Final report

Improving political communication

Results of a pilot field experiment in Pakistan

Miriam Golden
Saad Gulzar
Luke Sonnet

October 2017

IGC

International
Growth Centre

DIRECTED BY

LSE

UNIVERSITY OF
OXFORD

FUNDED BY

UKaid
from the British people

Improving Political Communication:
Results from a Pilot Field Experiment in Pakistan[*]

Miriam Golden[†]

Saad Gulzar[‡]

Luke Sonnet[§]

October 13, 2017

Abstract: This paper presents results from a pilot field experiment where a low-cost and scalable technology — Interactive Voice Response (IVR) — improves political communication in Pakistan. We study the effects of a robocall recorded by a politician asking specific questions that is sent out to large number of his constituents on respondents' political opinions. We find that constituent take-up is high: 31 percent of those called answer at least one question that the politician asked. There is also suggestive evidence that the call improves perceptions of government competence and evaluations of the incumbent. A follow-up call by the politician cross-randomizes whether the politician exhibits responsiveness to the feedback from citizens or instead responds without acknowledging citizen input. We find that a responsive follow-up does the most of all treatment variations to increase the weight of incumbent performance in citizens' voting calculus. The main takeaway of the pilot is that communication from politicians using IVR can encourage constituents to engage more effectively with the democratic process.

[†]Professor of Political Science, University of California, Los Angeles, golden@ucla.edu.

[‡]Assistant Professor of Political Science, Stanford University, gulzar@stanford.edu.

[§]Ph.D. candidate, Department of Political Science, University of California, Los Angeles, lukesonnet@ucla.edu.

1 Introduction

Information and Communications Technology (ICT) is shaping the modern political land-scape in important ways: social media has transformed the speed and method through which mass protests mobilize around the world (Tucker et al., 2016; Enikolopov, Makarin and Petrova, 2016);[1] governments routinely rely on new ICT tools to improve how they interact with citizens;[2] political parties are using the power of mass communication in their campaigns;[3] and there is even a proliferation of social enterprises that aim at 'hacking' democracy.[4] In the face of these new developments, conventional conceptions of what it means to govern and be governed may need to be modfied. High cellphone penetration in the developing world means both elites and non-elites face new incentives when deciding whether and how to communicate with one another.

We report results from a field experiment that evaluates the democratizing potential of a new technology — Interactive Voice Response (IVR). In the field experiment, we record a message by a politician in his own voice and deliver the call to his constituents. Using IVR, citizens respond to the specific questions posed in the message by pushing buttons on their cell phones. Responses are then delivered to politicians at an aggregated level, allowing the politicians to take appropriate action based on this new information. The IVR technology allows for a wide variety of questions and for the process to iterate several times.

IVR improves upon business-as-usual communication between politicians and citizens in three critical ways. First, the communication is direct, rather than relying on traditional intermediaries like village headmen, and therefore potentially improves democracy by flatten-ing citizen access to politicians (Grossman, Humphreys and Sacramone-Lutz, 2014). Because the message is recorded in the voice of the politician, it may mitigate citizens' concerns with the efficacy of political communication. Second, the communication is cheap and scalable, which allows for easy deployment even in periods of low political salience, such as between elections. Third, IVR is specific in the sense that politicians can ask direct and particular questions (such as how to spend a newly-acquired fund) and can receive responses from citizens to those specific questions. Together, these elements change the incentives and

---

[1] The Umbrella Movement in Hong Kong was coordinated and captured on Twitter (https://blogs.wsj.com/chinarealtime/2014/10/01/hong-kong-protests-how-they-unfolded-on-twitter/ last accessed Aug 17, 2017).

[2] In Punjab, Pakistan the Citizen Feedback Model robocalls citizens to report instances of corruption and fraud (Callen and Hasanain, 2011) (http://www.punjabmodel.gov.pk/ last accessed Aug 17, 2017).

[3] See http://www.telegraph.co.uk/news/worldnews/asia/india/10803961/Magic-Modi-uses-hologram-to-address-dozens-of-rallies-at-once.html, last accessed August 21, 2017.

[4] Democracy Earth (http://democracy.earth/, last accessed August 17 2017) aims to create a peer-to-peer democratic system where people choose to vote on issues they care most about.

costs faced by citizens and politicians. Using IVR, a large group of political principals can communicate with a difficult-to-reach elected agent who can now cheaply collect preferences.

We test whether improved communication between politicians and citizens using an IVR channel has any bearing on crucial components of the democratic process. Specifically, we focus on the reception of citizens to this technology through their take-up, changes in expressed favorability to the incumbent, and perceptions of government performance. We also estimate whether this intervention can strengthen the building blocks of the "accountability chain" (Lieberman, Posner and Tsai, 2014) by providing a platform through which citizens think about politicians' responsibilities and make demands directly of their elected officials.

To do this, we design an experiment with one Member of a Provincial Assembly (MPA) in Pakistan.[5] The experiment proceeds as follows. First, the MPA records a message in his voice that includes an update about his recent activities, a question about how to spend development funds, and another question about how to allocate effort across parliamentary and other duties. Second, we randomly assign 792 out of 1,218 sampled citizens to receive a call. Third, we aggregate the answers by geographic area, party of support, and throughout the constituency for the MPA. Fourth, the MPA records a response that either (1) directly acknowledges citizen preferences expressed in the first call and reports how he intends to respond or (2) is a generic credit-claiming message with no responsive content. Fifth, we randomize citizens who answered the first phone call to receive either the responsive follow-up call or the generic follow-up call.

Implementation of the experiment proved successful. Take-up of the treatment is high. Around 86 percent of the people we called engaged in some aspect of treatment, and 31 percent of those assigned to receive the MPA's call answered at least one of the questions posed by the MPA. This response rate to an ICT intervention is much greater than those reported in previous studies (Blair, Littman and Paluck, 2017; Grossman, Humphreys and Sacramone-Lutz, 2015). Furthermore, the percentage of individuals who answer at least one of the MPA's questions doubles when the question is about development priorities rather than MPA time-use in the legislature, suggesting that the perceived value of the communication may be pivotal to the success of an ICT intervention.

