizer (Endline Survey)	(5) Information (Endline Survey)	(6) Fertilizer (Tracking Exercise)
Centrality (R)	−0.003 (0.003)	−0.004 (0.003)	−0.004 (0.003)	0.005 (0.007)	0.000 (0.005)	0.002 (0.005)
Age (R)	0.000 (0.003)	−0.001 (0.002)	0.000 (0.002)	0.001 (0.004)	−0.001 (0.003)	0.002 (0.003)
Male (R)	0.002 (0.002)	0.000 (0.002)	0.000 (0.002)	0.002 (0.003)	−0.001 (0.003)	0.003 (0.004)
Literate (R)	−0.003 (0.002)	−0.001 (0.002)	−0.003 (0.002)	−0.004 (0.004)	−0.002 (0.003)	−0.004 (0.003)
Migrant (R)	0.003* (0.002)	0.004* (0.002)	0.003 (0.002)	0.001 (0.003)	0.000 (0.003)	0.005 (0.003)
Ethnic Majority (R)	−0.001 (0.001)	0.000 (0.001)	0.002 (0.002)	−0.001 (0.002)	0.000 (0.001)	0.004*** (0.001)
War Exposure (R)	−0.001 (0.002)	0.001 (0.002)	−0.002 (0.002)	0.003 (0.004)	0.003 (0.003)	−0.001 (0.003)
Wealth (R)	0.004* (0.002)	0.002 (0.002)	0.001 (0.002)	0.004 (0.005)	0.004 (0.004)	−0.001 (0.005)
Farm Size (R)	0.002 (0.003)	0.000 (0.002)	0.002 (0.002)	0.006 (0.006)	0.004 (0.005)	0.002 (0.004)
Family	0.008*** (0.003)	0.009*** (0.003)	0.011*** (0.003)	0.010** (0.005)	0.011*** (0.004)	0.009* (0.005)
Field Neighbors	0.010** (0.004)	0.009*** (0.003)	0.009*** (0.003)	0.012* (0.007)	0.010* (0.005)	0.007* (0.004)
Agriculture	0.008*** (0.002)	0.004** (0.002)	0.004* (0.002)	0.006** (0.003)	0.004** (0.002)	0.001 (0.003)
Physical Distance	−0.015*** (0.002)	−0.010*** (0.002)	−0.013*** (0.002)	−0.021*** (0.005)	−0.013*** (0.004)	−0.015*** (0.003)
Social Distance	0.000 (0.003)	−0.003 (0.003)	−0.002 (0.002)	0.008 (0.009)	0.002 (0.006)	−0.003 (0.008)
Treatment = village has central entrypoint				−0.010 (0.006)	−0.003 (0.004)	0.001 (0.004)
Treat * Centrality (R)				−0.015* (0.008)	−0.007 (0.006)	−0.013** (0.006)
Treat * Age (R)				−0.002 (0.005)	−0.001 (0.004)	−0.004 (0.004)
Treat * Male (R)				0.001 (0.005)	0.003 (0.004)	−0.006 (0.005)
Treat * Literate (R)				0.003 (0.005)	0.002 (0.004)	0.001 (0.005)
Treat * Migrant (R)				0.005 (0.004)	0.007** (0.004)	−0.005 (0.004)
Treat * Ethnic Majority (R)				−0.002 (0.004)	−0.001 (0.004)	−0.007** (0.003)
Treat * War Exposure (R)				−0.007 (0.005)	−0.004 (0.004)	−0.001 (0.004)
Treat * Wealth (R)				−0.001 (0.005)	−0.003 (0.004)	0.003 (0.005)
Treat * Farm Size (R)				−0.008 (0.007)	−0.007 (0.005)	0.001 (0.005)
Treat * Family				−0.006 (0.006)	−0.005 (0.005)	0.004 (0.007)
Treat * Field Neighbors				−0.004 (0.008)	−0.003 (0.007)	0.003 (0.006)
Treat * Agriculture				0.003 (0.004)	0.000 (0.004)	0.008 (0.005)
Treat * Physical Distance				0.010* (0.006)	0.006 (0.005)	0.003 (0.004)
Treat * Social Distance				−0.011 (0.009)	−0.008 (0.007)	0.001 (0.009)
Observations	5763	7783	5329	5763	7783	5329

Notes: Standard errors clustered at the village and sender level. Randomization fixed effects included. Based on data from visit 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variable definitions can be found in Table A.2.1.

2.7.9 Distribution Fixing for Counterfactual Ambassadors

This section explores whether results changes when fixing for the existence of counterfactual ambassadors. Counterfactual ambassadors are those who would have been chosen as ambassadors had the village been assigned the opposite treatment. So in a central (isolate) treatment village the counterfactual ambassadors are the three households with the lowest (highest) eigenvector centrality. Any transfer to a counterfactual ambassador is replaced with a transfer to the household with the fourth lowest (highest) eigenvector centrality.



(a) All villagers (b) Receivers (From isolate ambassadors) (c) Receivers (From central ambassadors)

(d) Fertilizer (survey) (e) Information (survey) (f) Fertilizer (tracking)

Figure A.2.5 – Distribution and Receiver Centrality by Treatment Excluding Counterfactual Ambassadors

Notes: This set of graphs replaces transfers to counterfactual ambassadors to those just below. Counterfactual ambassadors would have been ambassadors have been had the village been assigned the opposite treatment (and would then not have been able to receive). Transfers to these ambassadors are replaced with a transfer to an individual one step lower (higher) in the centrality ranking in central (isolate) villages. Panel (a) displays a histogram of receiver centrality for all villagers (unchanged from Figure 2.3), panel (b) for the receivers in villages with isolate ambassadors, and panel (c) for receivers in villages with central ambassadors. In panels (d) to (f), the y-axis is the probability of a transfer, the x-axis is the centrality of the receiver, isolate (central) ambassadors are indicated by the solid (dashed) line, and bounds are 95% confidence intervals

Table A.2.4 – Distribution Behavior by First-Stage Ambassadors, transfers to Counterfactual Ambassadors replaced

	(1) Fertilizer (Endline Survey)	(2) Information (Endline Survey)	(3) Fertilizer (Tracking Exercise)
Centrality (R)	0.017***	0.010***	0.010**
	(0.003)	(0.004)	(0.004)
Age (R)	0.002	0.004	0.009**
	(0.004)	(0.005)	(0.004)
Male (R)	0.000	0.001	0.012***
	(0.005)	(0.004)	(0.003)
Literate (R)	0.009***	0.005	0.009**
	(0.003)	(0.004)	(0.004)
Migrant (R)	0.001	−0.003	0.001
	(0.004)	(0.004)	(0.005)
Ethnic Majority (R)	0.003	0.002	−0.007**
	(0.002)	(0.002)	(0.004)
War Exposure (R)	0.001	−0.001	0.001
	(0.005)	(0.005)	(0.004)
Wealth (R)	−0.003	−0.003	0.002
	(0.004)	(0.005)	(0.006)
Family	0.013	0.013	0.009
	(0.013)	(0.013)	(0.015)
Field Neighbors	0.007	0.024***	0.014
	(0.007)	(0.009)	(0.010)
Agriculture	0.002	0.006	0.009
	(0.008)	(0.009)	(0.009)
Physical Distance	−0.022***	−0.023***	−0.014**
	(0.006)	(0.006)	(0.005)
Social Distance	0.003	0.003	−0.001
	(0.004)	(0.004)	(0.004)
Treatment = village has central entrypoint	−0.007	−0.006	−0.003
	(0.008)	(0.007)	(0.009)
Treat * Centrality (R)	−0.031***	−0.023***	−0.015*
	(0.006)	(0.006)	(0.008)
Treat * Age (R)	−0.005	−0.009	−0.004
	(0.008)	(0.008)	(0.009)
Treat * Male (R)	0.006	0.000	0.005
	(0.008)	(0.007)	(0.007)
Treat * Literate (R)	−0.011*	−0.005	−0.001
	(0.006)	(0.006)	(0.007)
Treat * Migrant (R)	0.000	0.005	0.001
	(0.008)	(0.007)	(0.009)
Treat * Ethnic Majority (R)	0.002	−0.003	0.012***
	(0.002)	(0.005)	(0.004)
Treat * War Exposure (R)	0.016**	0.018***	0.010
	(0.007)	(0.007)	(0.008)
Treat * Wealth (R)	0.006	0.005	−0.001
	(0.006)	(0.006)	(0.007)
Treat * Family	−0.009	−0.006	−0.005
	(0.014)	(0.014)	(0.016)
Treat * Field Neighbors	−0.004	−0.021**	−0.013
	(0.009)	(0.010)	(0.011)
Treat * Agriculture	0.009	0.005	0.003
	(0.009)	(0.010)	(0.010)
Treat * Physical Distance	−0.010	0.000	−0.015**
	(0.008)	(0.010)	(0.007)
Treat * Social Distance	−0.010	−0.008	−0.008
	(0.006)	(0.007)	(0.008)
Observations	4614	4818	4717

Notes: This set of regressions replaces transfers to counterfactual ambassadors to those just below. Counterfactual ambassadors would have been ambassadors had the village been assigned the opposite treatment (and would then not have been able to receive). Transfers to these ambassadors are replaced with a transfer to an individual one step lower (higher) in the centrality ranking in central (isolate) villages. Standard errors clustered at the village and sender level. Randomization fixed effects included. Based on data from visit 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficients for farm size not shown (all insignificant). Variable definitions can be found in Table A.2.1.

2.7.10 Deviations from Pre-Analysis Plan

This study is preregistered at EGAP (www.egap.org, ID= 20151202AA). Below we discuss deviations from the pre-analysis plan, additional analyses not preregistered and items that were preregistered but not undertaken.

First, we list the deviations from the pre-analysis plan. These deviations do not change the results of the study.

- Table 2.2: Initially, we specified our balance check across treatment arms using a t-test. We changed this to a linear regression with clustered standard errors to correct for within-village correlation.
- Table 2.4: Initially, we had specified two additional dependent variables: the width of distribution and the speed of diffusion. We dropped the first because it is a village-level outcome. We dropped the latter because data quality was low. We planned to use a fixed and random effects model, and compare the two. For simplicity, the main text uses a simple difference-in-difference model instead.
- Table 2.3: Initially, we specified a Chow test to compare across coefficients across treatment arms, instead we interact with the coefficients treatment status. We also did not pre-specify that we would standardize all variables. We pre-specified a logit model, instead we use a linear probability model as this makes the interpretation of the coefficient easier. Given the Congolese context, we added three explanatory variables: ethnic majority, war exposure and farm size. We also removed two explanatory variables: whether villagers were in the same agricultural group or whether they went to the same church. Both variables would be constructed based on the matching of names, which proved not feasible.

The following was not initially pre-registered:

- Figure 2.2 was added to illustrate the relationship between centrality and political marginalization.
- Figure 2.3 was added to illustrate the effects found in Table 2.3.
- Table 2.6 in the discussion section was added to illustrate the relationship between a villager's centrality and characteristics related to farming, and wealth.

Finally, we list analyses we preregistered but have not included in the document.

- Given that we undertook village censuses, we proposed to examine differences between those individuals that were interviewed and those that were not, the latter based on information from the village chief or neighbors. Because response rates are very high we did not do so.

Chapter 3

Social Networks and Social Preferences: A Lab-in-Field Experiment in Eastern DRC

Considerable research points to social ties, and the social networks underlying these ties, as the underlying driver of pro-social behaviors upon which large-scale societal organization is based. However, little is known about the empirical relationship between social network position and pro-social preferences. Based on very detailed original network data and a lab-in-the-field experiment, we explore the relationship between social networks and social preferences of trust and trustworthiness. We explore whether an individual's observed social preferences are correlated with an individual's centrality within the network structure. Our results indicate that individuals with high centrality are more trusting and more trustworthy than individuals with lower centrality. We also find that measures that explore the type of relationship between players are predictive of trusting behavior. Being directly linked in the social network increases trusting and trustworthy behavior. Having a lot of mutual connections to others with the other player increases trusting behavior as well – despite these others not being present during the game. We take this as evidence that individuals use their social network connections as 'social collateral'.

Publication status: Hofman, P., Larson, J.M., Ross, M., Van der Windt, P. and Voors, M., 2020. Social Networks and Social Preferences: A Lab-in-Field Experiment in Eastern DRC. Working Paper.

3.1 Introduction

Individuals do not make economic decisions in a vacuum. Instead, individuals are embedded in rich social environments comprised of both formal and informal institutions that shape preferences and drive behavior (Douglas North, in Mwabu et al. (2001)). The relevance of these social institutions has been well documented: individuals both inside and outside the lab contribute to societal investments even at a direct personal wcost. People assess risky investments in part based on pre-existing feelings of trust between the investor and investee (Berg et al., 1995). Informal markets emerge and persist due to the interpersonal ties connecting their participants (Greif, 1993; Landa, 1981). Individuals in these situations exhibit trust and trustworthiness, the same pro-social behaviors that underpin the proper functioning of societies by generating gains from group living even absent strong formal governing institutions. While much research points to social ties, and the social networks comprised of these ties, as the underlying driver of these trust-based interactions, little is known about the empirical relationship between social network position and pro-social preferences. Based on original network data and a lab-in-the-field experiment, we explore the relationship between social networks and social preferences of trust and trustworthiness. We explore whether an individual's observed social preferences are correlated with an individual's centrality within the network structure.

Departing from standard economic theory which holds that preferences are exogenous and stable, laboratory and lab-in-the-field experiments have shown that preferences may co-evolve with the social context (Bowles, 1998; Bowles and Polanía-Reyes, 2012; Brosig et al., 2007; Volk et al., 2012; Voors et al., 2012). Within a social context, social preferences can vary: while there is substantial cross-cultural variation, there is also substantial within-culture variation of social preferences (Bowles, 1998; Croson and Gneezy, 2009; Fehr and Hoff, 2011; Henrich et al., 2005, 2010). One source of variance is salient social cleavages. When individuals interact with in- and out-group members, individuals tend to exhibit more pro-social behavior toward the in-group. This dynamic is shown for cleavages defined by gender (Croson and Gneezy, 2009), ethnicity and race (Benjamin et al., 2007; Fershtman and Gneezy, 2001), as well as for exogenously imposed or self-selected groupings (Ben-Ner et al., 2009; Chen and Li, 2009; Halevy et al., 2008). Studies exploring why this is the case, suggest that social networks that interconnect members of an in-group are part of the answer (Habyarimana et al., 2007; Miguel and Gugerty, 2005).

Social networks may bear on pro-sociality for several reasons. We test several predictions from Karlan et al. (2009)'s Social Collateral theory, which explores the determinants of trust between individuals. The theory posits that individuals use their relations in a network as social collateral to facilitate trust-based exchanges (such as loans). The idea is that social relations carry value and that breaking a promise in a trust-based exchange destroys social relations. If the value of these relations is higher than the gain from betrayal, the exchange can take place. This generates testable hypotheses which we apply on real-world network data in a developing country context.

The standard approach to studying the effect of networks on behavior is to assign participants to an artificial 'social network' in a lab, vary the connections in that network to vary who could punish and be punished by whom, and observe behavior. This approach can cleanly identify a causal effect of a network characteristic; however, the network is artificially-imposed and thus cannot illuminate whether participants draw on their own social network for these purposes. Our study differs in that we focus on the relationship between individuals' positions within their real social networks and their behavior in interactions with one another. This design sacrifices some precision in our causal effects – we cannot perfectly disentangle the mechanisms by which the network affects outcomes – in exchange for greater external validity with real-life networks. Specifically, we begin by measuring the social network among individuals in rural villages in the Democratic Republic of Congo. We then recruit individuals to play a lab-in-the-field experiment (a trust game) based on their eigenvector centrality in the village network, including some with low, middle, and high values. Individuals observe the other participants in the games, and so are free to condition their behavior on real information, including social network information. Finally, we relate individuals' eigenvector centrality in their village's social network to their observed social preferences in this game.

Eigenvector centrality is a network statistic that captures the extent to which a person is highly-connected to others in the network, and the extent to which a person's connections are themselves highly-connected to others (Bonacich, 2007; Borgatti, 2005). Eigenvector centrality thus bears on one's access to information flowing through a network, as well as those whom one could access easily to report a bad reputation or coordinate social sanctioning. Consequently, our design admits both post-experimental punishment mechanisms and reputation effects, a perfect context for testing whether individuals use their social network connections as collateral.

The game played is a one-on-one trust game that tests trusting and trustworthy behavior. In this way, we have information on a person's general level of trust and trustworthiness (how they behave overall in the trust game) and a person's selective level of trust and trustworthiness (how they behave when playing certain other people in the trust game).

Our results indicate that network eigenvector centrality ('centrality' for short) matters in one-on-one interactions. Individuals with high centrality are more trusting and more trustworthy than individuals with lower centrality. We also find that dyadic relationships between players predict cooperation. Being directly linked in the network increases trusting and trustworthy behavior. Also, when players have a lot of network links in common (e.g. links to others, not present during the game) this increases their trusting behavior. We also see that individuals choose to send more when the other player has a lower centrality than them, both when sending and returning during the trust game. We take these results as evidence for the validity of the social collateral theory.

The existing literature on social preferences and social networks is discussed in Section 3.2. Section 3.3 offers a detailed overview of the study design. Section 3.4 describes the sample, and our results are presented in Section 3.5. Section 3.6 concludes.

3.2 Social Networks and Social Preferences

Our study focuses specifically on the relationship between the social network in which an individual is embedded and their observed social preferences.

3.2.1 Social preferences

While social networks capture the external relationships that define daily social interaction and access to information and resources, social preferences are the internal set of preferences determining rank order of different allocations of material benefits between oneself and others. This set of preferences includes interpersonal values of altruism, fairness, cooperation, trust, and inequality aversion. Social preferences are most often measured through laboratory experiments designed to elicit each preference under varying conditions and have become a focal point within development economics (Camerer and Fehr, 2002; Cárdenas and Carpenter, 2008). Significant research within developing country contexts has been undertaken with experimental designs that vary the anonymity, group composition, information, and monitoring and punishment mechanism effects on observed social preferences through lab-in-field experiments (for a review see Cárdenas and Carpenter (2008)). Social preferences have been studied across cultural contexts (Cardenas et al., 2000; Henrich et al., 2010, 2001; Jakiela, 2011) and linked to economic outcomes of labor markets (Barr and Serneels, 2009; Carpenter and Seki, 2011; Fehr et al., 1993), financial decisions (Karlan, 2005), markets for goods and services (List, 2005), tax systems (Alesina and Angeletos, 2003), and environmental resource management (Bouma et al., 2008) amongst others.

3.2.2 Eigenvector centrality

Numerous measures of centrality for capturing the importance of a network member for the flow of information and resources exist. Each has its own functional form that uniquely captures the varying traits or characteristics inherent within flow patterns of different resources or information (Borgatti, 2005). We use eigenvector centrality, where an individual is considered more influential the better connected they are to central individuals within the network, making it a recursive measure of connectedness (Bonacich, 2007). This recursive function is solved by using degree centrality weightings for an iterative estimation approach. While eigenvector centrality is popularly considered a measure of influence, it can also be thought of as the probability of an individual participating in any resource flows that travel via unrestricted walks. In other words, for resources (such as information) that can take any possible path throughout the network, eigenvector centrality provides an approximate likelihood of a given node participating in this path. Information is the most common resource that flows through networks both divisibly and unrestricted, making eigenvector centrality an appropriate measure for capturing the extent to which an individual participates, influences, and potentially controls, information flows within their network (Borgatti, 2005).[1]

3.2.3 Trust and Social Collateral

Karlan et al. (2009)'s theory of social collateral explores what determines trusting behavior between individuals. The model centers around the value of relationships between individuals. If person A wants to borrow a good from person B, this requires that person

[1] Banerjee et al. (2013) devise an information-diffusion specific measure of centrality that they term diffusion centrality. This measure is closely related in mathematical form to eigenvector centrality and so we persist with eigenvector centrality as the determining network parameter of interest within our study design.

B trusts A to return the good after the specified period.[2] If the value of the relationship between A and B exceeds the value of the borrowed good A will return the good, as not returning the good would leave A with the good but also with the relationship destroyed, resulting in a net loss. But not just the relation between A and B matters, relations that sit between A and B matter as well. If there is a person C who has a relation with both A and B, the value of this relation can additionally serve as social collateral. If person A betrays person B, person B will share this knowledge with all his relations, who will likely end their relations with person A. Therefore, the level of trust between two individuals in a network is determined by the entire network's structure. The theory is based on earlier work by sociologists such as the Structural Embeddedness Theory by Granovetter (1985) and the Network Closure Theory by Coleman (1988).

Our data is well suited to test this theory. To overcome not knowing the value of a specific relationship we assume that all relations are valued equally. This generates several testable hypotheses:

1. Paired individuals with a direct social network connection will exhibit higher trust
2. Paired Individuals that share more mutual connections will exhibit higher trust
3. Individuals with a higher eigenvector centrality will show higher trust
4. Paired individuals with a lower social distance will exhibit higher trust[3]

The first two predictions follow directly from the theory. The third and fourth are based on an extension of the theory when information flow is imperfect. It centers around eigenvector centrality's main attribute: access to information. Individuals with a high eigenvector centrality have access to more individuals, who themselves have access to more individuals. They can use this access to help infer the value of relationships that they are not directly a part of themselves. This allows them to put a non-zero value on relations they do not know, increasing the total collateral available.[4] Similarly, a lower social distance between players also allows them better knowledge on the value of others' relationships. Karlan also alludes to these hypotheses in the paper by referring to high/low closure nodes, a similar concept to eigenvector centrality.

[2]This assumes the absence of formal enforcement mechanisms, which is not unusual given our rural and poor context.

[3]Social distance is the length of the shortest path, along social network lines, between two individuals. For directly linked individuals the distance is 1.

[4]We assume that when this information is unknown individuals will err on the side of caution and assume a zero/low value of the relation.

3.2.4 Contribution to the Literature

We contribute to this literature in two ways. First, we use full network data at the community level. By utilizing community social network data we are able to explore how an individual's position within their most relevant social network is related to observed social preferences. By using real-world networks we contextualize decisions to capture both within-game and ex-post social forces that potentially relate to social preference behavior. While this compromises the identification of causal relationships, it indicates how social preferences and social networks interact within real-world networks.

Second, the use of full network data allows us access to a large number of social network metrics, allowing us to directly test hypotheses generated by the Social Collateral theory. By testing this we gain insight into the wider applicability of the theory in a developing-country context. Besides testing hypotheses stemming from this theory we also engage in a more exploratory exercise to determine what other (network) variables determine trust.

3.3 Experimental design

To test our claims we combine social network data of household heads with a lab-in-the-field experiment. To collect data for this study we visited forty villages in Eastern DRC twice. The first visit entailed a household survey with all household heads to collect social network information as well as socio-economic data. The second visit occurred approximately one month later, during this visit we implemented the lab-in-the-field experiment.

3.3.1 Obtaining network information

Our research assistants first conducted a full village census of all heads of household in the village, during which the head of household's full name, age, and gender was recorded as well as whether other adults were present in the household. Upon completion of the census, each household head was individually interviewed.

The survey consisted of two parts. The first part collected information about basic socio-economic characteristics, such as demography as well as income and related questions. In the second part we collected social network data. Specifically, we aimed to obtain data on three types of networks. The first is the family network: whether the head of the household is biologically related to any member within the other households.[5] Second, we ask about the field-neighbor network. Whether the head of the household's fields borders a field owned by any member of the other households. Third, we measure the agriculture network. Whether the head of the household discusses agricultural-related topics with anybody else in the other household. We focus on these three networks because they are most closely associated with our interest. That is, Kendzior et al. (2015) found that these three networks are the three predominant channels via which agricultural resources were shared in the same geographical area. We conducted several pilots that included more networks.[6] Pilot results showed that these three dimensions were the most distinct from one another and thus captured maximum variation while minimizing the number of network survey questions.

To elicit network ties, each relationship type was its own survey question in which the interviewer informed the respondent that each name from the census would be read aloud. For each name read aloud the respondent would indicate 'yes' or 'no' on whether there existed between them the relationship in question.[7] Household heads had their names read off the roster individually, but all other household adult members were grouped into a single category and phrased 'Are you [connection type] with any other adults within this household?'

Once network data along the three connections were collected for all household heads the data was first collapsed to the household level. If any connection exists between a household head and a member of another household, the two households are said to be linked. The result was three separate network graphs with each node representing a single household and each connection being unweighted and undirected. Secondly, we collapsed these three to a single network graph combining all three connection types. The result is a single network graph for each community in which each node is a household and each link represents the presence of any measured connection being declared from one or both of the household nodes. The eigenvector centrality score for each household was calculated using the respective community's composite social network graph. To

[5] Specifically, we use whether the other person is biologically related to a maximum of the third degree (this is a well-understood term in Congo). This does not include the wife's family; it has to be through descent.

[6] In total we conducted three pilots. The other networks were friends (the problem was that everyone was everyone's friend), and work on another person's farm (that overlapped with the other networks).

[7] This approach was considerably more time-intensive than alternative strategies in which names were offered up by the respondent for each relationship but overcame potential measurement error arising from memory-based recall. This approach also offered the benefit of reducing the misidentification of individuals in a context with frequently repeated names.

increase comparability of a household's eigenvector centrality score across communities, eigenvector centrality scores are transformed at the community level, with scores ranging from 0 (lowest eigenvector centrality) to 1 (highest eigenvector centrality) in each community.

The participants in our lab-in-the-field experiments were selected based on the rank order of households' eigenvector centrality scores. We selected the six household heads with the highest centrality score, whom we call 'Centrals'. We also selected those with the lowest centrality score, the 'Isolates'. Finally, we also selected the six heads of households with the median centrality scores, the 'Middlings'. Figure 3.1 shows a picture of the social network in one village, with the centrals, middlings, isolates and non-players highlighted in different colors. Clearly, Centrals are clustered near the center of the graph and have more connections, while isolates are near the fringes with few connections.

About one month after the first visit, we revisited each village to conduct the trust game with selected participants. We discuss this now.

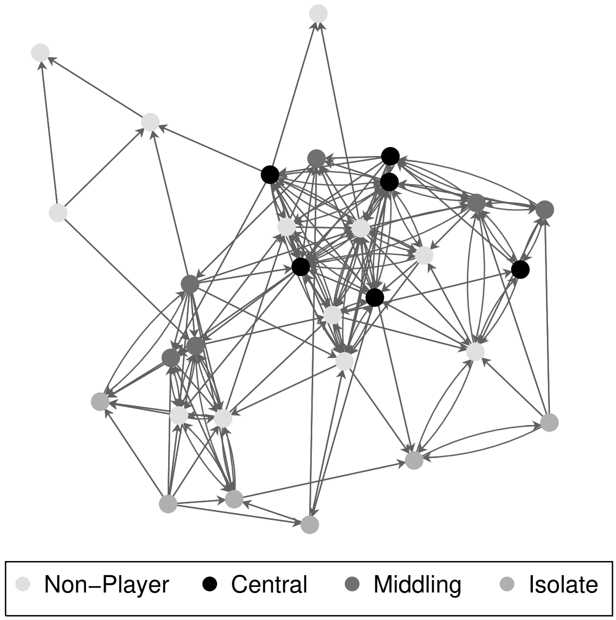


Figure 3.1 – Players within their network in one village

Graph shows combined, directed network for one village, based on survey data in visit 1. Non-respondents are excluded. The network is the unity of the Blood Family, Field Neighbors and Agricultural Discussion Partners networks, which our centrality measures are calculated from. Typology of players is based on their eigenvector centrality. Networks elicited using the list method.

3.3.2 Measuring trust and trustworthiness

To measure trust and trustworthiness we make use of the trust game, following Berg et al. (1995). In this game, two participants are partnered together as Player 1 and Player 2. From an initial endowment of 15 tokens Player 1 decides how many, if any, they wish to share with their partner. Each token was valued at 100 Congolese Franc. The total endowment (1,500 FC) is equivalent to a day's work. Any token amount that is shared is tripled and given to Player 2. Player 2 is then given the opportunity to return any amount of this received money back to Player 1. Both Player 1 and Player 2 are informed of the identity of their partner. The number of tokens shared by Player 1 captures the level of trust Player 1 has in Player 2. The number of tokens returned by Player 2 is a measure of Player 2's trustworthiness. We implement this experiment following a round-robin approach, where each participant is paired one time with the other participants. In other words, each participant is 17 times Player 1, and 17 times Player 2. The order with whom they played was randomized to avoid ordering effects.

After the trust game, each participant received compensation based on their decisions made with one randomly selected partner-pairing from the trust game (paid for rounds played with that partner as both Player 1 and Player 2). Participants were paid out at the end of the day so participants were not aware of their realized earnings at any point during gameplay.[8]

3.3.3 Estimation strategy

We first explore the correlation between individual characteristics and trusting / trustworthy behavior. This allows us to test hypothesis 3 on the relation between centrality and behavior. We estimate the following equation:

$$\boldsymbol{Y}_{ij} = \beta_0 + \beta_1 Centrality_i + \beta_2 Centrality_j + \boldsymbol{\beta_3 X}_{ij} + \gamma_r + \alpha_z + \varepsilon_i \qquad (3.1)$$

where $\boldsymbol{Y}_{ij}$ is the number of tokens contributed (returned) by individual i with partner j. $Centrality_i$ is an indicator variable for the centrality type of individual i, where $Centrality_i \subset [central, middling, isolate]$. $Centrality_j$ is the same for their partner. $\boldsymbol{X}_{ij}$ is a vector of individual characteristics for both Player 1 and Player 2. Included in this vector is age, literacy, migrant status, an index of income,[9] highest education achieved (none, primary, secondary, tertiary). Finally, γ_r are round fixed effects, α_z are village fixed effects to control for localized social norms that may affect general expectations of social behavior, and ε_i is the residual error term clustered at the individual decision-maker level (Player 1 for outcomes of trust and Player 2 for outcomes of trustworthiness).

[8]In addition, to minimize sharing of private information between participants, each player was incentivized not to share information regarding their decisions throughout the day with 'silence tokens'. Silence tokens represented additional bonuses of 50 Congolese francs to be earned by not discussing private information and were earned for each round of play. If players were overheard sharing information that was meant to remain anonymous or private, silence tokens for that round were confiscated and noted on record sheets. This rarely happened.

[9]Included variables are: # of chickens owned, # of goats or sheep owned, # cows owned, land size, land access (score). The index is made using the Stata program WMEANEFFECTS, based on the approach by Kling et al. (2007)

Note that in the regressions related to trustworthiness, we also control for the number of tokens received by Player 2 from Player 1.

To test whether individuals indeed use their social connections as social collateral we also run the following dyadic equation, which is at the pair-level:

$$\begin{aligned} \boldsymbol{Y}_{ij} = \beta_0 + \beta_1 \Delta CentralityHigher_{ij} + \beta_2 \Delta CentralityLower_{ij} \\ + \beta_3 DirectLink_{ij} + \beta_4 SharedConnections_{ij} + \beta_5 ShortestPath_{ij} \\ + \boldsymbol{\beta_6 X}_{ij} + \gamma_r + \alpha_z + \varepsilon_i \end{aligned} \tag{3.2}$$

This equation includes several dyadic relational network characteristics between the two partners. Specifically, we measure the relationship between two partners in four ways. First, the difference in eigenvector centrality score between individual i and partner j. This is split up into two parts: a variable for when individual i is more central than j, and a variable for when individual i is less central than j.[10] Second, an indicator for whether i and j are directly connected through one of the network links. Third, the proportion of direct network connections shared by i and j within their overall community social network. Fourth, the shortest path, along network lines, from player i to player j (This distance is 1 when direct link is true).[11] We once again include the same set of control variables for both player 1 and 2.

3.4 Data and Sample

Before moving to the results in the next section, we first introduce our data and the participants.

