Identifying communication spillovers in lab in the field experiments

Alexander Coutts[*][†]
Nova School of Business and Economics

October 15, 2018

Abstract

The use of lab in the field experiments has increased dramatically, given benefits of studying relevant populations. Conducted in environments where researchers must relinquish the control a standard laboratory offers, they raise the specter of communication from past to future participants, posing problems for inference. In rural villages participating in public goods games in Rwanda, I recover estimates of these spillovers by matching villages on all available pre-study observables, comparing those with and without communication opportunities. I find communication led to substantial unanticipated increases in cooperation, driven by conditional cooperators. I conclude with advice to manage potential bias from spillovers.

JEL classification: C92, C93, D83, H41, O10, O12.

Keywords: Lab Experiments, Public Goods Games, Field Experiments, Development, Rwanda, East Africa, Spillovers, Information Transmission, Social Learning.

---

[*]Nova School of Business and Economics, Faculdade de Economia, Campus de Carcavelos, Rua da Holanda, n.1, 2775-405 Carcavelos, Portugal; alexander.coutts@novasbe.pt

[†]Acknowledgements: I would like to thank Hunt Allcott and David Cesarini for providing constant guidance and expert advice for this paper. I am additionally grateful for helpful comments from Isaac Baley, Canh Thien Dang, Deborah Goldschmidt, Nicole Hildebrandt, Elliot Lipnowski, Molly Lipscomb, David Low, Joseph Mullins, Yaw Nyarko, Rossa O'Keeffe-O'Donovan, Giorgia Romagnoli, Andrew Schotter, Emilia Soldani, Tobias Salz, Pedro Vicente, Christopher Woolnough, and seminar participants at CSAE, New York University, and Nova SBE. I am indebted to Zachary Clemence and Kris Cox at Innovations for Poverty Action for their help with management and implementation, and to Lea Stuff for assistance running the experiments. All errors are my own.

1 Introduction

Laboratory experiments to measure preferences and behavior have become a standard component of the economist's toolkit, and now a significant number of experiments are conducted in field contexts: outside of a university classroom or computer lab, and with non-student populations, following the definition of Gneezy and Imas (2017). The utilization of such lab in the field experiments has increased dramatically in applied work in the social sciences. In development economics researchers are putting increasing weight on the importance of understanding relationships between preferences and development outcomes.[1] Such work is increasingly finding outlets in leading journals of economics and political science.[2]

The internal validity of such studies hinges on unbiased identification of preferences, which among other things, requires thoughtful experimental design. There is significant awareness of the effects of individual learning both within and across games, however there has been significantly less study of social learning in these contexts: communication across participants that affects decision making.[3] Failure to account for spillovers from such communication in experiments can bias the identification of preferences.

To a certain extent, the fact that individuals may communicate socially and learn from one another has always been possible in standard lab experiments. A typical participant in an economics lab experiment is often a student, and may have friends or colleagues who previously participated in a particular experiment. Thus, it is conceivable that these students may discuss outcomes or strategies of a particular experiment with friends who are about to participate.

However, many experimental labs at universities are in fact less susceptible to this issue than corresponding field settings, such as rural Rwanda, the setting for this paper. First, many experimental labs have multiple experiments occurring in any given week, and often subjects are not aware which experiment they will be participating in. Second, economics experiments are not typically considered highly noteworthy events in a subject's semester of study. Comparatively, for participants in locations such as rural Rwanda,

[1] See Gneezy and Imas (2017) for a discussion of the definition of lab in the field.

[2] Recent examples include Avdeenko and Gilligan (2015), Casaburi and Macchiavello (2018), Enos and Gidron (2018) , Jakiela and Ozier (2015), and Kosfeld and Rustagi (2015). Use of lab in the field experiments is also on the rise in other fields such as sociology, see Baldassarri (2015).

[3] One can many discussions of learning from repeated play in various games, a good starting point is Kagel and Roth's *Handbook of Experimental Economics*. See Bednar et al. (2012) for learning across games.

these events are perceived as being more out of the ordinary. Beyond this, financial stakes can be substantially greater in lower income countries; in the current experiment average earnings were greater than a typical full day of earnings. Combined, these reasons hint that communication may be a bigger issue in field contexts.

I study one such large implementation of a lab in the field experiment in Rwanda involving the participation of 150 rural villages which was implemented over a period of three months. Theoretically, communication was not expected to alter behavior, yet observations from the field suggested otherwise. In one instance, the survey team visited a village that appeared similar to others in the region. Standard protocol was followed, however in this specific village, all (12) participants contributed the maximum possible amount in the public goods game. Because this was so exceptional, the team stayed behind to ask the villagers what led to such high levels of cooperation. A woman explained that she was friends with some of the women in a neighboring village, and one of her friends had participated in the same game only two days prior. Her friend told her that she should contribute the maximum amount, and she had shared this information amongst these villagers before the team had arrived.

To uncover unbiased estimates of this social learning, I utilize propensity score matching, which is well suited for this context. The reason is that in planning the order of visits of the 150 villages, the logistical planner for this study only had access to very few village characteristics, i.e. his ordering could only be conditioned on a small and finite set of observables. The matching strategy splits villages according to whether or not they had opportunities to communicate with past participants, using timing and GPS data, and then matches them on this key set of observables. The result is that one can compare villages which appear ex-ante identical to the planner, but for idiosyncratic reasons some were treated, i.e. they had neighbors which previously participated, and others were not. Having the full set of conditioning variables automatically fulfills the key selection on observables assumption underlying matching techniques, that treatment is independent of outcomes, conditional on observables. This in fact takes all the guesswork out of which variables need be included in the propensity score.

Implementing this matching strategy I indeed find large effects of communication. I find that being located near past participating villages increased contribution rates in the entire sample by 11-14% depending on the matching estimators used. Beyond this, in placebo tests, I consider how exceptional it would be to observe estimates of this magnitude for different counterfactual orderings of villages. The estimates of this paper exceed 99%

of 10,000 simulated random orderings, and additionally always exceed those based on hypothetical deterministic orderings based on available observables such as the village's distance to the study base.

While the evidence strongly suggests that communication had large impacts on cooperation, it is more difficult to identify the precise mechanism. One plausible explanation involves communication shifting beliefs about levels of cooperation upward, which would increase the contributions of so-called conditional cooperators: those who increase their contributions when they expect others to do the same. For example, Chaudhuri et al. (2006) find that providing future participants with advice from previous participants in a public goods lab experiment increases average contributions. A similar mechanism appears to be at work in the field data: there is evidence that the treatment effect is driven entirely by those who are identified as conditional cooperators in the sample.

Evidence on unstructured social learning in lab or lab in the field games is extremely rare. To my knowledge only Cardenas and Carpenter (2005) and Bernal et al. (2016) provide related evidence, from common pool resource lab in the field games in Colombia, where they conduct follow-up games with the same communities some months later. A critical difference is that these followup games included past participants, making up 30% of participants in Cardenas and Carpenter (2005) and 86% of participants in Bernal et al. (2016).[4] While their focus is different from this paper, on the dynamics of learning within groups, the results are consistent: villages became more cooperative in the second visit. Importantly, Cardenas and Carpenter (2005) find that the results are not simply driven by past participants - novice participants are more likely to contribute more in the second visit compared with novice participants in the first visit.[5]

The result that communication can have important impacts on behavior in lab in the field experiments has important implications for the design of experiments as well as the broader interpretation of results. The finding is especially significant, since these considerations exist in a context where standard theory makes the prediction that social information should not change behavior in public goods games.[6] Randomized control trials

[4]Cardenas and Carpenter (2005) visited three villages, and the number of past participants varied substantially from 6% to 67%.

[5]Both papers highlight the potential pedagogical effect of the experimental games, which could therefore have positively affected cooperative norms. These may be particularly salient in their setting, as in both papers the groups were selected specifically because common pool resources were important within the communities. Importantly, the mechanisms outlined in this paper cannot be excluded as alternative explanations.

[6]Standard theory posits a role for learning, in which case the learning should be about the Nash Equi-

that recognize the importance of spillovers through social learning may nonetheless fail to account for such spillovers when ex-ante theoretical predictions rule it out. If researchers fail to account for social learning this can weaken the external validity of the study.[7]

Traditionally, less attention has been given to social learning or spillovers in lab in the field experiments when compared with broader field experiments. This needs to change, and the results and discussion of this paper lead to a number of practical suggestions for researchers either designing or analyzing projects involving lab in the field data. The first is that care needs to be taken in design, as spillovers may arise even when theoretically unanticipated. The second is that techniques such as ex-ante randomization within geographical regions can be applied to generate ex-post tests for whether spillovers have occurred. And finally, attention must be paid to logistical planning. Even if logistics cannot be altered, the implementation plan can be key to uncovering estimates of potential spillovers, following the methodology of this paper.

To summarize the remainder of this paper, the next section outlines details of the public goods games and the data. This is followed by the matching analysis that demonstrates the effects of social learning on behavior. I next spend time examining the role of conditional cooperators, and discussing potential alternative mechanisms, followed by a concluding discussion.

---

librium strategy to contribute zero, the opposite effect to what is observed.

[7]Consider an example of a randomized controlled trial, where a new product (e.g. bed net) is being sold at a discount to households. Suppose individuals in the study area happen to communicate extensively. Those that have purchased these nets may communicate with others, and this may alter these individuals' willingness to purchase. For example they might give advice, "I slept under this net, and it was (un)comfortable. I think you should (not) purchase it." The external validity of such a study is compromised, unless information environments are similar. If part of the success (or failure) of a treatment reflects the role of advice, then in a different setting where communication is less prevalent, these effects may be substantially different. For example, Dupas (2009) conducts a field experiment with the purchase and usage of insecticide treated nets (ITNs) in Kenya. Randomly selected households were visited and given a voucher to purchase an ITN, over a period from April to October 2007. One possibility is that households visited early in the project gave advice to friends or family that could be visited later. In fact, in the data from Dupas (2009), being visited one month later in the study increases probability of purchase by approximately 5% (author's calculations). In a follow-up study Dupas (2014) does in fact find that being in an area where there was a high density of treated households had a positive effect on the likelihood of purchasing an ITN.