---

[5]MPAs sit in one of Pakistan's four provincial assemblies whose constituencies range in size from a quarter to half a million individuals. MPA responsibilities include drafting, debating, and voting on provincial legislation, serving as ministers if they are part of the provincial government, and directing provincial budgets and spending. Elections for MPAs are held every five years in general elections, and each province selects the majority of its MPAs from single-member constituencies. The province we study has above average turnover of MPAs and smaller than average constituency populations. A scale-up experiment with 40 MPAs is now underway.

We study outcomes of treatment in three pre-registered domains: (1) evaluations of the incumbent, (2) government performance, and (3) prospects for accountability. On the first, we find a positive but statistically insignificant effect of increased communication on an index of evaluations of the incumbent and his party. We find an intent-to-treat (ITT) effect of 0.058 standard deviations (SDs) (0.048 standard error), with a local average treatment effect (LATE) among full compliers — those who answer at least one of the MPA's questions — of 0.189 SDs (0.155 s.e.). These positive but imprecisely estimated effects indicate that citizens may value any and all contact from the MPA or they may value hearing about the MPA's recent activities in the call. We also conduct an endorsement experiment, results of which produce clearer and statistically significant evidence that receiving the IVR call has positive effects on support for bills that the MPA endorses.

On the second (government performance), we find evidence for an effect of IVR communication on citizen evaluations of government performance. We create an index of questions measuring perceptions of provincial government competence and whether various state entities look after people like the respondent. The ITT effect of IVR on this index is 0.077 SDs (0.038 s.e.), while the LATE among full compliers is 0.252 SDs (0.125 s.e.). Results thus suggest that the IVR intervention increases perceptions of government competence. Inconclusive results from the second stage follow-up calls suggest that the main effects on citizen evaluation of government performance come from receiving an IVR call and not from responsive feedback.

On the third (prospects for accountability), we find a positive but statistically insignificant effect of IVR communication on an index measuring prospects for political accountability. We construct the index from questions tapping the building blocks of accountability: evaluations of politician and citizen efficacy and the importance of performance in citizens' vote-making decision. We provide suggestive evidence that contact-seeking input into the MPA's decision making process promotes certain building blocks of electoral accountability (ITT on the index of 0.053 SDs; 0.045 s.e., with a LATE of 0.170 SDs; 0.145 s.e.). The responsive follow-up does the most to increase the importance of incumbent performance in our measure of the future vote decision-making process.

Taken together, these outcomes indicate that an IVR intervention undertaken with an MPA can produce high response rates, and is likely valued by the MPA's constituents. We base the second inference on the (admittedly weak) evidence that treated citizens tend to evaluate the incumbent and the government's competence more highly than control citizens. Importantly, contact through IVR, and especially contact that both opens and closes a feedback loop — by soliciting feedback and then directly responding to it — can be effective in promoting

prospects for accountability. Future work in a scale-up will aim to test these hypotheses in a larger sample and with the addition of measuring downstream effects on MPA knowledge of constituent preferences, MPA effort in treated areas, and how an IVR intervention mediates how MPA effort and development projects are responded to at the ballot box.

The first contribution of our work is to the literature on democratizing technology. Grossman, Humphreys and Sacramone-Lutz (2014) present positive results from a study that uses a platform to allow citizens to express their preferences to elected representatives by sending them text messages from their phones. Similar to our case, politicians have the ability to respond to citizen input. The authors find however that citizens do not make use of the platform when the technology is scaled up to a hundred political constituencies (Grossman, Humphreys and Sacramone-Lutz, 2015). The authors suggest that citizens do not engage in the scale up because they do not expect the politicians to listen to their communications expressed via SMS. Issues of citizen efficacy are mitigated with IVR because communication is initiated by the politician, and citizens respond to specific questions posted to them rather than initiating communication. Politicians are likely to initiate contact because they perceive it as useful for future vote-getting. Furthermore, we provide experimental evidence that the perceived value of the input provided by IVR can double participation rates in ICT interventions.

The second contribution of our work is to studies of electoral accountability. Lieberman, Posner and Tsai (2014) outline the various conditions along the "accountability chain" that must be met in order for electoral accountability to be possible. Recent interventions providing information to citizens in the hopes of promoting electoral accountability have produced mixed results, indicating that some of the conditions for electoral accountability are difficult to meet (Adida et al., 2016; Banerjee et al., 2010; Bidwell, Casey and Glennerster, 2015; Chong et al., 2015; Dunning et al., Forthcoming.; Grossman and Michelitch, 2016; Lieberman, Posner and Tsai, 2014; Bidwell, Casey and Glennerster, 2015). Our study provides evidence that IVR is a platform that may affect several steps in the "accountability chain," such as the importance of the incumbent evaluation in the voting decision, and the belief that the provincial government has the capacity to improve livelihoods.

Our study also builds on a literature that shows that political messaging regarding the effort and competence of an incumbent can be effective in changing voter evaluations of elected officials (Kendall, Nannicini and Trebbi, 2015). Prior work specifically on credit-claiming by politicians demonstrates that the quantity of messages shifts opinion on elected representatives more than the exact content of the messages (Grimmer, Messing and Westwood, 2012). The results of our endorsement experiment indicate that citizens value this communication

from their MPA and reward them for it. We hypothesize that citizens take "information-seeking" behavior — meaning efforts by the incumbent to learn about the preferences of citizens — as a signal of effort and good will, and as a sign that the politician seeks to improve his competence.

2   Context and Experimental Design

We conduct the experiment in partnership with one Member of the Provincial Assembly in one Pakistani province. This provides a highly appropriate research setting for our purposes. First, politicians in this province face stiff competition in retaining their seats from one election to the next, and are therefore eager to identify and utilize new channels of political communication with citizens. Second, Pakistan is in the process of consolidating its democratic institutions. The first democratically elected government to finish its complete term left office in 2013. The current incumbents are the first in the country's history to follow a democratically elected government. Improving communication between politicians and citizens contributes to the democratization of the country.