3.4.1 Data and attrition

For the census upon which we base our network measures we took every effort to ensure a high level of response in our census, which translates into a very high response rate of

[10] This is because motivations for giving up can be very different for giving down (e.g. gaining favor versus charity). The variables remain continuous to capture scale effects. For each pairing one variable is zero while the other is the absolute difference between centrality scores.

[11] Due to high correlation between direct link and share of network connections/shortest path, we split these regressions out separately, controlling for difference in eigenvector centrality within each. This circumvents concerns of collinearity within our estimation results.

97%.[12] In total, we aimed to collect trust and trustworthiness data from 720 participants (18*40) and 12,240 pairings (18*17*40). In total, this study builds on data from 11,810 trust game pairs, played by 707 participants. Lost data was due to three participants being incorrectly selected to participate in the game (they were absent in the first round). Another reason was when participants refused to answer or didn't know the answer to a question we use as a control (such as age).

3.4.2 Manipulation check: Does network position actually mean something?

We are interested in differences in social preferences due to different positions of participants in the social network. We now check if being a central, a middling or isolate, corresponds to other characteristics. Table 3.1 presents the differences. In exploring the variation in socio-economic variables across centrality types, we conduct a difference in means test between each of the three centrality groups. We find several large differences and some similarities. Centrals are more likely to be literate than Middlings, and Middlings are more literate than Isolates. This monotonous sloping trend is consistent for all differences we find. Lower centrality is associated with being a migrant, being a female-headed household and a lower income. We also see that more central individuals have a higher education level, though this only holds significantly for primary and secondary education. This is probably because for higher levels we have very little variation (only very rarely had people followed more than secondary education). Centrals are also much more likely to be the village chief, and speak to the village chief much more often, indicating that they might have a larger influence in village-level decision-making.

[12] When a head of household was absent, we returned to the village for a second time (most often that weekend). If the head of household was still absent, a replacement within the household was asked to stand-in for the household head. Representatives were asked to respond to all questions from the perspective of the household head. This allowed us to draw the full social network. Only household heads (so never representatives) were asked to participate in the games, to ensure comparability.

Table 3.1 – Description Game Participants

	Mean (SD)				Difference (SE)		
	Overall	Central	Middling	Isolate	Central V Middling	Central V Isolate	Middling V Isolate
Age (years)	47.84	48.06	48.77	46.68	-0.703	1.378	2.081
	(17.70)	(16.62)	(17.78)	(18.65)	(1.382)	(1.916)	(1.808)
Literate (=1)	0.42	0.52	0.42	0.33	0.100**	0.188***	0.088**
	(0.49)	(0.50)	(0.49)	(0.47)	(0.042)	(0.048)	(0.043)
Migrant (=1)	0.32	0.13	0.31	0.51	-0.180***	-0.379***	-0.199***
	(0.46)	(0.34)	(0.46)	(0.50)	(0.035)	(0.038)	(0.051)
Female (=1)	0.32	0.16	0.31	0.47	-0.151***	-0.312***	-0.161***
	(0.47)	(0.37)	(0.47)	(0.50)	(0.045)	(0.051)	(0.056)
Income Index	0.05	0.33	0.00	-0.18	0.332***	0.512***	0.180*
	(1.03)	(1.07)	(1.00)	(0.97)	(0.084)	(0.080)	(0.091)
Completed Primary (=1)	0.31	0.36	0.29	0.29	0.074**	0.073	-0.001
	(0.46)	(0.48)	(0.45)	(0.45)	(0.033)	(0.044)	(0.039)
Completed Secondary (=1)	0.21	0.26	0.21	0.16	0.053	0.107***	0.054
	(0.41)	(0.44)	(0.41)	(0.36)	(0.037)	(0.037)	(0.033)
Completed Tertiary (=1)	0.01	0.02	0.01	0.01	0.004	0.008	0.004
	(0.11)	(0.13)	(0.11)	(0.09)	(0.009)	(0.008)	(0.004)
Village Chief (=1)	0.04	0.10	0.01	0.00	0.087***	0.100***	0.013*
	(0.19)	(0.30)	(0.11)	(0.00)	(0.017)	(0.013)	(0.007)
# times spoken to chief in previous month	7.79	11.61	7.30	4.45	4.311***	7.163***	2.852***
	(9.74)	(11.22)	(9.26)	(6.87)	(1.036)	(0.892)	(0.704)
N	717	240	239	238			

Source: survey data visit 1. Difference column is from a simple regression comparing the two types with standard errors (in parentheses) clustered at the village level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5 Results

This section presents our results. We first discuss what characteristics predict trust / trustworthiness, before examining which dyadic relations predict behavior in the trust game.

Trust is measured as the proportion of the initial endowment that is shared by Player 1 to Player 2. We find that as Player 1s, Centrals share, on average, 40% of their initial endowments with Player 2 while Middlings and Isolates share, on average, 36% of their initial endowments with Player 2. Trustworthiness is measured as the percentage of the initial investment returned by Player 2. We find that players returned on average 45% of the amount received from Player 1. We see no difference between centrality types. If we ignore the typologies and instead look at the correlation between eigenvector centrality and trust and trustworthiness we see a clear upward sloping trend, presented in Figure 3.2.

For trustworthiness the trend slopes upward, moving from about 39% returned to 43% returned, though this result is not significant. We see this as evidence for Hypothesis 3: that individuals with higher eigenvector centrality have more access to information and therefore know who to trust. That the effect is weaker for trustworthiness is unsurprising, as the chance of being betrayed is no longer relevant as the game ends immediately after this decision.

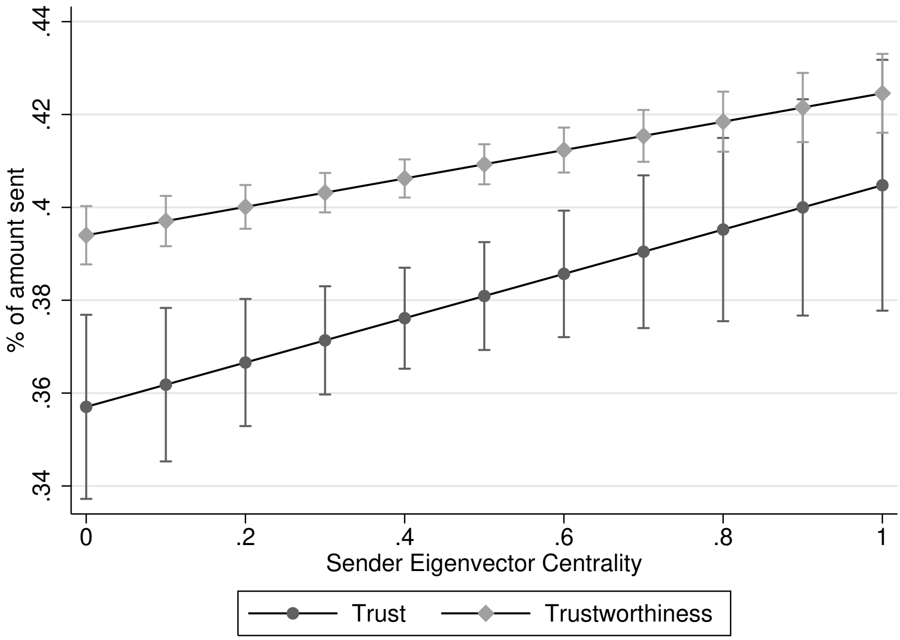


Figure 3.2 – Trust, Trustworthiness and Centrality

Estimates are based on a regression of Trust (Trustworthiness) on sender eigenvector centrality. Marginal effects are calculated at 0.1 intervals with 95% confidence intervals. Standard errors clustered at the sender level. Control variables are receiver centrality types, age, literacy, migrant status, sex, income index, primary education achieved, secondary education achieved and tertiary education achieved for both players.

Our regression results are reported in Table 3.2. Columns (1) and (2) present results from equation 3.1. We find again that central individuals are on average sending four percentage points more of the initial endowment than middling and isolate participants within the trust game. However, there is no significant difference in trust behavior between Isolates and Middling individuals. It might be that this effect is not pronounced enough to be established statistically, or that the effect of increased eigenvector centrality is not monotonously increasing. We also see that the centrality status of the receiver does not matter for the initial investment. Next, we look at the predictive power of characteristics of both the sender and receiver. Richer households are on average less trusting. This is surprising as usually richer households are more trusting because of the endowment effect. Furthermore, we see that higher education levels and being the village chief increase trust behavior. The village chief might behave more charitably to protect his reputation in the village. Regarding receivers' characteristics, there seems to be a clear effect of education: literate, educated players are trusted more. This might be because participants expect these players to understand the game better and act more trustworthy as a result.

Next, we examine what factors predict trustworthy behavior: how much players return of the tripled amount sent by Player 1. This is shown in column 2. We see no effect of centrality on trustworthiness, reflecting the flatter curve in Figure 3.2. As mentioned, this is unsurprising as the improved access to information that individuals with high eigenvector centrality enjoy is not useful for trustworthiness. Regarding characteristics, we first focus on P2 characteristics, which in this case are the characteristics of the decision-maker. For trustworthiness we do not see an effect of income, though the effect of education remains. Having a secondary education increases trustworthy behavior by 5%, compared to individuals without an education. Characteristics of the receiver (P1 for this column) that are important are once again education, though the effect is not very strong. If the receiver is the village chief this also increases trustworthy behavior, likely to gain or keep favor with the village chief. This might be individuals trying to build connections to the influential in the village.

Table 3.2 – Trust and Trustworthiness Characteristics Results

	(1) Trust		(2) Trustworthiness	
Central Sender	0.036***	(0.014)	0.016	(0.012)
Isolate Sender	0.003	(0.014)	−0.009	(0.011)
Central Receiver	0.000	(0.003)	−0.004	(0.003)
Isolate Receiver	0.003	(0.003)	−0.005	(0.003)
P1 Age (years)	0.000	(0.000)	0.000***	(0.000)
P1 Literate (=1)	−0.015	(0.019)	−0.004	(0.005)
P1 Migrant (=1)	0.004	(0.014)	0.004	(0.003)
P1 Female (=1)	0.006	(0.015)	0.002	(0.003)
P1 Income Index	−0.013**	(0.006)	0.001	(0.001)
P1 Completed Primary (=1)	0.023	(0.017)	0.002	(0.005)
P1 Completed Secondary (=1)	0.095***	(0.025)	0.011*	(0.006)
P1 Completed Tertiary (=1)	0.000	(0.059)	0.001	(0.016)
P1 Village Chief (=1)	0.057*	(0.033)	0.016**	(0.007)
P1 # times spoken to chief in previous month	0.001	(0.001)	0.000***	(0.000)
P2 Age (years)	0.000	(0.000)	0.000	(0.000)
P2 Literate (=1)	−0.007*	(0.004)	−0.005	(0.015)
P2 Migrant (=1)	−0.004	(0.003)	0.015	(0.011)
P2 Female (=1)	−0.002	(0.003)	0.002	(0.011)
P2 Income Index	0.000	(0.001)	0.001	(0.005)
P2 Completed Primary (=1)	0.007**	(0.003)	0.024*	(0.014)
P2 Completed Secondary (=1)	0.008	(0.005)	0.053***	(0.021)
P2 Completed Tertiary (=1)	0.019*	(0.012)	0.008	(0.056)
P2 Village Chief (=1)	0.017**	(0.007)	0.030	(0.027)
P2 # times spoken to chief in previous month	0.000	(0.000)	0.000	(0.001)
Initial Endowment (# tokens)			0.001	(0.000)
Constant	0.393***	(0.041)	0.451***	(0.032)
Observations	11 810		11 788	
# clusters (individuals)	707		706	
R-squared	0.214		0.132	
Village Fixed Effects	Y		Y	
Round Fixed Effects	Y		Y	
P-value Central sender = Isolate sender	0.0431		0.0341	
P-value Central receiver = Isolate Receiver	0.539		0.873	

Standard errors clustered by Player 1 for trust regressions and Player 2 for trustworthiness regressions are reported in parentheses. Trust and Trustworthiness measured as proportion of total possible amount sent
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Next we explore outcomes on trust based on relational network parameters between Player 1s and Player 2s, using Equation 3.2. Table 3.3 shows that relationship variables are important. In column 1 we see that if the other player is less central, this increases trusting behavior. A 1 SD lower eigenvector centrality increase trusting behavior by 1%, a small but precisely estimated effect. This result holds consistently for trustworthiness as well. In column 1 we also see that when players are directly linked through the social network, this increases trusting behavior by about 2%, a small but once again precisely estimated effect. In column 2 we examine the effect of sharing links. A 1 SD increase in the number of shared connections increases trusting behavior by 1%. This confirms the hypothesis that shared connections can also be used as social collateral to improve trusting behavior. We see no effect of the path length between the two players on trusting behavior. These results provide evidence for the social collateral theory, where individuals use their relations as collateral in trust-based interactions.

Next, we move to trustworthiness. Results are reported in Columns 3 and 4 of Table 3.3. We again see that if the other is less central, the players return more, though the effect remains small. If players are directly linked this also increases trustworthy behavior, with similar magnitude. We see no significant effect of sharing more connections on trustworthy behavior. This is not so surprising. Interestingly, the path length between players now enters significantly, with 1 SD increase in path length increasing trustworthy behavior by 1%, which is precisely estimated. For trustworthiness there is less evidence for the social collateral theory, though it should be noted that the theory makes no predictions for trustworthiness.

Table 3.3 – Trust and Trustworthiness Dyadic Results

	(1) Trust	(2) Trust	(3) Trustworthiness	(4) Trustworthiness
Other more Central (Diff Centrality, standardized)	-0.003	-0.003	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Other less Central (Diff Centrality, standardized)	0.011^{**}	0.011^{**}	0.008^{**}	0.007^{*}
	(0.005)	(0.005)	(0.004)	(0.004)
Direct Link (=1)	0.018^{***}		0.015^{***}	
	(0.007)		(0.005)	
Proportion Shared Connections (standardized)		0.009^{**}		0.002
		(0.004)		(0.004)
Shortest path (Steps) to other (Standardized)		-0.001		-0.007^{***}
		(0.003)		(0.003)
Initial Endowment (# tokens)			0.001	0.001
			(0.000)	(0.000)
Constant	0.401^{***}	0.398^{***}	0.443^{***}	0.445^{***}
	(0.039)	(0.040)	(0.032)	(0.032)
Observations	11 810	11 742	11 788	11 720
# clusters (individuals)	707	705	706	704
R-squared	0.213	0.213	0.132	0.132
Socio-Econ Indicators	P1 P2	P1 P2	P1 P2	P1 P2
Village Fixed Effects	Y	Y	Y	Y
Round Fixed Effects	Y	Y	Y	Y

Standard errors clustered by Player 1 for trust regressions and Player 2 for trustworthiness regressions are reported in parentheses. Trust and Trustworthiness measured as proportion of total possible amount sent. Control variables for both players: age, literacy, migrant status, sex, income index, primary education achieved, secondary education achieved, tertiary education achieved.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We identify several important correlations of network position (centrality) and localized network characteristics on trust and trustworthy behavior. The magnitude of these effects is modest, with effects increasing trust/trustworthiness by about 2-3%. An important caveat is that within these villages there appear to be strong norms on sharing windfalls. Within the trust game 25% of players sent exactly half of their endowment. Conversely, 25% of players return 1/3rd of the tripled amount the receive (or exactly what Player 1 sent them). This explains the modest effects sizes we find: the social norm is so strong that these network characteristics can only move this norm by a small margin.

3.6 Conclusion

Building on literature exploring origins and drivers of social preferences, we explore relationships between social network positioning and social behaviors in a trust game. We test several hypotheses from Karlan et al. (2009)'s Social Collateral theory. We employ a round-robin styled trust game to explore how behaviors of trust and trustworthiness are tied to first and Player 2 centrality measures and dyadic characteristics. Using household network data collected from 40 communities game participants are selected as the most central, least central, and those in the middle, based on calculated eigenvector centrality score.

Our results indicate that centrality is positively but not uniformly correlated with social preferences of trust and trustworthiness. Central individuals displayed higher levels of trust and trustworthiness in the trust game. These results indicate that there are shaping forces tying an individual's social preferences and positioning within their community social networks. Determining whether more pro-social individuals become more central or whether more central individuals develop more pro-social preferences is beyond the scope of this paper. However, it is clear that the two are linked, providing evidence supporting existing theoretical literature on relationships between pro-social behaviors and centrality within social networks.

Expanding beyond traits of an individual to exploring mutual network traits between paired individuals in social transactions highlights the more substantial role of relational network indicators, consistent with the social collateral theory. In the bilateral trust game experiment, partnered individuals exhibited higher trust levels when they shared a greater proportion of mutual connections within their social network. A direct network tie increased both trusting and trustworthy behavior. All this is consistent with individuals using their network ties as social collateral. A shorter social path length increased trustworthy behavior. Players with a lower centrality also received higher contributions, both in a trust and trustworthiness setting.

Taken together, our results provide evidence for the Social Collateral theory. When more central players exhibit more trusting behavior (and to a lesser extent trustworthy behavior) we see this is evidence of them using their greater access to information to determine the trustworthiness of players. That players react strongly (especially in a trust setting) to being directly linked to the other player or sharing a larger number of mutual connections is evidence of them using these connections as social collateral to realize trust-based interactions. Individuals therefore choose to behave more trusting and trustworthy when faced with players with whom they have a stronger connection.

Chapter 4

Local Economy effects of Large-Scale Agricultural Investments

The last decade has seen a surge in land acquisitions in developing countries by foreign companies. To date there has been little rigorous quantitative evidence on the impacts of such investments on local communities. We examine the economic impacts of a large-scale biofuel plantation in Sierra Leone - a major investor target. We conduct a difference in difference analysis using three waves of a large n survey in both communities directly affected by the plantation and those outside the catchment area. We find a large average drop in incomes, mainly driven by lower revenues from agricultural activities. These findings are consistent with a labor demand shock, caused by a clash between the private and commercial agricultural calendar, increasing the local price of labor. A spillover analysis confirms that the impacts are at least partially transmitted by a shock to the local economy. Within land leasing communities, households that are employed at the plantation see their incomes and assets increase. However, as a result, village-level inequality increases. Finally, we also see a decrease in access to land, indicating that investments can cause land shortages and are not just using unused land.

Publication status: Hofman, P., Mokuwa, E., Richards, P. and Voors, M., 2020. Local Economy effects of Large-Scale Agricultural Investments. Working Paper.

4.1 Introduction

Foreign investments in African agriculture have increased dramatically. Driven by the 2007-8 price spike of key primary commodities in conjunction with the world financial crisis, commercial investment companies increasingly sought out new investment ventures (Arezki et al., 2013; Koning and van Ittersum, 2009). The Land Matrix, which documents all transnational land acquisitions, to date has recorded 1694 'concluded' agricultural investments, in total covering about 50 million hectares.[1] In some African countries over 30 percent of arable land is foreign-owned (Landmatrix, 2020; Nolte et al., 2016). These investments often take the form of large scale plantations, with land rights acquired for a long period (typically 49 or 99 years). This trend is likely to increase due to the projected rise in demand for food, animal fodder and energy crops.

Some herald this new wave of investment by commercial parties as an important vehicle to achieve poverty reduction, highlighting the potential benefits of scale economies in agricultural production (Collier and Dercon, 2014; Ellis, 2005), inducing innovation (Borensztein et al., 1998), enabling access to finance (Alfaro et al., 2010) and the organization of production and marketing (Reardon et al., 2003). On the other hand, there are arguments against land consolidation that stress important potential negative impacts on distributional, social and institutional outcomes. First, while large scale investments may create new opportunities for some (through land rents and employment), they exclude others (Peters, 2004). Such effects may be particularly strong in the African context characterized by strong social dependencies (Townsend, 1994). Investments may deepen social divisions, possibly contributing to conflict (Peters, 2013; De Schutter, 2011; Baxter, 2013; Scott, 1998). Second, large-scale land acquisition by foreign companies often amounts to 'land grabbing' (Liversage, 2010), generating benefits for foreign investors (and domestic elites). Land rights are impacted as investors obtain leases and clear land for industrial monoculture plantations. For many households this implies a change in access to land (in extreme case even forced migration), and nutritional security, thereby impacting family livelihoods (Liversage, 2010). Some global analyses show that foreign investments are greater where property rights regimes are weakest (Alfaro et al., 2010; Arezki et al., 2013). One national analysis in Liberia points out that this relationship might in fact be the other way around: more secure property rights increase investments (Christensen et al., 2020). This suggests an important role for institutions as a mediating factor in determining potential development outcomes (Sokoloff and Engerman, 2000;

[1] Date of access: November 2019.

Herbst, 2014; Dorward et al., 2009). Often, land investment deals are made between companies and elites and exclude local people from the negotiations, increasing corruption (Peters, 2013; De Schutter, 2011).

Despite the scale of foreign investments in agriculture, local economic impacts have to date failed to receive rigorous quantitative investigation. Exceptions are Herrmann and Grote (2015), who assess a sugarcane plantation and outgrower scheme in Malawi and find positive economic returns for laborers. The plantation attracts labor from nearby villages typically from the poorest households. Compared to non-laborers, incomes nearly double. A similar paper, by Herrmann (2017) examines rice and sugar plantations in Tanzania. He finds for both sectors an increase in per capita income for plantation laborers compared to other households in the same villages. There is however, no significant effect on agricultural or total household income.[2] A key limitation of these papers is that they rely on post-intervention data, requiring strong assumptions to prove causal effects. Investors typically do not select concession sites at random and take important ecological, political and economic characteristics into account such as agricultural potential, distance to input and output markets, local institutions and labor availability. Failing to adequately control for such variables may severely bias results. Below, we improve on this work and use data from before and after the creation of a large scale agricultural plantation. Baseline data, pre-dating the plantation allows us to control for such selection effects. In addition, the analysis compares those hired by the plantation to those that are not hired but are from the same village. This incorporates household economic impacts through community-wide channels such as increased competition over land and labor. It is an open question whether incomes should increase on average in the local village economy. Theoretical work by Kleemann and Thiele (2015) and Dessy et al. (2012) show how the net effects of such investment projects crucially depend on the intermediate impacts on labor and land markets. If labor and land are abundant, increased demand for labor and land should not impact local economies. However, this is rarely the case. For instance in rural Sierra Leone (where the investment we examine is located), there is severe competition over labor (Mokuwa et al., 2011; Bulte et al., 2018). In such cases increased employment opportunities outside the village may cause a decrease in labor input for private farms, undermining income and food security.

We examine the impact of a large scale agricultural sugarcane investment project in Sierra Leone. The country is an appropriate choice for investigating the impact of foreign agricultural investments, as it has been one of the larger recipients of these types of

[2]Other papers include Shete and Rutten (2015) and Jiao et al. (2015). Both use a matching algorithm on post-intervention data, and suffer from a low number of observations/clusters.

investments. Sierra Leone is a poor country, characterized by rotational fallow agriculture and limited access to financial and output markets. The majority of the population is engaged in the agricultural sector. Farms are very small: average farm size is about 0.5 hectares. To a large degree farm output is determined by labor rather than land or capital (fertilizer application and improved seed varieties are rare) (MAFFS, 2011). There has been a surge in commercial investments in agriculture. Since 2000, foreign companies have acquired over 25% of the country's arable land (Baxter, 2013; Landmatrix, 2020).

We use a difference-in-difference approach allowing us to correct for important time-invariant characteristics, such as agricultural potential, distance to input and output markets, local institutions and labor availability, all of which are crucial selection criteria for investors. We assess impacts on several key outcomes: household income (stock and flow), access to land, food security, health and village level inequality. Our data allow us to examine effects over shorter (2 years) and longer (5 years) periods. We find that average income drops substantially, by about 0.4 standard deviations. We also see a small drop in access to land and some improvements in health outcomes in villages where the company works. We find that the labor demand shock, caused by a clash between the private and commercial agricultural calendar, increases the local price of labor. As a result, average farm productivity and agricultural incomes decrease. In contrast, households that have a member working for the company compensate for this drop with salaried income. As a result village inequality increases. The hypothesis that the main impacts work through local markets is bolstered by a spillover analysis that shows income changes are smaller further away from the investment. As a robustness check, we provide some evidence that the parallel trends assumption holds. We also examine attrition and find that our main findings are robust when examining bounds on the treatment effect.

This fits in a larger literature that aims to move from examining small-scale impacts towards looking a the effects on the local economy (Taylor and Filipski, 2014; Cust and Poelhekke, 2015). One example of this is Aragón and Rud (2013) who examine the impacts of an exogenous expansion of a gold mine in Peru on the local economy. They find that this expansion increases local labor prices and local income, and this effect declines when moving away from the mine. We find similar results: in our case reductions in labor availability reduce household production.

Closest to our work is Bottazzi et al. (2018) who examine the same investment project. They use a matching algorithm to match 592 respondents in 34 villages where the company leased land and compare this to 290 respondents in 21 control villages. They use retrospective cross-sectional data to establish the match. They find that on average incomes and food expenditure increase, as well as labor prices. They also see improvements in food and water security. They also note that positive economic effects are mainly for landowners and men that are employed. While we examine the same investment and a similar period, we find an opposite impact for incomes: we find a large and substantial drop. This is likely because our identification strategy allows us to correct for pre-existing differences: we note a strong imbalance in pre-investment data in incomes, villages that end up leasing land to the company are on average richer than comparison villages. As a result, the positive economic results Bottazzi et al. (2018) find may be due to different initial conditions, rather than due to the impact of the investment project.

We go further than this work by including pre-investment data, greatly improving the identification strategy and using a substantially larger sample, improving statistical power.

The rest of this article is structured as follows. Section 4.2 introduces the research context. Section 4.3 presents our data and empirical strategy and section 4.4 contains our results. We present some robustness analyses in section 4.5 and conclude in section 4.6.

4.2 Large Scale Investments in Agriculture in Sierra Leone

We focus on Sierra Leone, which has received a lot of attention from land investors. Since 2000, 24 deals have been concluded, covering 1 million hectares (25% of total arable land) (Landmatrix, 2020). The country ranks low on the Human Development Index (UNDP, 2016) and has high poverty levels and low food security. Most Sierra Leoneans are smallholder farmers, especially in rural areas. Farm productivity is low and access to productive inputs, such as fertilizer and high-yielding seeds is minimal. As a result, agricultural production is limited by labor availability. A 2011 survey found that 65% of households experienced a shortage of labor in the agricultural season. Farm production to a large extent relies on family labor. About one-third of households hire labor (MAFFS, 2011). Land is communally owned by extended families, who redistribute the land for long-term use within these extended families. These extended families are free to give land in use-right to individuals, but to lease land they need permission from the paramount chief (a local leader), and the national government.

Outside investments can potentially improve this low productivity by bringing in improved technologies and large-scale production that achieve economies of scale. The Government of Sierra Leone aspires to 'promote an attractive business environment based on fair and responsible investments in land for both small and large scale businesses' (GoSL, 2015, pp 7).

We assess the impacts of a large-scale plantation in the north of Sierra Leone. In 2010, a commercial investor acquired 24'000 hectares of land for a 49-year lease. Landowners received 8.90 US$ in compensation per hectare per year, half of which (i.e. 4.45 US$) goes to the landowners and the other half to various local elites. This is according to national standards for land payments. Landowners also receive an additional payment of 3.46 US$/Ha/year, making the total payment for landowners 7.91 US$/Ha/year. In 2014, a peak year, the company leased land from 52 villages, amounting to 10-60% of total village land. The investor employed local and international staff to grow sugarcane using center-pivot irrigation. In 2014, the company employed 3'500 people, half of whom were on fixed-term contracts. The company aimed to recruit unskilled labor from communities supplying land to the plantation. The main labor demand of the plantation overlaps with peak periods for smallholder production. Smallholder labor demand is greatest in February-April when land is prepared ('cleared') and planted according to a rotational cycle (Richards, 1986), matching the plantation's peak labor demand. Besides providing benefits in terms of employment for laborers from nearby villages, and surface rents for landowners, the company established a health clinic, ran health outreach programs, provided several farming training programs and had a compensation program for destroyed tree crops. The investment was funded by a consortium of ten Western development banks. This means that besides a business project it was also explicitly aimed to be a development project.

The plantation has received considerable attention in the media and has been the focus of several policy reports and journal articles. We provide a summary in Table 4.1. Most reports critique the investment and describe how it was forced through by politicians and local elites without involving communities other than through superficial consultation, and conclude average incomes decreased. Some cite improved incomes, especially for specific landowners. Some also point to increases in social disharmony due to the plantation creating conflicts over access to land and surface rents and in other cases exacerbating existing tensions over land claims. A key drawback is that most of these studies rely on qualitative and small sample case studies. While these sometimes provide rich insights into relevant local dynamics for selected localities, they fall short in assessing average differences.

Table 4.1 – Studies on investment under study

Author	Type	Cluster N	N	Methods	Findings
Anane and Abiwu (2011)	PR	12	NA	SI, SSI and FG	The development programs were slow to start and did not cause tangible benefits. Food production has gone down because the company is using fertile land. Access to water has gone down. Working conditions for the company are poor: irregular contracts, no safety gear or food provided
Baxter (2011)	PR	NA	NA	SSI	The land leased was under use and fertile, despite contrary claims by the company. Women were not consulted in the decision-making process. Wages for casual labourers are too low to cover daily food needs
Baxter (2013)	PR	10	84	FG and SSI	Food security has gone down, poverty went up. Benefits for job-holders and landowners (though jobs are reported to be low-paying). Higher school dropout, teenage pregnancy, broken marriages, theft, social tensions. Breakdown of traditional social structures
SiLNoRF (2014)	PR	NA	NA	SSI and FG	Increase in income in villages close to the factory. Working conditions for employees are good. The company's development programs are improving local food security. Individuals do not feel that they had a choice in accepting the project. Landowners do not agree with the land rent split (only 50% of rent accrues to them). There are several cases of water shortages because of the company's actions. There were several strikes for higher wages, conditions and discrimination
Fielding et al. (2015)	PR	9	459	SI, SSI and FG	Increased labour scarcity, especially during the growing season. Increased in-migration by individuals looking for work. Improved infrastructure: more roads and houses. Reduced land availability. Lower agricultural productivity (or production). Higher incomes because of wage labor
Millar (2015a)	J	12	55	SSI	Most participants had high hopes for economic improvement because of the investment. Many farmers stopped farming to work for the company. Salaries are lower than income from subsistence farming. Land-lease payments are distributed to three people per village, who do not always distribute further. Economic benefits are concentrated with village elites
Millar (2015b)	J	12	26	SSI	Women were excluded in the decision to accept or not accept the project. Women are rarely employed by the company and have no say in deciding how the land-lease payments are spent. This is in line with persistent disempowering gender norms in Sierra Leone. The company was not aware of these norms and took no measures to correct for this
Bottazzi et al. (2016)	J	NA	54	SSI and FG	Land has become more 'monetized': is now a means to earn money rather than produce food. Migrants do not get any benefits. Monetization of land and 'hard' boundaries create new types of land conflicts. The investment exacerbates existing social cleavages
Marfurt et al. (2016)	J	2	180	SSI, FG and PO	Direct payments do not compensate for the negative effects of the company. Labour contracts are very insecure and wages are low. The company leases fertile land, decreasing agricultural production and income
Millar (2016b)	J	12	115	SSI and PO	Land has become more 'monetized' and families feel they have to defend their claim to it. This requires more formal land titles which causes conflicts over (a.o.) exact borders. Jobs are mainly given to individuals part of landowning families. There are tensions around labour provision: many want work but the company cannot provide. There are also tensions between local (not employed) youth and employed youth from outside the project area. Another source of tension is between generations: youth did not get a say in the decision to accept the company, and do not have control over the land-lease payments
Millar (2016a)	J	12	55	SSI and PO	There is a disconnect between how the company and the inhabitants view land. The company uses technology to 'control' the land, which inhabitants were not able to protest against as this requires literacy. Most land is regularly used, even though it is not under constant cultivation
Millar (2017)	J	NA	NA	SSI and PO	Regional elites, who used to function as conflict-solving institutions are now using their influence to acquiesce the local population to ensure their access to company-provided benefits. This makes it almost impossible for the local population to voice grievances. In the long run this led to feelings of marginalization and increased conflict
Bottazzi et al. (2018)	J	55	882	SI	Farmers around the plantation use less agricultural land, attain lower yields and pay more for labour. In contrast, they also find increased incomes, improved food security and more food expenditures.These improvements were largest for landowners and men.