2 Implementation and Theory of the Public Goods Games

The experimental games were conducted as a component of a broader impact evaluation of community health programs.[8] In parallel with household surveys for that evaluation, public goods experiments were conducted in 150 villages in the Rusizi district in Rwanda. These villages were chosen randomly from a total of 598 villages in the district. For the purposes of this paper, the evaluation of the community health programs is not relevant.

The experimental games were conducted over a three month period in 2013. All 150 villages that were part of the larger evaluation participated in these games. 12 individuals were randomly selected from the household survey list, and given a ticket to participate in the games the following working day. At the time of the games, the 12 individuals were checked-in by the survey team and completed a brief questionnaire.[9] Local survey staff then explained the game in the local language of Kinyarwanda. A significant amount of time was spent explaining the game, including the tradeoffs between private and public benefits, providing a demonstration, and conducting a full practice session. It was important that individual decisions were completely private and anonymous; at no time were individual contributions revealed, a fact emphasized to participants.

The experimental design followed a standard public goods game format. Individuals were given an endowment of 4 x 100 RWF coins.[10] They were given real money to ensure that the stakes were salient, and to minimize confusion. One by one participants were instructed to leave the room, go to a completely private area, and decide how much to contribute to the common fund by depositing this in a small change purse, henceforth referred to as the contribution purse. The remainder of their endowment was kept on their person. This had the added advantage of making it clear for individuals that the money they kept was theirs. Since all individuals carry money on their person, there was no reason for concern about having decisions accidentally revealed.

After allocating their money, the participant would then place the contribution purse in a designated location. After all 12 participants had made their decisions, each individual

[8] The official project name is: Impact Evaluation of Community-Based Health Programs in Rwanda (CBEHPP); ClinicalTrials.gov Identifier: NCT01836731; PI: James Habyarimana, Georgetown University.

[9] Eight individuals were randomly selected for a wait list, in case individuals did not show up at the specified time.

[10] At the time of the study 400 RWF was approximately 0.60 USD. From the Integrated Household Living Conditions Survey 2010-2011, 400 RWF comprises of more than an average day's income for 45% of the district population. Earnings in the experiment were larger than 1800 RWF on average, which greatly exceeds a day's income for the majority of the population.

amount was recorded, using anonymous ID numbers located inside the contribution purse, to prevent identification of individuals by the survey team. After recording, all the purses were emptied publicly one by one, and in a transparent manner the coins were counted, tripled, and divided equally among all 12 participants.[11] Subjects played two rounds of the public goods game with real stakes, receiving income directly after each round. The second round consisted of one of four different versions of the game.[12] Subjects were aware that there would be a second round, but were not given any information about the specific variation that would be used, ensuring comparability across villages. For this reason only the first round is used for the primary analysis of the paper, while the second round will be considered when examining the role of conditional cooperators.

Individual payoffs v_i were thus:

$$\begin{aligned} v_i &= 400 - c_i + \frac{3}{12} \cdot \sum_{j=1}^{12} c_j \\ &= 400 - \frac{3}{4} c_i + \frac{1}{4} \cdot \sum_{j \neq i} c_j \end{aligned} \tag{1}$$

Notice that contributing $c_i = 0$ is the unique optimal strategy absent social preferences, with the marginal per capita return (MPCR) to cooperation set at 0.25. A long history of public goods experiments have shown that individuals tend to contribute non-negligible positive amounts in these type of games, despite the dominant incentives to free-ride. Chaudhuri (2011) provides a detailed overview of cooperation in public goods experiments, and places particular emphasis on the presence of individuals that play as conditional cooperators (CCs). CCs are so named because they contribute positive amounts in public goods games conditional on their beliefs about what others will be contributing. In their survey of the literature they find that CCs make up the largest group within experiments, representing between 35% to 81% of subjects.

CCs are important to be aware of as their optimal contribution depends on their beliefs about what the contributions of others will be, which will be an important potential mechanism for the present paper. Since this paper examines the potential effects of communication across participants in the public goods games, it is important to briefly review

[11] When necessary, the amount was rounded to the nearest 50 RWF.

[12] The four versions included the baseline game repeated, one game with the ability to punish, one with the ability to reward, and a unique game meant to measure uncertainty in public goods investment.

the evidence of how such communication could alter contribution decisions.

Chaudhuri et al. (2006) showed that CCs presented with public advice upwardly adjust their beliefs about average contributions, and thus increase their own contributions. They find that a substantial fraction of past participants advise future participants to contribute the full amount. When this advice is common knowledge, it has a significant impact, increasing contributions in the first round of play by over 18%.

Isaac and Walker (1988a) find that pre-game communication leads to higher contributions in public goods games.[13] Importantly, when decisions are private, CCs are essential for pre-game communication to have any impact on behavior. Because such communication is cheap-talk, any promises or announcements are non-binding. Hence of all the observed behavioral types of participants in public goods games, see Chaudhuri (2011), only CCs would be responsive to such communication, if it can successfully shift beliefs.[14]

To summarize, CCs make up the largest proportion of subjects in lab experiments and there is evidence that they are likely to be affected by either cheap talk or advice. Hearing a promise from an individual about contributing a large amount may lead one to increase their expectation of what that individual will contribute; receiving advice to contribute a large amount, and knowing that others have heard the same advice may have a similar effect. In order to identify CCs, in this paper I will use data on contributions in the second, final round, which occurred after participants had already observed first round contributions. Identifying CCs will be important when it comes to understanding the mechanisms at play.

3 Overview and Empirical Strategy

3.1 Overview of the Games

Table 1 presents summary statistics of village level variables. The sample of participants was randomly selected from the population in the larger survey which included one person from *every* household in a selected village with at least one child under the age of 5. Women

[13]There is a sizeable literature on the effects of pre-game communication or cheap-talk in experiments. Sally (1995) examined the results of over 30 years of public goods experiments and found that the effects pre-play communication on contributions are indeed a robust finding in the literature. Crawford (1998) additionally surveys a number of experiments with cheap talk. More recently Bochet and Putterman (2009) and Brosig et al. (2003) have examined communication in public goods games.

[14]In the summary of Chaudhuri (2011), CCs are the most common type, followed by free-riders, then unconditional cooperators, and finally others who cannot be classified.

were over-represented in the larger survey, and subsequently are 74% of the participants in the experimental games.[15] The average years of completed education is 4.5, which corresponds to partially completing primary school, while the average age was 35.

Additionally, subjects were asked how many of the other 12 participants they knew participating in the current session, the average number known was approximately 2.5. Of the 150 villages visited, in only two villages were we unable to find the full 12 participants. These villages have been dropped from the analysis. The remaining sample is of 147 villages, with one additional village being unable to be matched.

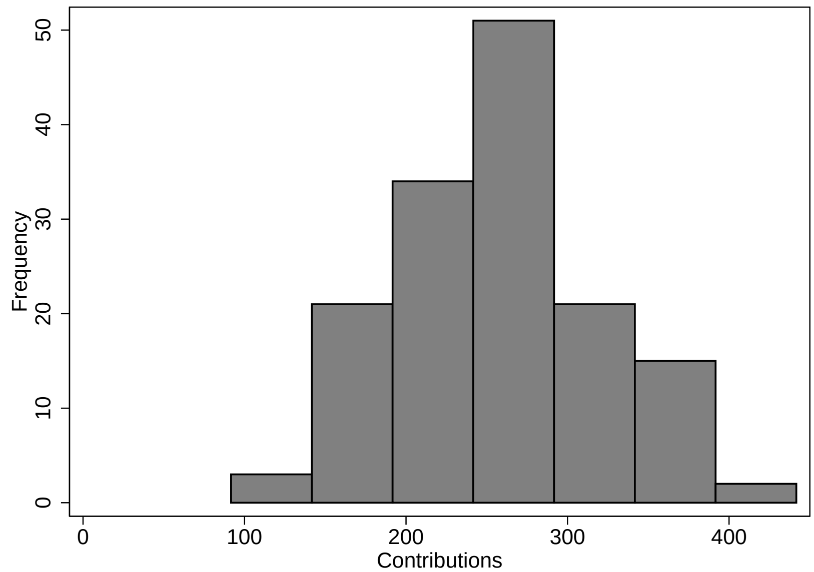
[15]Women were over-represented in this larger survey because the respondent was not randomly selected from within the household. Of note is that the gender ratio in Rusizi district is highly skewed, so the over-representation is not as large as it appears. According to the 2013 Rusizi District Gender Statistics Report, the percentage of females aged 25-64 is 55.4%. This age group represents 89% of those participating in this study.

Table 1: Summary of Village Level Variables

	Mean	Std. Dev.	Min.	Max.
Average Contribution	254.92	65.89	91.7	400.0
Average Contribution (Round 2)	247.58	69.33	75.0	400.0
Proportion Female	0.74	0.14	0.4	1.0
Average Age	35.39	3.38	27.0	45.3
Average Years of Education	4.47	1.29	0.8	8.5
Average Number current Participants Known	2.56	1.09	0.7	6.8
Community Cooperation Index	0.81	0.13	0.3	1.0
Community Effort Index	0.79	0.16	0.3	1.0
General Trust Index	0.73	0.15	0.3	1.0
Distance to base (km)	13.14	7.98	1.0	38.8
Village Size (# HHs)	131.68	42.84	43.0	345.0
# Villages $\leq$ 1.75 km	1.86	1.33	0.0	5.0
Distance to paved road (km)	4.73	3.81	0.0	14.1
Observations	147			

Figure 1 presents the distribution of contributions in the public goods games. The possible levels ranged from 0 to 400 RWF, in 100 RWF increments. The average level contributed to the group fund is 255 RWF, which is about 64% of the socially optimal level of contributing the maximum 400 RWF. As is typical in public goods experiments, the standard theoretical prediction of contributing zero is rejected.

Figure 1: Distribution of Contributions (RWF)



Contributions in the public goods experiments. Possible values ranged from 0 to 400 RWF, in 100 RWF increments. N = 147.