2.1   Sampling

We use multistage cluster sampling in the partner politician's electoral constituency. Within the constituency, we sample Village Councils (VC), which are local government bodies; within Village Council-demarcated areas, we sample settlements, which are clusters of households ranging from 70 to 500 households; within settlements we sample individual households and enroll the male head of household. We sample 11 of the 25 Village Council areas within our partner MPA's constituency and we sample 4 of between 6 and 12 settlements within each Village Council area. Within each settlement, we sample 25 to 28 households. We enroll a total of 1,218 male heads of household in this study.

Detailed gazettes enumerating all settlements within Village Council areas are not available. The list of names of the Village Councils that we have references only one to three of the largest villages. There are unknown numbers of additional smaller settlements within each Village Council. To generate a more complete list of settlements, we sent teams of enumerators to each of our sampled Village Councils to map all settlements that they were sure fell within the boundaries of the Village Council. We sample four settlements from this more complete list.

We use a random walk to select households within settlements. Within each sampled settlement, enumerators go to the center of the settlement and fan out in two directions (if there were only two directions) or four directions (if there were more), surveying every other household on the right hand side. In each sampled household, we survey the most senior male with a cell phone.[6] Enumerators conduct a baseline survey to collect phone numbers and information on pre-treatment covariates, as well as to administer informed consent.

Because we use simple random sampling at each stage of the design, more populous Village Councils and more populous settlements have the same probability of being sampled as less populous Village Councils and less populous settlements. In our pre-specified analysis we planned to weight all of our analyses by sampling weights that account for this discrepancy. However, because our second stage analysis is not on this same sample, we decided to be consistent and only report estimates of sample average treatment effects in the main body of the paper for both the first and second stage analyses. We report the first stage analyses with weights as a robustness check in Appendix B. The signs are all unchanged except for one component measure that is very close to zero. The effects on the government performance index are larger while the effects on the prospects for accountability index are smaller.

2.2 Main Treatment Arms

Our project has three treatment arms, represented by the household numbers that appear in Table 1. Stage One assigns households into *contact* and *control* conditions. Households that comply with the *contact* treatment in Stage One (that is, answer the phone) are then randomized into either a *generic* or a *responsive* treatment in Stage Two. Stage Two occurs in time after the completion of Stage One.

Table 1: Number of Households Assigned to Main Experimental Conditions

Stage One	Control		Contact	
	426		792	
Stage Two			Generic Follow-Up	Responsive Follow-up
			340	345

Note: The number of households in Stage Two does not equal the number of treated households in Stage One because the 107 Contact households that did not comply with the Stage One treatment were dropped for Stage Two.

Of the 11 Village Council areas in our study, we randomly place one in a pure control condition. We assign all sampled households in the single pure control Village Council

---

[6]Most adult males in the study area possess cell phones.

to the control condition, as shown in Figure 1b. We call these households *pure control* households. Within the ten treated Village Council areas, we block on settlement. We assign households in the ten treated Village Council areas to either treatment or control, as the numbers reported in Figure 1a show.

Figure 1: Main experimental stages and treatement conditions by settlement

(a) Treatment in 40 treated settlements (in 10 Village Councils)

(b) Treatment in 4 control settlements (in 1 Village Council)
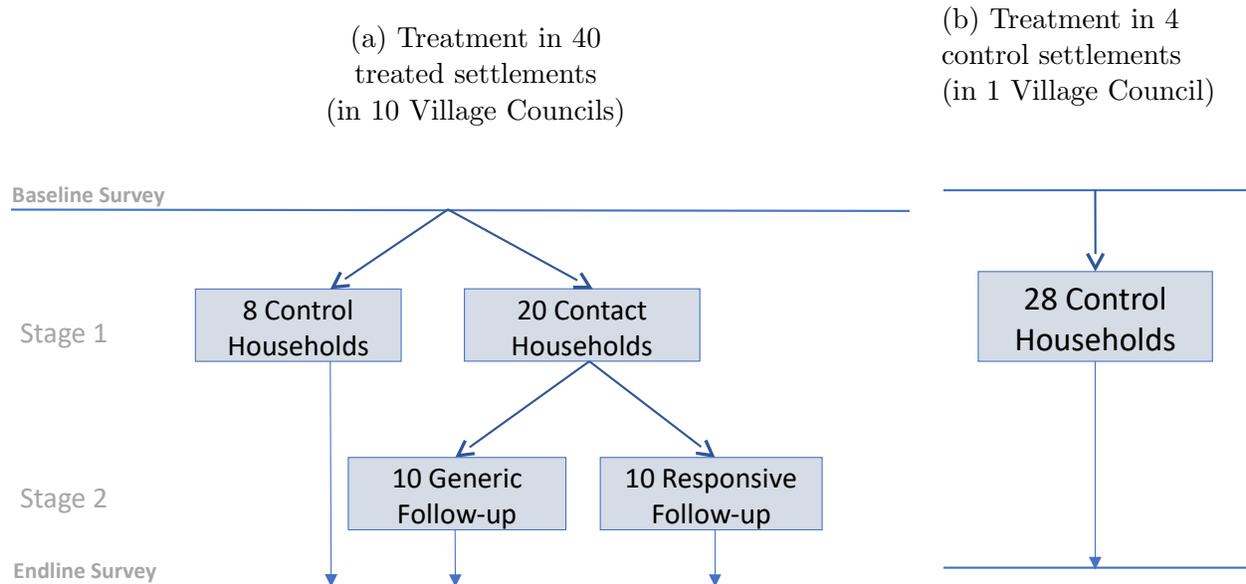
Note: This represents the ideal treatment assignment that took place in 37 of our 44 settlements. In the other 7, we sampled 25, 26, or 27 households instead of 28 and the randomization schemes are adjusted slightly to account for this. We randomize Stage Two only among compliers from Stage One and not among all treated households.