PR=Policy Report, J=Peer-reviewed journal. SI=Structured interviews, SSI=Semi-structured interviews, FG=Focus Groups, PO=Participatory observation
NA=not specified

4.3 Data and Empirical Strategy

4.3.1 Sample

We use data from multiple rounds of survey work. Table 4.2 summarizes the sample sizes for each round. Baseline data were collected prior to any plantation activities in 2010 by a research team of the University of Cape Town, at company request, to comply with reporting requirements. In total, baseline data encompasses 78 villages and 4'233 households, comprising a census of all households in these villages. The plantation then started operations in 41 of these villages, creating a natural comparison group. In 2012, a second survey was implemented (again by the University of Cape Town), this time in 118 villages and with 4'824 households. In the meantime, the company scaled up operations to 47 villages. Figure 4.1 shows the locations of all villages where we have access to panel observations for the 2010-2012 survey waves. In 2015, a team from Njala university led by the researchers collected an additional wave of data. We returned to all villages included in the 2010 dataset and interviewed 25 randomly selected households per village, selected from the 2010 interviewees. The survey instrument was designed to closely match the earlier rounds of data collection. If there were less than 25 people from the 2010 survey present additional households were randomly invited to participate.[3] For the 2015 survey round we have data on 1'767 people in 75 villages. In the meantime, the company relinquished land from some villages ending with 36 villages from the original pool.[4]

Table 4.2 – Sample sizes over time

	2010		2012		2015	
	Control	Land Leased	Control	Land Leased	Control	Land Leased
Cross-section						
Observations	1415	2818	1790	3034	649	1118
Villages	37	41	71	47	39	36
Panel						
Observations			915	2240	99	529
Villages			28	40	15	34

Number of panel observations. Participants were matched based on company-assigned ID code (Matching on names leads to a lower number of observations but similar conclusions). Some subsequent analyses have a lower number of observations and/or clusters. This is because in those cases some participants did not answer that specific question. Source: survey data

To examine how people are affected over time we ideally rely on data from the same people

[3] We do not use these observations in our analyses.

[4] The next section describes why this happened.

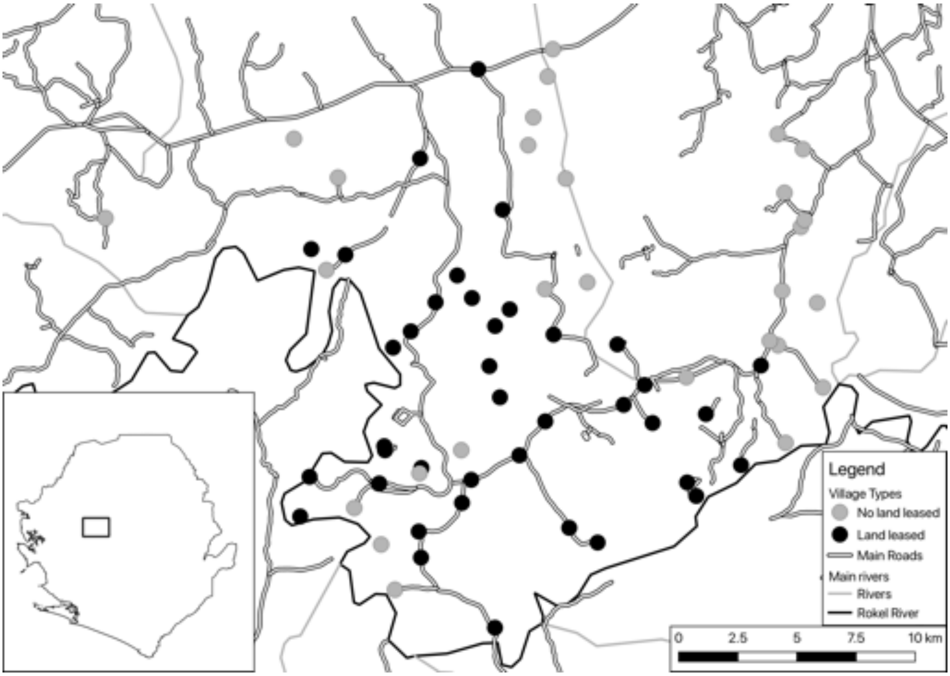


Figure 4.1 – Village Locations

Shows location of all study villages. Source: survey data

across each survey wave. Fortunately, the company assigned all households ID codes and identification cards. We used these ID codes to match respondents across waves. In total we have 3'155 respondents in both 2010 and 2012, and 628 observations for both 2010 and 2015. We examine our main outcomes both using the 2010 and 2012 data to assess short-run effects and compare across 2010 and 2015 for longer-run effects.[5]

[5]If we are more stringent and also match on village name, participant name, years in the area and GPS location the number of matched participants drops. In this study we use the match on ID codes, though as a robustness check we examine whether the direction of coefficients holds for the more restrictive match. These results are shown in Table A.4.3 and Table A.4.4 and are qualitatively similar to our main results.

4.3.2 Identification Strategy

Our identification strategy relies on a difference-in-difference approach. We estimate the differences in outcomes over time for both the villages that rented land to the plantation and control locations. This corrects for all time-invariant characteristics (observable or not). The main identifying assumption is that in the absence of investment, the villages would have developed in a similar pattern. This assumption is of course fundamentally untestable. However, using data on forest loss and vegetation (EVI) available from satellite images, we can show that deforestation trends were parallel before the investment started (See Figures 4.5 and 4.6). The control group is a set of villages that the company was originally planning to work in but decided not to. This was for various reasons: villages decided not to join, the villages could not provide enough land and most importantly the distance to the Rokel river (darker in Figure 4.1) was too great to pump water for the center pivots used by the company for irrigation. Therefore, they are similar in characteristics that are likely to be predictive of yield. Furthermore, since all smallholder agriculture is rain-fed distance to the Rokel river is unlikely to correlate with local agricultural production.

4.3.3 Empirical Model

To assess impacts of this investment we estimate the average treatment effect on the treated for original households using a standard difference-in-difference specification. Specifically, we estimate:

$$\boldsymbol{Y}_{ij} = \beta_0 + \beta_1 treat_j + \beta_2 post_{ij} + \beta_3 post_{ij} * treat_j + \varepsilon_{ij} \tag{4.1}$$

Where $\boldsymbol{Y}_{ij}$ refers to our set of outcome variables (such as income, land access, see section 4.3.4), $treat_j$ refers to the villages where the company leased land and $post_{ij}$ refers to the later time period. β_3 is our coefficient of interest. i indexes the household level, while j indexes the village level. We cluster standard errors at the village level.

Furthermore, as a plausibility check to see if labor shortages are driving our results, we examine if our outcomes taper off further away from the plantation. We estimate:

$$\boldsymbol{Y}_{ij} = \gamma_0 + \gamma_1 distance_j + \gamma_2 post_{ij} + \gamma_3 post_{ij} * distance_j + \varepsilon_{ij} \tag{4.2}$$

γ_3 is our coefficient of interest and we again cluster standard errors at the village level. This model is estimated only for the subsample of control villages. That is, we compare villages close to the investment (but not directly affected by it) with villages further away.

Finally, we assess if individuals employed by the company benefit. In the 2015 survey we asked respondents if they had worked for the plantation. We examine the extensive margin and regress our main outcome variables on a dummy indicating if a household member at any time worked for the plantation during the 2010-2015 period. We then estimate a triple differences model:

$$\begin{aligned} \boldsymbol{Y}_{ij} = \eta_0 + \eta_1 laborer_{ij} + \eta_2 treat_j + \eta_3 post_{ij} \\ + \eta_4 post_{ij} * treat_j + \eta_5 treat_j * post_{ij} * laborer_{ij} + \varepsilon_{ij} \end{aligned} \tag{4.3}$$

Our coefficient of interest is η_5, how laborers differ from non-laborers in treatment villages in the later time period. Note that laborer status is endogenously determined which makes this effect less well-identified. Laborers are likely selected from landowning families as a favor to these families. This means that they likely had higher incomes initially and more means to expand their incomes. Level differences drop out in our DiD estimation, but we cannot net out differences in trends.

4.3.4 Outcome variables

Our main outcome indicators relate to incomes, land access, food security and health. Our variables are defined in Table A.4.1 and descriptive statistics at baseline for both treatment and control villages are shown in Table 4.3. Average household monthly income is 60’000 Leones (200’000 in Treated), or 13 USD (36 USD), far below the World Bank international poverty line of 1.25 USD per day.[6] This measure includes only cash incomes and does not account for self-consumption or in-kind contributions. Figure 4.2 shows the relative components of traditional income.[7] Agricultural income accounts for the majority of income, with 60% for the control group and 80% for the treatment group

[6] As income is highly sensitive to outliers we use the inverse hyperbolic sine transformation to correct for this. These numbers are calculated back from the inverse hyperbolic sine transformation.

[7] We split our income into two measures: traditional income and total income. Total income also includes all kinds of payments by the company. See Table A.4.1 for the definition

before treatment. The income differences between treatment and control villages are large. Given our difference-in-difference set up, these drop out. The number of assets in a list of what farmers owned is 4, which might mean a household owned its house, a mosquito net, an iron pot and a bed mattress, but no mobile phone, tv, iron kettle or generator. Housing quality averages 5, which is the rating for a house with a mud floor, wattle and daub walls and thatch or tarpaulin as roof. For the livestock index the value is around 0.25, comprising (for example) 2 goats and 5 chickens. Almost all households have access to arable land for cultivation, though almost all households have faced a seasonal food shortage in the previous year. 92% of households had a mosquito net in their house (80% in control villages). The participants are very poor, have few assets and low food security.[8]

Table 4.3 – Descriptive Statistics

		Control			Treatment		
	n	mean	sd	n	mean	sd	Diff
Traditional Income (Leones, IHS)	1004	10.98	4.23	1504	12.21	3.33	1.231**
# Assets	1415	3.94	1.48	2818	3.86	1.52	-0.086
House quality (Score, 1-33)	1098	5.13	2.16	1529	5.28	2.05	0.146
Tropical Livestock Unit	1028	0.27	1.67	2144	0.22	0.38	-0.049
Access to Land (=1)	1351	1.00	0.06	2714	0.99	0.08	-0.003
Food shortage (=1)	1374	0.99	0.11	2700	0.99	0.10	0.003
Bed net in household (=1)	1412	0.92	0.27	2811	0.80	0.40	-0.120***

Table shows averages for 2010 (before any land was leased from communities). The final column shows the coefficient of a simple regression of treatment status on the variable, with clustered standard errors at the village level. Stars indicate whether the treatment - control difference is statistically significant, with $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Results

We first estimate model 4.1, to assess the short-run effects of the large-scale agricultural investment. Table 4.4 presents the results. Our main variable of interest is the interaction term, which shows the effect of the treatment over time, correcting for initial differences in levels. This shows a big drop in traditional income of over 0.6 standard deviations. For total income this is lower (0.4 SD) but still substantial and significant. This drop is largely driven by a large drop in agricultural income (See Table A.4.5 and Table A.4.6 for the effect on the four components of traditional income). We hypothesize that this is caused

[8]We use the following user-written computer programs in preparation of the data, tables and figures: Jann (2005, 2007, 2012, 2016); Van Kerm (2009); Gallup (2012); R Core Team (2017); Hijmans (2017); Wickham et al. (2017); Højsgaard and Halekoh (2018); Müller and Wickham (2018); Gorelick et al. (2017)

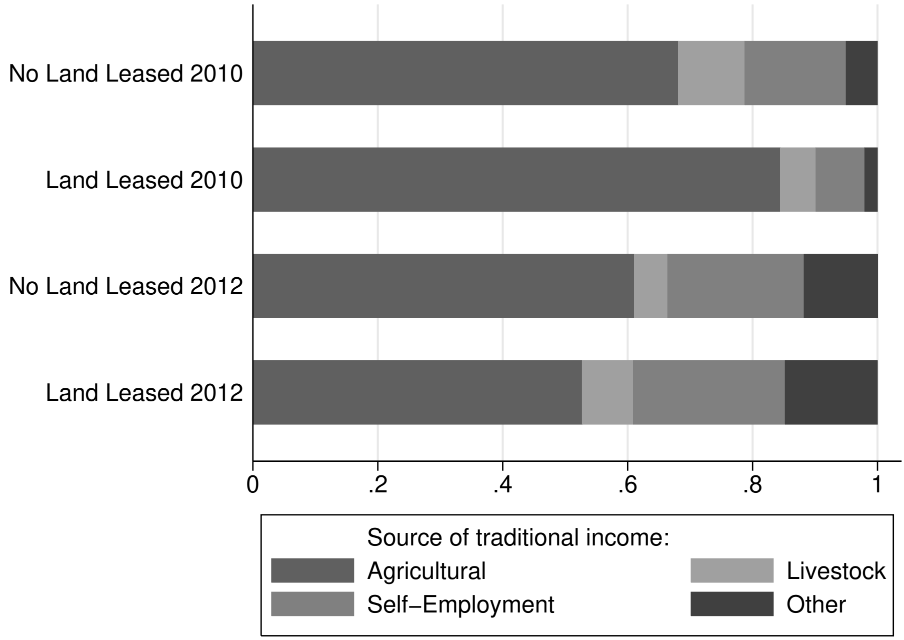


Figure 4.2 – Income Proportions (2010-2012)

Shows proportions of traditional income (that is, excluding 'new' income sources like land lease payments and salaried income). Other income includes remittances, self-declared other revenues and pension income. Source: survey data

by an increase in the local labor price which makes it more difficult for households to hire in local labor, reducing agricultural production and thus sales. Our spillover analysis (see Table 4.6) and laborer analysis (see Table 4.7) confirm this hypothesis. Furthermore, in 2015 we asked households whether the price of labor had gone up after the company started working. 87% of farmers said that it did. The drop in income (a flow variable) partially translates to a change in stock variables (ie assets). There is a substantial drop (0.11 SD) in housing quality. On the other hand, the TLU score increases by 0.3SD, though neither of these are significant.

Table 4.4 – Short Run (2010-2012) effects of a Large-Scale Agricultural investment

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Treated	0.389***	0.389***	−0.031	0.106	−0.015	−0.003	0.003	−0.118***
	(0.143)	(0.143)	(0.094)	(0.079)	(0.072)	(0.003)	(0.005)	(0.040)
Short Run	−0.256	0.124	0.036	0.314***	−0.091	−0.031**	−0.101***	−0.082**
	(0.229)	(0.181)	(0.126)	(0.060)	(0.067)	(0.012)	(0.013)	(0.040)
Treated * Short Run	−0.625**	−0.424**	−0.047	−0.110	0.315	−0.049***	−0.006	0.135***
	(0.262)	(0.197)	(0.141)	(0.080)	(0.192)	(0.018)	(0.018)	(0.050)
Constant	0.000	0.000	0.000	0.000	0.000	0.997***	0.988***	0.920***
	(0.110)	(0.110)	(0.073)	(0.048)	(0.038)	(0.002)	(0.004)	(0.022)
Observations	3762	3762	6310	3914	3470	6082	6068	6302
# Clusters	67	67	68	68	67	68	68	68

OLS regressions. Standardized and centered on control group at baseline (not columns 6-8). Robust standard errors in parentheses clustered at the village level. IHS is inverse hyperbolic sine transformation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Access to land goes down 5% more than in the control group which is small but precisely estimated. The often-used narrative that these investments are utilizing unused land and thus not affecting land availability of productive assets does not hold here. When we examine this group that has lost access to land separately using a similar approach as Model 4.3 we find a large decrease in full income of 0.8SD ($p<0.01$). The incidence of food shortages drops by 10% in the short run. Treated households had a lower rate of bed nets before the investment, and in treated households this has gone up in the short run. This could be linked to health outreach programs run by the company.

We dig a little deeper into the drop in income by examining how the proportions of income evolve over time. This is shown in Figure 4.2. Before treatment the treated group relied more on agricultural income, accounting for almost 80% of total income. In the mid-term data set (2012) this had dropped to around 55%. There is also a drop for the control group, though this is much smaller. This is largely driving the income effect we find. However, it is possible that 2010 was a better than average year for agriculture. Since the treated group relies on agriculture more, they would be more affected when returning to normal harvest levels. Figure 4.6 suggests, however, that 2010 was a normal year for agriculture.

Next we estimate Model 4.1 for the longer (5-year) period, shown in Table 4.5. Generally, results are similar in the longer run, though a much lower number of observations makes our estimates noisier. In the long run there is again a substantial drop in income in treated villages, again signifying a negative income effect of the plantation. Looking at the stock variables there are no significant differences. House quality now has a positive coefficient on the interaction term (opposite to before), but this is not significant. Access to land remains lower in treated villages, and the effect is now slightly larger (7% lower). We again see no effect on food security. We again see a higher presence of bed nets, but this is not significant.

Table 4.5 – Long Run (2010-2015) effects of a Large-Scale Agricultural investment

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Treated	0.221	0.210	−0.070	0.298**	0.214	−0.002	0.002	−0.042
	(0.204)	(0.200)	(0.169)	(0.144)	(0.144)	(0.002)	(0.011)	(0.068)
Long Run	0.128	0.220	0.279	0.733***	0.361	−0.021	−0.115**	−0.061
	(0.240)	(0.220)	(0.196)	(0.196)	(0.280)	(0.013)	(0.046)	(0.070)
Treated * Long Run	−0.615**	−0.427*	0.104	0.333	−0.424	−0.072**	0.072	0.052
	(0.264)	(0.232)	(0.208)	(0.222)	(0.305)	(0.033)	(0.047)	(0.081)
Constant	0.000	0.000	0.000	0.000	0.000	1.000***	0.990***	0.847***
	(0.176)	(0.168)	(0.157)	(0.116)	(0.104)	(0.000)	(0.010)	(0.058)
Observations	748	748	1256	796	990	1202	1190	1242
# Clusters	44	44	49	47	45	48	48	48

OLS regressions. Standardized and centered on control group at baseline (not columns 6-8). Robust standard errors in parentheses clustered at the village level. IHS is inverse hyperbolic sine transformation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The main effects we find for both the longer and shorter run analyses are a drop in income, lower access to land and some health improvements. For the latter two the link to the plantation is clear: the company is using village land and is providing some health services. In terms of the income effect we have hypothesized that this is caused by an increase in the labor price. Whether there is such a local effect can be tested, which we do next.

Within our control group there is substantial variation in distance to the plantation (defined as the distance to the closest treated village). The mean is 3.5 km with a standard deviation of 2.6 km. We can exploit this variation by repeating our previous analysis, but now taking distance to the plantation as the treatment variable and examining only control villages, as in Model 4.2. The results of this analysis are in Table 4.6. For both measures of income, we see that before the investment, places further away from the plantation had lower average incomes. The interaction shows that control villages further away increased their income more than control villages closer to the plantation. Being 1 SD further away from the investment results in a 0.39 SD higher full income. If higher labor prices are indeed locally determined and spill over partially to neighboring villages we would expect to find these results. For assets we see an increase in the number of assets further away over time, which also holds for house quality (though the effect is much smaller and only marginally significant). Access to land is higher further away from the investment, but it is not very high (1 SD distance leads to 2% higher access) and only significant at the 10% level. This is unsurprising as the treatment is defined by having land leased. That there is some small effect might indicate that households start farming in neighboring villages. There is again no effect on food security, and no effect on mosquito net presence either. Overall, these results provide evidence that there are some local market effects (or spillovers) that are driving the effects we found in Tables 4.4 and 4.5. This could mean that our results in those tables are biased. However, we consistently find that the treatment effect is weaker the further away from the investment. This means that our main estimates are a lower bound of the actual effect.

Table 4.6 – Spillover effects (2010-2012)

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Distance	-0.225^{***}	-0.225^{***}	-0.106^{*}	0.066	-0.011	0.001	0.003	0.033^{**}
	(0.075)	(0.075)	(0.056)	(0.039)	(0.027)	(0.001)	(0.003)	(0.014)
Short Run	-0.236	0.140	0.036	0.304^{***}	-0.092	-0.031^{**}	-0.101^{***}	-0.082^{**}
	(0.166)	(0.124)	(0.103)	(0.057)	(0.065)	(0.011)	(0.013)	(0.039)
Distance * Short Run	0.482^{***}	0.389^{***}	0.246^{***}	0.072^{*}	-0.059	0.016^{*}	-0.010	0.020
	(0.138)	(0.102)	(0.088)	(0.041)	(0.044)	(0.008)	(0.014)	(0.031)
Constant	-0.009	-0.009	0.000	-0.009	0.000	0.997^{***}	0.988^{***}	0.920^{***}
	(0.085)	(0.085)	(0.068)	(0.048)	(0.038)	(0.002)	(0.003)	(0.021)
Observations	1288	1288	1830	1444	980	1750	1784	1828
# Clusters	28	28	28	28	27	28	28	28

OLS regressions. Only control villages included. Normalized variables (not columns 6-8). Distance is euclidean distance (in km) to nearest treated village, standardized. IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

So far, we have been examining these effects across all households in a village. Next, we examine the effect separately for laborers and non-laborers. We lack information on company employment for 2012, thus limiting our analysis comparing over the 2010 - 2015 period. In treated villages, about 40% of the households supplied labor to the plantation, with an average length of employment of 14 months. We examine effects on laborer households in Table 4.7, using Model 4.3. For traditional income, ie excluding wage earnings, we see a substantial drop of 0.33 SD. However, this is not the case when examining full income (which includes salaries). When comparing laborers to the other households in the village, there is an increase in income of 0.34 SD. When we move to the stock variables there is a consistent increase of about 0.2 SD, for the number of assets, housing quality and livestock. Clearly, laborers were able to transform their additional earnings into tangible assets. Furthermore, laborers do not have lower access to land or better health access compared to others in their village. This is unsurprising, as these are effects enjoyed by all households (the health clinics and programs are accessible to all villagers). There is one small effect, in that laborers are slightly more likely (4%) to have had food shortages in the previous year, though it is not very significant. It might be that laborers now working on the plantation are not producing their own food anymore, causing some domestic shortages. This also implies that there is not enough food available on the local food market.

Table 4.7 – Effects on laborers (2010-2015)

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Laborer	0.194	0.186	−0.116	−0.526***	−0.299**	0.003	−0.003	−0.059*
	(0.141)	(0.146)	(0.069)	(0.116)	(0.117)	(0.003)	(0.009)	(0.032)
Treated	0.146	0.138	−0.027	0.449***	0.330**	−0.003	0.004	−0.018
	(0.230)	(0.229)	(0.171)	(0.149)	(0.152)	(0.003)	(0.012)	(0.072)
Long Run	0.128	0.220	0.279	0.733***	0.361	−0.021	−0.115**	−0.061
	(0.240)	(0.220)	(0.196)	(0.196)	(0.281)	(0.013)	(0.046)	(0.070)
Long Run * Treated	−0.485*	−0.560**	0.028	0.259	−0.555*	−0.054*	0.091*	0.052
	(0.262)	(0.240)	(0.209)	(0.222)	(0.310)	(0.029)	(0.048)	(0.083)
Laborer * Long Run	−0.336**	0.344**	0.231**	0.242*	0.336**	−0.045	−0.044*	0.000
	(0.152)	(0.129)	(0.087)	(0.143)	(0.162)	(0.029)	(0.025)	(0.039)
Constant	0.000	0.000	0.000	0.000	0.000	1.000***	0.990***	0.847***
	(0.176)	(0.168)	(0.158)	(0.116)	(0.104)	(0.000)	(0.010)	(0.058)
Observations	748	748	1250	794	990	1202	1190	1242
# Clusters	44	44	49	47	45	48	48	48

OLS regressions. Standardized and centered on control group at baseline (not columns 6-8). Laborers are all households who claimed to work for the company at some point in the 2015 survey. IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, as a logical consequence, we can assess whether the investment has affected within-village inequality. In Figure 4.3 we draw Lorenz curves for both traditional and full income for the treated and untreated group separately for both 2010 and 2012. Panels a and b show the results for traditional income. We see that in 2010 the curve for the treated group is closer to the line of unity, indicating higher equality. After the company has started work this is reversed, shown in panel b, suggesting that inequality has increased for traditional sources of income. When we add company payments (panels c and d) this effect is weakened. Figure A.4.1 shows the same analysis for the long run, with qualitatively similar results. We present a more formal analysis in Table A.4.2 where we analyze the village-level Gini coefficient. For both short and long run, the interaction term is positive, albeit larger for traditional sources of income. This shows that inequality has increased as a result of the company's activities.

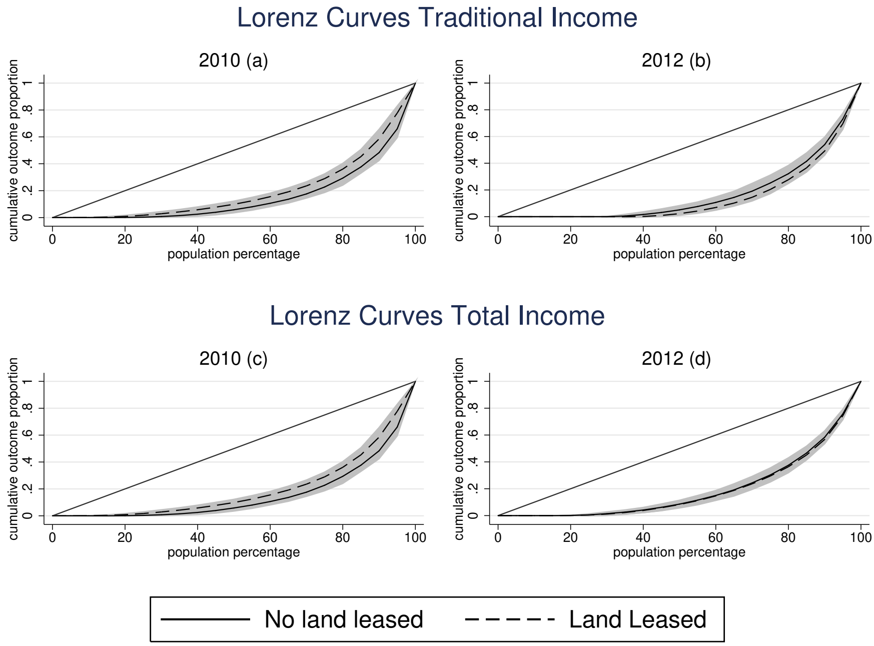


Figure 4.3 – Inequality (2010-2012)

Lorenz curves based on income (not IHS) for panel observations only. Shaded area are confidence intervals, with standard errors clustered at village level. Source: survey data

4.5 Robustness

The previous section showed evidence that this large-scale agricultural investment has had strong effects on local incomes, access to land, health and inequality. In this section we provide a test for our main identifying assumption of parallel trends. Furthermore, we test whether attrition is systematic, examine some bounds on the treatment effect under alternative attrition assumptions and provide some evidence that agricultural conditions are similar in treated and control villages.

We first examine the parallel trends assumption. This is fundamentally untestable, but we gain some reassurance from showing that pre-treatment trends are parallel: they would likely be parallel afterward as well. We do this by examining changes in forest loss for treated and control villages. Agricultural production in Sierra Leone is closely linked to forest loss: most agriculture is rotational bush fallowing, a highly labor-intensive form of production. Under rotational fallowing, forest loss is likely to correlate with increased agricultural production and income in the shorter term - one of our main outcome variables. We examine whether trends in forest loss are parallel using forest loss data from Hansen et al. (2013). Their worldwide dataset contains extremely detailed (30m resolution) data on forest loss for the period 2001-2018. To assess whether forest is lost in a specific year we draw circles with a 1km radius around each village, and then count the number of pixels that were lost in the circle for that village in a certain year (see Figure 4.4). We convert these pixels to the number of hectares lost per village per year and plot this out in Figure 4.5. The vertical black line represents the year that the company started its activities. Trends are very similar across treatment and control pre-2010 but diverge after 2010. For most years the amount of forest loss is significantly higher in treated villages, and the amount of forest lost is always higher than in control villages. Pre-2010 they were almost always equal. We surmise that the divergence after 2010 is partially caused by activity by the company, and partially by farmers having to move to new plots after leasing their land to the company. This figure provides some evidence that pre-treatment trends are parallel, and also that examining forest losses in this context is a relevant variable.

Next, we examine attrition. Attrition in the short run (2 years) is somewhat high at 25% (35% in the control group and 20% in the treated group). For the long run comparison we only interviewed a randomly selected subsample, dropout therefore is much higher: 85% (93% for the control group, and 81% for the treated group).[9] In Table A.4.7 we

[9]We do not consider this attrition but examine the differences for completeness' sake.

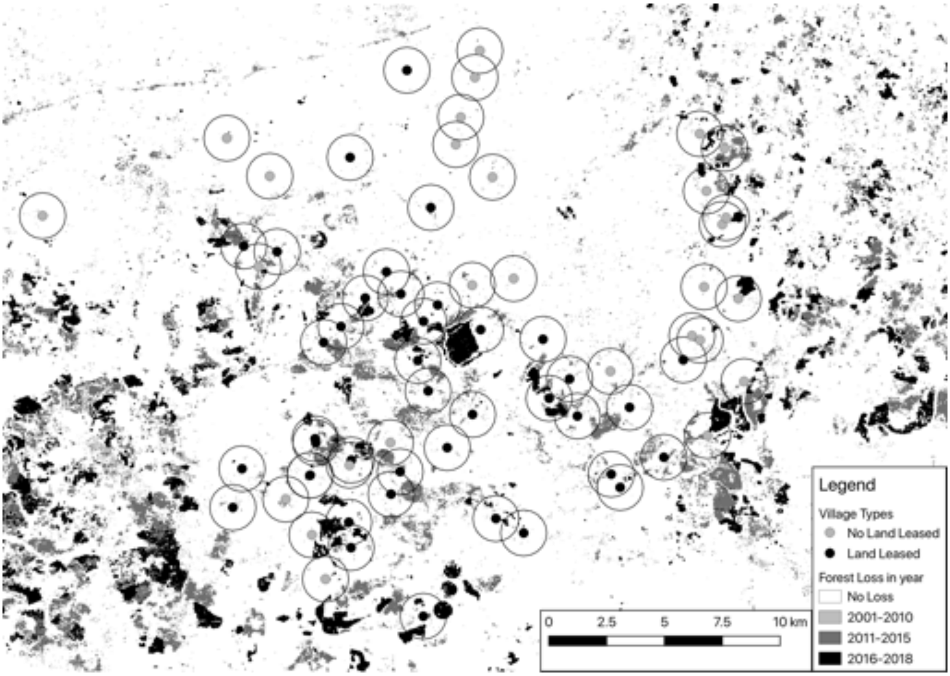


Figure 4.4 – Forest Loss Map

This map shows forest loss from 2001-2018 around the sample villages (Pixel resolution is 30x30m). The circles have a radius of 1km. Forest loss within one of these circles is considered forest loss for that village. Source: Hansen/UMD/Google/USGS/NASA

examine what pre-treatment variables determine attrition, and crucially, whether this differs between the treatment and control group. We see some differences in dropout in the short run (the non-interacting variables), but these are not worrying as this does not indicate differential dropout. There is one worrying finding: a higher traditional income before treatment leads to a lower chance of dropping out in the short run – but in the treatment group only. This could mean that richer households are overrepresented in the short run in treated villages. This means that the negative effect we find is a lower bound of the actual effect. For the long-run dropout we find no significant predictors that differ between the treated and control group.

To further dig into the effect of attrition on the impacts we find we employ a bounds analysis as suggested by Manski (1990) using the approach by Blattman et al. (2014).