The 64% contribution rate is on the higher end of contributions in experimental public goods games. Typically, contribution rates range between 40-60%, though with a range of different MPCR. Another difference between these results and previous experiments is that the proportion of "free-riders" or those contributing nothing, is lower than commonly found.[16]

3.2 Strategy to Identify Communication Spillovers

The identification of the effects of communication on cooperation involves a matching strategy, which pairs otherwise similar villages which only differed on whether there existed opportunities for such social learning to occur. In identifying the effects of social learning on contribution levels, the primary threat to identification is that the order of visits was not explicitly randomized. In practice, a study planner observed some available characteristics

[16] See Ledyard (1995) and Chaudhuri (2011). A MPCR of 0.25 is on the lower end, suggesting that the observed contribution rates are indeed quite high, see Isaac and Walker (1988b).

of the 150 villages that were in the study, and had to determine an ordering.[17] While not random, there was also no explicit strategy conditioned on these observables. Figure 2 presents a graphical representation of the order of visits, with villages shaded progressively darker to indicate later dates of visit.

Figure 2: Location and Timing of Village Visits

Each circle represents one village. The shading constitutes the date of visit, with lighter circles representing earlier visits in the study, and darker later visits. The size of the circle corresponds to the average amount contributed by that particular village: larger circles correspond to greater average contributions.

The primary concern of the planner was logistical convenience, as well as ensuring "difficult" villages were spread evenly throughout the study. Difficult villages were those that were located far from the study's base location, and/or those that had large numbers of

[17]The planner was a local who had been involved in supervising an earlier census which tallied the number of households per village.

households, which required longer working hours.[18] This in fact helps with the identification strategy, since it helps to re-balance these characteristics among villages visited earlier and later, which will correlate with opportunities for potential social learning.

The key strategy is to exploit the exogenous variation in the planner's decision making, to find otherwise identical looking villages, based on all observables available to the planner, but by chance some had neighbors who previously participated in the public goods games, while others had no such neighbors. Those with previously participating neighbors thus had potential opportunities to communicate with past participants, while those without did not. In the analysis it will also be important to control for total number of neighbors, this is discussed in further detail in the upcoming analysis.

For this estimation, it is most useful to consider an exercise where one frames this in terms of the treatment effects literature. In particular, certain villages are "treated" with exposure to previous participants of the public goods games, while others are not. This setting is particularly amenable to propensity score matching techniques, see Rosenbaum and Rubin (1983), in order to recover causal effects of communication on behavior. The reason is that the order of visit could only be conditioned on observables known to the planner, *before* the games were conducted. Thus one can make use of this full set of observables to generate propensity scores for the "treatment": having neighbors who previously participated. By matching treatment and control villages with similar values of the propensity score, i.e. villages who are similar based on all observables available to the planner, one is able to recover causal estimates of treatment effects.

The key assumption in using matching techniques is the selection on observables assumption, i.e. that after controlling for observables, treatment assignment is independent of the outcome of interest: contributions in the public goods game. In most empirical studies, this assumption is difficult to satisfy, since often the treatment of interest, e.g. job training, is self-selected into by individuals, and the econometrician may not observe all relevant variables that determine this selection. In the case of the present paper, the selection on observables assumption is satisfied automatically, as I have the small but complete set of possible conditioning variables.

[18] All staff and enumerators were based in a central location. There were no overnight stays during the study, with the exception of 1 out of 18 political sectors. Removing this sector does not alter the results. The rationale behind spreading difficult villages evenly across the study was to minimize staff fatigue and ensure reasonable daily working hours.

3.3 Defining the Treatment

The treatment of interest for this study is whether a village had opportunities for communication with past participating village neighbors. The treatment is not the effect of communication directly, since this is not observed. Two villages are defined as neighbors if they are within 1.75 kilometers of one another. This specific distance is chosen as it reasonably captures opportunities for cross-village communication, but additionally because it creates balance between villages with neighbors, and those without. As will be shown, this is important for power in the matching strategy.

However, since this research question and distance was *not* identified prior to the study, it is important to demonstrate that the results are consistent for other distances, and that 1.75 kilometers was not chosen after an ex-post comparison of different distances. To demonstrate this, Appendix Section 5.4 shows robustness checks which verify that the results are broadly consistent for other distances. Moreover, all of the results of this paper are presented for the distance of 2 km in Online Appendix Section A, which was the initially chosen distance for this paper. The results remain significant and similar in magnitude, though some of the matching estimates are noisier, due to the greater levels of imbalance and fewer number of observations.

There is a further dimension through which treatment definition may be interpreted differently, and that is the threshold for the number of neighboring villages. The primary concern is in identifying reasonable opportunities for communication. Using one neighbor is the most intuitive starting point. Moreover, it is important to note that choosing the threshold incorrectly should only work to bias the estimates downwards *against* finding effects. Choosing too low a threshold runs the risk of downward bias due to the inclusion of villages which didn't have opportunities to communicate in the treatment group. On the other hand, choosing too high a threshold runs the risk of downward bias due to control villages having opportunities to communicate.[19]

According to the treatment classification of having at least one neighboring village that previously participated in the games, 74 out of 147 of villages are not-treated, while the remaining 73 are treated, resulting in a balanced 50% distribution, which will be important

[19]There is also an element of timing. For example, one could further restrict the definition of treatment to only apply to villages that had neighboring participants in the last week or month. It is not clear what the relationship should be between information and time. On one hand, more time might allow information to be disseminated across villages. On the other hand, more time could allow information to deteriorate. As the relationship may be non-monotonic, I remain agnostic by not conditioning on time.

for the matching strategy. Figure 3 shows the number of villages in control and treatment groups, broken down into exactly how many past participating neighboring villages they had. Of note is that the vast majority (88%) of villages have 1 or fewer neighbours that participated. The maximum number of neighbours was 5, this occurred for only one village in the study.

Figure 3: Defining the Treatment

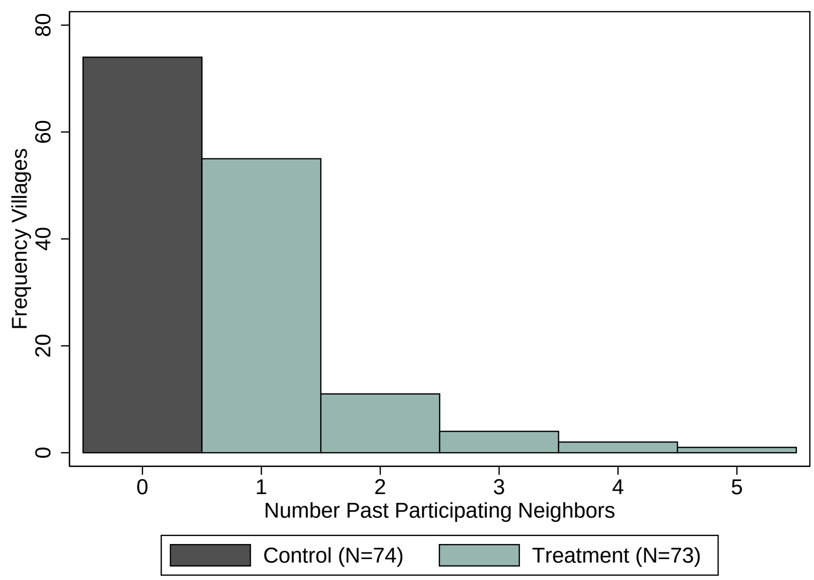


Figure shows the frequency of observing villages in the sample with given number of past participating villages within 1.75 kilometers as neighbors.

3.4 Determinants of visit order

The key variables available to the planner were the following. Note that the planner additionally knew the locations of these villages, however this will be used only indirectly in the matching strategy.[20] With regards to political sectors, Figure 5 in the appendix presents a map of these sectors in Rusizi district.

[20]The reason is that while ideally villages would be matched on distance from one another, this is not directly possible in the matching strategy. Moreover, distance to the base location and to a paved road will already convey much of this information, as will the addition of political sectors. Beyond this, adding latitude and longitude into matching may mis-characterize similarities (e.g. villages which share the same longitude but differ in their latitude or vice-versa).

1. Distance to village (from study base location).
2. Distance to paved road.
3. Number of households in the village.
4. Number of total villages in the study located within 1.75 km (village density)
5. Sector (political region; there are 18 sectors in the district).

Note that some of these variables contain information which may be used in ways that are difficult to capture in simple linear regressions, particularly regarding village locations. However, Table 8 in the Appendix examines such a regression of these variables on the order of visits. It can be seen that a simple regression with the four variables mentioned above entering linearly and sector fixed effects accounts for 96% of the variation in visit order. Thus one can conclude that a simple linear weighting of these variables is sufficient to capture nearly all of the decision making which uses these variables by the planner.

3.5 Summary Statistics on Treatment and Balance

Table 2 examines a logit regression village characteristics on the treatment indicator for potential communication, i.e. 1 if the village had at least one neighboring village within 1.75 km which previously participated in the public goods games, and 0 otherwise. This provides an overview of which variables are important for the determination of treatment.

From Table 2 it is possible to see that among the variables that the planner had access to, the only significant variable is the number of total neighboring villages within 1.75 km, independent of date of participation, i.e. the village density. This is not surprising, since villages with more neighbors that participated in the sample at any point in time are mechanically more likely to have neighbours at an earlier point in time.

Given this relationship, an issue could potentially arise if differences in contributions between treatment and control arise because of differences in village density. There are a number of ways I will address this possible issue with identification. First, one thing to note is that a small proportion, 16% of villages, had no neighbors within 1.75 km. Mechanically, these villages cannot be in the treatment group. As such, I remove them from the matching analysis.[21]

[21] If one includes these villages, 88% of them end up being removed anyways to due extremely low propensity score values that fall outside of the common support. As such, including them does not alter the main results of this paper.

For the most part, matching will alleviate this issue as only similar villages across treated and control groups will be matched, and number of neighbors is a key variable in the matching estimation. Beyond this, in Appendix Section 5.4.3 I examine exact matching of villages by density, i.e. comparing only villages with the same number of neighbors within 1.75 km in the study, and show that the results continue to strongly hold. As an additional sanity check there is no statistically significant relationship between contributions in the public goods games and the number of sampled villages within 1.75 km.[22]

Table 2: Logit regression for treatment: Past participating neighbors within 1.75 km

Logit Regression	**Treatment**
	(1)
Distance to base (km)	0.041
	(0.110)
Village Size (# HHs)	-0.005
	(0.006)
# Villages $\leq$ 1.75 km	1.084***
	(0.219)
Distance to paved road (km)	0.059
	(0.153)
Observations	143
Sector Fixed Effects	✓

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parenthesis.