We sample 28 households from most settlements. Within settlements where we sample 28 households, eight households are sorted into the *control* condition while the remaining 20 households receive a *contact* treatment in Stage One, as depicted in Figure 1a. (We refer interchangeably to these as households and as respondents.) In total, we have 317 control households within the 10 treated Village Council areas. There are an additional 109 control households in the one pure control Village Council area. In the 10 Village Council areas that are assigned to the treatment condition, there are 792 treated households. These numbers do not directly map onto Figure 1a because in 7 of the 44 settlements in the sample, we sampled between 25 and 27 households rather than 28.

We randomize the Stage Two treatment among households that were both assigned to *contact* in Stage One and that partially complied (i.e. answered the phone when called). Of the 20 treated units per settlement, we intended to call 10 with a *generic* follow-up call whereas the other 10 were slotted to receive a *responsive* follow-up call. Because we only randomized

among compliers, these groups instead contained eight or nine respondents each rather than 10. In total, 340 households received the *generic* follow-up call and 345 households received the *responsive* follow-up call. The latter follow-up call directly addresses the demands of citizens in their Village Council area, acknowledging their input, whereas the generic follow-up simply mentions future plans. Drafts of the scripts for the follow-up calls appear in Section 2.3.2.

2.3 Treatment Content

Within the Stage One *contact* condition, we implement our design to explore whether it is more effective for politicians to communicate with citizens rather than not to communicate with them, and if they do communicate, whether the nature of the question matters to citizens. We do this by having the MPA record two separate questions: one that asks how the MPA should allocate constituency development funds and a second that asks how the MPA should allocate time between legislative debate and other (i.e. constituency) services. We hypothesize that spending questions will engender larger treatment effects because we believe that citizens put greater priority on the local development that spending produces than on their representative's parliamentary activities. We randomize which question the MPA asks in the robocall to assess whether this is true. The questions that we use were determined working with our partner MPA, and reflect his informational priorities.

2.3.1 Stage One Treatment (Contact) and Secondary Arms (Questions)

Our Stage One treatment is a recorded call by our partner MPA that asks a question about upcoming decisions and also gives him the opportunity to credit-claim for other activities. The main treatment of theoretical interest is whether a voter receives a call asking a question (*contact*) compared to no direct contact by phone from the MPA (*control*). We also randomize whether respondents receive a development question, a time-use question, or both.

Stage One Scripts

Assalam U Alaikum. I am [MPA NAME], your elected MPA. I am calling you as your elected representative in the provincial assembly. This phone call is a part of a new effort to reach out to my constituents to get their opinions and inform them about my activities. This phone call will not take long.

Contact development question: As you know, many development works in the constituency

---

have been completed and some are underway. Recently, I have approved some more funds for the constituency. I would like to know your opinion on how to spend this money. I am going to read a list of options to you. Please press the number that corresponds to the option you would like to recommend. Please wait until I have finished the list before choosing.

- If you prefer roads, press 1
- If you want to have a funeral services area, press 2
- If you prefer electricity infrastructure like transformers and electric poles, press 3
- If you want to have paved streets, press 4

Contact legislation question: As your MPA, I have participated actively in the [PROVINCE NAME] assembly's discussions. Besides other successes, I have passed a bill to prohibit interest on loans. Secondly, I have also passed a bill to do away with the housing tax in the rural areas of the province. Do you want me to continue spending time in the assembly discussions or to concentrate more on your other issues and concerns?

- If you prefer I spend more time participating in the assembly discussions, press 1
- If you prefer I spend more time looking at other issues, press 2

Thank you for your feedback. In the end I would like to bring to your notice that in terms of education, I have opened two degree colleges in our constituency, one for boys and another for girls; different schools have been upgraded; and new schools are being constructed. In terms of health care, we are going to build a large hospital in the constituency, and many BHUs [Basic Health Units] have been upgraded to RHCs [Rehabilitation Hospital Units]. In terms of electricity infrastructure we have done a record number of works; installed transformers and electric poles and overhauled the old transformers. In terms of roads, the construction of [MAIN ROAD LOCATION], has been approved and the work is already underway. For the farmers of this constituency, we are building watercourses. To prevent [TWO AREAS IN CONSTITUENCY] from flooding, we are strengthening the banks of River [river name]. Similarly, we are making progress in other areas such as paving the streets and sewers. As you know I am the only MPA in this constituency who is trying to bring the general public and MPA closer. I have started different ways of communicating with my constituents. I promise you that in coming days I will do everything for the prosperity and development of my constituency. I look forward to your support. Good Bye.

2.3.2 Stage Two Treatment (Follow-up Contact)

After we collect the results from the first treatment and aggregate them at the village level for anonymity and clarity, we share them with the MPA and craft his responses with him. His responses are given two different formats: a follow-up phone call that provides *responsive* feedback or one that provides *generic* feedback. The two follow-up messages were written working with the MPA.

3   Data

We conduct baseline and endline surveys of 1,218 households in the partner MPA's constituency. We ask a short battery of demographic and political knowledge questions in the baseline and also collect phone numbers. In both the baseline and endline, we ask a series of questions about political preferences and attitudes. We group our outcomes into three domains: incumbent evaluation, government performance, and prospects for accountability. All the baseline surveys were conducted in person by lingistically-competent enumerators, whereas roughly half the endline surveys were conducted in person and the other half over the phone in order to explore whether these two data collection methods were equally effective. Within the 11 Village Council areas in our study, we randomize settlements into two main endline data collection methods. We assign 22 of the 44 total settlements (606 of the 1,218 households) to receive the endline in-person and 22 of the 44 settlements (612 of the 1,218 households) to receive the endline over the phone. Because the endline data collection method is randomized at the settlement level, any effects of this method on the outcomes is absorbed by settlement fixed effects. In addition, we validate the phone data collection method by administering the exact same endline survey over the phone to half of

the in-person endline households (11 settlements nesting 302 households) several days after the in-person endline was completed. In Appendix A, we demonstrate that the endline data collection method is not predicted by treatment, that in-person and phone responses are fairly stable for the 302 households that we collected both from, and that our main results are qualitatively similar when we replace the in-person responses with the phone responses for those 302 households that received the endline a second time over the phone.