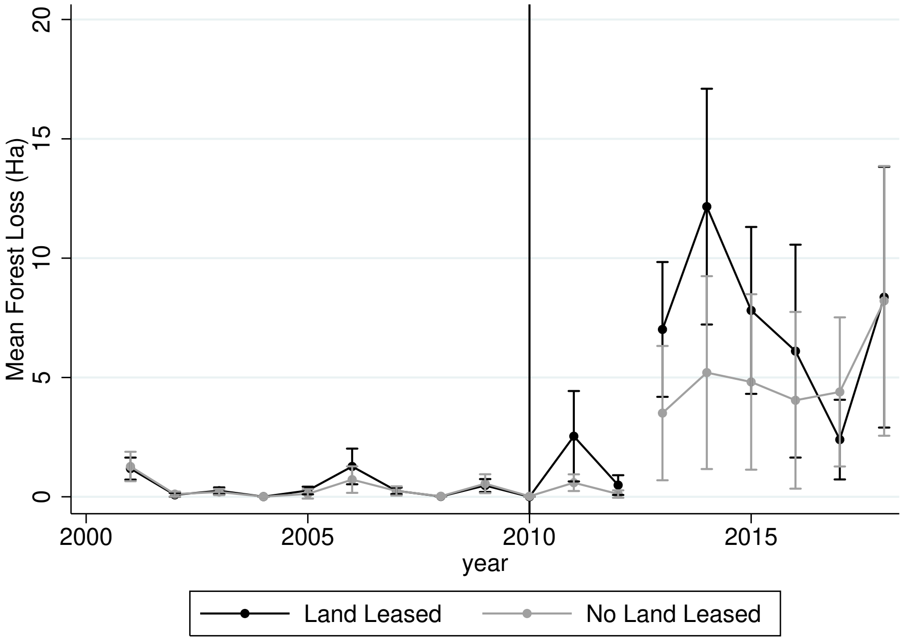


Figure 4.5 – Forest Loss 2001-2018

Graph shows average yearly forest loss in circles with 1km radius around villages. Graph shows 95% confidence intervals. The break in trend lines denotes the inclusion of data from a more precise satellite (Landsat 8). Source: Hansen/UMD/Google/USGS/NASA

For this analysis we make alternative assumptions about those who leave the sample. Values for missing observations are filled in to zero out the treatment effect we find. By doing so we can calculate lower bounds for our treatment effects. We examine four bounds, the Manski worst possible bound, and 3 deviations from the mean. In case of a negative treatment effect, control drop-outs are assigned a low value, while treatment drop-outs are assigned a high value, thus zeroing out the negative treatment effect. For the Manski worst case the high (low) value is that group's maximum (minimum) value. For the SD deviations the high (low) value is the group mean plus (minus) X SD, with X being 0.5, 0.25 and 0.1. These results are shown in Table 4.8 for the short run only. We only show continuous variables as the SD adjustments do not make sense when examining dummies. Column 1 shows the original treatment effect of the treat*post interaction as in Table 4.4. Columns 2-5 show the bounds on this effect. For column 2, the Manski worst possible bound, the treatment effect is opposite to our original effect and highly significant for all outcome variables. This is unsurprising when attrition is high (Blattman et al. (2014) find this also) and is shown here for completeness. When we examine our main effect on income the results from columns 3-5 are reassuring. In most cases the signs of all coefficients are the same, and for the 0.1 and 0.25 SD deviation these effects are significant as well. Note that these deviations represent large, systematic deviations on the characteristics of drop-outs for which we found no evidence in Table A.4.7.

Table 4.8 – Bounds Analysis

	(1) Original	(2) Worst	(3) + (-) 0.5SD	(4) + (-) 0.25SD	(5) + (-) 0.1SD)
Traditional income (IHS)	-0.625**	0.526*	-0.194	-0.377*	-0.488**
	(0.262)	(0.310)	(0.212)	(0.198)	(0.192)
Full Income (IHS)	-0.426**	0.760**	-0.036	-0.182	-0.270*
	(0.197)	(0.298)	(0.160)	(0.145)	(0.139)
# Assets	-0.047	1.946***	0.284**	0.126	0.031
	(0.141)	(0.372)	(0.122)	(0.106)	(0.100)
House Quality (Score)	-0.110	1.753***	0.291**	0.125	0.025
	(0.080)	(0.377)	(0.112)	(0.087)	(0.075)
Tropical Livestock unit	0.315	-3.004***	-0.386**	-0.084	0.098
	(0.192)	(0.753)	(0.185)	(0.155)	(0.144)

OLS Regressions, standard errors in parentheses clustered at the village level. Table reports coefficients of interaction terms of treatment and later time period. Column 1 reports the same coefficients as in table 4 and ignores attrited households. Columns 2-5 give alternative values to attrited households, depending on whether the original coefficient is positive (or negative). For column 2 (worst-case) attrited households in the treatment group get the minimum (maximum) in the treatment group and households in the control group get the maximum (minimum) in the control group. Column 3-5 assigns attrited households in the treatment group the treatment mean minus (plus) X SD, and attrited households in the control group the control mean plus (minus) X SD.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, since our identification strategy partially relies on the distance to a large river, there might be differences in the agricultural suitability of the available farmland. We explore this by examining the EVI (Enhanced vegetation index), which is a measure of live green vegetation based on satellite imagery. The EVI ranges from -1 (water bodies) to 0 (desert) to 1 (mature forest). It is often used to examine fertility/crop success in a certain year. We plot trends in EVI in Figure 4.6 and examine the maximum yearly EVI in the same circles around villages with a 1km radius. By using the maximum we automatically filter out clouds and looking for the maximum within a year means we examine the entire agricultural season. We coarsen the pixels from 30x30m to 150x150m to reduce spatial autocorrelation. The trends are extremely similar. This shows that treated and control villages are subject to very similar agricultural conditions. Interesting to note is that there is no difference between treatment and control after 2010, the company's start year. It appears that while the company did contribute to significant forest loss, live green vegetation was unaffected.

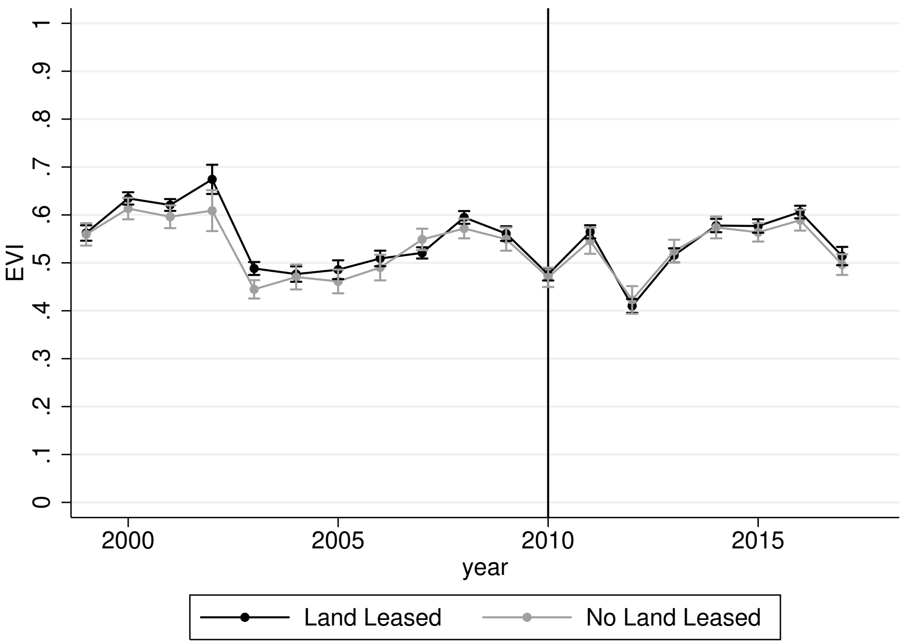


Figure 4.6 – Greenness (EVI) 1999-2017

Graph shows average maximum yearly EVI (greenness or vegetation) in circles with 1km radius around villages. Original pixel size was 30x30m, coarsened to 150x150m to reduce spatial correlation. Graph shows 95% confidence intervals. Source: USGS

4.6 Conclusion

This paper is one of the first to provide empirical evidence for the impact of large-scale agricultural investments. This allows us to examine how rural communities respond to land and labor shocks. While there might be positive effects (higher incomes, better infrastructure and access to new farming technologies), most research so far has pointed to negative effects: loss of land, increased marginalization and exploitation by powerful foreign companies (Baxter, 2013; De Schutter, 2011; Liversage, 2010). Our case is a large-scale agricultural investment in Sierra Leone, a country which has received a lot of interest from investors in land during the past decade, consequent on an opening to international capital following a decade of civil war. A for-profit company leases 24'000 hectares of land and uses this to grow sugarcane for biofuel. The company pays landowners yearly for the land and employs local labor on the farm.

We use a difference-in-difference analysis to compare outcomes for communities within and outside the catchment area of the plantation investment. We find a large drop in average incomes for treated communities, almost half a standard deviation compared to the control group at baseline. This is mainly driven by lower agricultural income. We surmise that this is because the increased labor demand increases the labor price, making it too expensive to hire in labor, the most important factor of production. A spillover analysis confirms this. We see mixed effects on physical assets. It might be that households are holding on to (some) of their assets to weather future shocks. We also see a drop in access to land, which runs counter to the argument that land is plentiful and not a relevant constraint. Lower access to land also likely contributes to lower agricultural income. Food security is largely unaffected, surprisingly. This suggests that the cause of seasonal hunger is related to storage issues and local market failures. We also see some improvements in health, which plausibly can be attributed to the company health program. These effects hold in the longer run (5 years) as well. When we examine company laborers specifically, we see that they benefit relative to non-laborers in their village. Their incomes rise and this translates into more tangible assets.

We have hypothesized and given evidence that a portion of the impact is transmitted through local markets, especially the labor market. This is likely to hold for most external agrarian investors; by definition they are looking to acquire land, and often seek to hire local labor to lower transportation costs and to obtain goodwill from the local community. This shows that to examine the full impact of one of these kinds of investment the full village economy should be examined. To improve this, local economy models as in Taylor and Filipski (2014) should be developed to gain more insight into the functioning of local markets and social welfare institutions. We leave this for future work.

Taken together, the results from this paper paint a bleak picture. While an increase in income for laborers is positive, only 40% of households provide laborers, and their gains do not outweigh the losses by households without income from plantation laboring. It is likely that this inequality and marginalization increases social conflict, a scenario suggested in previous work. The investment as a whole appears to be a poor deal for recipient communities. Indeed, that over one-third of land lease payments go to political elites rather than the landowners themselves is a warning sign that local benefit is not the priority in these kinds of investments.

4.7 Appendix

Table A.4.1 – Variable definitions

Variable	Variable Definition
Traditional Income	Sum of Agricultural and livestock sales, self-employment and other income (including remittances) in January of that year. Winsorized at the 95% level, then transformed with Inverse Hyperbolic Sine
Total Income	Traditional income, plus company's land payments (2012 only) and salaried income (2015 only). Winsorized at the 95% level, then transformed with Inverse Hyperbolic Sine
# Assets	Sum of how many of the following assets they owned: house, car, bicycle, tv, radio, satellite, sewing machine, fridge, iron pots, iron kettle, mobile phone, bed mattress, motorcycle, plastic chairs, mosquito nets, tractor, generator
House Quality	Score based on the average quality of their houses. Floors: No floor 0p, Mud 1p, Cement 5p. Walls: Wattle & Daub 1p, Reeds & Thatch 2p, Mud bricks 3p, Mud bricks and plaster 4p, Wooden 4p, Concrete 5p. Roof: None 0p, Thatch 1p, Tarp 2p, Zinc 5p. Maximum score: 33
Livestock	Tropical livestock unit on number of livestock owned, based on cattle, goats, sheep, pigs, rabbits and chickens. One tropical livestock unit is often equated to a 250 kg animal (Jahnke, 1982).
Access to land	Answer to question 'Do you currently have access to land for cultivation?' (yes/no)
Food Security	Answer to question 'Was there a shortage of food in the household at any time last year?' (yes/no)
Bed net	Whether a bed net is present in the household (yes/no)

Table A.4.2 – Inequality Short and Long Run

	(1) Gini Short Run Traditional Income	(2) Gini Short Run Total Income	(3) Gini Long Run Traditional Income	(4) Gini Long Run Total Income
Treated	−0.051	−0.051	−0.051	−0.040
	(0.034)	(0.034)	(0.061)	(0.051)
Post	0.154***	0.038	0.006	−0.004
	(0.042)	(0.040)	(0.071)	(0.062)
Treated * Post	0.114**	0.036	0.162**	0.087
	(0.055)	(0.049)	(0.078)	(0.070)
Constant	0.170***	0.170***	0.149**	0.133***
	(0.029)	(0.029)	(0.057)	(0.046)
Observations	96	96	54	54

OLS regressions. Gini Index is calculated only for villages with at least 5 observations with income data.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

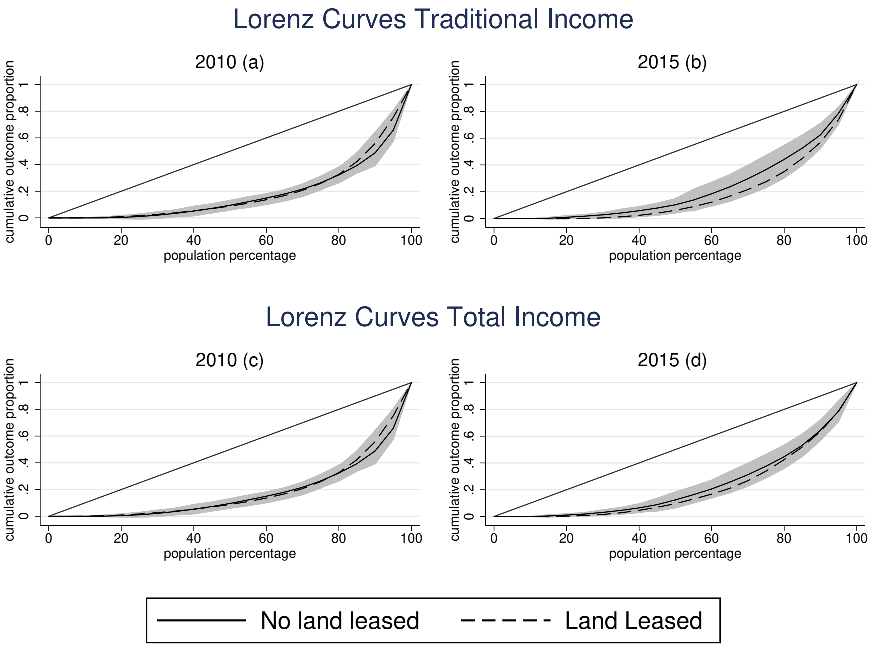


Figure A.4.1 – Inequality (2010-2015)

Lorenz curves based on income (not IHS) for panel observations only. Shaded area are confidence intervals, with standard errors clustered at village level. Source: survey data

Table A.4.3 – Short Run (2010-2012) effects of a Large-Scale Agricultural investment: stricter merge results

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Treated	0.387^{***}	0.387^{***}	-0.028	0.115	0.039	-0.004	0.002	-0.120^{***}
	(0.146)	(0.146)	(0.095)	(0.082)	(0.089)	(0.002)	(0.005)	(0.042)
Short Run	-0.241	0.150	0.066	0.326^{***}	-0.089	-0.030^{**}	-0.103^{***}	-0.074^{**}
	(0.234)	(0.180)	(0.122)	(0.062)	(0.073)	(0.012)	(0.014)	(0.035)
Treated * Short Run	-0.622^{**}	-0.452^{**}	-0.069	-0.127	0.358	-0.046^{**}	0.000	0.133^{***}
	(0.267)	(0.198)	(0.138)	(0.083)	(0.261)	(0.019)	(0.019)	(0.047)
Constant	0.000	0.000	0.000	0.000	0.000	0.997^{***}	0.989^{***}	0.921^{***}
	(0.113)	(0.113)	(0.072)	(0.051)	(0.032)	(0.002)	(0.004)	(0.024)
Observations	3428	3428	5764	3578	3200	5570	5540	5758
# Clusters	65	65	67	67	65	67	67	67

OLS regressions. Standardized and centered on control group at baseline (not columns 6-8). IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level. Sample is based on a more restrictive merge which also checks name, village name and number of years lived in the area.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4.4 – Long Run (2010-2015) effects of a Large-Scale Agricultural investment: stricter merge results

	(1) Traditional income (IHS)	(2) Full Income (IHS)	(3) # Assets	(4) House Quality (Score)	(5) Tropical Livestock unit	(6) Access to Land (=1)	(7) Food shortage (=1)	(8) Bed net in household (=1)
Treated	0.243	0.233	−0.085	0.290*	0.175	−0.002	0.002	−0.075
	(0.205)	(0.200)	(0.193)	(0.148)	(0.144)	(0.002)	(0.012)	(0.075)
Long Run	0.177	0.275	0.273	0.760***	0.321	−0.022	−0.100**	−0.087
	(0.259)	(0.233)	(0.205)	(0.189)	(0.290)	(0.014)	(0.045)	(0.075)
Treated * Long Run	−0.658**	−0.482*	0.086	0.222	−0.400	−0.062**	0.055	0.085
	(0.275)	(0.242)	(0.219)	(0.214)	(0.315)	(0.027)	(0.047)	(0.088)
Constant	0.000	0.000	0.000	0.000	0.000	1.000	0.989***	0.870***
	(0.179)	(0.170)	(0.181)	(0.123)	(0.103)	(.)	(0.011)	(0.065)
Observations	690	690	1118	716	890	1076	1062	1108
# Clusters	43	43	47	46	45	47	47	47

OLS regressions. Standardized and centered on control group at baseline (not columns 6-8). IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level. Sample is based on a more restrictive merge which also checks name and village name.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4.5 – Short Run (2010-2012) of a Large-Scale Agricultural investment: Income splits

	(1) Agricultural Income (IHS)	(2) Livestock Income (IHS)	(3) Self-Employment Income (IHS)	(4) Other Income (IHS)
Treated	0.470^{**}	-0.184^{**}	-0.192	-0.087
	(0.181)	(0.087)	(0.124)	(0.060)
Short Run	-0.280	-0.068	0.034	0.579^{***}
	(0.223)	(0.097)	(0.072)	(0.097)
Treated * Short Run	-0.695^{***}	0.064	0.167	-0.004
	(0.250)	(0.114)	(0.127)	(0.120)
Constant	0.000	0.000	0.000	0.000
	(0.139)	(0.077)	(0.077)	(0.048)
Observations	3762	3762	3762	3762
# Clusters	67	67	67	67

OLS regressions. Standardized and centered on control group at baseline. IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level. Traditional income variable in main tables are sum of these main components of income.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4.6 – Long Run (2010-2015) of a Large-Scale Agricultural investment: Income splits

	(1) Agricultural Income (IHS)	(2) Livestock Income (IHS)	(3) Self-Employment Income (IHS)	(4) Other Income (IHS)
Treated	0.176	0.174	0.347	0.180
	(0.255)	(0.147)	(0.241)	(0.135)
Long Run	−0.414	0.261*	1.148***	0.909***
	(0.306)	(0.141)	(0.209)	(0.222)
Treated * Long Run	−0.497	−0.128	−0.310	−0.274
	(0.332)	(0.174)	(0.299)	(0.249)
Constant	0.000	0.000	0.000	0.000
	(0.211)	(0.122)	(0.112)	(0.097)
Observations	748	748	748	748
# Clusters	44	44	44	44

OLS regressions. Standardized and centered on control group at baseline. IHS is inverse hyperbolic sine transformation. Robust standard errors in parentheses clustered at the village level. Traditional income variable in main tables are sum of these main components of income.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4.7 – Attrition

	(1) Dropout SR	(2) Dropout LR
Traditional Income (Leones, IHS)	0.057**	0.004
	(0.028)	(0.018)
# Assets	0.035	-0.055
	(0.064)	(0.072)
House quality (Score, 1-33)	-0.037	0.003
	(0.050)	(0.034)
Tropical Livestock Unit	0.026	0.246
	(0.025)	(0.302)
Access to Land (=1)	-0.293	0.000
	(0.811)	(.)
Food Shortage (=1)	-0.025	0.000
	(0.451)	(.)
Bed net in Household	-0.198	0.558*
	(0.274)	(0.295)
Treated	0.608	0.112
	(0.909)	(0.530)
* Traditional Income (Leones, IHS)	-0.096**	0.008
	(0.042)	(0.021)
* # Assets	-0.048	0.011
	(0.074)	(0.082)
* House quality (Score, 1-33)	0.012	-0.051
	(0.056)	(0.041)
* Tropical Livestock Unit	-0.021	-0.293
	(0.138)	(0.316)
* Access to land (=1)	0.000	0.000
	(.)	(.)
* Food Shortage (=1)	-0.096	0.000
	(0.602)	(.)
* Bed net in household (=1)	0.088	-0.624*
	(0.309)	(0.359)
Constant	-0.440	0.899**
	(1.022)	(0.429)
Observations	1223	1198
# Clusters	76	75
% Dropped out	0.25	0.85

Probit regressions. Robust standard errors in parentheses clustered at the village level. Data is all 2010 data with indicators for being absent in later rounds. Some dummy variables are dropped from the model because of low variation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 5

Productive Spillovers of Foreign Land Investments

There is very little known about the effectiveness of foreign agricultural investments on development. One possible mechanism, stemming from the literature on FDI, is that investments can cause productive spillovers, increasing productivity on local farms. This paper empirically tests this at the micro-level, by examining productive spillovers of a foreign agricultural investment on local production. We use Difference-in-Difference design over three years with systematically selected control observations. We find that the presence of the investment leads to lower productive losses, both at the extensive and intensive margin. We find no evidence that this is caused by increased labor effort. Welfare is unaffected, though this unsurprising as the investment is still in the startup phase. We conclude that productive spillovers can be an important positive externality of foreign investments.

Publication status: Hofman, P. and Voors, M., 2020. Productive Spillovers of Foreign Land Investments. Working Paper.

5.1 Introduction

There has been a marked increase in foreign investments in agriculture in African countries. The Land Matrix, a watchdog group that tracks all foreign-funded investments worldwide has recorded 585 concluded deals (which means deals are signed and work has begun on construction), totaling over 15 million hectares of land (Landmatrix, 2020).[1] This increase started during the 2007-2008 price spike of commodities and continued with interest in biofuels to reach global climate goals (Koning and van Ittersum, 2009; Arezki et al., 2013). Furthermore, many Western governments are remodeling their development aid practices to focus on 'Business for Development': stimulating national companies to invest in developing countries to spur development, for example through agricultural investments (Engström and Hajdu, 2019; Kolk et al., 2008). Foreign-funded agribusiness schemes are a possible yet controversial method to overcome rural poverty. They can be seen as a form of agricultural foreign direct investment (FDI). There are several advantages to this approach: risks are borne by large companies that can shoulder the burden. Investment funds are plentiful in the North, while investments lag in West-Africa. Investments are more likely to be long-term than conventional development projects which have fixed end dates. Finally, North-based companies have extensive knowledge of developing large-scale agricultural projects and can operate more efficiently (Ellis, 2005; Collier and Dercon, 2014; Godfray et al., 2010).

The literature on foreign direct investments can shed insight on agricultural investments' effectiveness as a development tool. There is a long-standing debate in economics on the relation between foreign direct investment (FDI) and economic growth in developing countries (de Mello Jr., 1997). The most important channel theorized to increase growth is through technology spillovers that boost local productivity (Crespo and Fontoura, 2007; Liu, 2008). Most papers test this hypothesis using cross-country datasets of African economies and find zero to weakly positive effects. This depends on the time period (Gui-Diby, 2014), the econometric estimator used (Herzer et al., 2008), including remittances and foreign aid as covariates (Nwaogu and Ryan, 2015), examining input accumulation as a channel (Makiela and Ouattara, 2018), including financial fragility as a covariate (Hagan and Amoah, 2019) and examining the role of institutions (Li and Tanna, 2019). Country-level analyses point to backward linkages as a key positive impact but are similarly inconclusive (Liu, 2008; Jordaan, 2008). There is scant micro-level evidence on technology spillovers through FDI (Deininger and Xia, 2016; Ali et al., 2019; Lay et al., 2018).

[1] Date of access: February 2020.

We add to this literature in several ways. We examine the micro-level productivity spillovers of an agricultural FDI. The micro-level allows for a more robust identification strategy (Difference-in-Difference) compared to macro papers. It also allows us to examine the direct effect of the investment on local productivity, rather than aggregate measures that are sensitive to measurement error. Furthermore, there is very little evidence on this new form of agricultural FDI, both on welfare and technology spillovers, both of which we examine. We use a Difference-in-Difference analysis with data from before the investment was started. We use a novel and systematic approach to select control villages: they are selected using a matching algorithm (Coarsened Exact Matching, CEM (Iacus et al., 2011)) based on satellite data, allowing us to select from all villages in the area.

The agribusiness investment we examine is located in Sierra Leone, which has been one of the larger recipients of foreign investments. Indeed, in one estimate over 25% of the total arable land in Sierra Leone is under contract with a foreign company (Landmatrix, 2020). We examine a 750-hectare cocoa plantation/outgrower scheme. For the investment the costs are borne by the foreign company, while the local population supplies the labor and receives a wage. In Sierra Leone labor is the main constraint to production, with over 65% of households experiencing a labor shortage during the peak agricultural season (MAFFS, 2011). The investment is located in Eastern Sierra Leone, which has had smallholder cocoa farmers for many years, though production is very low. This is because during the brutal Civil War (1990-2002) most cocoa farms were abandoned for many years and became overgrown. Average yields are around 100 kg/ha, much lower than yields in, for example, Indonesia which range from 600-1500 kg/ha (Tothmihaly and Ingram, 2017). Therefore, there is a large scope for yield increases through productive spillovers.

We see some evidence that knowledge on cocoa farming from this investment spills over: There is a marked decrease in losses to a common fungal disease. We see no strong economic effects, indicating that the investment is not making the local population (much) worse off. We also see so no clear effects on local food expenditures and the local labor price, opposite to earlier results (Hofman et al., 2020).

This paper is structured as follows: we examine key literature in section 5.2, section 5.3 details the investment and our study design, while section 5.4 provides results, before concluding.

5.2 Literature Review

First, we examine theoretical predictions on the effect of plantations on welfare and productive spillovers. Kleemann and Thiele (2015) focus on the effect of plantations on the local food market. They predict that local food supply will decrease as the land available for staple crops is reduced. As food markets are often dysfunctional in Sub-Saharan Africa, this will increase the price of food, which negatively affects laborers and possibly subsistence farmers (depending on whether they are net buyers or sellers of food). However, following the literature on FDI, a plantation can have spillover effects: knowledge gains, lower prices of inputs and availability of inputs can improve local farming practices. Furthermore, they predict that larger plantations will lead to lower wages for day labor, as larger plantations displace more laborers, leading to a larger supply of potential laborers, and thus lower wages. However, more labor-intensive production will demand more labor and thus lead to a higher wage. Net welfare crucially depends on how food and labor prices respond, and the size of the spillovers. We expect a high likelihood of productive spillovers in this case as the crop and intensity of farming are closely aligned with local farming. Effects on the local food price and local labor market are also possible, and we examine these as well.

There are a few papers that are closely aligned with this paper and examine the productive spillovers of foreign investments. Deininger and Xia (2016) examine this using a nation-wide survey on large and small farms in Mozambique. They argue that distance to a large, same-crop farm is an important determinant of spillovers, and exploit variation in these distances for their identification strategy. With a long-running panel on 6000 small-scale farmers and locations of large-scale farms throughout the country, they use a difference-in-difference approach to examine the effect on agricultural practices, modern input use and yields. They find that within 25km of large-scale farms, small farms are more likely to adopt agricultural practices and use more modern inputs (e.g. chemical fertilizer), though the effect is modest. Crucially however, this does not affect yields, nor did it increase access to local output markets. Ali et al. (2019) use the same approach for Ethiopia. They find mixed improvements in fertilizer use, depending on the crop, and small increases in yields for maize and wheat.

Lay et al. (2018) use a similar but less spatially disaggregated approach: they compare Zambian wards with foreign investments with wards without. Using a long-running (10 year) nationwide (27'000 households) survey on small-scale farms. They do not have a panel dataset so add additional controls on household characteristics to correct for selection bias. They find an increase in farm size in wards where large-scale farms work, though they point out that this may be driven by medium-scale farmers following in the wake of large-scale farms. There is an effect of large-scale farm presence on fertilizer use on small farms as in Deininger and Xia (2016), though this might be driven by state-driven fertilizer distribution with a focus on larger farms. They also find increases in maize yields, but this holds mostly for larger small-scale farmers, so this might again be caused by newly established medium-scale farms. Overall, they find some evidence of productive spillovers, but their identification strategy does not allow them to disentangle whether the effect is driven by spillovers to small-scale farms, or from new entrants into the market.

We contribute to these strands of literature in several ways. We make an important addition to the literature on micro-level effects by using an empirical design where we can examine effects over time, which only a small subset of papers can examine. Our difference-in-difference approach allows us to net out all time-invariant effects such as soil suitability and local market access. We also use matching to select control villages, which is an improvement over the baseline ad-hoc approach. Finally, we have a range of outcome variables that allows us to examine the effect on knowledge/technology spillovers explicitly, which has been understudied at the micro-level.

5.3 Study Design

5.3.1 Setting

The agribusiness scheme we examine is located in Sierra Leone. Sierra Leone has seen a large number of large-scale land acquisitions: since 2000 24 deals have been completed, covering over 1 million hectares of land (25% of total arable land) (Landmatrix, 2020). Land in Sierra Leone is owned by extended families within which land is distributed to individual families. Families can lease (not sell) land for 25-50 years, subject to permission by the local authorities. Sierra Leone is a poor country: it is in the bottom 5% on the Human Development index (UNDP, 2016), the poverty rate is high and food security is low. Most Sierra Leoneans are smallholder farmers, especially in rural areas. Farm productivity is low. The main agricultural input is labor (besides land). Fertilizer and high-yielding seeds are virtually unavailable. In 2011 the government conducted a nation-wide survey and found that 65% of households experienced a shortage of labor in the agricultural season. Farm production to a large extent relies on labor from within the family. About one-third of households hire labor (MAFFS, 2011). The likely constraint to production is therefore likely to be labor, not land.

5.3.2 Treatment Description

The treatment revolves around the establishment of 'Block Farms' around eight villages in Eastern Sierra Leone. 'Block Farms' is a local name for internationally very common outgrower schemes, though the implementation differs in several aspects. The eight villages provide unused land that is property of one or more of the landowning families in the community. This results in 8 plots of plantation land which together are around 750 hectares in size. They establish a 15-year partnership: the Implementing Partner (IP, which is a joint venture between an international and local cocoa buyer) provides inputs and the families provide land. The IP establishes cocoa farms on this land. Labor is communally organized within the village and the laborers are paid a daily wage (In most outgrower schemes plots of land are assigned or owned individually and the IP provides inputs. This is the long-term aim). The company hires laborers to clear the land, plant cocoa seedlings and plant intercrops. After establishment the IP oversees farm management: proper pruning, underbrushing and shade management of the cocoa trees. The landowning families own any harvests from the intercrops. After 4-5 years, the first crop will be ready. When the farms start producing, the landowning families will sell the harvest exclusively to the IP.[2] Note that none of the farms are producing yet, so we cannot examine the effects of cocoa sales yet. After the 15 years the ownership of the farm, including the trees, transfers back to the landowning families and they face no more selling restrictions.

5.3.3 Sample

To examine the effect of the agribusiness scheme on the local population we collected survey data in 2016 and 2019. The treatment was implemented in 2016, just after our first round of data collection. Negotiations about the village land had been ongoing at

[2]The price will be 70% of FOB (Free on Board) price, which more or less equals the local market price. 20% of the proceeds go to the IP: this is for project costs (eg the labor costs). 50% goes to landowning families. The intent is to distribute directly to the farmers that worked on the land rather than just the heads of the family. 10% is for fees/transport costs. 20% is for the local cocoa buyer. Of course, the Cocoa Buyer also makes a profit when selling the cocoa to an international buyer. As the IP knows the farm size it can figure out the extent of side-selling: selling to another trader.

this point and the contracts were nearly finalized. After three years we came back and spoke to the same people interviewed in 2016. This section details how our sample was selected, and the number of observations over time.