Next Table 3 examines the balance across treated villages: those with at least one neighboring village which previously participated in the public goods games, and control villages. Only one statistically significant imbalance is found for variables which the planner had available, namely the number of total study villages located within the 1.75 km radius, echoing the earlier result. Distances to the study's base location or to the nearest paved road are not significantly different across the groups, nor is the village size. Regarding the variables unavailable to the planner, the only statistically significant variable is in fact the main outcome of interest of this paper, village average contributions in the public goods game. In particular contributions are 30 RWF *greater* in treatment villages, i.e. villages

[22]They are positively correlated. The p-value on a test for whether village density can explain contributions is 0.165. Adding the treatment dummy reduces this p-value to 0.868. This relationship is investigated further using OLS regressions in Online Appendix Table C2. One additional reason this is less of an issue than first appearances may suggest is that the variable is defined as neighboring villages *in the study*. Since only 150 villages out of 598 participated in the study, the variable itself is only correlated with the actual number of neighbors.

Table 3: Balance of treatment: Past participating neighbors within 1.75 km

	Treatment	Control	Difference
Available to planner			
Distance to base (km)	12.17	14.09	-1.92
Village Size (# HHs)	128.65	134.64	-5.98
# Villages $\leq$ 1.75 km	2.48	1.24	1.24***
Distance to paved road (km)	4.96	4.50	0.45
Unavailable to planner			
Average Contribution	270.22	239.83	30.40***
Proportion Female	0.74	0.74	0.01
Average Age	35.39	35.38	0.00
Average Years of Education	4.53	4.40	0.13
Community Cooperation Index	0.81	0.82	-0.01
Community Effort Index	0.79	0.80	-0.01
Average Number current Participants Known	2.53	2.60	-0.07
General Trust Index	0.72	0.74	-0.02
Observations	73	74	147

which had neighbors that were past participants.

Since treatment and control villages are similar on all observable characteristics related to demographics and preferences, one might in fact interpret the difference in contributions as the unbiased treatment effect, that is the true impact of opportunities for communication on cooperation. However, it will be important to ensure that potential differences in selection of visit order, based on the variables observable by the planner, are controlled for in a more systematic way. The next section presents the matching strategy.

3.6 Propensity Score Matching

This setup is well suited to matching strategies, due to the fact that all observables that the treatment could have been conditioned on are available in the data. While control and treatment villages may on average have differed on some characteristics, matching enables one to compare similar groups of villages who either received or did not receive the treatment respectively.

Here I follow the notation of Imbens and Rubin (2015), with some slight adaptations. Let $C_i(1)$ denote the outcome of interest, village level contributions, if village i had at least one neighboring village within 1.75 km which previously participated in the games (treated, $W_i = 1$), and $C_i(0)$ be the contribution of a village with no previous neighbours (untreated, $W_i = 0$). In an ideal world, one could observe both outcomes (treated and untreated) for the same village, and hence could calculate the average treatment effect τ. In the real world, the classic problem is that one cannot obtain an unbiased estimate of the treatment effect by naive comparison of the average outcomes of the two groups ($\bar{\tau} = \bar{C}(1) - \bar{C}(0)$) because these groups may have different characteristics.

In practice, randomization can solve this problem, by creating comparable treatment and control groups. Here, randomization did not occur. Instead, following Rosenbaum and Rubin (1983) and a number of others, the strategy is to find a set of observable covariates X, which are known to be not affected by the treatment, such that:

$$W_i \perp C_i(1), C_i(0) | X_i. \tag{2}$$

This assumption is referred to as unconfoundedness or selection on observables. It means that the outcomes are uncorrelated with treatment, conditional on covariates X_i. In the current context this assumption is likely to be satisfied. The reason is that, unlike most observational studies, the treatment W_i (being exposed to villages who previously participated) could only have been conditioned on observables. This is because, as stated earlier, the planner determined the order of visits, in advance, with a limited number of pre-visit observables. In particular, it would be impossible for the planner to condition the treatment on features of data which had not yet been collected.

Denote the propensity score, $e(x)$ by:

$$e(x) = Pr(W_i = 1 | X_i = x), \tag{3}$$

i.e. the probability that a village receives the treatment conditional on having characteristics $X_i = x$. This is also equivalent to the expectation of the treatment, $\mathbb{E}[W_i = 1 | X_i = x]$.

We can thus define the average treatment effect as:

$$\tau = \mathbb{E}[\mathbb{E}[C_i|W_i = 1, X_i] - \mathbb{E}[C_i|W_i = 0|X_i]] \tag{4}$$

As Imbens and Rubin (2015) note, in addition to unconfoundedness, a second key assumption is required for the analysis. This involves a requirement that there is overlap in the distribution of covariates across treatment and control villages. Intuitively speaking, one needs to be able to find similar villages in control and treatment groups, in order to make valid comparisons.

Regarding this second assumption of overlap, as noted earlier, in Table 3, average characteristics on variables observed by the planner are reasonably balanced across treatment and control villages. As the next section will show in more detail, the assumption of overlap is broadly supported in the data.

3.7 Estimating the Propensity Score

The propensity score needs to be estimated from the covariates which may potentially have had an impact on which villages received the treatment (having neighbors that previously participated). In the case of this study, these variables can only come from the set of all observables available to the planner at the time the order of visits was determined. In determining the propensity score, I do not include sector dummies, as sectors will be conditioned on using an exact matching strategy, which is detailed in the next section.

Regarding the key variables available to the planner, outlined in section 3.4, one cannot assume that the planner used these variables in a linear way. Thus it is also important to take into account potential higher order interactions between these variables and the treatment. To determine the optimal specification, I follow the algorithm outlined in Imbens and Rubin (2015), which involves selection of these higher order terms based on their added value in terms of predicting treatment assignment. The algorithm involves step-wise regression estimation of the propensity score, to select only those covariates that add value in determining treatment status.[23]

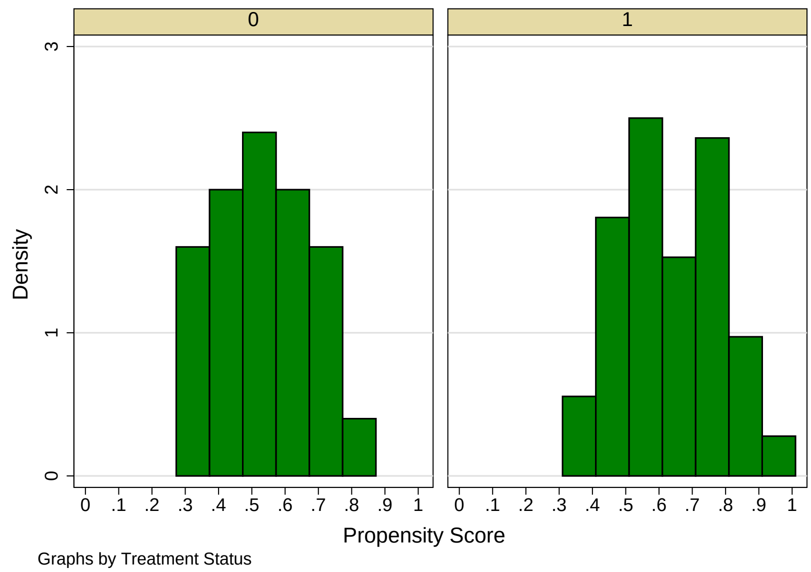
In fact, the algorithm does not select any additional higher order terms. Thus the final terms selected for estimation of the propensity score are solely the four variables corresponding to distance from base, distance from paved road, number of households in

[23] I follow Imbens and Rubin (2015), and set the threshold value for second order terms to be $C_{qua} = 2.71$.

village, and village density within 1.75 km.

Figure 4 presents the distribution of the propensity score by treatment status. There is significant overlap over the sample, with the exception of values close to 1.[24] In the analysis I will impose restrictions that matching must occur in regions with common support.

Figure 4: Propensity Scores by Treatment Status



Distribution of propensity score. $N = 123$.

3.8 Results on Average Treatment Effect (ATE)

I focus on two main specifications for computing average treatment effects. Both include in the calculation of the propensity score the four key variables that the planner had available. The first specification does not utilize the 18 political sectors in the propensity score, in order to not dramatically increase the dimensionality. To account for the possibility that matching on sectors is important, the second specification forces exact matching on these political sectors. That is, I require that in addition to villages being similar across treatment and control according to available variables to the planner, I also require that these villages

[24] If one were to include villages with no neighbors, this would also present additional imbalances in the left tail.

be located in the same political sector.

Table 4 presents these two main empirical specifications. The first column presents standard matching estimates using two neighbors, with replacement. Alternative estimates for 1 or 3 neighbors are found in Appendix Table 9, with similar results. The second column presents exact matching on political sectors, again with matching on two neighbors. The average treatment effect ranges from 28.4 to 36.9 RWF, each significant at the 1% level. These results are quite similar in magnitude, despite the specification differences, and correspond to a 11-14% increase in contributions over the entire sample.

Each approach has its advantages. Appendix Figure 9 presents the balance of covariates after matching. Covariates are balanced quite evenly across treatment and control after matching. In column 2 where exact matching on sectors is implemented, balance is only marginally affected. However, this comes at a cost of observations, as some sectors have few villages in either treatment or control groups. Regardless, it is re-assuring that both estimates are similar in magnitude.

In the Appendix, Section 5.4 examines these specifications for different treatment definitions that involve different distances. There Figures 6 and 7 present graphically ATE estimates for distances between 1 km and 3 km. The results are broadly consistent, with a slight pattern of shorter distances being associated with larger treatment effects, though there also arise issues of sample size, due to few treatment villages when distances are shorter, and few control villages when distances are longer. As noted, 1.75 km corresponds to the distance where treatment and control groups are most balanced.