Our outcomes are grouped into three domains: (1) evaluations of the incumbent, meant to capture how the intervention influences attitudes towards the individual sending the messages; (2) government performance, measuring whether increased communication that requests feedback is taken as a sign of government competence; and (3) prospects for accountability, measuring whether our intervention creates a focus on aspects of politician performance that are important components of electoral accountability

First, in the incumbent evaluation domain, we ask respondents a feeling thermometer question (ranging from 1 to 10) about their MPA ("MPA thermometer") and about their party ("inc party thermometer") as well as asking them about their preferred party, a measure we transform to a binary variable for whether they support the incumbent MPA's party ("prefers MPA party").[7]

Second, in the government performance domain, we ask people whether they agree that the provincial government is competent at providing goods ("prov govt competent provider"), whether having democratically elected officials is important to them ("elections important"), and whether they agree that the state ("state looks after"), the incumbent's party ("party looks after"), and the MPA ("MPA looks after") look after them.

Third, in the prospects for accountability domain, we ask three questions. We ask whether or not the MPA is likely to follow through on his promises ("MPA likely to follow through"), a measure of perceived MPA capacity, and whether the respondent can affect change in politics ("political efficacy"). In this domain, we also ask respondents to rate six inputs into their voting decision on a scale of one to six: one of these inputs is MPA performance ("MPA performance important in voting decision"), and the others are the party of the incumbent, the promises the incumbent makes, the political preferences of the respondent's family's, and the state of the Pakistani economy. We use the distance between the mean rating of these six inputs and the rating of MPA performance as the outcome measure.

---

[7]In our pre-analysis plan we intended to include questions about support for policies of the MPA. However, because we used these questions for a endorsement experiment, we chose to omit them from the index as half of our respondents did not see the MPAs name next to the policy and thus they were not evaluations of the incumbent.

For each of our three outcome domains, we construct an index out of the relevant questions, using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of standardized variables, where each standardized variable is constructed by subtracting the mean of the control group and dividing by the standard deviation within the control group. Following Kling, Liebman and Katz (2007), we impute all missing values using the mean of the treatment group. For the incumbent evaluation index and the government performance index, we collect all of our measures at both baseline and endline. For these two indices we use difference-in-differences estimation, constructing the indices by standardizing measures to the post-treatment control mean and standard deviation and imputing missing data by mean of the treatment group *and* time period. We construct the prospects for accountability index using answers to questions asked exclusively in the endline survey, and thus results compare treated and control respondents.

In Panel A of Table 2, we present summary statistics of the baseline data. Throughout Table 2, we split the data into three treatment conditions: the main contact condition, control units in the same villages as those in the contact condition, and control units in the pure control condition. We split responses this way because we hypothesize that results for respondents from the pure control village council will be different than those for the rest of the sample.[8] The reason is that we expect there may be some spillover within settlements from treated to control respondents. Panel A1 shows the breakdown of basic covariates by treatment status and Panels A2 and A3 show summary statistics for two sets of outcomes from the baseline survey. Comparing the treatment and control units (after dropping the pure control units), evaluations of provincial government competence in providing goods is the only pre-treatment covariate with a statistically significant difference at the 0.1 level.[9] For outcomes we collect in the baseline, we prefer a difference-in-differences estimation as outlined in Section 4.1.

Panel B shows summary statistics for our three categories of outcomes on the endline. The most important information to emerge in the summary statistics is the overall downward trend in evaluations of the incumbent, government and in prospects for accountability. This contextualizes the positive findings of treatment that we detect. Our treatment did not generate absolute increases in perceptions of the MPA, his party, or government performance,

---

[8]While we include these pure control units in the main analysis as control units, they have no influence on the effect estimates because settlement-level fixed effects absorb all of the pure control units. Furthermore, the pure control households are not perfectly balanced because they were cluster randomized into that condition at the Village Council area. Again the settlement-level fixed effects will absorb these units and render the imbalance inconsequential.

[9]Kolmogorov-Smirov tests also fail to reject the null that the treatment and control groups have different distributions for all non-binary pre-treatment covariates and baseline outcomes.

but instead reduced the drop in perceptions of these individuals and institutions.