Control villages are selected to be similar to the treatment villages, based on satellite imagery.[3] The following characteristics were mentioned by the IP as being important for their selection process:

1. Enough Suitable land. Cocoa needs to be planted in existing forest to have enough shade. We proxy this by examining the amount of forested land and the slope of the land.
2. Enough land available. The IP wanted bigger plots rather than many scattered plots across many villages. We proxy this by drawing Voronoi polygons for all villages and then calculating the size of these polygons.
3. Easily accessible from the local town. This makes it possible for the IP to make regular visits. All of the treatment villages are on tarmac roads or good dirt roads close to the local city. We proxy this by looking at distance to the local city and distance to the nearest road (as the crow flies).
4. Village size. This was not explicitly mentioned by the IP, but we expect enough labor supply to be important. This is proxied with the number of houses in a village.

We collect this data for all villages in 2016 (including treatment) within a 30 km radius around the local town, the basis of operations of the IP. This leads to a list of 464 villages. Then, we matched these villages to the treatment villages. The most common matching algorithm is Propensity Score Matching (PSM). King and Nielsen (2019) suggests a better matching algorithm: Coarsened Exact Matching (CEM). CEM divides the variables of interest into bins of a (user-chosen) width and then looks for exact matches within these bins. This allows for more user-choice and requires fewer assumptions than PSM. It also ensures that the resulting balance applies to all variables. We match the villages where the company works to the villages in the 30km radius, based on the variables explained above. For each treatment village this leads to a list of potential control villages. We hand-check the matches and drop villages that are much further away from the local city by road rather than as the crow flies. Then we randomly select two villages from the list that remains. The exact matching bins are described in Table A.5.1.

[3]We also conducted interviews in 'spillover' villages (8 in total) which are closely situated to the treated villages. We find no evidence of spillovers and add this group to the control sample.

Within villages we take the following sampling approach. We first made a roster of all households within the village. From this roster, we randomly selected 45 households. We then talk to the head of household of the selected households. If the head of household was unavailable and would not be around on later days we spoke to the wife or the eldest son. If neither of those was available a household was selected from a backup list, which was also randomly drawn from the population of households.

Three years later we returned to interview the same individuals. We used printed photographs that were taken at the end of the survey in 2016 to identify the same individual in 2019 (with their explicit permission). Once again, if the head of household would not be present in the 2-3 days we were present in the village we spoke to the wife or eldest son. When households are absent or moved out they are replaced with other households, randomly drawn from the 2016 census. These observations are excluded from our main analyses. Attrition is relatively low at 10%, and equal between control and treated groups.

5.3.4 Empirical Model

This section defines our empirical model. For the unconditional treatment effect we estimate the following model:

$$\boldsymbol{Y}_{ij} = \beta_0 + \beta_1 agrib_j + \beta_2 post_{ij} + \beta_3 post_{ij} * agrib_j + \boldsymbol{\beta_4 matchgroup_j} + \varepsilon_{ij} \tag{5.1}$$

Where $\boldsymbol{Y}_{ij}$ refers to our set of outcomes, $agrib_j$ is a dummy for villages where the agribusiness scheme was set up, $post_{ij}$ is a dummy referring to the later time period. $\boldsymbol{matchgroup}_j$ is a set of dummies for each group of matched villages (consisting of treatment and control villages). β_3 is our coefficient of interest. i indexes the household level, while j indexes the village level. We cluster standard errors at the village level.

5.4 Results

This section details our results. We first do a manipulation check, where we examine what is happening in these villages as a result of the investment, then we discuss descriptive statistics of our sample, before moving on to our main results on welfare and knowledge spillovers

5.4.1 Manipulation Check

Table 5.1 is a short manipulation check, where we examine the direct effects of the investment. In treatment villages 35% of the sample has worked for the company at least one day in the previous year, versus a much lower rate (5%) in control villages. This 35% worked on average 20 days in the previous year. This is a substantial amount of work, especially as most work takes place in the period when labor demand is highest on smallholder farms.[4] 28% of households gave some of their land to the investment, which means that eventual payments should accrue to a substantial portion of the village population.[5] This shows us that this investment is already affecting inhabitants through labor requests, and future revenues will accrue to a substantial portion of the village.

Table 5.1 – Manipulation Check

	Treatment		Control		
	mean	sd	mean	sd	Diff
Worked at least one day on block farm in previous year (=1)	0.35	0.48	0.05	0.22	0.216***
# days worked for block farms in previous year	20.43	22.20	14.37	15.48	5.079
Gave land to block farms project (=1)	0.28	0.45	0.03	0.16	0.231***
N	305		1003		

Source: 2019 survey data. Diff column is a simple regression comparing treatment and control villages, with standard errors clustered at the village level * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4.2 Descriptive Statistics

Table 5.2 shows some descriptive statistics of our sample, based on the 2016 data (so before activities by the IP). Only a small portion of households is led by a female: 88% of the households are male-headed. The average age of the head of household is 42, and 42% is literate (36% in control). 33% of households are migrants. Households consist of 2.3 adults on average. 64% of households in treated villages have a cocoa farm,

[4] The 14 days worked on average in control villages only applies to the 5%, so total labor effort remains low. A low number of observations in the control group explains the lack of significant effect.

[5] Rate is 3% in control villages. This can likely be attributed to participants misunderstanding the question.

which is significantly lower in control villages (48%). It might be that matching villages on forest cover and the slope is not enough to control for differences in suitability for cocoa, and this is driving that difference. The average farm size is small, at 2.5 Ha. Cocoa farms are somewhat smaller at 1.5 Ha. We look at two forms of expenditures: irregular, larger expenditures (for example school fees) and monthly consumption (food) expenditures. Both are very low: 325'000 leones monthly equates to about 56 dollars in monthly expenditures. Spread out over at least 2 adults in the household this means that per capita expenditures are lower than a dollar daily.

Table 5.2 – Descriptive Statistics

		Treatment			Control		
	n	mean	sd	n	mean	sd	Diff
Gender (1=male)	323	0.88	0.32	1023	0.84	0.37	0.027
Age (years)	322	42.38	12.59	1021	43.70	14.17	-1.434
Literate in English or Arabic (1=yes)	323	0.42	0.49	1023	0.36	0.48	0.015
Migrant (1=yes)	323	0.33	0.47	1023	0.34	0.47	-0.020
# Adults in household	322	2.29	2.13	1023	2.40	2.16	-0.067
Belongs to landowning family (1=yes)	323	0.77	0.42	1023	0.75	0.43	0.029
Has cocoa farm (1=yes)	323	0.64	0.48	1023	0.48	0.50	0.162**
Farmsize (Ha), excluding cocoa farm	304	2.52	2.21	922	2.15	1.88	0.735
Cocoa farm size (Ha)	206	1.57	1.73	496	1.60	1.84	-0.028
Yearly expenditures (in 1'000 Le)	323	1173.65	854.34	1023	1004.16	851.56	115.788
Monthly expenditures (in 1'000 Le)	323	325.78	187.16	1023	308.99	168.48	23.426

Source: 2016 survey data. Diff column is a simple regression comparing values of Treatment to Control, with standard errors clustered at the village level * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4.3 Effects

The literature on FDI points to knowledge spillovers as one of the main mechanisms through which foreign investments impact local producers. The crop of the agribusiness scheme, cocoa, has a very long history in Sierra Leone and is the most important cash crop in the region. Indeed, in our sample between 64 and 48% own a cocoa farm. Results on cocoa farm management and performance are shown in Table 5.3. We see clear effects on the prevalence of black pod, a fungal disease that is prevalent in the region. Both at the extensive margin (incidence down by 12%) and the intensive margin, where the effect is substantial: losses go down by 17 Kg, though losses were initially higher. These are substantial and important impacts. Local NGOs dub black pod the largest problem for the local cocoa sector. We don't see these gains translated into significantly improved yields, though this self-reported measure has very high variability. We also expect that the loss measure is more precise than the yields measure as people are more sensitive to losses and are thus more likely to precisely remember these. Overall yields do improve substantially in the later time period by 21 Kg/Ha (a 27% increase). We see no evidence that farmers have increased the size of their cocoa farm, though self-reported farm sizes are highly sensitive to measurement error and this outcome therefore has high variation. Improved farm management might have contributed to the reduction in black pod losses, though we find no evidence for this. We do not see farmers increasing the number of measures they use against black pod (such as regular checks or using fungicides). Farmers do not 'brush' (remove undergrowth) their farms more regularly, which is one of the main methods to reduce the fungal growth. Shade cover reduces by 5% though not different in the block farm villages. Lower shade cover can also reduce fungal growth by reducing humidity. Interestingly, the number of days worked on the farm drops substantially by 29 days in the later time period (a 36% reduction). That this not leads to lower yields shows that more effective techniques may be used. This does not differ between block farm and untreated households though. Overall we see some potential effects of productive spillovers on the incidence of black pod and losses, though this does not appear to be caused by improved management or increased labor effort.

Table 5.3 – Productive Spillover effects

	(1) Suffered from black pod in previous year (1=yes)	(2) Total losses to black pod in previous year (Kg)	(3) Cocoa Yield (Kg/Ha)	(4) Cocoa farm size (Ha)	(5) # of times brushed in previous year	(6) % shade cover	(7) # days worked on cocoa farm in previous year
Block Farm	−0.017	12.456***	15.890	−0.032	0.302	−2.508	2.690
	(0.047)	(3.974)	(15.772)	(0.243)	(0.369)	(2.377)	(14.019)
Post	0.067*	6.245	21.547***	−0.206	0.015	−4.578*	−21.831
	(0.034)	(3.706)	(7.441)	(0.183)	(0.247)	(2.604)	(15.765)
Block Farm * Post	−0.121*	−14.490 ***	9.986	−0.052	−0.552	5.205	−5.726
	(0.070)	(4.614)	(10.097)	(0.277)	(0.456)	(3.815)	(16.573)
Mean control group BL	0.75	26.0	72.6	1.68	2.60	52.2	73.9
SD control group BL	0.43	29.8	78.8	2.01	2.11	16.4	80.7
Observations	922	784	700	962	966	940	956
# Clusters	28	28	28	28	28	28	28

OLS regressions. Dummies for matchgroups included. Robust standard errors in parentheses clustered at the village level. Cocoa yield and black pod losses winsorized at the 95% level. Sample includes only cocoa farmers, and in some cases only producing cocoa farmers.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 5.4 we examine the economic effects of the investment. One of the main criticisms of land investments is that by utilizing land it reduces available land for others, reducing farm incomes. In Hofman et al. (2020) the increased labor demand of a plantation drives up the local labor price, which also reduces farm incomes. Our aggregate welfare index sees no differential effects between block farm villages in the later time period. We see a relatively large reduction of 70% in yearly farm earnings, though this is not significant. Initial earnings were higher in the treated group already. It might be that labor shortages only arise when investments are of a much larger scale. As incomes are notoriously unstable (Meyer and Sullivan, 2003) we also examine expenditures. For monthly (food) and yearly (irregular) expenditures we find no effect either, and the coefficients are modest with reductions of 7-18%. This is counter to other papers and theoretical predictions (e.g. by Kleemann and Thiele (2015)) which expect increases in food prices. Another measure of wealth is assets, which we examine through a score of goods and the tropical livestock unit. Again, we find no evidence of reduced wealth, though there is a large but imprecise increase in the tropical livestock unit of 0.05 (that equates to 5 chickens). We also find no effect on the extensive margin of savings. We examine labor prices in Table A.5.2. We see some indication of increased labor prices (of about 1000 leones per day, an 11% increase), but this is not significant although we suffer from low power for this village-level outcome. Overall, we find no effect on economic outcomes. Crucially, the net effect is neither positive nor negative, indicating a potential null effect on welfare. This might be because of the modest scale of the investment, or that larger effects will only arise further in the future.

Table 5.4 – Economic effects

	(1) Livelihoods index	(2) Yearly earnings from farm (IHS)	(3) Monthly (Food) expen- ditures (IHS)	(4) Yearly expen- ditures (IHS)	(5) Total as- sets score (assets + house)	(6) Tropical Livestock Unit	(7) Have savings (1=yes)
Block Farm	0.257**	1.185***	0.146	0.288***	0.643	0.029	−0.053*
	(0.113)	(0.351)	(0.115)	(0.093)	(1.617)	(0.023)	(0.028)
Post	0.251**	0.294	0.216*	0.301***	2.125***	0.043	−0.044
	(0.108)	(0.229)	(0.110)	(0.107)	(0.695)	(0.034)	(0.030)
Block Farm * Post	−0.002	−0.751	−0.072	−0.180	0.293	0.051	0.054
	(0.189)	(0.580)	(0.227)	(0.208)	(1.219)	(0.072)	(0.046)
Mean control group BL	0	4.81	6.18	7.19	34.2	0.080	0.20
SD control group BL	1.00	3.15	0.97	1.13	14.2	0.17	0.40
Observations	2432	2432	2432	2432	2408	2406	2414
# Clusters	31	31	31	31	31	31	31

OLS regressions. Livelihoods index is a composite index of all subsequent variables in this table, normalized and centered on control group at baseline. Dummies for matchgroups included. Robust standard errors in parentheses clustered at the village level. Earnings and expenditures winsorized at the 95% level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.5 Conclusion

In this paper we examine the impact of a foreign-funded agribusiness scheme on the local population in Sierra Leone. We improve on previous papers on this subject by a more robust identification strategy. We have a large sample, data from before the agribusiness scheme got started and select our control group systematically. Furthermore, we examine a diverse set of outcomes on knowledge spillovers and economic effects. We use a Difference-in-Difference design to examine these effects.

We find effects on productive spillovers, with reduced production losses from a pervasive local fungal disease. This is probably caused by knowledge on-farm management spilling over from the agribusiness farms to the local farms. This is consistent with the literature on FDI and theoretical predictions on foreign investments which states that knowledge spillovers are the main driver of improved welfare because of foreign investments (Crespo and Fontoura, 2007; Kleemann and Thiele, 2015). This is the first paper (to our knowledge) that examines these types of productive spillovers at such a detailed level.

In this paper we can only examine short-term (3 years) welfare effects, and prior to the start of agricultural sales from the investment, which is expected to bring in a lot of money for the local economy. We look at the short-term economic effects in several ways, to overcome problems with measuring wealth for subsistence farmers. We examine earnings, expenditures, assets and savings, and find no significant effects. A composite index shows the same result. This is in itself interesting, as the prevalent narrative states that effects are large and far-reaching. We find no evidence for this. Interestingly, we see some weak evidence that the labor price increases, which corroborates the hypothesis made in Hofman et al. (2020), though we suffer from low power for this village-level variable.

Overall, we see evidence for productive spillovers of foreign agricultural investments, a subject that has been rarely studied. This shows a potential positive channel for impacts of foreign investments, whereas most current research has focused on the negative impacts. We see no evidence on improved welfare, though this may be caused by the investment still being in the startup phase. A longer-run analysis will shed more insight on the net effects of the investment.

5.6 Appendix

Table A.5.1 – Match Specification

Variable	Narrow Match	Crude Match
Village Land size (Ha)	Three equally sized bins	Two equally sized bins
% village land forested	Three equally sized bins	Two equally sized bins
Mean slope of land	Two equally sized bins	Two equally sized bins
Distance to local town (Km)	0, 1, 5, 10, 20	0, 5, 10, 20
Distance to closest road (Km)	0, 0.5, 1, 5	0, 2, 5
Number of houses	0, 20, 50, 100, 150	0, 50, 100
Number treated villages matched	6	2

Table describes the bins for the coarsened exact match used. To be matched to a treatment village a subject village needed to be in each of the exact same bins as the treated villages. Bins are described in this table. For example, a treated village at a 15 Km distance to the local town would be matched to all villages between 10-20 Km from the local town (that also matched bins for the other variables). Two villages did not get a suitable number of potential matches in the narrow match, so these were matched using more crude bins, shown in the right-most column.

Table A.5.2 – Labour price

	(1) Mean labour price (1'000 leones)
Block Farm	−0.328
	(1.747)
Post	1.957
	(1.255)
Block Farm * Post	1.133
	(2.471)
Mean control group BL	11.9
SD control group BL	3.60
Observations	62
# Villages	31

OLS regressions. Labour price is the average of seven common agricultural activities spread throughout the agricultural season (Cutting overgrowth on upland/cocoa/swamp farms, clearing overgrowth, weeding, ploughing and harvesting). Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 6

Conservation Impacts of REDD+: Evidence from Sierra Leone

The climate is in the global spotlight generating substantial interest in interventions to protect forests and biodiversity and stabilize the global climate. One prominent approach to reduce deforestation in the Global South whilst ensuring affected communities are not disadvantaged are REDD+ programs. We evaluate the five-year impact of a REDD+ program surrounding Gola Rainforest National Park, a global biodiversity hotspot in Sierra Leone. This park outperforms other protected areas in the region in terms of within-park deforestation but is vulnerable to pressure from its buffer zone. The REDD+ program provides development interventions to communities in this buffer zone to reduce pressure on the park. We apply a difference-in-difference approach using satellite imagery and find that the REDD+ program reduced deforestation by 1 percentage point relative to control communities, translating into a reduction of deforestation of 30%. We use survey data to explore mechanisms underlying this result and find suggestive evidence that communities moved to forest-friendly activities. We find no evidence of large changes in livelihood outcomes, indicating that the REDD+ program did not make communities worse off.

Publication status: Malan, M., Carmenta, R., Gsottbauer, E., Hofman, P., Kontoleon, A., Swinfield, T. and Voors, M., 2020. Conservation Impacts of REDD+: Evidence from Sierra Leone. Working Paper.

6.1 Introduction

Conserving forested areas is crucial for reducing carbon emissions, preserving biodiversity, and securing nature's benefits to people. This global concern is undersigned by worldwide support for the UN Convention on Biological Diversity and the UN Sustainable Development Goals 13 (climate action) and 15 (life on land). The approaches championed by these international conventions focus on the creation of 'protected areas' in which the natural environment must be left mostly undisturbed. The success of protected areas to reduce deforestation has been limited as deforestation is increasing worldwide, including in important protected areas and biodiversity hotspots (Heino et al., 2015; Butchart et al., 2010; DeFries et al., 2005; Hansen et al., 2013). Moreover, such approaches are often criticized for harming communities relying on the forest or residing in its crucial buffer zone. It is for these reasons that Reducing Emissions from Deforestation and forest Degradation (REDD+) programs aim to improve conservation outcomes while simultaneously mitigating the costs to local communities.

There is substantial evidence on the effect of protected areas on: (i) local deforestation (Okumu and Muchapondwa, 2020; Heino et al., 2015; DeFries et al., 2005; Butchart et al., 2010; Geldmann et al., 2019; Jayachandran et al., 2017; Herrera et al., 2019) (ii) livelihoods (Okumu and Muchapondwa, 2020; Jayachandran et al., 2017) and (iii) deforestation in buffer zones (Lui and Coomes, 2016; Heino et al., 2015; DeFries et al., 2005; Herrera et al., 2019). But to maximize the effectiveness of protected areas, all three of these aspects should show improvements. Furthermore, studying all aspects allows for examination of trade-offs between livelihoods and conservation. There are very few papers that study this. Those that do rely on matching techniques to determine a counterfactual or lack a counterfactual at all. In this paper, we aim to contribute to this research gap through the rigorous evaluation of a REDD+ program in Sierra Leone.

This accredited REDD+ program is aimed specifically at improving conservation and livelihood outcomes in the crucial buffer zone of the Gola Rainforest National Park (GRNP), a global biodiversity hotspot. Communities residing in the 4-kilometer buffer zone surrounding the national park received a range of development activities including agricultural extension, marketing support and access to (co-managed) financial services.

Following the evaluation design set up before the start of the program, and a detailed published pre-analysis plan, we evaluate the five-year impact on livelihoods and deforestation in the buffer zone using a robust difference-in-difference approach. Crucially, we provide evidence for the central assumption of parallel trends. We also explore deforestation within the national park and examine the mechanisms that drive the observed impact.

Based on satellite data we find that that deforestation within the park is low, outperforming all other protected areas within Sierra Leone. Furthermore, we find a reduction in deforestation of 30% in the buffer zone for REDD+ communities compared to non-REDD+ communities. This is a large effect given that the intervention, primarily focused on training, was relatively light. We look at survey data to uncover whether the livelihoods of the local population are affected. We find a null effect, indicating that the local population was not worse off, but not better off either. We do not find significant changes in conservation attitudes but do find suggestive evidence that REDD+ communities switched to more forest-friendly activities, demonstrated by a negative effect on labor access and positive effects on income from non-timber forest product collection. This provides a potential explanation of the observed reduction in deforestation.

6.2 Previous work on Deforestation, Livelihoods and Encroachment

Our work contributes to an emerging literature on the causal effects of REDD+ interventions on environmental and social outcomes. While many have argued that scientific evidence on REDD+ implementations is critical to provide insights if the core purpose of REDD+ in terms of reduction of carbon emissions and its impact on social well-being are fulfilled (Wunder et al., 2014; Sills et al., 2017; Wiik et al., 2019), only few impact evaluations have adopted rigorous experimental designs to evaluate how REDD+ affect forest conservation and the well-being of the local population. An overview of these is shown in Table 6.1.

Table 6.1 – Conservation impact evaluations

Method	Country	Period	Program	Cash/non cash	Condit.	Deforest. indicator	Deforest. outcome	% avoided deforest.	Liveli. indicator	Liveli. outcome	Mechanism	Cost	Reference
RCT	Uganda (1 site)	2 years	PES	Cash	Yes	Satellite data: tree cover loss	+ (tree cover loss)	4.90%	Income	No impact	Income & expenditure	Average cost per household: $37.80 per participant	Jayachandran et al. (2017)
RCT	Sierra Leone (1 site)	2 years	Uncond. PES	Cash	No	Satellite data: land cover change. Self-report: land conversion, labor input for logging	- (land clearing)	Note: 3.5% more land clearing in treatment	Not assessed	Not assessed	Income & expenditure. Farm size/additional labor/agricultural production	Average payment per household: $15	Wilebore et al. (2019)
BACI (DID)	Bra Per Cmr Tza Idn Vnm (23 sites)	2 – 9 years	REDD+ initiatives	Incentives + Disincentives	Mixed	Satellite data: tree cover loss	+ (tree cover loss)	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed	Bos et al. (2017)
BACI (DID)	Bra Per Cmr Tza Idn Vnm (17 sites)	2 years	REDD+ initiatives	Incentives + Disincentives	Mixed	Self-reported: forest clearing; tenure security	+/- (forest clearing) +/- (tenure security)	Not assessed	Subjective well-being	+/-	Not assessed	Not assessed	Duchelle et al. (2017)
BACI (DID)	Bra Per Cmr Tza Idn Vnm (22 sites)	2 years	REDD+ initiatives	Incentives + Disincentives	Mixed	Not assessed	Not assessed	Not assessed	Subjective well-being. Income sufficiency.	No impact	Not assessed	Not assessed	Sunderlin et al. (2017)
BACI (DID)	Brazil (1 site)	1 year	REDD+ pilot with a PES-component	Cash	Yes	Satellite data: tree cover loss	+ (tree cover loss)	5.40%	Not assessed	Not assessed	Additional wage labor income, Intensification of livestock, Proportion of cropland/pasture	Average payment over study period per participant: $626	Simonet et al. (2019)
BACI (synt. Matching)	Guyana (nationwide)	5 years	National REDD+	Cash/Non-cash	Yes	Satellite data: tree cover loss	+ (tree cover loss)	0.03%	Not assessed	Not assessed	Not assessed	Cost: $19.53 per averted tCO2	Roopsind et al. (2019)

+ positive impact (i.e. reduced deforestation; increase in social welfare), - negative impact (i.e. increased deforestation; decrease in social welfare)
+/- mixed impact (i.e. increase/decrease of relevant indicator), Self-report means that information was collected via household surveys.
Condit. = Conditional, Deforest. = Deforestation, Liveli. = Livelihoods

Of recent studies that evaluate the causal effects of REDD+ program components, a randomized control trial (RCT) on conditional payments in Uganda by Jayachandran et al. (2017), stands out as the only study where payments were experimentally allocated to treatment and control households. Here, forest-owning households received annual cash transfers, conditional on conserving the forest. Using satellite imagery, they find that payments reduced deforestation over a two-year time period. However, assessed welfare effects were insignificant vis-à-vis control households. Despite the rigorous randomized design, a caveat remains for its external validity. The study was carried out with private forest owners, which is a highly specialized setting as most land tenure arrangements in Africa are based upon communal- or community-oriented land. Wilebore et al. (2019) also use a randomized control trial to study an unconditional payment over a contract period of two years in Sierra Leone. Payments are made to local communities due to communal land rights. They find that unconditional payments increase land clearance. Note that while RCTs are considered the 'gold standard' to evaluate conservation interventions, both aforementioned studies are limited to payments made in cash and effects on short-run changes in deforestation. Here, we focus on the truly long-run impacts of a REDD+ program which provides benefits without a cash component.

There are very few studies that have studied the performance of actually implemented REDD+ schemes as we do. Simonet et al. (2019) evaluate a REDD+ pilot project in Brazil including a Payment for Ecosystem Service component that offers payments conditional on forest conservation. Using a difference-in-difference approach, they find a 50% decrease in deforestation after the first contract year. It should be noted, however, that the study relies upon self-reported survey data to assess changes in deforestation. Similarly, Duchelle et al. (2017) and Bos et al. (2017) evaluate several sites using self-reported forest outcomes (i.e. forest clearing, tenure security). Self-reported measures are potentially biased because respondents may, out of social desirability, underreport forest clearing.

Out of the existing evaluations, only two studies measure deforestation using remote sensing techniques. Roopsind et al. (2019) is the only nationwide REDD+ evaluation in the literature, analyzing deforestation outcomes in Guyana. Using a synthetic matching approach, they find that the program reduced tree cover loss by 35%. Notably, they also find a sharp increase in deforestation after the conditional program has come to an end. Bos et al. (2017) provides a global comparative quasi-experimental analysis of multiple sites and find small reductions in deforestation rates. These studies all evaluate conditional schemes, whereas many conservation programs, as in our case, are unconditional.

Apart from REDD+, we also contribute to a nascent literature on the evaluation of payments for ecosystem service programs for forest conservation (See e.g. Arriagada et al. (2012); Robalino and Pfaff (2013); Alix-Garcia et al. (2015); Börner et al. (2017), the latter a review). The most recent research also examines spillover and leakage effects from conservation programs (Le Velly et al., 2017; Pfaff and Robalino, 2017; Herrera et al., 2019; Lui and Coomes, 2016).

Relative to these papers, we make the following contributions. First, we study the long-term impact of an actual REDD+ scheme in Sierra Leone and thus provide one of the first studies that goes beyond short-term measurements after 1 or 2 contract years. Second, we discuss the effectiveness of an unconditional program which is potentially the more common contract set-up in real-life REDD+ implementations. Third, unlike existing impact evaluations on actual REDD+ schemes, we can use remote sensing data to evaluate deforestation outcomes and connect survey measures of our participants to welfare consequences and help fill that knowledge gap in the literature.

6.3 Deforestation and REDD+ in Sierra Leone

The GRNP is a 71,000-hectare remnant of the upper Guinean moist tropical forest, on the border with Liberia. It is part of the Upper Guinean Forests, classified as a global biodiversity hotspot by Conservation International, a global conservation watchdog. The GRNP was officially established in 2011 and the managing company, Gola Rainforest Conservation (GRC) has been engaged in GRNP conservation efforts for over twenty years. Satellite images and field observations suggest that forest cover within the uninhabited GRNP has largely remained intact throughout this period (Figure 6.1). To protect the park, GRC imposes restrictions on logging, hunting and mining within the park and employs forest guards to enforce these rules. From this figure, it also becomes apparent that the park is surrounded by land that has been substantially deforested. This so-called buffer zone should provide two crucial functions: it protects the park from encroachment due to population pressures but also serves as a corridor between the park sections. In 2014, the national park received REDD+ accreditation aimed at a.o. safeguarding these buffer zone functions. GRC has since then been implementing the REDD+ program.

As part of the REDD+ program, GRC offers several compensation programs to local communities to compensate them for direct losses of income from land usage restrictions. Part of the benefits are given to communities throughout the seven Chiefdoms

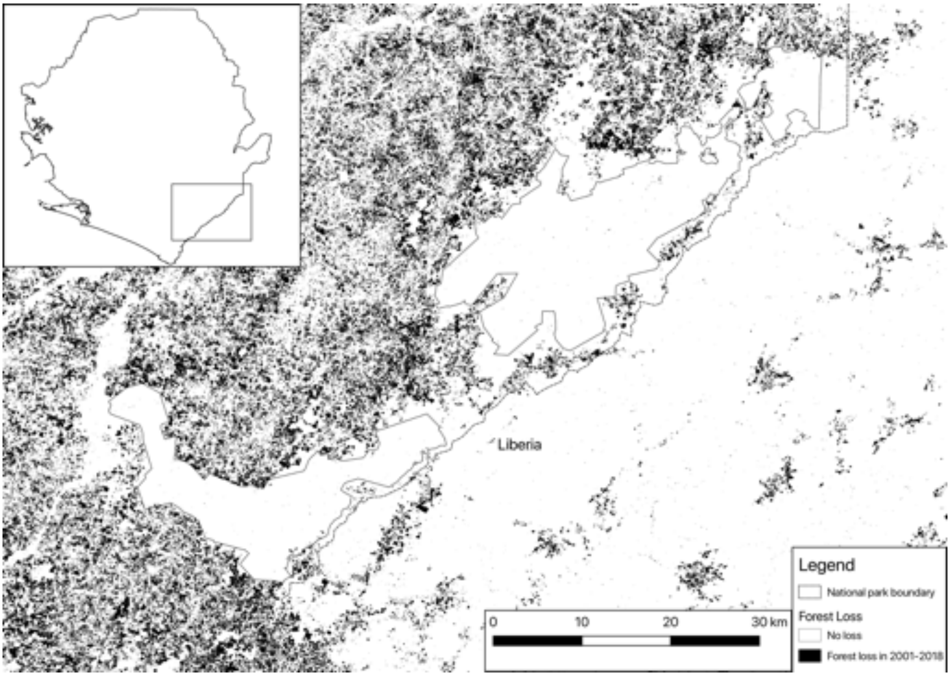


Figure 6.1 – Yearly forest loss in the Gola Rainforest National Park area in Sierra Leone

This figure shows for each pixel whether any deforestation has taken place from 2001 until 2018. Source: Hansen 2013/UMD/Google/USGS/NASA.

in which the GRNP lies, including educational scholarships, surface rents to landowners and a Chiefdom Development Fund. In addition to these more general benefits, several activities are directed specifically to communities located in the 4km buffer zone. These REDD+ activities focus mostly on reducing extensive agriculture (i.e. upland rice farming) and moving towards forest-friendly crops (e.g. cocoa), thereby aiming to reduce pressure on the buffer zone, which in turn reduces pressure on the GRNP. The three main REDD+ activities exclusively in the buffer zone are:

Agricultural programs: Agricultural programs consist mainly of trainings on crop production. Generally, a demonstration plot is established and farmers are invited to observe and learn new methods to improve yields. This is done specifically for wetland rice, groundnuts and several vegetables. Crucially, the programs do not include upland rice, the most commonly produced crop, as it requires slash-and-burn agriculture and large amounts of land. In fact, GRC actively discourages upland rice farming.