Table 4: ATE of Presence of Past Participating Neighbors Within 1.75 km

	(1) Standard Matching	(2) Exact Matching
Contribution	28.368***	36.868***
	(10.677)	(11.332)
Observations	117	101

Analysis uses nearest neighbor propensity score matching, with 2 neighbors, with replacement. Significantly different from zero at * 0.1; ** 0.05; *** 0.01. Abadie-Imbens Robust Standard Errors in parentheses. Values of propensity score outside common support range are dropped. Exact matching excludes sectors with only 0 or 1 village in either treatment or control groups.

3.9 OLS

It is also useful to examine an OLS specification, particularly as the matching strategy reduces the number of observations quite substantially. Table 5 presents a simple OLS specification of the impact of having neighboring villages within 1.75 km who previously participated on contributions. It is possible to see that the impact is positive and significant, consistent with the matching estimates. The coefficient is approximately 30 RWF, a 12% increase from average contributions. Of note is that the estimate hardly varies at all when additional controls are added, including sector fixed effects. This is consistent with earlier evidence from Table 3 which demonstrated that the treatment was not significantly correlated with most observables. Beyond this, it is reassuring that the only variable significantly associated with treatment, the number of neighboring villages within 1.75 km, is not significant in the regression. Taken together, these results further suggest that treatment is indeed exogenous. Beyond this, the results obtained are quantitatively similar to those obtained using matching.

In addition to these results, Online Appendix Section B considers alternative dependent variables which capture opportunities for communication with past participants. These variables are (i) the order a village was visited, within its political sector; and (ii) the number of villages in the sector which previously participated. Both are found to statistically significantly increase contributions, in line with the results found here. Additionally the Online Appendix Section C considers placebo style regressions for the number of total villages within the sector, as well as the density: number of villages within 1.75 km. Reassuringly, no statistically significant associations are found.

Table 5: Effect of Presence of Past Participating Neighbors Within 1.75 km

	(1)	(2)	(3)
Treatment Status	30.395***	30.876***	29.773**
	(10.626)	(10.604)	(12.032)
Distance to base (km)		−0.133	−1.312
		(0.850)	(3.188)
Distance to paved road (km)		3.771**	−0.307
		(1.786)	(4.520)
Village Size (# HHs)		0.001	0.094
		(0.115)	(0.142)
# Villages $\leq$ 1.75 km		−0.846	2.298
		(4.770)	(5.676)
Years of Education		0.280	−3.146
		(5.502)	(6.881)
Female		91.854**	76.834*
		(40.884)	(45.056)
Age		10.552	15.083
		(26.628)	(25.444)
Age^2		−0.135	−0.206
		(0.368)	(0.354)
Controls		✓	✓
Sector Fixed Effects			✓
R^2	0.05	0.20	0.30
Observations	147	146	146

Analysis uses OLS regression. Significantly different from 0 at * 0.1; ** 0.05; *** 0.01. Robust standard errors in parentheses. Controls includes all remaining variables found in Table 3.

3.10 Counterfactual Planner

The results point to statistically significant sizeable effects of communication from past to future participants on contributions. Further, it appears that based on observables not available to the planner, villages that had neighbors which previously participated do not appear different from those that did not. Beyond this, matching and OLS strategies generate estimates which are similar in magnitude, suggesting that initial imbalances in the variables observed by the planner (namely the density of villages) do not appear to directly impact contributions. If any unobservables are correlated with these observables, this additionally assuages concerns that differences in unobservables drive the results.

Nonetheless, it is possible to get a statistical sense of how unlikely it would be to observe these results. I therefore conduct an exercise where I generate different logistical paths for the planner, run the same statistical analyses, and compare these estimates with the main estimates of the paper. I do this in two ways. The first is by simulating 10,000 random paths, defining the analogous treatment, and then examining how often an estimate is observed that exceeds in magnitude the estimates of this paper. I find that the main estimates in the paper exceed 99.0% of those estimates resulting from these simulated paths. This analysis provides a type of non-parametric test of significance.[25] It provides a measure of how "lucky" the planner would have to have been to by chance choose villages in such a way that treatment villages contributed the extent to which they did more than control villages.

The second way is that I consider a number of "worst case scenarios". I imagine a scenario where the logistical planner chooses the ordering in an extreme way - by simply ordering villages according to values of the available variables. For example, one scenario involves the planner choosing villages in order from the nearest to the study base to farthest. I do this for all variables available to the planner, ranking variables from either low to high and vice-versa (randomizing the order with ties). The results of this exercise are shown in Table 6. As one can see, even in these extreme scenarios, the treatment effects are never significant and vary widely and inconsistently across the matching and OLS specifications. Taken together these results suggest that it is highly unlikely that the main results of this paper are driven by unobservables correlated with the order of visits.

[25] By estimation strategy this is (1) 98.1% for standard matching, (2) 99.5% for exact matching and (3) 99.5% for OLS. Note that this is not identical to the permutation test where treatment is randomly re-assigned and exact tests of significance can be conducted. The reason is that treatment is not a monotonic function of visit order, and in particular depends on the spatial relationship between villages.

Table 6: Counterfactual Tests

	(1) Matching	(2) Exact Matching	(3) OLS
Distance to base (km)	−12.067	10.924	−25.220
	(21.959)	(30.465)	(15.870)
Reverse order	6.171	−8.690	−7.741
	(15.979)	(20.871)	(14.860)
Distance to paved road (km)	−5.454	−15.880	−8.503
	(13.408)	(17.958)	(12.831)
Reverse order	−17.304	19.884	−13.258
	(14.027)	(16.970)	(14.159)
Village Size (# HHs)	−4.757	−17.919	−18.411
	(20.977)	(17.959)	(14.615)
Reverse order	−18.468	−9.538	−6.399
	(15.604)	(15.521)	(15.246)
# Villages $\leq$ 1.75 km	9.842	−2.081	−9.923
	(15.745)	(14.322)	(17.245)
Reverse order	−18.153	−1.931	−9.984
	(13.199)	(16.411)	(11.922)

Average Effect of Presence of Past Participating Neighbors, for counterfactual orders of village visits. Order of visit is simulated for low to high values for odd rows, and the reverse (high to low values) for even rows. Analysis uses OLS regression and matching, following the empirical strategy in the main paper. Significantly different from 0 at * 0.1; ** 0.05; *** 0.01. Abadie-Imbens or standard robust standard errors in parentheses respectively. Observations vary.

3.11 Mechanisms

I have shown in the previous analysis that individuals tend to contribute more in public goods games when they have neighbors who previously participated in the games. The next step is to more closely examine why this is the case. Most plausible, is that previous participants communicated with individuals who were yet to participate. In a standard theoretical model, pre-game communication should have no impact on behavior. Equilibrium contributions are expected to be zero, since subjects have full information about the game and the determination of their payoffs. I now discuss four broad reasons why the theoretical predictions failed to materialize.

The first is a potential failure of the full information assumption if subjects do not understand the game. The second reason is that the theoretical game may be different from the actual game, as individuals may sanction outside of the scope of the game. The final two reasons relate to participants not behaving as standard rational agents. Specifically, the third is that playing the game may have altered cooperative norms, which were then passed on to future participants. Finally, the fourth and I argue most likely reason, is that individuals behave as conditional cooperators, and communication altered their beliefs about anticipated levels of cooperation. I now address these four reasons in detail.

If individuals do not completely understand the public goods game, this could lead to off path equilibrium behavior, i.e. contributing positive amounts. If pre-game communication improved individuals' understanding of the game, this should lead to decreasing contributions, not increasing as seen in the data. A possible explanation is that communication actually decreased understanding, leading to more confusion, and greater contributions. This seems highly implausible, and moreover, contributions in the second round of the game are statistically significantly (at the 10% level) lower, suggesting that learning may in fact lead to lower contributions.

Next, the second reason is that there may be components of play that are not incorporated into the theory. A relevant example is social sanctions. If community members can punish one another outside of the framework of the game, it may be possible to sustain other equilibria. Collusion may lead to higher contributions, sustained by the threat of costly sanctions for deviators. Learning about the game in advance could also provide individuals with more time to devise a collusive strategy, which would explain the relationship between communication and increased contributions.

One convincing argument against collusive behavior is that there is no clear way to enforce such collusion. The implementation of the game placed a very strong and serious emphasis on privacy. Individuals were told that their decisions were private, and that they should not reveal their actions at any time during or after the game. While I do not observe post-game interactions between players, anecdotal evidence suggests that individuals were not willing to reveal their decisions after the games. Collusion requires knowledge of individual decisions in order to credibly punish. Since this information was kept private, it is unlikely that pre-game communication led to better opportunities to collude and hence higher levels of contributions. Moreover, some opportunities to collude are present even without pre-game communication, as participants are free to talk during the instruction period of the games. As mentioned, contributions were significantly lower in the second round. Playing a second round should generate additional opportunities to collude, implying contributions should increase or stay the same.

The third reason following Cardenas and Carpenter (2005) and Bernal et al. (2016), is that participating in the games may have positively altered community norms around cooperation. In the current context, this appears unlikely since past participants would need to convey these norms to future participants in other villages. This possibility can also be examined using the data, since the questionnaire included a question about whether people in the community were generally cooperative about issues that affect the community. Online Appendix D investigates whether having past participating neighbors within 1.75 km altered attitudes about cooperativeness in the community, and finds no effects. Hence it seems unlikely that cooperative norms were altered.[26]

Finally, the fourth reason is that the rational model may not be an accurate predictor of behavior, due to the presence of CCs. I focus on the two related behavioral phenomena introduced earlier: the effect of advice, and pre-play communication. Regarding the former, Chaudhuri et al. (2006) demonstrated in public goods experiments that when past participants gave public advice to future participants, this advice led to significantly higher contributions.[27] Regarding pre-play communication, Isaac and Walker (1988a) documented a significant role of such communication in increasing contribution rates in public goods games.

Not surprisingly, I am unable to separately identify these two mechanisms. Presumably, when past and future participants communicated, some description of the experiment was given. But I do not know whether that description was bundled with advice. Nor is it reasonable to assume that every instance of communication followed the same structure. What I do in this section, is to identify which subjects are CCs, and examine whether villages with more CCs are more likely to increase their contributions as a result of having neighbors who previously participated.