Table 2: **Summary Statistics**

	Whole Sample		Contact Treatment			Control			Pure Control		
Panel A: Baseline Data	Min	Max	Obs.	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD
Panel A1: Covariates											
Prefers MPAs Party	0	1	792	0.201	0.401	316	0.174	0.38	109	0.138	0.346
Owns Motorized Vehicle	0	1	792	0.433	0.496	317	0.432	0.496	109	0.523	0.502
Knows PML-N out of KP Govt.	0	1	789	0.492	0.5	316	0.462	0.499	109	0.44	0.499
Knows Who President Is	0	1	782	0.469	0.499	317	0.495	0.501	108	0.657	0.477
Panel A2: Evaluation Outcomes											
Incumbent evaluation index	−0.811	2.255	792	0.093	0.913	317	0.052	0.891	109	−0.152	0.751
MPA thermometer (1-10)	1	10	792	3.554	2.793	317	3.423	2.644	109	2.743	2.234
Inc. party thermometer (1-10)	1	10	792	4.537	3.067	317	4.536	3.134	109	3.789	2.579
Prefers MPAs party (0/1)	0	1	792	0.201	0.401	316	0.174	0.38	109	0.138	0.346
Panel A3: Performance Outcomes											
Government performance index	−1.296	2.251	792	0.006	0.614	317	0.01	0.588	109	−0.03	0.489
Prov. govt. competent provider (1-4)	1	4	651	2.647	0.769	267	2.753	0.735	102	2.725	0.966
Elections important (1-4)	1	4	761	3.106	1.103	304	3.145	1.099	106	3.415	0.904
State looks after (1-5)	1	5	791	1.53	0.879	316	1.579	0.941	109	1.541	0.845
Party looks after (1-5)	1	5	789	1.976	1.327	314	1.857	1.244	108	1.648	1.008
MPA looks after (1-5)	1	5	789	2.129	1.422	317	1.987	1.355	109	1.725	1.088
	Whole Sample		Contact Treatment			Control			Pure Control		
Panel B: Endline Data	Min	Max	Obs.	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD
Panel B1: Evaluation Outcomes											
Incumbent evaluation index	−0.811	2.255	792	0.008	0.971	317	−0.072	0.891	109	−0.296	0.785
MPA thermometer (1-10)	1	10	785	3.134	2.895	314	2.981	2.616	108	2.241	2.174
Inc. party thermometer (1-10)	1	10	785	4.186	3.314	314	3.933	3.196	108	3.046	2.522
Prefers MPAs party (0/1)	0	1	785	0.21	0.408	314	0.175	0.381	108	0.139	0.347
Panel B2: Performance Outcomes											
Government performance index	−1.296	2.251	792	−0.055	0.535	317	−0.099	0.503	109	−0.24	0.434
Prov. govt. competent provider (1-4)	1	4	722	2.53	0.819	289	2.502	0.821	107	2.206	1.007
Elections important (1-4)	1	4	743	3.207	1.001	302	3.248	1.012	104	3.394	0.897
State looks after (1-5)	1	5	780	1.533	0.858	312	1.612	0.949	108	1.509	0.717
Party looks after (1-5)	1	5	783	1.731	1.149	311	1.585	0.966	108	1.333	0.723
MPA looks after (1-5)	1	5	782	2.073	1.382	314	1.85	1.281	108	1.667	0.986
Panel B3: Accountability Outcomes											
Accountability prospects index	−1.631	1.889	754	0.055	0.583	302	−0.003	0.559	107	−0.006	0.626
MPA likely to follow through (1-4)	1	4	677	2.455	0.769	267	2.434	0.76	93	2.419	0.771
Political efficacy (1-4)	1	5	778	2.35	1.369	311	2.315	1.383	108	2.398	1.427
MPA performance important in voting decision (1-6)	0	5	697	2.395	1.898	288	2.115	1.939	86	2.058	2.111

3.1 Take-up

Overall, 86.5 percent of households answered the Stage One contact call, and 35.6 percent of those who answered responded to at least one question the MPA asked. Attrition was just below 1 percent and is not predicted by treatment. This means about 30.8 percent of households in our sample not only answered the call but answered some of the MPA's questions. This is a high rate of compliance. We define "Stage One partial compliance" as answering the phone whereas we define "Stage One full compliance" as listening long enough to answer at least one IVR question. We carried out Stage Two among the 86.5 percent of households that answered the Stage One contact call. Of these households, 88.2 percent of them answered the Stage Two phone call, and 38.9 percent of those that answered also listened all the way to the end of the call. This means that about 34.5 percent of the households assigned to receive the follow-up call listened all the way to the end. This is also

14

a high rate of compliance.

Full compliance varies widely across settlements, with average rates of full compliance ranging from 5 percent to 55 percent. We present a regression analysis of Stage One full compliance in Table 3, controlling for settlement fixed effects and several baseline covariates.

Table 3: **Question Type and Compliance**

	Answered First IVR Question
Intercept	0.099***
	[0.033]
Asked dev. q. first (0/1)	0.109***
	[0.032]
Prefers MPAs party (0/1)	0.032
	[0.038]
Owns motorized vehicle (0/1)	0.048
	[0.042]
Knows maj. party not in prov. govt. (0/1)	−0.014
	[0.031]
Knows president name (0/1)	0.075**
	[0.038]
Settlement Fixed Effects	X
N	779

Notes: *p<0.1; **p<0.05; ***p<0.01. OLS estimates; standard errors accounting for clustering at the settlement level in square brackets.

There are two strong predictors of full compliance, one of which is causally identified through random assignment. We randomized whether respondents heard the question about development or about MPA time-use as the first question. The treatment effect of being asked about development priorities first rather than about MPA time-use ("Asked dev. q. first") leads to a 10.9 percentage point increase in the probability of answering the first question you hear (s.e. 3.2). This represents a increase of more than 100 percent in the response rate and demonstrates the strong effect of the importance of the question asked on the willingness of citizens to engage with their MPA. A non-causal result of interest is the positive relationship between political knowledge and participation. The regression suggests that individuals who know the president's name — the president is not as politically salient a figure as the prime minister — comply at a rate 7.5 percentage points higher than those with lower political knowledge. We also conducted an SMS encouragement experiment that

indicates that messages about the participation of peers and shorter messages do the most to encourage participation, although the effects are not as large. We report the results of the endorsement experiment in Appendix D.

There are three possible explanations for this high rate of take-up compared with the lower rates reported in previous studies (Blair, Littman and Paluck, 2017; Grossman, Humphreys and Sacramone-Lutz, 2015; Leo et al., April 2015). First, most ICT interventions provide citizens a platform that is open-ended, requiring them to craft messages; this generates a large number of messages with politically irrelevant or unusable content (Blair, Littman and Paluck, 2017). By contrast, politician-initiated IVR provides citizens a limited set of possible responses, making communication simpler, although more limited. Second, the IVR phone call is an incoming call that voters do not have to pay for, whereas SMS interventions can be costly unless subsidized. Third, the IVR call is recorded in the voice of an important politician, increasing voters' expectations that the interaction may worth their time. Previous evidence exists that low expectations hurt participation in ICT governance interventions (Grossman, Humphreys and Sacramone-Lutz, 2015).