Cocoa programs: Within the cocoa programs, GRC provides trainings on production, farm management, and post-harvest processing. Cocoa in Sierra Leone is considered a 'forest-friendly' crop: cocoa farms are often created in secondary forests with very minimal land clearing. The shade provided by trees (which are rarely cut down) reduces the need for extensive weeding. GRC's activities are mainly run through farmer field schools. In addition, GRC established farmer associations of which community members of REDD+ communities can become a member if they own a cocoa plantation. These farmer associations are equipped with buying stations and trained buying officers. These buying officers are responsible for sourcing the cocoa from the REDD+ communities at a somewhat higher price. This price is based on the market price in the regional cocoa hub minus a transportation fee (the final price for the farmer is typically higher than the net local price). The cocoa is used to produce high-quality single-origin niche chocolate which is sold at a premium. Some of this premium is returned to the cocoa farmers. Farmers are still free to sell cocoa to any other trader.

Savings and Lending Associations: GRC also established Village Savings and Lending Associations (VSLA) as part of the REDD+ program. The aim is to improve financial access and facilitate investment, thereby increasing resilience. Participation is voluntary and participants can either save money or take out a loan from the saved money. The size of the loan depends on how much was contributed. The VSLA is run by a trained committee that decides on the interest rates for saving and lending and on membership. Besides, the VSLA has a separate fund for emergency loans. Members also receive business training and financial literacy training through the VSLA. GRC's role has been to establish the VSLAs and provide trainings on their functioning. They are currently only involved in monitoring and providing support when necessary.

As part of the REDD+ accreditation process, an evaluation strategy was set up before the start of the program. Following this strategy, we evaluate the impact of the above-described interventions on livelihood and conservation outcomes. Also, we pre-specified all outcomes and the analysis strategy in a Pre-Analysis Plan (EGAP id: 20190711AA) before analysis. As the intervention was not randomly assigned to communities, we employ a difference-in-difference analysis in which our counterfactual group consists of the communities that lie within the GRNP chiefdoms, but outside of the buffer zone. Within the buffer zone, the assignment of the three interventions was also not random, and often communities received multiple interventions, we are therefore unable to test the impact of each intervention separately but do so for the REDD+ program as a whole. Moreover, we do not examine the REDD+ activities that were targeted to communities beyond the buffer zone (educational scholarships, surface rents, and the Chiefdom Development Fund) for two reasons. Firstly, the main goal of the project was to improve outcomes in the buffer zone specifically and secondly, we lack the data to construct a convincing counterfactual for the area beyond the buffer zone. Refer to Table A.6.1 for the percentages of village receiving the different intervention components.

6.4 Methods

The analyses in this paper rely on three main sources of data: Satellite data using a publicly available dataset by Hansen et al. (2013), border definitions (polygons) of all protected areas in Sierra Leone and survey data collected over 3 rounds in villages surrounding the GRNP (2010, 2014 and 2019). We discuss these sources of data in turn, before moving to the estimation strategy.

6.4.1 Satellite Deforestation data

The dataset by Hansen et al. (2013) gives worldwide, yearly data on forest loss over the 2001-2018 period. The dataset is very high-resolution with pixel size at 30x30m. This allows us to get detailed information and allows us to also recognize small-scale deforestation (as is likely with slash-and-burn agriculture). Forest is defined as an area with >50% vegetation taller than 5 meters. Forest loss is defined as a change from a forested to a non-forested state. We disaggregate forest loss to year and village level. To assign forest loss to specific villages we use the approach and dataset by Wilebore and Coomes (2016). The approach works as follows: first simple voronoi polygons are drawn for all villages in the surrounding areas (454 villages in total). Then, some of these polygons (189 in total) are adjusted in size based on population size (Results are similar when using unweighted polygons, see Table A.6.7). Larger villages are thus assigned a larger polygon (see Figure A.6.3 for the polygon map). Village land is communally owned by extended families in Sierra Leone. Larger villages contain more families and are likely to have more land as well. The data on locations and population sizes are based on a survey of 189 villages in 2010-11 by the researchers (see Section 6.4.3 for more details). Finally, we count the number of pixels lost in a village's polygon in a given year and calculate the percentage of forest lost by dividing the area deforested by the size of the polygon. By using the percentage we can compare villages with different-sized landholdings.

6.4.2 Protected areas definition

We place the (lack of) deforestation in the GRNP in context by examining other protected areas in Sierra Leone. This is based on a map provided by the Sierra Leonean Ministry of Agriculture, which we use to infer the exact borders. We examine all existing National Parks, forest reserves and game sanctuaries with a legal status protecting them. The satellite deforestation data is used to examine forest loss over the entire period for each protected area separately. We also examine forest loss in 4km buffer zones which provide important corridors for endangered species and prevent encroachment on the protected area. We use a 4km distance from the border to define this buffer zone, to be consistent with the GRNP's buffer zone. We only consider buffer areas that fall within the national borders of Sierra Leone. Deforestation results for these national parks individually are shown in Figure A.6.1. We also use national park boundary data from the World Database on Protected Areas to extend our analysis to neighboring Guinea and Liberia, shown in Figure A.6.2 (UNEP-WCMC and IUCN, 2020).

6.4.3 Survey data

We use data collected in Sierra Leone during three waves. During March/April 2010, Wageningen and Cambridge University researchers collaborated with GRC to implement a baseline survey in villages in the seven chiefdoms surrounding the GRNP. GRC selected 200 villages that were closest to the National Park and most likely to have community forests with high biodiversity value. From this list 11 did not exist (anymore) and the survey was implemented in 189 communities. This survey is also the source of village locations and sizes, which are used in the voronoi polygon definition by Wilebore and Coomes (2016). 15 households were randomly sampled and interviewed regarding demographics, economic outcomes, hunting and gathering behavior, and attitudes towards conservation. We implemented a second survey in April 2014, just before the start of REDD+ activities. From the villages included in the 2010 survey wave, we randomly selected 30 Forest Edge Communities (FEC), i.e. those eligible for REDD+ benefits. These communities all lie within a 4 km buffer zone around the National Park. We also selected 30 non-FECs which were randomly selected from villages 4-25 km from the National Park boundary. The sampling was stratified by regional quadrants to ensure the representation of villages between the GRNP boundary and the border with Liberia.

One of the FEC villages was removed from the sample as it no longer existed, bringing our full sample down to 59. The same households as in 2010 were interviewed. During this survey wave, in total 841 households were surveyed across the 59 villages, with an average of 14 households per village (some villages had fewer than 15 households). For the follow-up survey during April 2019 we revisited each household included in the 2014 survey. If the head of household was not available we selected a representative of the household. Our attrition rate for the 2019 sample is 19% (15% in non-REDD+ villages and 23% in REDD+ villages).

6.4.4 Survey Outcomes

We assess two main survey outcomes: a family of outcomes relating to livelihoods and a family relating to conservation attitudes. By grouping our variables into families we reduce the number of statistical tests necessary. We use the approach by Kling et al. (2007) to combine variables with different units into families. This works by first normalizing all variables, and then taking the row mean of these z-scores. If some variables are missing for observations these are imputed at the own-group mean (by survey round and treatment status).

The first family is related to the livelihoods of farmers that are likely affected through the REDD+ program. It consists of data on income, expenditures, resilience, productive loans and assets. Income is the sum of a very broad range of income categories which includes almost all sources of income and increases our precision. We ask this question over the previous year. We also look at two forms of expenditures as a more robust estimate of incomes. We ask about expenditures in the previous month on a set of common consumption items. We also ask about yearly expenditures that are more irregular. Resilience is a dummy on whether individuals were able to cope with an emergency in the previous year. Productive loans are the sum of loans in the previous year for productive activities. Assets is the sum of a common set of assets owned, like tables, beds and housing materials. Outcomes that are expressed in monetary terms (income, expenditures and productive loans) are transformed using the Inverse Hyperbolic sine which is similar to taking the natural logarithm and reduces the variance of the outcome.

The second family is about conservation attitudes, which consists of stated attitudes, knowledge of conservation rules, sustainable farming and perceptions of human-wildlife conflict. Stated attitudes are responses on a five-point likert scale to four questions related to the GRNP and conservation in general. Knowledge is assessed by asking five questions about what is allowed and not allowed in the national park (on mining, gathering, fishing, logging and hunting). Sustainable farming is the number of sustainable farming practices used, for example on lower land use. Finally we ask how big of a problem human-wildlife conflict is (on a 0-3 scale). Increased human-wildlife conflict is often associated with the creation of the national park, which might have increased animal populations.

We also explore several mechanisms, mainly related to labor changes. Labor is one of the main seasonal constraints for agricultural production in Sierra Leone, with over 65% of households reporting labor shortages in the agricultural season in a nation-wide survey (MAFFS, 2011). To assess labor shortages we ask respondents how much of a problem it is to get labor (scale 0-3) for the three main types of farms and calculate the average value. We also assess income from farm wages in the previous year, and finally also look at yearly income from NTFPs (Non-Timber Forest Products). NTFPs are an important alternative form of income associated with the creation of the national park, as these are explicitly allowed to be collected and will be more plentiful if the park is well-preserved.

6.4.5 Empirical strategy

To estimate the average treatment effect of REDD+ we estimate a standard Difference-in-Difference model:

$$\boldsymbol{Y}_{ijt} = \beta_0 + \beta_1 REDD_j + \beta_2 post_t + \beta_3 post_t * REDD_j + \varepsilon_{ijt} \tag{6.1}$$

Where $\boldsymbol{Y}_{ijt}$ refers to our normalized set of outcomes (as a family or individually), $REDD_j$ is a dummy for Forest Edge Communities (ie. REDD+ eligible communities), $post_t$ is a dummy referring to the second survey wave (2019). β_3 is our coefficient of interest. i indexes the household level, j indexes the village level and t the survey wave. We cluster standard errors at the village level. Only households where we have panel data (e.g. they were interviewed in both rounds) are included in this regression. This estimator gets us an unbiased estimate of the treatment effect if we can assume that without the project, the villages would have trended similarly (parallel trends assumption). We explore this in the next section.

6.4.6 Parallel trends

Our main identifying assumption is that of parallel trends. That is, that the REDD+ villages would have *trended* similarly, had the REDD+ project not been implemented. This is fundamentally untestable, but if trends are similar before the implementation of the project, they would likely have trended similarly had there been no project. For conservation behavior (deforestation), we have many rounds of data available before the start of the REDD+ activities, and these are shown in Figure 6.3. The break in trend lines in this figure shows the launch of a new, more accurate satellite (Landsat 8). As can be seen, trends (and levels) were very similar before the break, which gives confidence that this would continue after the implementation of the REDD+ activities. There is one observation with the new satellite, but before the activities, which still show very similar levels between the two groups of villages. This reassures us that the parallel trends assumption likely holds.

For our survey outcomes, we make use of the unique opportunity provided by having access to two rounds of pre-REDD+ data (the 2010 and 2014 rounds of data), to investigate parallel trends in our data before the commencement of REDD+ activities. We run the Difference-in-Difference model above for the 2010-2014 data on our main outcomes and mechanisms. This is shown in Table A.6.8. In no case is the post*REDD+ coefficient significant: we find no different trends between the two groups. This reassures us that the parallel trends assumption is likely to hold.

6.5 Results

Since 2000, Sierra Leone lost 25% of its tree cover, primarily for (slash-and-burn) agriculture (Curtis et al., 2018). Figure 6.2 shows forest loss in the GRNP (Panel A), as well as its buffer zone (Panel B), compared to other parks in Sierra Leone. In panel A GRNP stands out with a much lower rate of forest loss compared to other protected areas. Panel B shows forest loss in the 4-kilometer buffer zone, where the GRNP buffer zone is more in line with the national forest loss trends. We conducted the same analysis using alternative data on protected areas for Sierra Leone and extended the analysis to the neighboring countries Guinea and Liberia and find very similar trends (see Figure A.6.2).

We examine the effects of the REDD+ program on deforestation in the buffer zone by looking at all 454 communities in the seven Chiefdoms, of which 117 lie in the 4km buffer zone and receive the REDD+ program. We examine conservation behavior (forest

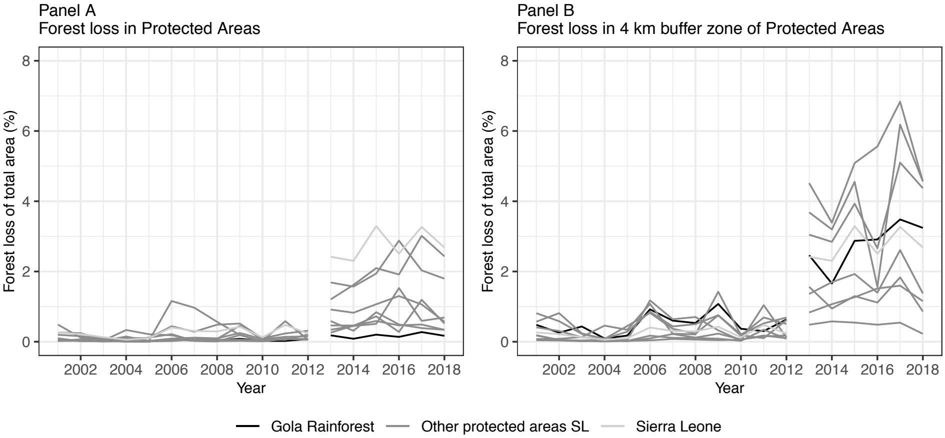


Figure 6.2 – Total forest loss in Gola Rainforest National Park, other protected areas in Sierra Leone, and Sierra Leone as a whole

The left panel shows total forest loss from 2001 to 2018 in protected areas of Sierra Leone. The right panel shows total forest loss from 2001 to 2018 in the 4km buffer zones of these parks. The break in the lines in 2013 denotes the launch of a more precise satellite (Landsat 8). Source of data: Hansen 2013/UMD/Google/USGS/NASA.

loss) for the REDD+ and non-REDD+ communities. Based on community location and population size, we draw Voronoi polygons around these communities to assign forest loss, disaggregated by year, to each community. Mean deforestation rates are shown in Figure 6.3. This shows the trend in percentage forest loss in the REDD+ communities (black line) and non-REDD+ communities (grey line), with shaded areas representing 95% confidence intervals. The vertical line indicates the start of the REDD+ activities. While before the start of the REDD+ program the two groups trended very similarly, after 2014 percentage forest loss is significantly and substantially higher in non-REDD+ communities. Deforestation rates are up to two percentage points higher in some years, a nearly 40% increase. In Table 6.2 we use a difference-in-difference analysis to test this result on deforestation with a regression, shown in column 1. Pre-intervention deforestation rates are low at 0.7% annually (constant), with no differences in REDD+ and non-REDD+ areas. In the later time period this increases strongly, by 3.3% (post coefficient). This increase can largely be attributed to higher precision in the deforestation dataset, caused by the inclusion of additional satellite data (Landsat 8). However, for REDD+ communities this rate is one percentage point lower, which is precisely estimated. This amounts to a 30% reduction in annual deforestation rates, a substantial slowing down of deforestation in this crucial buffer zone.

Table 6.2 – 5-year REDD+ impacts

	Forest Loss	Livelihoods family	Conservation family
Post*REDD+	-1.032^{***}	0.022	-0.257
	(0.114)	(0.132)	(0.210)
Post	3.314^{***}	0.222^{**}	-0.689^{***}
	(0.066)	(0.103)	(0.117)
REDD+	-0.052	-0.144	0.126
	(0.033)	(0.118)	(0.129)
Constant	0.740^{***}	0.000	-0.000
	(0.017)	(0.089)	(0.077)
Years	18	2	2
Villages	454	59	59
Num. obs.	8172	1320	1320

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regressions for forest loss (satellite data) and livelihood and conservation norms families (survey data). Forest loss is the percentage loss of forest (primary and secondary). The livelihood family outcome is a summary index (average of z-scores) of an income index, an assets index, a durable loan size measure, and a measure for resilience. The conservation family outcome is a summary index (average of z-scores) of a conservation attitudes index, an awareness of conservation norms index, the number of sustainable farming practices practiced, and an index for human wildlife conflict perception. Family outcomes are standardized and centered on control group at baseline. For survey outcomes, standard errors are clustered at the village level. Robust standard errors in parentheses.

From the exploratory analysis we show that, compared to other parks, the GRNP has been largely kept intact. Note, that we do not have a counterfactual for within-park deforestation and only compare trends. We do however find that the REDD+ program resulted in substantially lower deforestation rates in buffer zone communities compared to areas beyond the buffer zone. To measure how well-being in the REDD+ communities was impacted by the program, and to gain insight into the mechanisms behind the sizable effect on deforestation, we exploit collected household survey data and use a similar difference-in-difference approach.

We examine livelihoods and mechanisms using survey data on a subset of 59 communities (30 non-REDD+, 29 REDD+), where 15 households were randomly selected and interviewed in 2014 (before REDD+ activities) and in 2019. Results are shown in Table 6.2. These outcomes are standardized with respect to the control group at baseline and can thus be interpreted as standard deviation changes. We see that the livelihoods index (column 2) was 0.14 SD lower in REDD+ communities before the project, though this is

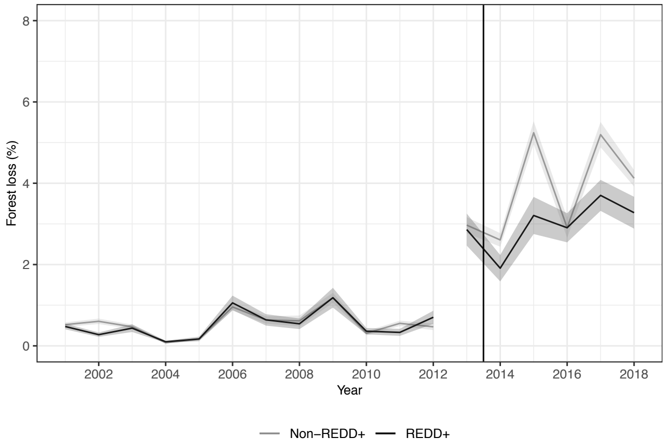


Figure 6.3 – Total forest loss in REDD+ and non-REDD+ villages

This graph shows total forest loss from 2001 to 2018 in REDD+ versus non-REDD+ villages. The village polygons are estimated using population weighted Voronoi estimations. The shaded areas in the graph denote confidence intervals and the vertical black line indicates the start of REDD+. The break in the lines in 2013 denotes the launch of a more precise satellite (Landsat 8). Source of data: Hansen 2013/UMD/Google/USGS/NASA.

not significant. We see an increase in the livelihoods index of 0.22 SD over the five-years of the program, which is a substantial and significant improvement. However, there is no difference between REDD and non-REDD+ communities, the coefficient for that difference is low at 0.02 SD. Therefore, households were not worse off in terms of economic well-being due to the REDD+ project. No shift in livelihoods indicates that the deforestation effect is not linked to any REDD+ induced change in livelihoods (e.g. higher incomes reducing the necessity for cutting down trees for cultivation, firewood, timber, etc). The improvements in conservation behavior may be caused by improved conservation attitudes (column 3). Conservation attitudes were 0.13 SD higher before in the REDD+ villages, though not significant. Attitudes lowered substantially by 0.7SD, though this is not different between REDD+ and non-REDD+ communities. This is therefore unlikely to contribute to the effect on deforestation we find. For the full results, as specified and published in the pre-analysis plan, refer to Tables A.6.3, A.6.4, A.6.5, A.6.6.

Earlier studies have pointed to changes in the labor market in Sierra Leone potentially contributing to increased deforestation, i.e. increased labor availability may increase forest clearing, as the process is labor-intensive (Wilebore et al., 2019; MAFFS, 2011). We explore several possible mechanisms in Table 6.3, and find some suggestive evidence. We first look at an index of labor availability for the three main types of farms (upland, swampland and plantation). In the later time period labor access increases substantially, by about 0.37 SD compared to the baseline. However, in REDD+ communities there is a sharp reduction in access to labor, of 0.55 SD. Secondly, we see that incomes from farm wages are substantially higher in REDD+ communities in the later time period, by 0.20 SD. We hypothesize that GRC's activities increased opportunity costs of labor by providing alternative income possibilities. Many farmers choose then to pursue these alternative income possibilities, leaving fewer laborers available for the local labor market (and thus reducing labor access). This leaves fewer laborers for conventional, labor-intensive slash-and-burn agriculture, which is associated with deforestation. This lower labor availability and higher opportunity cost increase the local labor price, which increases income from working on other people's farms (as we find evidence for). One possible alternative source is income from Non-Timber Forest Products (NTFP). These are collected in forested areas (including the GRNP). The collection is encouraged by the GRC, as collection is non-invasive and increases incentives for protecting the national park. We see a substantial increase of 0.34 SD in REDD+ communities, though this still represents only a minor proportion of total income in REDD+ communities (2.4%). Another option is that farmers switched to more forest-friendly crops, such as cocoa. We see a substantial increase in cocoa harvest size (0.2SD) in REDD+ communities, though this is measured with substantial noise and is therefore not significant.

6.6 Discussion

We examine a REDD+ program surrounding a national park in Sierra Leone. Protection of the park has been largely successful, evidenced by extremely low deforestation rates within the park borders, based on satellite imagery. Deforestation is also lower than all other protected areas within Sierra Leone for the same time period. We also examine deforestation in 4km buffer zones surrounding these parks and find that this REDD+ program performs similarly to the country-wide trend. However, when we examine the buffer zone in more detail and compare it to a control group we find substantial improvements, with 30% lower deforestation rates yearly compared to the area outside the buffer zone. This shows us that a relatively light program can have substantial beneficial effects on this environmentally important buffer zone.

However, REDD+ programs should not only reduce deforestation but also improve local livelihoods. We implement a survey and find no evidence of improved livelihoods, nor of

Table 6.3 – Mechanisms of REDD+ impacts

	Labor access index	Income farm wages	Income NTFP	Cocoa harvest
Post*REDD+	−0.545**	0.199*	0.343**	0.196
	(0.257)	(0.106)	(0.153)	(0.129)
Post	0.365**	0.037	0.021	−0.514***
	(0.160)	(0.083)	(0.106)	(0.103)
REDD+	0.120	−0.014	−0.152	−0.123
	(0.133)	(0.096)	(0.101)	(0.126)
Constant	−0.000	0.000	0.000	−0.000
	(0.091)	(0.063)	(0.087)	(0.102)
Years	2	2	2	2
Villages	59	59	59	59
Num. obs.	1150	1228	1320	1320

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regressions for mechanisms. Independent variables are standardized and centered on control group at baseline. Labor access index is an index of three farm labor access variables (upland rice, wetland rice, and plantation) indicating to what extent there is access to labor. Income farm wages is a continuous variable (IHS transformed) measuring the yearly household income from farm wages. Income NTFP is a continuous variable (IHS transformed) measuring the yearly income from Non-Timber Forest Products collection. Robust standard errors in parentheses clustered at the village level.

worsened livelihoods, indicating that the program was at least partially successful. Using this survey data we also examine through what mechanism deforestation was reduced in the buffer zone. We find no evidence that it is caused by improved conservation attitudes, but hypothesize that the REDD+ program affected the opportunity cost of labor, which increased the local labor price through alternative income possibilities. Some of these possibilities are sales of Non-Timber Forest Products and (forest-friendly) cocoa farming.

6.7 Appendix

Table A.6.1 – Interventions in sample of REDD+ villages

Intervention	**# REDD+ vil. with intervention**	**% of total sample**
Agricultural intervention	20	69%
Cocoa intervention	24	83%
Village savings and loans associations	18	62%
REDD+ villages in sample	29	

We only have data on which interventions were implemented in this randomly selected sample of REDD+ villages.

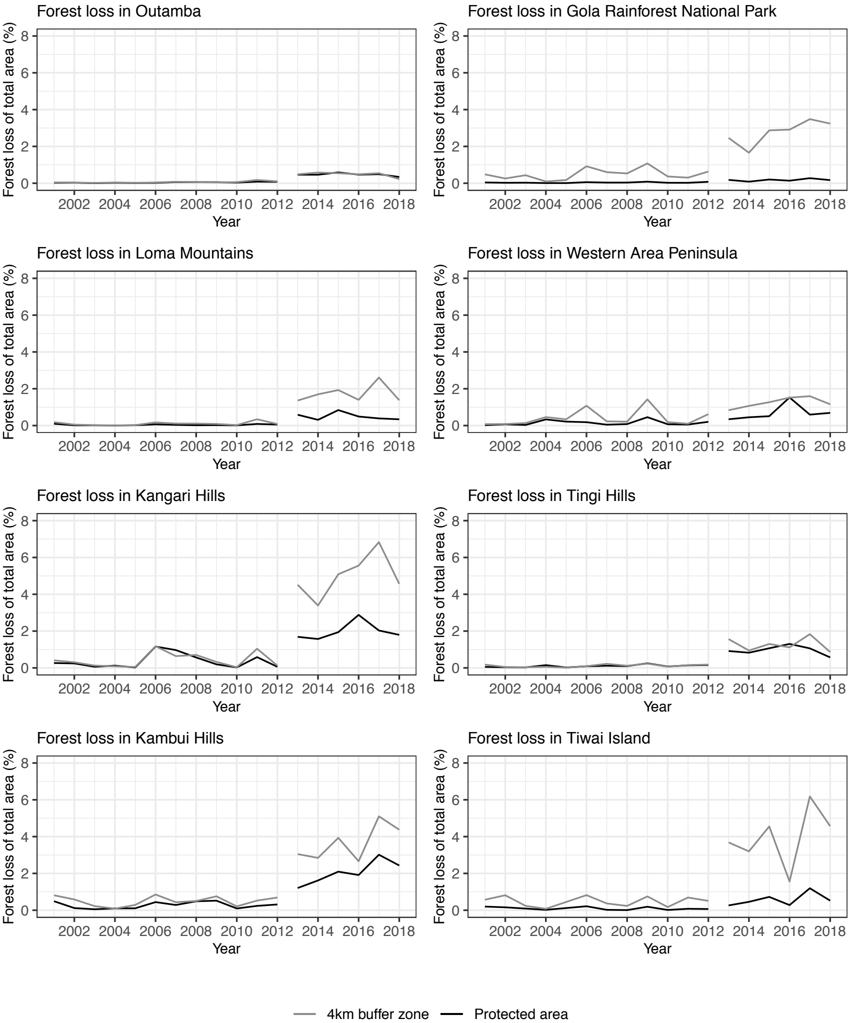


Figure A.6.1 – Yearly forest loss in the Protected areas in Sierra Leone

This graph shows total forest loss from 2001 to 2018 in protected areas of Sierra Leone and their 4 km buffer zones. The break in the lines in 2013 denotes the launch of a more precise satellite (Landsat 8). Protected area definitions come from the Sierra Leonean government. Source of data: Hansen 2013/UMD/Google/USGS/NASA.

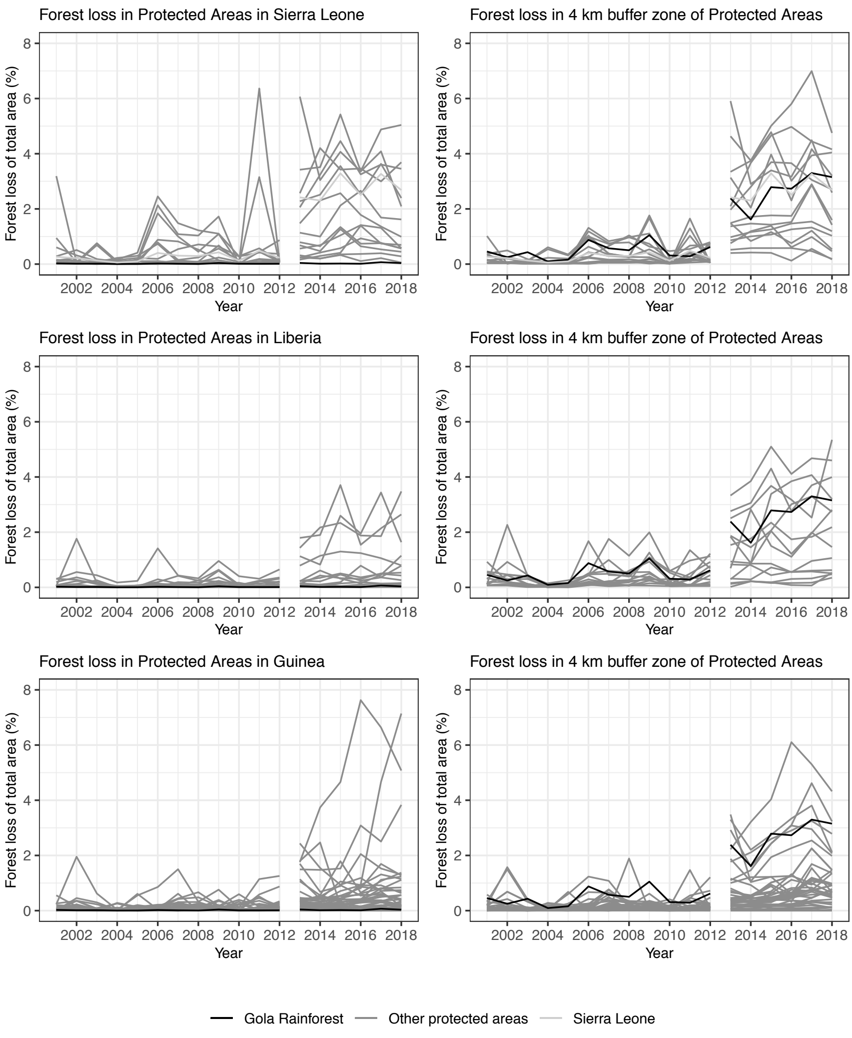


Figure A.6.2 – Yearly forest loss in the Protected areas in Sierra Leone, Liberia, and Guinea

This graph shows forest loss from 2001 to 2018 in protected areas of Sierra Leone, Liberia, and Guinea and their 4 km buffer zones. The break in the lines in 2013 denotes the launch of a more precise satellite (Landsat 8). Protected area definitions come from the Worldwide Database on Protected Areas. We exclude all polygons below a certain size (10.000 pixels) for the readability of the graph and because we are unsure of the reliability of these data. Source of data: Hansen 2013/UMD/Google/USGS/NASA/WDPA.

Table A.6.2 – Difference in means in outcomes in 2019

	non-REDD+			REDD+			
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Difference**
Total income (Leones, IHS)	364	6.702	2.477	296	6.487	2.535	-0.215
Monthly consumption expenditure (Leones, IHS)	364	5.846	1.552	296	6.135	0.838	0.289
Yearly irregular expenditure (Leones, IHS)	364	7.030	1.945	296	7.366	1.045	0.335
Durable loan size, (Leones, IHS)	364	0.257	1.415	296	0.338	1.491	0.081
Resilience (=1)	234	0.979	0.145	197	0.990	0.100	0.011
Yearly income farm wages (Leones, IHS)	334	1.385	2.356	282	1.786	2.567	0.401
Yearly income from NTFPs (Leones, IHS)	364	0.671	1.738	296	0.986	1.993	0.315*
Labor access index (0-3)	323	1.519	0.965	284	1.134	0.933	-0.385**
Conservation attitudes (4-20)	330	14.782	3.721	281	14.260	3.568	-0.522
Awareness of conservation norms (0-5)	301	3.150	0.375	277	3.350	0.493	0.201***
Sustainable farming practices (0-4)	354	0.370	0.783	294	0.565	0.879	0.195*
Human wildlife conflict (0-3)	354	-2.249	1.007	292	-2.517	0.932	-0.269**

N is the number of observations, SD is the standard deviation. Difference gives the difference in means. IHS means the variable is transformed using an inverse hyperbolic sine function. P-values are calculated for a clustered difference in means t-test where
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6.3 – Difference-in-difference: Conservation behavior

	Forest loss	Primary forest loss	Secondary forest loss
Post*REDD+	-1.032^{***}	0.058	-0.577^{**}
	(0.114)	(0.039)	(0.230)
Post	3.314^{***}	0.050^{***}	0.455^{***}
	(0.066)	(0.014)	(0.110)
REDD+	-0.052	0.212^{***}	-0.157
	(0.033)	(0.031)	(0.211)
Constant	0.740^{***}	0.082^{***}	2.527^{***}
	(0.017)	(0.011)	(0.096)
Years	18	6	6
Village polygons	454	434	434
Num. obs.	8172	2604	2604

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regressions for forest loss (satellite data). Forest loss is the percentage loss of forest (primary and secondary). Primary forest loss is loss of old growth forest and secondary forest loss measures conversion of fallow to production agriculture. Both are classified through extensive ground measurements, which were done in 2013. The number of observations is therefore lower for these two outcomes, as data ranges from 2013-2018. Robust standard errors in parentheses.