As previously stated, all villages participated in a second round of the public goods games, that took the form of one of four variations, which I do not distinguish in this

[26]The discussion below describes how it is likely that beliefs about contributions were shifted upwards when individuals had past participants as neighbors. This is not necessarily inconsistent with unchanged attitudes toward cooperativeness. Indeed, it would be surprising if simply the act of thinking about playing a game in the future, (until that point the game had not been introduced and even when introduced was couched in neutral language with no context), caused significant changes in how cooperative people believed their community was.

[27]Note that there are differences between their controlled setting and this field experiment. First, their subjects are incentivized to provide advice, as they receive a portion of the future subjects' earnings. Second, in the common knowledge treatment where they find the largest effects, advice is read out loud in front of all participants.

analysis. I code individuals as being a CC if they contribute in the second round the nearest allowable amount to the modal contribution in the first round. If they contribute any other amount, I code them as not being a CC. Because the purses were emptied one by one and counted in front of all participants, the complete distribution of contributions was public knowledge. This procedure makes the mode particularly salient, though it will also be important to compare results when alternative definitions of CC using the median or mean are used. Aggregating this variable to the village level, the average proportion of CCs is 0.35, ranging all the way from 0 to 1 in the sample of villages.[28]

To examine the effects of being classified as a CC, the matching strategy is not tenable for splitting the sample into further subgroups, due to insufficient observations.[29] Instead, I examine heterogeneous effects using OLS regressions. Given the similarities between the matching and OLS results, and that only one covariate (village density) is imbalanced across treatment and control groups, such an OLS specification is likely credible. Table 7 presents the results interacting the treatment with the proportion of individuals classified as CCs in the village. As one can see, the effect of the treatment disappears, and the interaction becomes large and positively significant. The theoretical maximum treatment effect for village consisting of 100% CCs is 78 RWF, while the average effect given that villages have on average 35% CCs is approximately 21 RWF. Additionally, CCs contribute significantly greater amounts in general, approximately 88 RWF, or 35% more than their non-CC counterparts. This is intuitive, as typically after CCs, the largest proportion of classifiable subjects are free-riders, see Chaudhuri (2011). Table 11 in the Appendix examines different definitions for CCs involving medians or means, where similar effects are found for medians, while for the mean the magnitude is similar but the coefficient on the interaction is not significant.

The differential effect for CCs lends some support to the hypothesis that pre-game communication altered their beliefs, leading them to increase contributions. However, I am unable to disentangle the relative importance of cheap-talk versus advice in altering beliefs. Overall, this result suggests that social learning led to increases in contributions, and that these increases were disproportionately driven by CCs. This finding, combined

[28]This is consistent, though on the lower end, with the survey of Chaudhuri (2011), which notes that experimental studies have found between 35% and 81% of subjects are CCs, but differences in classification procedures make it difficult to compare.

[29]Such an analysis is nonetheless conducted in Online Appendix E, however with sample size ranging from 19-57, there is not sufficient power to draw sharp conclusions, though the pattern of results is quite consistent.

with the previous discussion regarding alternative mechanisms, suggests that social learning did not lead to a better understanding of the game, and likely did not create opportunities for collusion due to the private nature of decisions. Rather, it suggests that social learning altered subjects beliefs, as a result of pre-play communication, advice, or some combination of the two.[30]

Additional OLS specifications which split the sample into higher and lower proportions of CCs are presented in the Appendix. While heterogeneity on propensity to be a CC is central to this paper, it is also of interest to examine heterogeneity along other dimensions. Appendix Section 5.4.6 examines whether there are similarly differences in treatment effects when median splitting villages by age, education, number of others known, and trust. However significant differences are only found for the specification for CCs.

Finally, one may wonder whether the outcome of the games for past participating neighbors matters for future participants. Perhaps poor outcomes lead to pessimistic communication, lowering beliefs about cooperativeness. On the other hand, poor outcomes could increase the importance of providing advice to future participants, e.g. "don't make the same mistake we did". This is difficult to answer with the data given, as only 75 villages have past participating neighbors within 1.75 km, and there may be unobserved similarities for neighboring villages. In fact, comparing contributions in the treatment group based on whether the nearest past participating neighboring village contributed above or below the median, the level of contributions is nearly identical: 270 and 269 RWF, respectively. Hence there is no evidence that the precise outcome of past games played an important role in altering future cooperation.

[30]Given the cultural context of these games, it is highly likely that if communication involved giving advice, that such advice became public and was widely shared amongst future participants. The woman described in the introductory anecdote explicitly stated that she had passed on this information to her co-participants. The common knowledge nature of this advice was shown to be a key component of increasing contributions in Chaudhuri et al. (2006).

Table 7: The Role of Conditional Cooperators (CCs)

	(1)	(2)	(3)
Treatment Status	−3.897	−3.332	−10.246
	(13.106)	(13.120)	(15.550)
Conditional Cooperator (CC)	106.185***	92.484***	88.166***
	(30.387)	(29.288)	(25.901)
CC × Treatment	71.983**	80.091**	88.669***
	(33.518)	(32.043)	(31.976)
Distance to base (km)		−0.410	−3.349
		(0.679)	(2.558)
Distance to paved road (km)		1.075	−0.680
		(1.356)	(3.862)
Village Size (# HHs)		−0.014	0.088
		(0.101)	(0.114)
# Villages $\leq$ 1.75 km		−3.558	0.672
		(3.355)	(4.052)
Years of Education		−3.066	−6.332
		(4.702)	(5.075)
Female		62.474*	33.779
		(32.701)	(33.382)
Age		−2.557	4.352
		(18.432)	(18.510)
Age^2		0.036	−0.060
		(0.255)	(0.259)
Controls		✓	✓
Sector Fixed Effects			✓
R^2	0.43	0.50	0.58
Observations	147	146	146

Analysis uses OLS regression. Significantly different from 0 at * 0.1; ** 0.05; *** 0.01. Robust standard errors in parentheses. Treatment defined as having past participating neighbors within 1.75 km. Controls includes all remaining variables found in Table 3.

4 Discussion

Lab in the field experiments have become increasingly important for research in economics and political science, as a means to study behavioral preferences and how they relate to broader economic outcomes. Yet field contexts may generate opportunities for communication between past and future participants to a greater extent than lab contexts. This can be problematic, when opportunities for communication are correlated with other unobserved characteristics of individuals or the environment, which themselves are related to outcomes under study. The results in this paper show convincing evidence that for a particular context in Rwanda, communication took place and changed behavior. Fortunately for identification, the opportunities for communication appear to be uncorrelated with other variables of interest.

Researchers are often aware of these problems, and may have experienced their effects firsthand. How should lab in the field experiments be conducted to minimize the effects of unintended communication on outcomes? The first relevant point involves the type of experiment and the context, and whether communication is likely to be an issue. This can be difficult to assess ex-ante. In the current paper, it was not anticipated that communication would take place. Moreover, since subjects were given full details about the rules of the public goods game, including a discussion about the tradeoff between private versus public payoffs, it was thought that even if there was communication, it would not change behavior. However there are some characteristics of Rusizi district of Rwanda which indicate communication may occur, namely that it is relatively small region, and has a high population density (420/km^2; greater than every US state except one). Moreover, the median village has 2 other neighboring villages within 1.75 km, and this only refers to villages in the study. This makes it clear, that communication should not be unexpected. Thus researchers must assess whether communication is likely to be an issue in their context.

Next, researchers can take steps to control or mitigate the effects of communication. At the study design stage, this can involve creating suitable distance between sessions. If sessions occur in the same place, this becomes more difficult to control, and it may be necessary to avoid overlap in times, recruit from disparate populations, or create slightly different game versions which make it clear that advice from past games may not apply.

Beyond this, a number of steps can be taken to test for and recover estimates of the effects of communication, as in the case of this paper. In the case that sessions are in the same location, one can randomize the order, and see whether the order of the session has any impact on behavior. In the case that sessions are conducted in different locations, one can similarly randomize the order. Clearly, this may sometimes not be feasible due to logistics. In this event, one can select smaller zones of randomization: e.g. randomize the order within 10x10 km grids, or randomize the order within smaller political regions. In this manner one can test whether the order of visit within zones has any significant impact on outcomes.

Finally, if the design and order cannot be altered by the researcher, or if the study has already taken place, one can conduct an analysis following the matching techniques in this paper. The key part of this strategy is to pay attention to how the order of visits is determined, and identify all variables which have been used in its determination.

In the case of this paper, I demonstrated that theoretically unanticipated communication can have large effects, representing increases of 11-14% in contributions to the public goods game. The key to identification was leveraging the fact that logistical restraints required that villages participated at different dates, and because of natural variation in geography this led to variation in opportunities for communication between past and future participants. The matching strategy ensured balance on all covariates available to the planner at the planning stages, though the results suggest that imbalances do not pose a problem to identification in this context.

While unexpected, these patterns suggest a link with lab experimental evidence that advice, Chaudhuri et al. (2006), and communication, Isaac and Walker (1988a), can alter behavior in important ways. Moreover, both mechanisms rely on conditional cooperators, since purely selfish or purely altruistic individuals are not responsive to the behavior of others. Such a role for conditional cooperators was indeed supported in the analysis. Overall these results have important implications for the design of lab in the field experiments, and the interpretation of results. Such spillovers may occur if past participants communicate with future ones, and this communication can change behavior in meaningful ways, biasing the measurement of preferences.

5 Appendix

5.1 Political Sectors in Rusizi

Figure 5: Rusizi District Sectors

Map presents 18 political sectors of Rusizi district. The district is bordered by both DRC and Burundi. Nyamasheke is the bordering district within Rwanda. Map is adapted from Fourth Population and Housing Census, National Institute of Statistics of Rwanda.

5.2 Determinants of Visit Order

Table 8: Determinants of Visit Order

	(1)
Distance to base (km)	0.281*
	(0.155)
Distance to paved road (km)	0.123
	(0.165)
Village Size (# HHs)	0.009
	(0.006)
# Villages $\leq$ 1.75 km	-0.132
	(0.199)
Sector Fixed Effects	✓
R^2	0.96
Observations	146

Analysis uses OLS regression. Significantly different from 0 at * 0.1; ** 0.05; *** 0.01. Robust standard errors in parentheses. Dependent variable indicates order of visit and varies from 1 to 53. Villages visited on the same day receive the same value of this variable.