4 Analysis

4.1 Estimation

We employ two different specifications to estimate treatment effects of our intervention on the main outcomes of interest. For both specifications, we present intent-to-treat (ITT) effects and local average treatment effects (LATEs) among partial compliers and among full compliers using instrumental variables. These specifications are as registered in our pre-analysis plan except we do not center and interact our treatment variable with our pre-treatment covariates as we prefer the specification with a simpler interpretation.[10]

Our main specification uses a small set of baseline covariate information as well as settlement fixed effects to improve the power of our tests. For outcomes where we do not have a baseline measure, our preferred specification uses simple pre-treatment covariate adjustments and settlement fixed effects to increase the power of our tests:

$$Y_{ij} = \tau D_i + \beta X_i + \theta_j + \epsilon_{ij}, \tag{1}$$

---

[10]The ITT estimates and standard errors for the indices all remain the same order of magnitude and the statistical significance of the estimated effects on the indices remains unchanged.

where Y_{ij} is the endline value of some outcome for household i in settlement j, where X_i is a vector of pre-treatment variables, and where θ_j is a settlement fixed effect. Note that when we analyze spillover effects, we remove the fixed effects because the pure control was randomly selected at the village level. We estimate the ITT effects using OLS for all outcomes of interest and use 2SLS for the LATEs.

For outcomes where we have a baseline measure, our preferred specification uses a difference-in-differences approach, continuing to adjust for pre-treatment covariates:

$$Y_{ijt} = \alpha_0 D_i + \alpha_1 T_t + \tau D_i T_t + \beta X_i + \theta_j + \epsilon_{ijt}, \tag{2}$$

where Y_{ijt} is the outcome at time $t = 0$, the baseline, or time $t = 1$, the endline, and T_t is a dummy for that is 1 at the endline and 0 at the baseline. Again we estimate coefficients using OLS for the ITT effects and 2SLS for the LATEs.

In all specifications, our control variables X_i include baseline measures of whether the respondent supports the incumbent's party, whether the household has a motorized vehicle, and a measure of the political knowledge of the respondent (based on answers to questions about whether the respondent knows that a major party in Pakistan is not in this provincial government and whether he knows the identity of the president). For some outcomes, we are able to use a difference-in-differences specification, whereas for others we use pre-treatment covariates in addition to the above covariates to boost our power. The additional pre-treatment covariates we use follow those specified in our pre-analysis plan. We report both heteroskedasticity consistent standard errors as well as cluster-robust standard errors at the settlement level.

Although we originally intended to fit all of our analyses using sampling weights, many of our analyses are on subsets of our full sample. Sampling weights do not make sense applied to those regressions. Therefore, we report all analyses without any weights. See Appendix B for results of the analysis of the Stage One treatment effects with sampling weights. The results are qualitatively similar, although effect estimates for the index of government performance outcomes is greater (and significant at the 0.01 rather than 0.05 level) and the estimates for the prospect for accountability outcomes are smaller. We also originally intended to weight our estimates by the inverse probability of attrition as a robustness check, but attrition is below one percent and not predicted by treatment status.

4.2 Stage One effects of any contact

We present the overall effects of the intervention — the Stage One *contact* treatment — in Table 4. Overall, respondents warmed to the MPA and his party, evaluated the government as more competent, and prospects for accountability improved, although many of these estimates are imprecisely estimated. All the indices are expressed in standard deviation units whereas the effects on constituent outcomes are in their natural units. Summary statistics including standard deviations can be found in Section 3. The first column presents the ITT effects, whereas the second and third columns present LATE effects on partial and full compliers, respectively. We define partial compliers as households that answered the Stage One *contact* call only and full compliers as households that listened to the call long enough to answer at least one of the IVR questions.

4.2.1 Evaluation of the incumbent

In the first panel of results reporting incumbent evaluations, although all point estimates are positive, none of the effects are statistically significant save for one specification for one constituent outcome. More specifically, receiving any *contact* from the MPA increases evaluations of the incumbent and his party by 0.058 standard deviation units (s.e. 0.048), although the effect is not statistically significant. We also see a larger effect of the intervention on support for the incumbent party than for the MPA, but these effects are also not statistically significant. Lastly, there little evidence that respondents intend to switch partisanship to the MPA's party following the intervention, although again the signs are positive.

There is clearer evidence of an overall treatment effect on support for the MPA when analyzing an embedded endorsement experiment. In our endline survey, we ask about the level of support for two bills, randomizing whether the MPA's name appeared next to the bill or not. We assign two settlements per Village Council to receive the endorsed version of the questions and the remaining two settlements to receive the control version.[11] The English translations of the questions are given below; we deliver the italicized phrase to respondents only in the settlements assigned to receive the endorsement treatment.

- A recent bill [*sponsored and led by [MPA NAME]*] did away with housing taxes in rural areas. Some argue that this is fair as it relieves a burden on rural citizens, but others argue it hurts

---

[11]We randomize the experiment at the settlement level rather than the household level to ensure that enumeration teams would have no difficulty administering the correct version. As a result, our analysis includes village rather than settlement fixed effects.