Table A.6.4 – Difference-in-difference analysis: Livelihoods

	Livelihoods family	Income	Assets	Durable loan	Resilience
Post*REDD+	0.022	0.017	−0.039	0.176	−0.080
	(0.132)	(0.143)	(0.094)	(0.107)	(0.070)
Post	0.222^{**}	−0.020	-0.120^{*}	−0.029	0.301^{***}
	(0.103)	(0.111)	(0.068)	(0.079)	(0.062)
REDD+	−0.144	−0.090	-0.275^{**}	−0.124	0.087
	(0.118)	(0.122)	(0.104)	(0.078)	(0.071)
Constant	0.000	0.000	0.000	0.000	0.681^{***}
	(0.089)	(0.085)	(0.079)	(0.068)	(0.063)
N clusters	59	59	59	59	58
N panel	660	660	660	660	416
Num. obs.	1320	1320	1320	1320	832

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regression for livelihood outcomes. The livelihood family outcome is a summary index (average of z-scores) of an income index, an assets index, a durable loan size measure, and a measure for resilience. The income index is a summary index (average of z-scores) of total household income, monthly consumption expenditure and yearly durable expenditure. Assets is the sum of all assets owned. Durable loan size is the amount borrowed for durable investments. Resilience is a conditional dummy (on whether the household suffered from an emergency) of whether households were able to deal with an emergency. All independent variables are standardized and centered on control group at baseline. Robust standard errors in parentheses clustered at the village level.

Table A.6.5 – Difference-in-difference analysis: Conservation norms

	Conservation family	Attitudes	Knowledge	Sustainable farming	HWC
Post*REDD+	−0.257	−0.300	0.287	0.113	−0.040
	(0.210)	(0.206)	(0.250)	(0.157)	(0.135)
Post	-0.689^{***}	-0.822^{***}	0.210	−0.012	0.124^{*}
	(0.117)	(0.113)	(0.170)	(0.119)	(0.067)
REDD+	0.126	0.132	0.427^{***}	0.102	-0.204^{**}
	(0.129)	(0.132)	(0.154)	(0.112)	(0.096)
Constant	−0.000	−0.000	−0.000	0.000	0.000
	(0.077)	(0.091)	(0.087)	(0.090)	(0.074)
N clusters	59	59	59	59	59
N panel	660	597	518	647	635
Num. obs.	1320	1194	1036	1294	1270

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regression for conservation norms outcomes. The conservation family outcome is a summary index (average of z-scores) of a conservation attitudes index, an awareness of conservation norms index, the number of sustainable farming practices practiced, and an index for human wildlife conflict perception (HWC). Conservation attitudes is an index of agreement with pro-conservation statements. Awareness of conservation norms is an index of knowledge on rules regarding conservation. Sustainable farming practices measures the number of practices used by a household. HWC measures how big of a problem crop-raiding is. All independent variables are standardized and centered on control group at baseline. Robust standard errors in parentheses clustered at the village level.

Table A.6.6 – Difference-in-difference analysis: Secondary outcomes

	Farm Income	Cocoa Income	Off Farm Inc	Upland Size	Wetland Size	Plantation Size	Health
Post*REDD+	0.220	0.078	0.099	0.014	−0.118	−0.006	−0.055
	(0.168)	(0.152)	(0.176)	(0.267)	(0.234)	(0.172)	(0.131)
Post	0.326***	0.278*	−0.115	0.178	0.333**	−0.064	−0.630***
	(0.109)	(0.135)	(0.148)	(0.169)	(0.156)	(0.152)	(0.074)
REDD+	−0.276**	−0.233*	−0.126	0.176	0.036	−0.017	−0.060
	(0.135)	(0.127)	(0.127)	(0.123)	(0.121)	(0.096)	(0.117)
Constant	0.000	0.000	−0.000	−0.000	0.000	−0.000	0.000
	(0.092)	(0.093)	(0.108)	(0.069)	(0.100)	(0.077)	(0.068)
N clusters	59	59	59	59	59	59	59
N panel	660	660	660	607	600	616	660
Num. obs.	1320	1320	1320	1214	1200	1232	1320

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Difference-in-difference analysis using OLS regression for secondary outcomes. Farm income is the total income from crop sales. Cocoa income is the total income from cocoa sales. Cocoa harvests is the total cocoa production. Off farm income is the total income from off-farm activities. Upland farm size is the total size of the upland (rice) farm. Wetland size is the total size of the wetland (rice) farm. Plantation size is the total size of the plantation area. Health is the number of household members with malaria and/or blood in stool and/or diarrhea in the previous month. All independent variables are standardized and centered on control group at baseline. Robust standard errors in parentheses clustered at the village level.

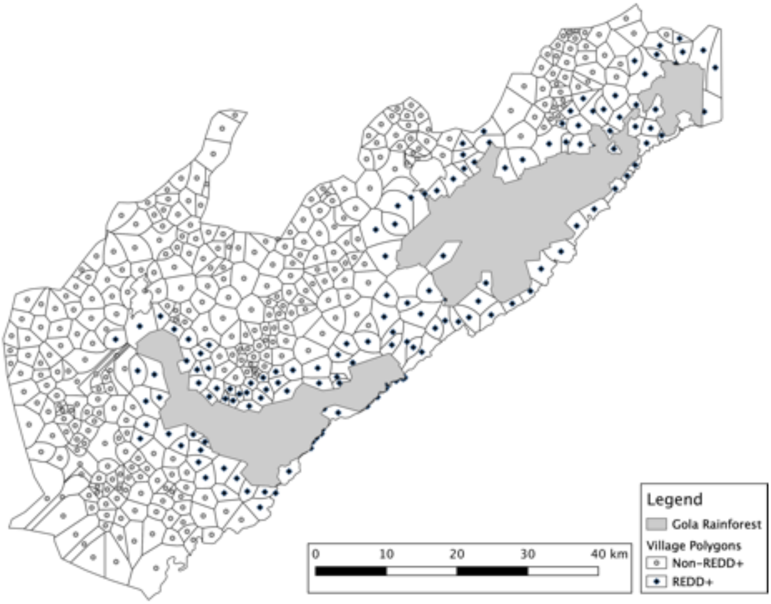


Figure A.6.3 – Village polygons for satellite analysis

This figure shows the village polygons used for the deforestation analysis. Polygons are Voronoi Polygons, weighted based on village population size. If population size was not available, the polygon was not weighted. REDD+ villages are defined as villages that were eligible for the REDD+ program. Non-REDD+ villages are villages that were not eligible for the REDD+ program and lie outside the forest edge. There are a couple of polygons excluded because they are part of another protected area (Tiwai island) or leased land by companies.

Table A.6.7 – Robustness check conservation behavior: only population-adjusted polygons

	Forest loss	Primary forest loss	Secondary forest loss
Post*REDD+	-0.972^{***}	0.073	-0.333
	(0.141)	(0.050)	(0.266)
Post	3.294^{***}	0.052^{**}	0.425^{***}
	(0.093)	(0.023)	(0.157)
REDD+	0.000	0.187^{***}	-0.390
	(0.040)	(0.041)	(0.240)
Constant	0.707^{***}	0.098^{***}	2.573^{***}
	(0.025)	(0.018)	(0.138)
Num. obs.	4158	1356	1356

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regressions for forest loss (satellite data). Forest loss is the percentage loss of forest (primary and secondary). Primary forest loss is loss of old growth forest and secondary forest loss measures conversion of fallow to production agriculture. Both are classified through extensive ground measurements, which were done in 2013. The number of observations is therefore lower for these two outcomes, as data ranges from 2013-2018. Robust standard errors in parenthesis. Sample is restricted to polygons that are weighted to population size. Robust standard errors in parentheses.

Table A.6.8 – Difference-in-difference analysis: parallel trends 2010-2014

	Livelihoods family	Income	Assets	Durable loan	Sustainable farming	Labor access	Income farm wages	Income NTFP
Post*REDD+	0.216	0.230	0.146	0.088	−0.157	0.219	0.088	0.221
	(0.152)	(0.284)	(0.126)	(0.096)	(0.158)	(0.189)	(0.124)	(0.145)
Post	0.291**	0.333	0.663***	−0.229***	0.120	−0.208	−0.397***	−0.470***
	(0.117)	(0.209)	(0.066)	(0.080)	(0.122)	(0.148)	(0.101)	(0.097)
REDD+	−0.370***	−0.298**	−0.350***	−0.128	0.307**	−0.125	0.013	−0.268**
	(0.109)	(0.122)	(0.121)	(0.084)	(0.120)	(0.108)	(0.112)	(0.123)
Constant	0.000	0.000	0.000	−0.000	0.000	0.000	−0.000	−0.000
	(0.076)	(0.089)	(0.087)	(0.055)	(0.081)	(0.088)	(0.089)	(0.082)
Years	2	2	2	2	2	2	2	2
Villages	56	56	56	56	56	56	56	56
Num. obs.	1312	1312	1312	1312	1225	1262	1310	1312

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Difference-in-difference analysis using OLS regressions to test for parallel trends between 2010 and 2014. The livelihood family outcome is a summary index (average of z-scores) of an income index, an assets index, a durable loan size measure. The income index is a summary index (average of z-scores) of total household income, monthly consumption expenditure and yearly durable expenditure. Assets is the sum of all assets owned. Durable loan size is the amount borrowed for durable investments. Sustainable farming practices measures the number of practices used by a household. Labor access index is an index of three farm labor access variables (upland rice, wetland rice, and plantation) indicating to what extent there is access to labor. Income farm wages is a continuous variable(IHS transformed) measuring the yearly household income from farm wages. Income NTFP is a continuous variable (IHS transformed) measuring the yearly income from Non-Timber Forest Products collection. Robust standard errors in parentheses clustered at the village level.

Chapter 7

Synthesis

7.1 Introduction

This thesis has examined three different approaches to development. None of the approaches can be considered a resounding success. It is not clear whether (agricultural) FDI or conservation programs can fill the gap the drop in Official Development Assistance (ODA) has created. The common thread of this thesis is economic development: higher incomes, more welfare, better lives for the world's poorest. This encompasses almost the entire field of development economics. What then, is the added value of combining these into one thesis? This chapter proposes three overarching insights: the effect of a strained labor market, how social networks matter for the distribution of resources and how development programs might exacerbate inequality.

7.2 Synthesis

7.2.1 Labor in Agriculture

Sub-Saharan Africa is generally considered to have a substantial labor force. Figure 7.1 shows the trend in the working-age and urban population over the past 30 years. The working-age population has more than doubled. Projections expect this trend to persist, with the total Sub-Saharan African population increasing to 2.5 billion by 2050 (Economist, 2020). Therefore, development practitioners consider labor surpluses to be one of the largest problems facing the continent in the coming decades. However, these average growth rates and population densities mask important regional variation. Figure 7.1 also shows that almost all of this growth has taken place in urban areas. This means that the agricultural sector will likely not see an increasing labor force.

This trend can be seen in Sierra Leone. A nation-wide survey by the Ministry of Agriculture, Forestry and Food Security found sharp labor shortages for agricultural labor, especially during times of peak labor demand (MAFFS, 2011). 65% of agriculture-focused households experience a shortage of labor sometime during the agricultural season. Chapters 4, 5 and 6 of this thesis have therefore assumed a strained labor market when explaining outcomes.

The main mechanism hypothesized in both Chapter 4 (on the sugarcane plantation in Northern Sierra Leone) and Chapter 6 (on the national park in Eastern Sierra Leone) are shifts in the labor market. In Chapter 4 a labor demand shock drives up the local

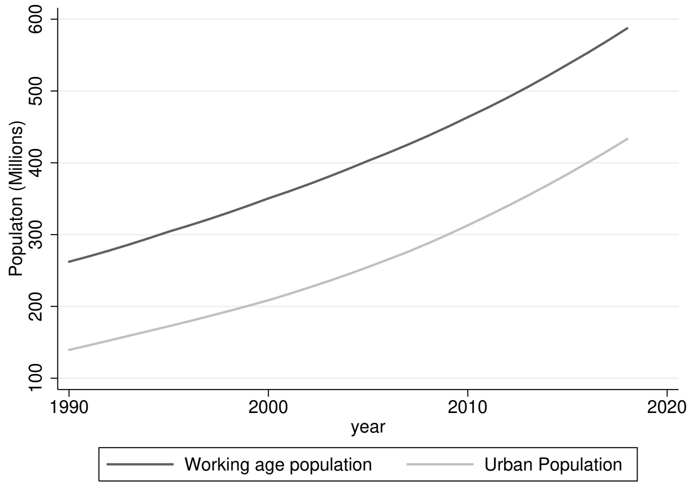


Figure 7.1 – Population trends in Sub-Saharan Africa

Source: World Bank (2020)

labor price which reduces local labor availability, in turn reducing local agricultural production and finally (greatly) reducing incomes. In Chapter 6 a shift to forest-friendly labor activities (Cocoa production and Non-Timber Forest Products collection) reduces pressure on the environmentally important buffer zone. Because of the labor shortage this shift to forest-friendly activities takes away labor from other activities such as slash-and-burn agriculture and logging. In both cases the outcomes are likely to be vastly different had the labor market not been strained. For Chapter 4, if the labor pool was larger labor prices would not have responded, leading to unchanged agricultural production. In Chapter 6 the creation and promotion of forest-friendly activities would not have reduced pressure on the national park had there been plenty of labor that could remain in forest-damaging activities.

Generally speaking, a focus on what factors are the main constraints for production (agricultural or otherwise) is useful to examine mechanisms. A constrained market in any factor, be it land, labor, capital or otherwise is likely to further transmit impacts. One problem with this approach is measurement. This thesis has used several survey-based approaches to quantify shifts in the labor and land market. Market prices are useful, but as they are usually determined at the cluster level (or even at a higher level when markets function better), statistical power is too low to find significant effects. Simple survey questions that ask about access to land/labor are subjective and cannot distinguish between shifts in supply and demand.[1] Labor diaries are time-intensive to gather and work best at short timeframes. Developing easier methods to assess market constraints is therefore very useful. Until these are found, taking the time to collect detailed market data is this thesis' recommendation.

This discussion fits into a larger existing discussion about the future of agriculture in Africa. Standard two-sector models predict that the marginal product of labor (and thus wages) will only rise when labor is moved from agriculture to other sectors (Lewis, 1954). If this is true, it can be expected that this trend of urbanization and a move out of agriculture will cause development. To keep producing enough food farming will need to become more intensive and/or productive (for example through the adoption of new technologies). This thesis contributes to this discussion by examining how these labor shortages in agriculture will affect development programs during this transition.

7.2.2 Networks and Distribution

There is a long history of using network analyses to model distribution and diffusion processes. For example, epidemiologists use network models (Bass models, simple contagion models) to predict the spread of viral diseases. This approach is very effective. This literature has shown that those with fewer connections are less likely to be 'infected', be it by a viral disease, a new piece of gossip or a marketing campaign. This thesis has shown a similar tendency for resource distribution: those unconnected are less likely to be on the receiving end.

A literature is emerging that attempts to use existing social networks to optimize processes of technology diffusion. This assumes that diffusion of resources will generally occur along existing social network lines (this thesis provides evidence for that in Chapters 2 and 3), and therefore taking this structure into account can optimize this process. Optimizing here means increasing the speed of diffusion, the adoption rate and even the overall spread. There is no consensus on the optimal way to approach this. Papers have used centrality measures (Kim et al., 2015; Banerjee et al., 2013), clustering (Chami et al., 2017) or model-based approaches (Beaman et al., 2018). However, one outcome that is often overlooked is the final distribution: who ends up on the receiving end?

[1] This thesis uses the question 'How much of a problem is it to get labor for your farm? 1. None 2. A little problem. 3. A problem. 4. A big problem'.

In Chapter 2 the two approaches to diffusion (one through 'central' ambassadors, one through 'isolate' ambassadors) lead to similar rates of adoption, knowledge and willingness to pay. However, there is an important effect on the final receivers. Central ambassadors give their resources to others that are similar to them (e.g. also very central) and isolate ambassadors give to others that are very central. The obvious interpretation is that central ambassadors use their access to information and influence to 'claim' the goods. Alternatively, isolate ambassadors use their resources to improve their position within the network, or optimize for efficiency. Regardless of the mechanism, the policy implication is that irrespective of who is targeted, in the end the resources will flow to those at the center of the network. Diffusion through the network is then not an inequality-neutral process but perpetuates existing inequalities.

Chapter 3 examines trusting behavior along social network lines. One interesting finding is that players trust others with a lower centrality *more*. Alternatively, individuals with a lower centrality are more likely to receive resources. This is opposite to the result from Chapter 2, where individuals with a high centrality were given the resources. What might cause this finding to be opposite? This fits into a larger discussion on whether lab results translate to real-life behavior (Levitt and List, 2007). One possible explanation is stakes. The value associated with an exchange in the trust game is lower than the value of the chemical fertilizer and information on its use. Spreading the information is also costly (opportunity costs of time). It might be that with low stakes other motivations (e.g. altruism) are relatively more important than other, strategic motivations. If this is the case, small-scale aid programs that use social networks might not increase inequality, but larger aid programs would.

There might also be researcher demand effects: the distribution of fertilizer in Chapter 2 happens mostly out of sight of the researchers, while the decision to share resources in the trust game was directly communicated to a research assistant. In other words, subjects in Chapter 3 face stronger scrutiny. If respondents believe that the goal of the research is to find evidence for altruism, they could fill this perceived demand by providing 'correct' answers: giving those worse off more. The context of the decision-making might also matter. The trust game was framed more as a game about giving/sharing resources with others,[2] while the experiment in Chapter 2 was more about teaching others a new technology. Again, in the lab approach altruistic motivations might be more important, while the field approach lends itself more to motivations of maximizing efficiency (which we find some evidence for).

[2] We never used the word trust during the game, not even to the research assistants. We referred to it as the 'triple game' instead.

Other potential causes that are unlikely to contribute in this case are anonymity of decisions made (no anonymity in both chapters), and selection (Subjects were from the same villages and were in both cases selected based on eigenvector centrality).

Overall, this adds to the discussion on how social networks play a role in determining the final distribution of resources. Most research so far has used lab approaches to determine the importance of networks on distribution and altruism. This discussion shows that it is unlikely that those results will translate to the field.

7.2.3 Development and Inequality

Piketty (2014)'s popular book has spurred a renewed interest in global inequality. There is evidence that high inequality negatively affects economic growth and causes social unrest (Cingano, 2014). Figure 7.2 shows that inequality in Sub-Saharan Africa has been constant and high over the past 30 years: around 55% of total income in Sub-Saharan Africa is earned by the top 10% richest. This has remained constant, despite substantial increases in GDP/Capita. Inequality is an oft-overlooked outcome when evaluating development programs, and is mostly analyzed using cross-country analyses (Ravallion, 2001). This thesis has examined how public (ODA) and private (FDI) approaches to development might affect inequality.

One of the main outcomes of Chapter 4 on agricultural FDI is an increase in inequality. Chapter 4 draws lorenz curves and calculates the village-level gini index to assess changes in local inequality, and finds that it does. Chapter 2 also finds suggestive evidence that development programs can increase inequality. By explicitly examining the pattern of distribution Chapter 2 argues that inequality is likely to remain through farmer-field-school approaches with a network focus. With Sub-Saharan Africa having such high rates of inequality, the opposite result would be preferred: that interventions *reduce* inequality.

But perhaps the increase in inequality is unsurprising: development interventions do not work in a vacuum and will be subject to local power dimensions. Attempts to bypass this (as in Chapter 2) are therefore unlikely to be successful. In Chapter 4 many of the stakeholders (the village inhabitants) were not consulted during the negotiations for the land lease. Instead, negotiations were limited to the village landowners and local elites. Had these other villagers been consulted additional income support programs could have been set up to compensate the losers of the investment.

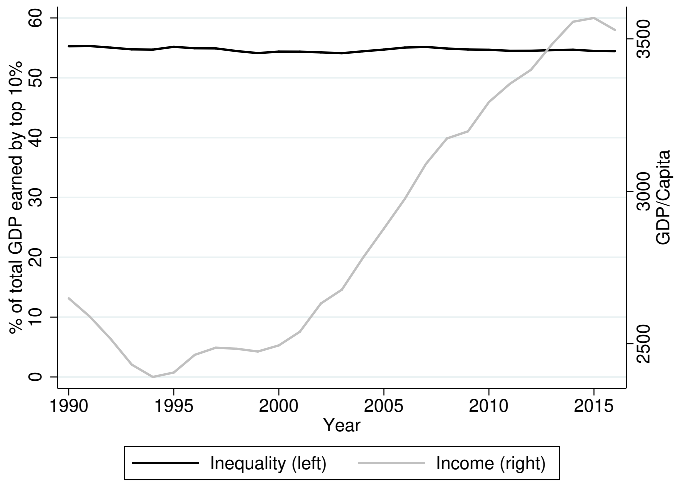


Figure 7.2 – Inequality and Income in Sub-Saharan Africa

Inequality is the % of total GDP earned by the top 10% richest. Income is GDP/Capita, PPP, constant 2010 US$. Sources: Alvaredo et al. (2018); World Bank (2020)

Linking back to the earlier discussion on labor in agriculture, it is likely that the productivity of the large-scale farm in Chapter 4 is higher than small-scale subsistence farming. If the future of agriculture is indeed large-scale, this means that increasing the efficiency of farming will also lead to higher inequality. This same tension applies to Chapter 2: technology diffusion is expected to be more efficient when taking a farmer-field school approach, but Chapter 2 shows that this approach is likely to perpetuate or increase existing inequalities. In both cases there appears to be a tradeoff between efficiency and inequality. When designing development programs or investments, policymakers should consider that a very efficient approach has the risk of perpetuating or increasing local inequality.

Bibliography

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies 72*(1), 1–19.

Abadie, A. and M. D. Cattaneo (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics 10*(1), 465–503.

Adloff, F. and S. Mau (2006). Giving social ties, reciprocity in modern society. *European Journal of Sociology 47*(1), 93–123.

Alesina, A. and G.-M. Angeletos (2003). Fairness and Redistribution: U.S. versus Europe. *NBER Working Paper Series 9502*(02).

Alfaro, L., A. Chanda, S. Kalemli-Ozcan, and S. Sayek (2010). Does foreign direct investment promote growth? Exploring the role of financial markets on linkages. *Journal of Development Economics 91*(2), 242–256.

Ali, D., K. Deininger, and A. Harris (2019). Does large farm establishment create benefits for neighboring smallholders? Evidence from Ethiopia. *Land Economics 95*(1), 71–90.

Alix-Garcia, J. M., K. R. E. Sims, and P. Yañez-Pagans (2015). Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's Payments for Ecosystem Services Program. *American Economic Journal: Economic Policy 7*(4), 1–40.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018). World Inequality Report 2018. Technical report.

Anane, M. and C. Y. Abiwu (2011). Independent Study Report of the ADDAX Bioenergy Sugarcane-to-Ethanol Project in the Makeni Region in Sierra Leone. Technical report, Sierra Leone Network on the Right to Food, Freetown.

Aragón, F. M. and J. P. Rud (2013). Natural resources and local communities: Evidence from a peruvian gold mine. *American Economic Journal: Economic Policy 5*(2), 1–25.

Aral, S., L. Muchnik, and A. Sundararajan (2009, sep). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences 106*(51), 21544–21549.

Arezki, R., K. Deininger, and H. Selod (2013). What drives the global "land rush"? *The World Bank Economic Review 29*(2), 207–233.

Arriagada, R. A., P. J. Ferraro, E. O. Sills, S. K. Pattanayak, and S. Cordero-Sancho (2012). Do payments for environmental services affect forest cover? A farm-level evaluation from Costa Rica. *Land Economics 88*(2), 382–399.

Bailard, C. S. (2009). Mobile phone diffusion and corruption in Africa. *Political Communication 26*(3), 333–353.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

Banerjee, A. V. and E. Duflo (2009). The Experimental Approach to Development Economics. *Annual Review of Economics 1*(1), 151–178.

Banerjee, R. and R. Maharaj (2020). Heat, infant mortality, and adaptation: Evidence from India. *Journal of Development Economics 143*, 102378.

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020, mar). Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial. *American Economic Review 110*(3), 824–859.

Barr, A. and P. Serneels (2009). Reciprocity in the workplace. *Experimental Economics 12*(1), 99–112.

Baxter, J. (2011). Understanding Land Investment Deals in Africa - Country Report: Sierra Leone. Technical report, The Oakland Institute, Oakland.

Baxter, J. (2013). Who is benefitting? Technical report, Action for Large-Scale Land Acquisition Transparency, Freetown.

Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2018). Can network theory-based targeting increase technology adoption? Technical report, National Bureau of Economic Research.

Ben-Ner, A., B. P. McCall, M. Stephane, and H. Wang (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior and Organization 72*(1), 153–170.

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2007). Social Identity and Preferences.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, Reciprocity and Social History. *Games and Economic Behavior 10*, 122–142.

Berry, J., S. Mehta, P. Mukherjee, H. Ruebeck, and G. K. Shastry (2020). Implementation and effects of India's national school-based iron supplementation program. *Journal of Development Economics 144*, 102428.

Bitzer, V., R. van Balen, and B. de Steenhuijsen Piters (2017). Aid & Trade in Dutch Development Cooperation. Technical report, KIT Royal Tropical Institute, Amsterdam.

Blackwell, M., S. Iacus, G. King, G. Porro, and Others (2010). cem: Coarsened exact matching in Stata. *The Stata Journal 9*(4), 524.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *The Quarterly Journal of Economics 129*(2), 697–752.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology 2*(1), 113–120.

Bonacich, P. (2007, dec). Some unique properties of eigenvector centrality. *Social Networks 29*(4), 555–564.

Borensztein, E., J. De Gregorio, and J.-W. Lee (1998). How does foreign direct investment affect economic growth? *Journal of International Economics 45*(1), 115–135.

Borgatti, S. P. (2005, jan). Centrality and network flow. *Social Networks 27*(1), 55–71.

Börner, J., K. Baylis, E. Corbera, D. Ezzine-de Blas, J. Honey-Rosés, U. M. Persson, and S. Wunder (2017). The Effectiveness of Payments for Environmental Services. *World Development 96*, 359–374.

Bos, A. B., A. E. Duchelle, A. Angelsen, V. Avitabile, V. De Sy, M. Herold, S. Joseph, C. De Sassi, E. O. Sills, W. D. Sunderlin, and S. Wunder (2017). Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters 12*(7).

Bottazzi, P., D. Crespo, L. O. Bangura, and S. Rist (2018). Evaluating the livelihood impacts of a large-scale agricultural investment: Lessons from the case of a biofuel production company in northern Sierra Leone. *Land Use Policy 73*(December), 128–137.

Bottazzi, P., A. Goguen, and S. Rist (2016). Conflicts of customary land tenure in rural Africa: is large-scale land acquisition a driver of 'institutional innovation'? *Journal* *of Peasant Studies 43*(5), 971–988.

Bouma, J., D. V. Soest, and E. Bulte (2008). Trust , Trustworthiness and Cooperation : Social Capital and Community Resource Management. *Journal of Environmental Economics and Management 56*(2), 1–30.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature 36*(1), 75–111.

Bowles, S. and S. Polanía-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature 50*(2), 368–425.

Brosig, J., T. Riechmann, and J. Weimann (2007). Selfish in the end?: An investigation of consistency and stability of individual behavior.

Bulte, E., K. Leuveld, E. Nillesen, and M. Voors (2015). Farm Households in Eastern Congo, Baseline Report. Technical report, Wageningen University.

Bulte, E. H., P. Richards, and M. Voors (2018). *Institutions and Agrarian Development: A New Approach to West Africa*. Springer.

Butchart, S. H. M., M. Walpole, B. Collen, A. Van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, and Others (2010). Global biodiversity: indicators of recent declines. *Science 328*(5982), 1164–1168.

Camerer, C. F. and E. Fehr (2002). Measuring Social Norms and Preferences using Experimental Games : A Guide for Social Scientists Measuring social norms and preferences using experimental games : A guide for social scientists. *Research in Economics* (97), 55–95.

Cárdenas, J. C. and J. P. Carpenter (2008). Behavioural Development Economics: Lessons from Field Labs in the Developing World. *The Journal of Development Studies 44*(3), 311–338.

Cardenas, J. C., J. Stranlund, and C. Willis (2000). Local environmental control and institutional crowding-out. *World Development 28*(10), 1719–1733.

Carpenter, J. and E. Seki (2011). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. *Economic Inquiry 49*(2), 612–630.

Carter, M. R., R. Laajaj, and D. Yang (2014). Subsidies and the persistence of technology adoption: Field experimental evidence from Mozambique. Technical report, National Bureau of Economic Research.

Ceballos, G., P. R. Ehrlich, and R. Dirzo (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences 114*(30), E6089—-E6096.

Centola, D. and M. Macy (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology 113*(3), 702–734.

Chami, G. F., A. A. Kontoleon, E. Bulte, A. Fenwick, N. B. Kabatereine, E. M. Tukahebwa, and D. W. Dunne (2017). Diffusion of treatment in social networks and mass drug administration. *Nature Communications 8*(1), 1929.

Chen, B. Y. and S. X. Li (2009). Group Identity and Social Preferences. *The American Economic Review 99*(1), 431–457.

Christensen, D., A. Hartman, and C. Samii (2020). Legibility and External Investment: An Institutional Natural Experiment in Liberia.

Cilliers, J., I. Kasirye, C. Leaver, P. Serneels, and A. Zeitlin (2013). Improving teacher attendance using a locally managed monitoring scheme: Evidence from Ugandan Primary Schools. *Rapid response paper for International Growth Centre*.

Cingano, F. (2014). Trends in Income Inequality and its Impact on Economic Growth. *OECD Social, Employment and Migration Working Papers* (163).

Coghlan, B., R. J. Brennan, P. Ngoy, D. Dofara, B. Otto, M. Clements, and T. Stewart (2006). Mortality in the Democratic Republic of Congo: a nationwide survey. *The Lancet 367*(9504), 44–51.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology 94*, S95—-S120.

Collier, P. and S. Dercon (2014). African Agriculture in 50 Years: Smallholders in a Rapidly Changing World? *World Development 63*, 92–101.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review 100*(1), 35–69.

Crespo, N. and M. P. Fontoura (2007). Determinant factors of FDI spillovers–what do we really know? *World Development 35*(3), 410–425.

Croson, R. and U. Gneezy (2009). Gender Differences in Preferences. *Journal of Economic Literature 47*(2), 448–474.

Curtis, P. G., C. M. Slay, N. L. Harris, A. Tyukavina, and M. C. Hansen (2018). Classifying drivers of global forest loss. *Science 361*(6407), 1108–1111.

Cust, J. and S. Poelhekke (2015). The local economic impacts of natural resource extraction. *Annual Review of Resource Economics 7*(1), 251–268.

de Mello Jr., L. R. (1997). Foreign direct investment in developing countries and growth: A selective survey. *The Journal of Development Studies 34*(1), 1–34.

De Schutter, O. (2011). How not to think of land-grabbing: three critiques of large-scale investments in farmland. *The Journal of Peasant Studies 38*(2), 249–279.

DeFries, R., A. Hansen, A. C. Newton, and M. C. Hansen (2005). Increasing isolation of protected areas in tropical forests over the past twenty years. *Ecological Applications 15*(1), 19–26.

Deininger, K. and F. Xia (2016). Quantifying spillover effects from large land-based investment: the case of Mozambique. *World Development 87*, 227–241.