5.3 Different Numbers of Neighbors for Matching Esimation

Table 9: Average Effect of Presence of Past Participating Neighbors: Varying Number Neighbors

	(1) Standard Matching	(2) Exact Matching
	1 Neighbor	
Contribution	26.342**	31.608***
	(11.041)	(11.422)
	3 Neighbors	
Contribution	30.814***	31.657***
	(10.298)	(10.894)
Observations	117	101

Analysis uses nearest neighbor propensity score matching, with 2 neighbors, with replacement. Significantly different from zero at * 0.1; ** 0.05; *** 0.01. Abadie-Imbens Robust Standard Errors in parentheses. Values of propensity score outside common support range are dropped. Exact matching excludes sectors with only 0 or 1 village in either treatment or control groups.

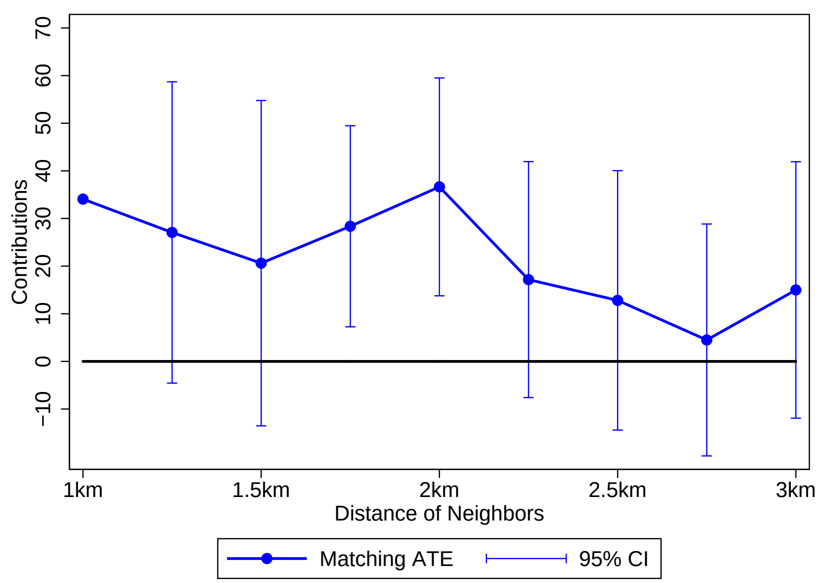
5.4 Estimates for Different Treatment Distances

5.4.1 Matching Estimates

Figure 6 additionally presents matching estimates for different distance cutoffs, using the specification with full controls analogous to column (1) in Table 4. The coefficient estimates range from 4.5 (not significant) to 36.6 (significant at 1%).

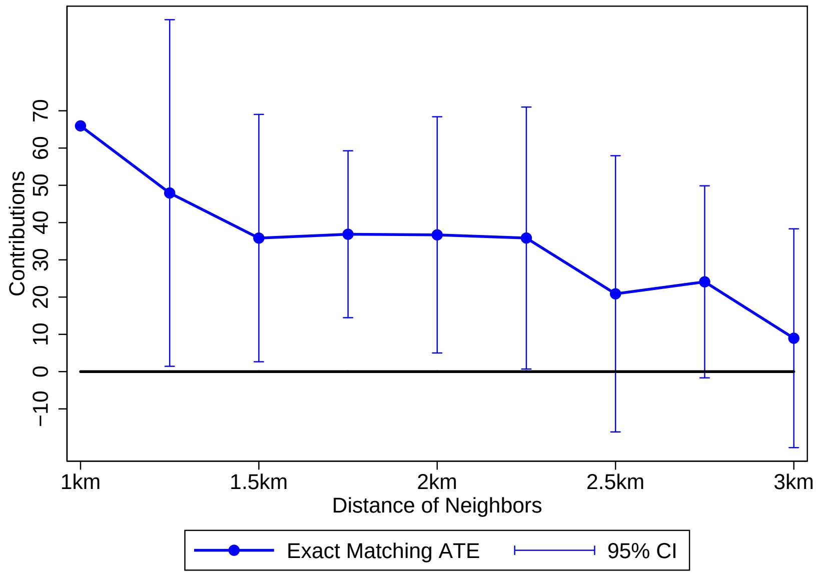
Figure 7 presents the analogous figure for exact matching on political sectors, the specification in Column (2) in Table 4. There one can see a pattern of diminishing effects as the distance cutoff is greater, where estimates range from 66.0 to 9.0 (neither significant). For the ranges 1.25km to 2.25km, estimates are significant at the 5% level. Note also that for these distances, treatment and control groups are also better balanced. Another point to note is that similarly decreasing patterns are observed for the OLS estimates in Figure 8, and one can see that OLS and matching estimates are similar for intermediate distances where standard errors from matching are moderate.

Figure 6: Matching estimates for different treatment cutoff distances



Each point corresponds to estimate of coefficient on treatment for specified distance cutoff for independent matching estimations, analogous to specification (1) in Table 4. Error bars suppressed for 1 km due to noise. Number of observations varies.

Figure 7: Exact matching on sector estimates for different treatment cutoff distances

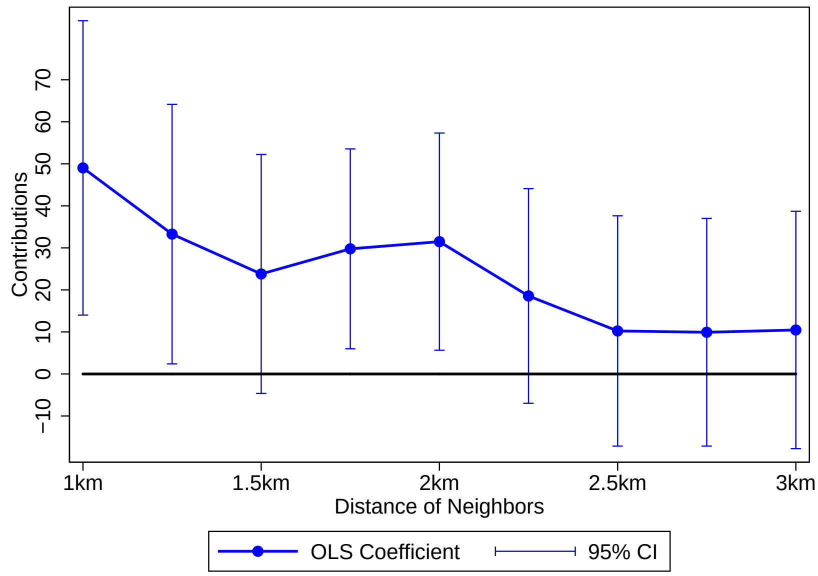


Each point corresponds to estimate of coefficient on treatment for specified distance cutoff for independent matching estimations, analogous to specification (2) in Table 4. Error bars suppressed for 1 km due to noise. Number of observations varies.

5.4.2 OLS Estimates

Figure 8 additionally presents OLS estimates for different distance cutoffs, using the specification with full controls analogous to column (3) in Table 5. These results show a pattern of diminishing effects as the distance cutoff is greater. The largest effects are found for villages with neighbors within 1km, increasing contributions by 49.0 significant at the 1% level, and the smallest at 2.75 km, increasing contributions by 9.9 RWF, not significant at conventional levels.

Figure 8: OLS estimates for different treatment cutoff distances



Each point corresponds to estimate of coefficient on treatment for specified distance cutoff for independent OLS regressions. N = 146.

5.4.3 Exact Matching on Village Density

In this section I mitigate concerns that treatment effects are in fact picking up differences in village densities, i.e. the number of neighboring villages in the study located within 1.75 km. I thus conduct an exact matching on the number of villages (in the study) that are within 1.75 km of a given village. By construction, the treatment variable, which is 1 whenever a village had neighbors located within 1.75 km that previously participated, is highly correlated with the number of total neighbors within 1.75 km in the study. Further, the initial balance checks indeed revealed that village density was not balanced across treatment and control villages.

To account for the possibility that village density may be correlated with unobserved variables, and that the earlier propensity score matching may have been unable to adequately control for this, I conduct a matching strategy where I require that matched treatment and control villages *must* have exactly the same number of neighbors in the study. Village density has a minimum value of 0 and a maximum of 6, though values of 0 were excluded as noted in the primary analysis. Further, as there are no control villages for densities with 5 or 6 neighbors, these 5 villages are necessarily dropped from this analysis as well. Table 10 presents this matching analysis. As can be seen, the effects are significant, and consistent with the main results.

Table 10: Average Effect of Presence of Past Participating Neighbors

	(1) Exact Matching on Number Villages within 1.75 km
Contribution	33.935***
	(11.568)
Observations	118