Deschenes, O., H. Wang, S. Wang, and P. Zhang (2020). The effect of air pollution on body weight and obesity: Evidence from China. *Journal of Development Economics 145*, 102461.

Dessy, S., G. Gohou, and D. Vencatachellum (2012). Foreign Direct Investments in Africa's Farmlands: Threat or Opportunity for Local Populations?

Development Indicators Unit (2016). Millennium Development Goals Indicators.

Dorward, A. R., J. F. Kirsten, S. W. Omamo, C. Poulton, and N. Vink (2009). Institutions and the agricultural development challenge in Africa. *Institutional Economics Perspectives on African agricultural Development 1*, 3–34.

Duchelle, A. E., C. de Sassi, P. Jagger, M. Cromberg, A. M. Larson, W. D. Sunderlin, S. S. Atmadja, I. A. P. Resosudarmo, and C. D. Pratama (2017). Balancing carrots and sticks in REDD+: Implications for social safeguards. *Ecology and Society 22*(3).

Economist (2020). Africa's population will double by 2050. *https://www.economist.com/special-report/2020/03/26/africas-population-will-double-by-2050*.

Ellis, F. (2005). Small farms, livelihood diversification, and rural-urban transitions: Strategic issues in Sub-Saharan Africa. In *The Future of Small Farms. Research Workshop Proceedings*, pp. 135–149.

Emerick, K., A. de Janvry, E. Sadoulet, and M. H. Dar (2016). Identifying early adopters, enhancing learning, and the diffusion of agricultural technology. *Working Paper*.

Engström, L. and F. Hajdu (2019). Conjuring 'Win-World'–Resilient Development Narratives in a Large-Scale Agro-Investment in Tanzania. *Journal of Development Studies 55*(6), 1201–1220.

FAO (2015). FAO Statistical Pocketbook 2015: World food and agriculture. Technical report, Food and Agriculture Organization of the United Nations.

Feder, G., J. R. Anderson, R. Birner, and K. Deininger (2010). Promises and realities of community-based agricultural extension. In *Community, Market and State in Development*, pp. 187–208. Springer.

Feder, G., R. Murgai, and J. B. Quizon (2003). *Sending farmers back to school: The impact of farmer field schools in Indonesia*, Volume 3022. World Bank Publications.

Fehr, E. and K. Hoff (2011). Introduction : Tastes , Castes And Culture : The Influence Of Society On Preferences. *The Economic Journal 121*(November), 396–412.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does Fairness Prevent Market Clearing ? An Experimental Investigation. *The Quarterly Journal of Economics 108*(2), 437–459.

Ferraro, P. J. and S. K. Pattanayak (2006). Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *PLoS Biology 4*(4), e105.

Fershtman, C. and U. Gneezy (2001). Discrimination in a Segmented Society : An Experimental Approach. *The Quarterly Journal of Economics 116*(1), 351–377.

Fielding, M., M. Davis, N. Weitz, I. Cummings-John, A. Hickey, F. X. Johnson, J. Senyagwa, L. Martinez, and M. Sun (2015). Agricultural investment and rural transformation: a case study of the Makeni bioenergy project in Sierra Leone. Technical report, Stockholm Environment Institute, Stockholm.

Foster, A. D. and M. R. Rosenzweig (2010). Microeconomics of technology adoption. *Annual Review of Economics 2*(1), 395–424.

Gallup, J. L. (2012). A new system for formatting estimation tables. *The Stata Journal 12*(1), 3.

Geldmann, J., A. Manica, N. D. Burgess, L. Coad, and A. Balmford (2019). A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America 116*(46), 23209–23215.

Godfray, H. C. J., J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin (2010). Food security: the challenge of feeding 9 billion people. *Science 327*(5967), 812–818.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

GoSL (2015). National land policy. Technical report, Government of Sierra Leone, Freetown.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology 91*(3), 481–510.

Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *The American Economic Review 83*(3), 525–548.

Grossman, G., M. Humphreys, and G. Sacramone-Lutz (2014). "I wld like u WMP to extend electricity 2 our village": On Information Technology and Interest Articulation. *American Political Science Review 108*(3), 688–705.

Gui-Diby, S. L. (2014). Impact of foreign direct investments on economic growth in Africa: Evidence from three decades of panel data analyses. *Research in Economics 68*(3), 248–256.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *The American Political Science Review 101*(4), 709–725.

Hagan, E. and A. Amoah (2019). Foreign direct investment and economic growth nexus in Africa. *African Journal of Economic and Management Studies*.

Halevy, N., G. Bornstein, and L. Sagiv (2008). "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm: Research article. *Psychological Science 19*(4), 405–411.

Hansen, M., P. Potapov, R. Moore, M. Hancher, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. Goetz, T. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. Justics, and J. Townshend (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science 342*(6160), 850–835.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, and Others (2013). High-resolution global maps of 21st-century forest cover change. *Science 342*(6160), 850–853.

Heino, M., M. Kummu, M. Makkonen, M. Mulligan, P. H. Verburg, M. Jalava, and T. A. Räsänen (2015). Forest loss in protected areas and intact forest landscapes: A global analysis. *PLoS ONE 10*(10), 1–21.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In search of Homo economicus: Behavioral experiments in 15 small scale societies. *American Economic Review 91*(2), 73–78.

Henrich, J., R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. Henrich, K. R. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton, and D. Tracer (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences 28*(6), 795–815; discussion 815–55.

Henrich, J., J. Ensminger, R. Mcelreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2010, mar). Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. *Science 327*(5972), 1480–1484.

Herbst, J. (2014). *States and power in Africa: Comparative lessons in authority and control*. Princeton University Press.

Herrera, D., A. Pfaff, and J. Robalino (2019). Impacts of protected areas vary with the level of government: Comparing avoided deforestation across agencies in the Brazilian Amazon. *Proceedings of the National Academy of Sciences of the United States of America 116*(30), 14916–14925.

Herrmann, R. (2017). Large-Scale Agricultural Investments and Smallholder Welfare: A Comparison of Wage Labor and Outgrower Channels in Tanzania. *World Development 90*, 294–310.

Herrmann, R. and U. Grote (2015). Large-scale Agro-Industrial Investments and Rural Poverty: Evidence from Sugarcane in Malawi. *Journal of African Economies*, 645—-676.

Herzer, D., S. Klasen, and Others (2008). In search of FDI-led growth in developing countries: The way forward. *Economic Modelling 25*(5), 793–810.

Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling. *https://cran.r-project.org/package=raster*. R package version 2.6-7.

Hofman, P., E. Mokuwa, P. Richards, and M. Voors (2020). Local Economy effects of Large-Scale Agricultural Investments.

Højsgaard, S. and U. Halekoh (2018). doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities. *https://cran.r-project.org/package=doBy*. R package version 4.6-1.

Iacus, S. M., G. King, and G. Porro (2011). Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Associa-*

tion 106(493), 345–361.

Jahnke, H. E. (1982). *Livestock production systems and livestock development in tropical Africa.* Kiel: Kieler Wissenschaftsverlag Vauk.

Jakiela, P. (2011). Social Preferences and Fairness Norms as Informal Institutions: Experimental Evidence and numerous seminar participants for helpful comments. *American Economic Review 101*(3), 509–513.

Jann, B. (2005). Making regression tables from stored estimates. *The Stata Journal 5*(3), 288–308.

Jann, B. (2007). Making regression tables simplified. *The Stata Journal 7*(2), 227–244.

Jann, B. (2012). ESTWRITE: Stata module to store estimation results on disk. *https://econpapers.repec.org/RePEc:boc:bocode:s450201*.

Jann, B. (2016). Estimating Lorenz and concentration curves. *The Stata Journal 16*(4), 837–866(30).

Jayachandran, S., J. D. Laat, E. F. Lambin, C. Y. Stanton, R. Audy, and N. E. Thomas (2017). Cash for carbon: A randomized trial ofpayments for ecosystem services to reduce deforestation. *Science 357*(6348), 267–273.

Jiao, X., C. Smith-hall, and I. Theilade (2015). Land Use Policy Rural household incomes and land grabbing in Cambodia. *Land Use Policy 48*, 317–328.

Jordaan, J. A. (2008). Intra-and inter-industry externalities from foreign direct investment in the Mexican manufacturing sector: New evidence from Mexican regions. *World Development 36*(12), 2838–2854.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and social collateral. *The Quarterly Journal of Economics 124*(3), 1307–1361.

Karlan, D. S. (2005). Using Experimental Economics to Measure Social Capital and Predict Financial Decisions. *The American Economic Review 95*(5), 1688 – 1699.

Kearney, M. S. and P. B. Levine (2015, dec). Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing. *American Economic Review 105*(12), 3597–3632.

Kendzior, J., J. P. Zibika, and M. Voors (2015). Social relationships, local institutions, and the diffusion of improved variety seed and field management techniques in rural communities: six case studies in South Kivu, DRC.

Kim, D. A., A. R. Hwong, D. Stafford, D. A. Hughes, A. J. O'Malley, J. H. Fowler, and

N. A. Christakis (2015). Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *The Lancet 386*(9989), 145–153.

Kindornay, S. and F. Reilly-King (2013). Promotion and partnership: bilateral donor approaches to the private sector. *Canadian Journal of Development Studies / Revue Canadienne d'études du Développement 34*(4), 533–552.

King, G. and R. Nielsen (2019). Why propensity scores should not be used for matching. *Political Analysis 27*(4), 435–454.

Kleemann, L. and R. Thiele (2015). Rural welfare implications of large-scale land acquisitions in Africa: A theoretical framework. *Economic Modelling 51*, 269–279.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007, jan). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Kolk, A., R. van Tulder, and E. Kostwinder (2008). Business and partnerships for development. *European Management Journal 26*(4), 262–273.

Koning, N. and M. K. van Ittersum (2009). Will the world have enough to eat? *Current Opinion in Environmental Sustainability 1*(1), 77–82.

Landa, J. (1981). A theory of the ethnically homogeneous middleman group: an institutional alternative to contract law. *The Journal of Legal Studies 10*(2), 349–362.

Landmatrix (2020). The Online Public Database on Land Deals.

Larson, J. M. and J. I. Lewis (2017). Ethnic networks. *American Journal of Political Science 61*(2), 350–364.

Larson, J. M., J. I. Lewis, and P. Rodríguez (2019). From Chatter to Action: How Social Networks Inform and Motivate in Rural Uganda.

Lay, J., K. Nolte, K. Sipangule, and Others (2018). Large-scale farms and smallholders: Evidence from Zambia. Technical report, GIGA Working Papers.

Le Velly, G., A. Sauquet, and S. Cortina-Villar (2017). PES impact and leakages over several cohorts: The case of the PSA-H in Yucatan, Mexico. *Land Economics 93*(2), 230–257.

Leonard, B., D. P. Parker, and T. L. Anderson (2020). Land quality, land rights, and indigenous poverty. *Journal of Development Economics 143*, 102435.

Levitt, S. D. and J. A. List (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, 153–174.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school 22*(2), 139–191.

Li, C. and S. Tanna (2019). The impact of foreign direct investment on productivity: New evidence for developing countries. *Economic Modelling 80*, 453–466.

Li, X. and X. Liu (2005). Foreign direct investment and economic growth: an increasingly endogenous relationship. *World Development 33*(3), 393–407.

List, J. A. (2005). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *National Bureau of Economic Research Working pa*, 1–51.

Liu, Z. (2008). Foreign direct investment and technology spillovers: Theory and evidence. *Journal of Development Economics 85*(1-2), 176–193.

Liversage, H. (2010). Responding to" land grabbing" and promoting responsible investment in agriculture. *IFAD Occasional Paper*.

Lui, G. V. and D. A. Coomes (2016). Tropical nature reserves are losing their buffer zones, but leakage is not to blame. *Environmental Research 147*, 580–589.

MAFFS (2011). Agricultural Household Tracking Survey (AHTS) Final Report. Technical report, Ministry of Agriculture, Forestry and Food Security, Freetown.

Makiela, K. and B. Ouattara (2018). Foreign direct investment and economic growth: Exploring the transmission channels. *Economic Modelling 72*, 296–305.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review 80*(2), 319–323.

Marfurt, F., F. Käser, and S. Lustenberger (2016). Local Perceptions and Vertical Perspectives of a Large Scale Land Acquisition Project in Northern Sierra Leone. *Homo Oeconomicus 33*(3), 261–279.

Mauss, M. (2002). *The gift: The form and reason for exchange in archaic societies.* Routledge.

Meyer, B. D. and J. X. Sullivan (2003). Measuring the Well-Being of the Poor Using Income and Consumption. *The Journal of Human Resources 38*, 1180.

Michaelsen, M. M. and P. Salardi (2020). Violence, psychological stress and educational performance during the “war on drugs” in Mexico. *Journal of Development Economics 143*, 102387.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics 89*(11-12), 2325–2368.

Millar, G. (2015a). Investing in peace: foreign direct investment as economic restoration in Sierra Leone? *Third World Quarterly 36*(9), 1700–1716.

Millar, G. (2015b). "We Have No Voice for That": Land Rights, Power, and Gender in Rural Sierra Leone. *Journal of Human Rights 14*(4), 445–462.

Millar, G. (2016a). Knowledge and Control in the Contemporary Land Rush: Making Local Land Legible and Corporate Power Applicable in Rural Sierra Leone. *Journal of Agrarian Change 16*(2), 206–224.

Millar, G. (2016b). Local experiences of liberal peace: Marketization and emergent conflict dynamics in Sierra Leone. *Journal of Peace Research 53*(4), 569–581.

Millar, G. (2017). For whom do local peace processes function? Maintaining control through conflict management. *Cooperation and Conflict 52*(3), 293–308.

Mokuwa, E., M. Voors, E. Bulte, and P. Richards (2011). Peasant grievance and insurgency in Sierra Leone: judicial serfdom as a driver of conflict. *African Affairs 110*(440), 339–366.

Müller, K. and H. Wickham (2018). tibble: Simple Data Frames. *https://cran.r-project.org/package=tibble*. R package version 1.4.2.

Mwabu, G., C. Ugaz, and G. White (2001). *Social provision in low-income countries: new patterns and emerging trends.* Oxford University Press.

Nolte, K., W. Chamberlain, and M. Giger (2016). International Land Deals for Agriculture Fresh insights from the Land Matrix : Analytical Report II. Technical report, Centre for Development and Environment, University of Bern; Centre de coopération internationale en recherche agronomique pour le développement; German Institute of Global and Area Studies; University of Pretoria; Bern Open Publishing, Bern, Montpellier, Hamburg, Pretoria.

Nwaogu, U. G. and M. J. Ryan (2015). FDI, foreign aid, remittance and economic growth in developing countries. *Review of Development Economics 19*(1), 100–115.

Okumu, B. and E. Muchapondwa (2020). Welfare and forest cover impacts of incentive based conservation: Evidence from Kenyan community forest associations. *World Development 129*, 104890.

Peters, P. E. (2004). Inequality and social conflict over land in Africa. *Journal of Agrarian Change 4*(3), 269–314.

Peters, P. E. (2013). Conflicts over land and threats to customary tenure in Africa.

African Affairs 112(449), 543–562.

Pfaff, A. and J. Robalino (2017). Spillovers from Conservation Programs. *Annual Review of Resource Economics 9*(1), 299–315.

Piketty, T. (2014). Capital in the 21st Century.

Putnam, R. D. and Others (2000). *Bowling alone: The collapse and revival of American community.* Simon and schuster.

Pypers, P., J.-M. Sanginga, B. Kasereka, M. Walangululu, and B. Vanlauwe (2011, feb). Increased productivity through integrated soil fertility management in cassava–legume intercropping systems in the highlands of Sud-Kivu, DR Congo. *Field Crops Research 120*(1), 76–85.

R Core Team (2017). R: A Language and Environment for Statistical Computing. *https://www.r-project.org/.*

Ravallion, M. (2001, nov). Growth, inequality and poverty: Looking beyond averages. *World Development 29*(11), 1803–1815.

Reardon, T., C. P. Timmer, C. B. Barrett, and J. Berdegue (2003). The rise of supermarkets in Africa, Asia, and Latin America. *American Journal of Agricultural Economics 85*(5), 1140–1146.

Richards, P. (1986). *Coping with hunger: Hazard and experiment in a West African farming system.* London: Allen & Unwin.

Robalino, J. and A. Pfaff (2013). Ecopayments and deforestation in Costa Rica: A nationwide analysis of PSA's initial years. *Land Economics 89*(3), 432–448.

Romero, M., J. Sandefur, and W. A. Sandholtz (2020, feb). Outsourcing Education: Experimental Evidence from Liberia. *American Economic Review 110*(2), 364–400.

Roopsind, A., B. Sohngen, and J. Brandt (2019). Evidence that a national REDD+ program reduces tree cover loss and carbon emissions in a high forest cover, low deforestation country. *Proceedings of the National Academy of Sciences of the United States of America 116*(49), 24492–24499.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed.* Yale University Press.

Sherry Jr, J. F. (1983). Gift giving in anthropological perspective. *Journal of Consumer Research 10*(2), 157–168.

Shete, M. and M. Rutten (2015). Land Use Policy Impacts of large-scale farming on local communities ' food security and income levels – Empirical evidence from Oromia Region , Ethiopia. *Land Use Policy 47*, 282–292.

Sills, E. O., C. de Sassi, P. Jagger, K. Lawlor, D. A. Miteva, S. K. Pattanayak, and W. D. Sunderlin (2017). Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being. *Global Environmental Change 43*, 148–160.

SiLNoRF (2014). Annual Monitoring Report on the operations of Addax Bioenergy. Technical report, Sierra Leone Network on the Right to Food, Freetown.

Simonet, G., J. Subervie, D. Ezzine-De-Blas, M. Cromberg, and A. E. Duchelle (2019). Effectiveness of a REDD1 project in reducing deforestation in the Brazilian Amazon. *American Journal of Agricultural Economics 101*(1), 211–229.

Simpson, B. M., S. Franzel, A. Degrande, G. Kundhlande, and S. Tsafack (2015). Farmer-to-farmer extension: Issues in planning and implementation. Technical report.

Smith, R. D. (2006, dec). It's not just what you do, it's the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics 15*(3), 281–293.

Sokoloff, K. L. and S. L. Engerman (2000). Institutions, factor endowments, and paths of development in the new world. *Journal of Economic Perspectives 14*(3), 217–232.

Taylor, J. E. and M. J. Filipski (2014). *Beyond experiments in development economics: local economy-wide impact evaluation.* Oxford University Press.

Tothmihaly, A. and V. Ingram (2017). How can the productivity of Indonesian cocoa farms be increased? Technical report, GlobalFood Discussion Papers, Göttingen.

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica: Journal of the Econometric Society*, 539–591.

Udry, C. (1996). Gender, agricultural production, and the theory of the household. *Journal of Political Economy 104*(5), 1010–1046.

UNDP (2016). Human Development Report 2016: Human Development for Everyone. Technical report, United Nations Development Programme, New York.

UNEP-WCMC and IUCN (2020). Protected Planet: the World Database on Protected Areas. *www.protectedplanet.net*.

van der Windt, P., M. Humphreys, L. Medina, J. F. Timmons, and M. Voors (2019). Citizen Attitudes Toward Traditional and State Authorities: Substitutes or Complements? *Comparative Political Studies 52*(12), 1810–1840.

Van Kerm, P. (2009). sgini: – Generalized Gini and Concentration coefficients (with factor decomposition) in Stata, v1.1 (revised February 2010). Technical report, CEPS/INSTEAD, Differdange, Luxembourg.

Volk, S., C. Thöni, and W. Ruigrok (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior and Organization 81*(2), 664–676.

Voors, M. J., E. E. M. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. Van Soest (2012). Violent conflict and behavior: a field experiment in Burundi. *The American Economic Review 102*(2), 941–964.

Wickham, H., R. Francois, L. Henry, and K. Müller (2017). dplyr: A Grammar of Data Manipulation. *https://cran.r-project.org/package=dplyr*. R package version 0.7.4.

Wiik, E., R. D'Annunzio, E. Pynegar, D. Crespo, N. Asquith, and J. P. G. Jones (2019). Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice 1*(2), 1–11.

Wilebore, B. and D. Coomes (2016, dec). Combining spatial data with survey data improves predictions of boundaries between settlements. *Applied Geography 77*, 1–7.

Wilebore, B., M. Voors, E. H. Bulte, D. Coomes, and A. Kontoleon (2019). Unconditional Transfers and Tropical Forest Conservation: Evidence from a Randomized Control Trial in Sierra Leone. *American Journal of Agricultural Economics 00*(0), 1–25.

World Bank (2020). World Development Indicators.

Wunder, S., A. Angelsen, and B. Belcher (2014). Forests, Livelihoods, and Conservation: Broadening the Empirical Base. *World Development 64*(S1), S1–S11.

Summary

This thesis examines three approaches to development. The first approach, Official Development Aid, is analyzed with a specific focus on the distributional effects: who ends up benefitting? Next, it examines whether Foreign Direct Investments can lead to development. It examines two cases of agricultural investments, or large-scale land acquisitions. Finally, this thesis examines how the creation of protected areas can contribute to development and environmental protection.

Chapter 2 examines how social network-based approaches to technology diffusion affect the final distribution of this new technology. It does this by allocating the new technology to either a group of highly connected individuals ('centrals'), or a group with very few connections ('isolates'). These groups are asked to spread this technology to others within their village. We find that both groups are similarly good at transferring knowledge and inducing others to adopt the new technology. However, it also finds important distributional effects: irregardless of who was targeted, the resulting resources ended with those most central. Furthermore, there was strong attenuation of the effect as it diffused throughout the network. This implies that sufficient initial recipients should be selected.

Chapter 3 tests the relationship between interpersonal trust and social network connections. Using a trust game with individuals whose social networks had been fully mapped out, we test whether individuals use their network connections as 'social collateral' when choosing who to trust. Individuals indeed trust more when they are directly connected through their social network, but they also trust others more if they shared a greater proportion of mutual connections with the other player. Individuals therefore choose to behave more trusting and trustworthy when faced with players with whom they have a stronger connection.

Chapter 4 examines the impact of a large-scale agricultural investment of 24'000 hectares

on the local population. The investment leads to a large drop in average incomes, mainly driven by lower agricultural sales. This is hypothesized to be caused by the large labor demand shock the investment represents, which decreases access to labor and increases the labor price. Indeed, households that work for the investment greatly improve their incomes, with increased village-level inequality as a result.

Chapter 5 examines another important aspect of agricultural foreign direct investment: productive spillovers. It examines a 750-hectare cocoa plantation, created in an area with substantial small-scale cocoa producers. It finds that economic welfare is largely unaffected, but there are some productive spillovers: small-scale cocoa farmers are substantially less affected by a local fungal disease. Productive spillovers might therefore be a positive externality of foreign investments.

Chapter 6 examines the environmental and developmental impact of the creation of a large protected area. The national park is very effective: over the past 20 years there has hardly been deforestation within its boundaries. However, the area around the park (the buffer zone) sees substantial deforestation. To reduce this, a program was set up to induce farmers to move to more forest-friendly cultivation, which was successful: deforestation in the buffer zone went down by 30%. At the same time, this did not make the inhabitants worse off.

Chapter 7 looks across the chapters and provides several overarching insights. The local factor markets appears to be crucial when determining the mechanism behind impacts of interventions. Also, social networks appear to have important bearing on how resources are distributed, though this depends on the context and stakes. Finally, it is possible that more efficient approaches to development might perpetuate or increase existing inequalities, which should be taken into account when designing these programs.

Acknowledgements

First I want to thank my supervisor Maarten Voors. Our collaboration goes back a long way: I originally started working as a research assistant all the way back in 2013. When Maarten asked me to do a PhD with him supervising I did not need to doubt long as our collaboration has always been fantastic. During our very regular meetings he was always able to suggest new interesting avenues to explore in my research projects. Furthermore, he also allowed me great independence in pursuing these projects, allowing me to learn tremendous amounts. He also allowed me to present my working at very interesting locations around the world by leveraging his impressive network. During writing and presenting he gave strong but fair feedback which helped me make great strides in the quality of my work. But most importantly, we also had a great personal relationship, with late nights over beer where we discussed more interesting research projects. I am sure that Maarten was the most important contributing factor to how much I enjoyed my PhD.

Next, I would like to thank my promotor Erwin Bulte. Back in 2014 while still working on my MSc he suggested doing a PhD, planting the seed for what has been an inspiring and fantastic journey. Erwin's guidance proved instrumental in keeping me focused on the right projects to finish my PhD with minimal delays. More importantly, Erwin allowed me the freedom to pursue those projects I found interesting which ultimately lead to a better thesis as I could work on what I was passionate about.

I also want to thank all my colleagues at DEC. It was fantastic having such a large group of smart, helpful and nice people to work with on a regular basis. I learned a lot from receiving comments, but also seeing what interesting projects others were working on. Special thanks to several colleagues I had the pleasure to collaborate with on papers. Martha Ross taught me a lot about the practicalities of doing fieldwork but also helped me write better, which the reader of this thesis has (hopefully!) benefitted from. Working with Peter van der Windt has also taught me about how to hone in on the main contribution of a paper and put that front and center. Working with Mandy Malan, our resident R wizard, was also great and she helped me get better at writing R code. We also had great times while in the field. Thanks also to my other colleagues at DEC. Rein Haagsma for always having wise advice and great stories from his time doing a PhD. Janneke Pieters for always having great constructive comments. Robert Sparrow for teaching me about the publishing culture. Nico Heerink for his inspiring stories about China. Marrit van den Berg for her advice on teaching. Robert Lensink for good comments and interesting stories. Fan Li for teaching me so much about China. Francesco Cecchi for everything he taught me about academic culture. Jeroen Klomp for teaching me that there is always a result. I also want to thank Marian Jonkers for all her help when I was sorting out finances. Big thank you to Dineke Wemmenhove as well for always helping out with complex administrative matters. Thank you to Betty de Haan for help with many things.

Thanks also to all others who worked on a PhD at DEC while I did. Niccolò Meriggi, for great laughs, tips on what to see in Rome and fantastic help while in the field. Koen Leuveld, for some great Stata tricks. Lan Nguyen, for some great discussions about Dutch culture. Zihan Nie, for our weekly seminar assessment and talks about terrible TV shows. Mark Treurniet for his great insights and fun chats over lunch. Maria Naranjo Barrantes for her great stories about Costa Rica. Trang Nguyen for having good comments. Chen Qian for always having a smile on his face. Michael Keenan for exciting stories. Betelhem Negede for always having kind words. Banawe Anissa for being so thoughtful. Minjie Chen for always offering an interesting perspective. Thanks also to all I could not mention here. Having such a wonderful community of academics to interact with professionally and personally made work so much better.

All papers in this thesis rely on data collected in the field, so I must thank everyone who played a role in that regard. Special thanks then goes to Esther Mokuwa, Paul Richards, Emmanuel Kandate, Remi van Balen, Pim van de Langenberg, John Jusu, Francis B. Johnson and Sellu Kallon. They all helped me to better deal with being in the field and/or taught me so much about the countries I work in. Big thanks here also to all research assistants who actually did the hard work of conducting interviews in difficult conditions. And of course to all research participants who gave their time to help us better understand their lives.

Besides the professional support and help I received, I also owe thanks to all the great people who have been in my life during my academic journey. Their support also helped me complete this thesis.

Thanks to all my housemates that I had the pleasure of living with over the last ten years. Pim for the great debates. Nicky for the great parties we went to. Lia for all the cool stuff we cooked. Alex for leaving such a great painting in her room. Robbert for the fun games. Marcus for the hikes, talks and introducing me to the world of whisky. Melanie for inspiring me to travel more. Martina for all the laughs. Remi for all the discussions, bike rides and academic collaboration. Djuna for being exactly who she is. Rita for all her funny comments. Roos for taking us to a great festival. Ines for the great times cooking, hiking and laughing. Julius for the great bike rides. Wouter for his dry humor, fun debates, bike rides and hikes. Madeleine for being one of the best people I know. Alex for all our fun fermenting experiments. Linda for her delightful stubbornness. Jeroen for all his cool gadgets. Hannah for the graphic novels, walks, talks and for designing a fantastic cover. Marnix for his help in making ice cream. I also have to thank the furry companions I had the pleasure of sharing a house with. Shimon for being the platonic ideal of a cat. Wiebe for his crazy climbing antics. Kofi for being a big loveable furball. Willy for being absolutely crazy. Annie for being my greatest supporter.

I also want to thank all my friends who have had to deal with me undertaking this large project. Especially to Menno, my oldest friend, who always had inspiring new ideas to implement in my work. A big thanks also to Kassandra who always had good ideas and words of encouragement when I was stuck on something. Her role in my thesis was instrumental. Thank you to Pieter for introducing me to the world of Belgian beers. Thanks to Remi for being a part of almost my entire academic journey and being a great friend throughout.

Finally I want to thank my family for their love and support. To my mother for always supporting me and always checking in if I disappeared off the rader. To my father for his practical tips and great cycling holidays. To my sister for her love and support. To Leon for always having a funny comment. To my grandfather Koos for all his stories. To my grandmother Jaapje for always being interested. To my niece Lara for teaching me so much. To my niece Noa for having such a great laugh. To Michelle for always offering thoughtful suggestions. To Paul for all his life advice. And to all my aunts, uncles and cousins for always being a great group to spend time with.

I could not have written this thesis without the support of all of you.

Paul Hofman

Wageningen School of Social Sciences (WASS)

Completed Training and Supervision plan

Name of the learning activity	**Department / Institute**	**Year**	**ECTS***
A) Project related competences			
Writing of Research proposal	-	2016	6
Transparent and Open Social Science Research	University of California, Berkeley; CEGA; BITSS	2018	0.8
Programming and Data Management	Tinbergen Institute, Amsterdam	2018	3
Social and Economic Networks: Models and Analysis	Stanford University	2020	0.9
Evaluating Social Programs	MITx	2020	1.1
From Poverty to Prosperity: Understanding Economic Development	OxfordX	2020	0.6
CS50 Introduction to Artificial Intelligence with Python	HarvardX	2020	5
B) General research related competences			
Teaching Assistant: Economics	WUR	2016	1
Teaching Assistant: Cost-Benefit Analysis	WUR	2016	1
Scientific Writing 10	WUR	2017	1.8
Teaching Assistant: Introduction to Development Economics	WUR	2017	1
Teaching Assistant: Impact Assessment of Policies and Programmes	WUR	2018	1
Organising the DEC PhD Club	WUR	2018	1
Workshop on Transparency and Data Management	WUR	2018	2
Workshop on Experimental Methods	WUR / Nanjing University	2020	2
C) Career related competences/personal development			
Introduction course	WASS	2016	1
Social Networks and Technology Diffusion in the Congo	HICN-FAO/ISDC Workshop, Rome	2016	1
Social Networks and Community Targeting	WUR Economics CIDER Seminar	2017	1
Social Networks and Social Preferences: A Lab-in-Field Experiment in Eastern DRC	SEEDEC Conference, Wageningen	2018	1
Local Economy Effects of Large-Scale Agricultural Investments	CSAE annual conference, Oxford	2019	1
Total			**33.2**

*One credit according to ECTS is on average equivalent to 28 hours of study load

The research described in this thesis was made possible through the financial support of the Dutch Research Council's Global Challenges Program (# W08.250.104) and NWO grant (# 451-14-001). Financial support for chapters 2 and 3 comes from the Economic and Social Research Council (ESRC) Growth Research Programme (grant ES/J009008/1). Chapter 4 was partially suppported by Addax Biofuel (SL) Ltd under a contract guaranteeing the researchers' independence in carrying out the research and publishing the findings. Chapter 5 was supported by the NWO GCP grant (# W08.250.104). Support for chapter 6 comes from the university of Cambridge's Research England Global Challenges Research Fund QR Funding (JPAG/267).

Financial support from the Development Economics Group (WUR) for printing this thesis is gratefully acknowledged.

Cover design by Hannah Van den Broeck