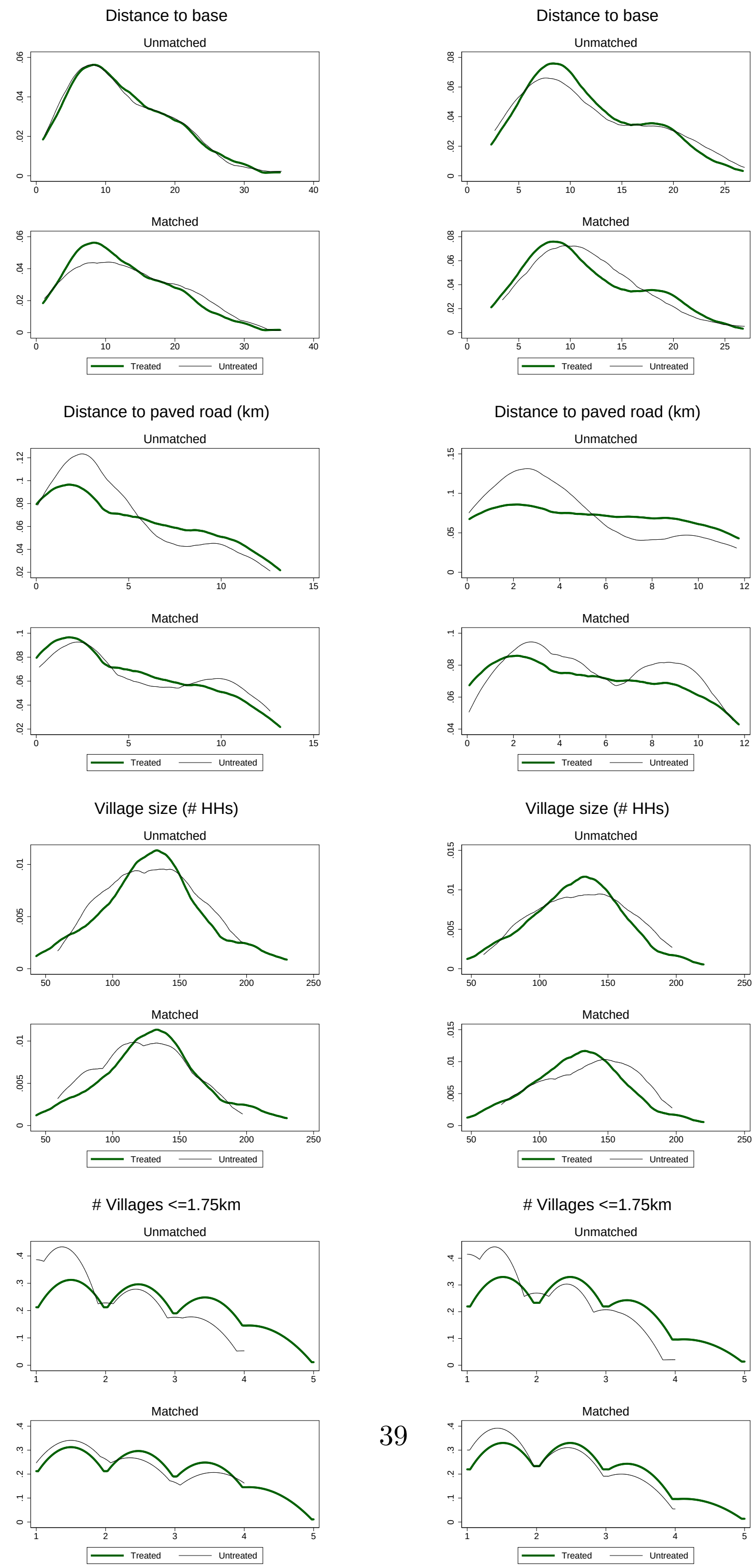
Analysis uses nearest neighbor propensity score matching, with 2 neighbors, with replacement. Significantly different from zero at * 0.1; ** 0.05; *** 0.01. Abadie-Imbens Robust Standard Errors in parentheses. Values of propensity score outside common support range are dropped. Exact matching excludes densities with only 0 or 1 village in either treatment or control groups.

5.4.4 Balance

Figure 9 presents graphically the distributions of matching covariates by treatment status, in order to evaluate balance of the matching strategy. For the most part there are not large imbalances across these variables, even in the unmatched stage. Matching improves balance, which is slightly better in standard matching (column (a)) rather than exact matching (column (b)), though these differences are not very substantial.

Figure 9: Balance of Matching Covariates

(a) Standard Matching (b) Exact Matching



Density of covariates in matching estimations by treatment status, before and after matching. **(a)** $N = 117$ **(b)** $N = 101$ observations.

5.4.5 Alternative definitions of CCs

Table 11 presents the effect of CCs for different classifications of CCs. Column (1) replicates the main results in the paper, defining an individual as a CC if they contribute in round 2 the mode of contributions in round 1. Column (2) uses the median, while column (3) uses the mean. From the table, one can see that mode and median results are quite consistent, while the mean results are not significant. Of note is that the coefficient on the interaction for the mean definition is of the same magnitude, though the estimates are quite noisy. This could be accounted for by the fact that participants would have found it easier to gauge the mode or median given the experimental protocol, which involved contributions being counted one by one. Additionally the definition using mode or median appear to provide a better fit to the data than the mean.

Table 11: Robustness: Different definitions of CCs

	(1) Mode	(2) Median	(3) Mean
Treatment Status	−10.246	−18.048	−4.613
	(15.550)	(17.118)	(19.818)
Conditional Cooperator (CC)	88.166***	95.128***	−16.900
	(25.901)	(31.976)	(61.003)
CC × Treatment	88.669***	92.109**	110.617
	(31.976)	(42.944)	(71.243)
Distance to base (km)	−3.349	−3.614	−2.359
	(2.558)	(2.760)	(3.098)
Distance to paved road (km)	−0.680	−0.956	−1.821
	(3.862)	(3.918)	(4.440)
Village Size (# HHs)	0.088	−0.069	0.055
	(0.114)	(0.120)	(0.132)
# Villages ≤ 1.75 km	0.672	−0.308	0.928
	(4.052)	(4.512)	(5.540)
Years of Education	−6.332	−1.875	−3.871
	(5.075)	(5.416)	(6.728)
Female	33.779	26.369	81.995*
	(33.382)	(39.261)	(45.921)
Age	4.352	13.983	10.391
	(18.510)	(18.949)	(23.840)
Controls	✓	✓	✓
Sector Fixed Effects	✓	✓	✓
R^2	0.58	0.53	0.35
Observations	146	146	146

Analysis uses OLS regression. Significantly different from 0 at * 0.1; ** 0.05; *** 0.01. Robust standard errors in parentheses. Specifications follow those in Table 5 column 3.

5.4.6 Conditional Cooperation and Heterogeneous Effects

Tables 12 presents further analysis of potential heterogeneous effects, using OLS. Online Appendix E presents the analogous table using matching strategies, though power is limited. Here villages are split according to the median value of the variable of interest, and average treatment effects are estimated for this subsample following the empirical strategy in Column (3) of Table 5. Additionally Chow Test statistics are reported to get a sense of whether the differences across these subsamples is significant. Of note is that along the dimensions: proportion of CCs, average age, average education, average number of others known, only for CCs is the difference in the two subsamples statistically significantly different (at the 10% level).

Table 12: Heterogeneous Effects of Treatment: Past participating neighbors within 1.75 km

	(1) OLS	(2) N
By Conditionally Cooperative		
More Conditionally Cooperative	39.027*	58
	(20.929)	
Less Conditionally Cooperative	−2.476	71
	(16.369)	
P-Value (Chow test)	[0.056]*	
By Age		
Older	39.729*	73
	(23.872)	
Younger	45.552**	71
	(20.785)	
P-Value (Chow test)	[0.780]	
By Education		
More Educated	43.306**	72
	(20.777)	
Less Educated	18.653	69
	(23.098)	
P-Value (Chow test)	[0.357]	
By Number Others Known		
More People Known	37.394*	70
	(21.665)	
Less People Known	38.522*	72
	(22.809)	
P-Value (Chow test)	[0.770]	

OLS regression of specification (3) in Table 5. Selected covariates split by median value. Significantly different from zero at * 0.1; ** 0.05; *** 0.01. Robust Standard Errors in parentheses.

References

Avdeenko, Alexandra and Michael J Gilligan, "International Interventions to Build Social Capital: Evidence from a Field Experiment in Sudan," *American Political Science Review*, 2015, pp. 1–45.

Baldassarri, Delia, "Cooperative Networks: Altruism, Group Solidarity, Reciprocity, and Sanctioning in Ugandan Producer Organizations," *American Journal of Sociology*, sep 2015, *121* (2), 355–395.

Bednar, Jenna, Yan Chen, Tracy Xiao Liu, and Scott Page, "Behavioral spillovers and cognitive load in multiple games: An experimental study," *Games and Economic Behavior*, jan 2012, *74* (1), 12–31.

Bernal, Adriana, Juan-Camilo Cárdenas, Laia Domenech, Ruth Meinzen-Dick, and Sarmiento Paula J., "Social Learning through Economic Games in the Field," *mimeo*, 2016, pp. 1–33.

Bochet, Olivier and Louis Putterman, "Not just babble: Opening the black box of communication in a voluntary contribution experiment," *European Economic Review*, 2009, *53* (3), 309–326.

Brosig, Jeannette, J Weimann, Axel Ockenfels, and Joachim Weinmann, "The effect of communication media on cooperation," *German Economic Review*, 2003, *4* (2), 217–241.

Cardenas, Juan Camilo and Jeffrey P. Carpenter, "Three themes on field experiments and economic development," in John A. List Glenn W. Harrison, Jeffrey Carpenter, ed., *Field Experiments in Economics*, 2005, pp. 71–123.

Casaburi, Lorenzo and Rocco Macchiavello, "Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya," *American Economic Review*, 2018.

Chaudhuri, Ananish, "Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature," *Experimental Economics*, 2011, *14*, 47–83.

— , **Sara Graziano, and Pushkar Maitra**, "Social Learning and Norms in a Public Goods Experiment with Inter-Generational Advice," *Review of Economic Studies*, apr 2006, *73* (2), 357–380.

Crawford, Vincent, "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory*, 1998, *78* (2), 286–298.

Dupas, Pascaline, "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?," *American Economic Review*, apr 2009, *99* (2), 224–230.

— , "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment," *Econometrica*, 2014, *82*, 197–228.

Enos, Ryan D. and Noam Gidron, "Exclusion and Cooperation in Diverse Societies: Experimental Evidence from Israel," *American Political Science Review*, jul 2018, pp. 1–16.

Gneezy, Uri and Alex Imas, "Lab in the field: Measuring preferences in the wild," *Handbook of Field Experiments*, 2017.

Imbens, Guido W. and Donald B. Rubin, *Causal inference: For statistics, social, and biomedical sciences an introduction* 2015.

Isaac, R. Mark and James M. Walker, "Communication and free-riding behavior: the voluntary contributions mechanism," *Economic Inquiry*, 1988, *26* (4), 585–608.

— **and James M Walker**, "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *The Quarterly Journal of Economics*, 1988, *103* (1), 179–199.

Jakiela, Pamela and Owen Ozier, "Does Africa Need a Rotten Kin Theorem ? Experimental Evidence from Village Economies," *Review of Economic Studies*, 2015.

Kosfeld, Michael and Devesh Rustagi, "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia †," *American Economic Review*, 2015, *105* (2), 747–783.

Ledyard, John O., "Public Goods: A Survey of Experimental Research," *Social Science*, jan 1995, *35* (12), 111–194.

Rosenbaum, Paul R and Donald B Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, apr 1983, *70* (1), 41.

Sally, David, "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society*, jan 1995, *7* (1), 58–92.