	ITT	Partial Compliers LATE	Full Compliers LATE	N
A. Incumbent evaluation index	0.058 (0.064) [0.048]	0.067 (0.074) [0.056]	0.189 (0.21) [0.155]	1192
MPA thermometer (1-10)	0.078 (0.183) [0.151]	0.09 (0.211) [0.174]	0.256 (0.597) [0.491]	1191
Inc. party thermometer (1-10)	0.331 (0.206) [0.204]	0.382 (0.237) [0.235]	1.084 (0.674) [0.648]*	1191
Prefers MPAs party (0/1)	0.013 (0.028) [0.016]	0.015 (0.032) [0.019]	0.042 (0.092) [0.053]	1191
B. Government performance index	0.077 (0.04)* [0.038]**	0.089 (0.047)* [0.044]**	0.252 (0.133)* [0.125]**	1191
Prov. govt. competent provider (1-4)	0.155 (0.075)** [0.08]*	0.179 (0.087)** [0.093]*	0.487 (0.237)** [0.254]*	959
Elections important (1-4)	0.033 (0.089) [0.041]	0.038 (0.102) [0.047]	0.108 (0.288) [0.133]	1102
State looks after (1-5)	−0.002 (0.072) [0.056]	−0.002 (0.084) [0.065]	−0.005 (0.236) [0.183]	1182
Party looks after (1-5)	0.051 (0.088) [0.097]	0.059 (0.102) [0.112]	0.166 (0.289) [0.316]	1180
MPA looks after (1-5)	0.065 (0.096) [0.054]	0.075 (0.11) [0.063]	0.211 (0.313) [0.178]	1186
C. Accountability prospects index	0.053 (0.037) [0.045]	0.061 (0.042) [0.052]	0.17 (0.119) [0.145]	1163
MPA likely to follow through (1-4)	0.044 (0.055) [0.064]	0.051 (0.063) [0.073]	0.148 (0.183) [0.213]	1002
Political efficacy (1-4)	0.021 (0.087) [0.091]	0.024 (0.1) [0.105]	0.068 (0.282) [0.295]	1181
MPA performance important in voting decision (-4.17, 4.17)	0.098 (0.104) [0.099]	0.113 (0.12) [0.115]	0.331 (0.352) [0.338]	1024

Notes: *p<0.1; **p<0.05; ***p<0.01. Each cell represents a model fit with either OLS or 2SLS. For the LATEs, *contact* instruments for partial compliance — answering the phone — in the second column and full compliance — answering an IVR question — in the third column. We present two standard errors: (HC2 robust standard errors) and [cluster robust standard errors] at the settlement level. Estimates for domains A and B are specified using the difference-in-differences specification in Equation 2 while estimates in panel C are fit using the simple specification reported in Equation 1. All estimates include settlement FEs. Most models include "supports MPA party," "owns motorized vehicle," "Knows maj. party not in prov. govt.," and "Knows president name " as covariates. Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "supports MPA party" is the outcome, we exclude the baseline variable "supports MPA party" as a covariate; for all outcomes in domain C except for "political efficacy," we add baseline values of "political efficacy" as a covariate; when "MPA likely to follow through" is the outcome, we include baseline values of "MPA thermometer" as a covariate. We build the index for each of our outcome domains using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of the constituent outcomes, standardized by the mean and standard deviation of that outcome for the control group in the baseline time period (except for domain C where we just use the control group). All missing values are imputed using the mean of the respective treatment and time group.

the ability of the [PROVINCE NAME] government to provide services. Do you agree with this bill?

- A recent bill [*sponsored and led by [MPA NAME]*] restricted lenders from charging interest. While this may make it harder to receive loans, it also prevents immoral lending practices. Do you agree with this bill?

We opt for a simple analysis of the endorsement experiment, interacting the main contact treatment with the endorsement treatment.[12] Table 5 contains the results of an OLS regression of the level of support for the bill, ranging from one to four, regressed on a dummy variable for whether the respondent received the endorsed version of the questions, a dummy variable for whether the household was in the main contact treatment, and the interaction of the two dummy variables. In the control group, the endorsement had a clear negative effect on overall support for the bill. However, the effect of the endorsement in the contact group is significantly more positive, changing the overall effect of the endorsement to around 0. The ITT on support for the incumbent as measured through his endorsement of the bill on charging interest is 0.252 (0.111 s.e.) on the original scale, or about 0.323 standard deviation units (0.142 s.e.). The ITT on support for the incumbent as measured through endorsement for the bill regarding the housing tax is 0.375 (0.150 s.e.) on the original scale and is roughly 0.423 standard deviation units (0.169 s.e.). This provides clear evidence that the intervention improves support for the incumbent.

One possible explanation for the precision of these results in contrast with the results from the straightforward feeling thermometers is that this endorsement experiment mitigated desirability bias in the control group, whose members might have been more likely to rate the MPA more highly when asked directly about him. Of course, it is also possible that the difference in results is due to different specifications and different questions being asked. Nonetheless, the sign of the effect in the endorsement experiment and in the main experiment is always positive and consistent with a shift towards the MPA in the *contact* group.

4.2.2 Government performance

We report results for the second domain of outcomes, outcomes covering government performance, in the second panel of Table 4. Rather than simple evaluations of the incumbent, these questions seek to grasp perceptions of the competence of several governing institutions and individuals. We again see all positive signs except for a very small negative effect on "state looks after." The treatment effect on the index of these outcomes is 0.077 standard

---

[12]Applying the method described in Bullock, Imai and Shapiro (2011) leads to the same conclusions.

Table 5: **Endorsement Experiment Effects**

	Support Interest Bill (1)	Support Tax Bill (2)
Intercept	3.294 [0.106]***	3.013 [0.196]***
Endorse	−0.265 [0.128]**	−0.209 [0.148]
Contact	−0.144 [0.077]*	−0.102 [0.098]
Endorse × Contact	0.252 [0.112]**	0.375 [0.151]**
Village Fixed Effects	X	X
Covariates	X	X
N	849	821

Notes: *p<0.1; **p<0.05; ***p<0.01. Both columns represent a distinct OLS model. Both outcomes range from 1 to 4, where 4 represents a higher level of support for the bill. Standard errors accounting for clustering at the settlement level are presented in brackets. Both models include Village Council area fixed effects (since the endorsement was randomized at the settlement level). Both models also adjust for baseline values of co-partisanship with the incumbent ("supports MPA party"), baseline assets ("owns motorized vehicle"), and baseline political knowledge ("Knows maj. party not in prov. govt." and "Knows president name").

deviation units (s.e. 0.038) and statistically significant at the 0.05 level with our preferred, clustered, standard errors and at the 0.1 level with heteroskedasticity-robust standard errors. The local average treatment effect of this index among full compliers is 0.252 standard deviations (s.e. 0.125). This positive effect is largely driven by the positive and statistically significant effect of the intervention on perceptions of provincial government competence. Given that in his recording, the MPA discusses several projects that he has already implemented, it is possible that when asked about competence in providing services, constituents are pleased by news of these projects. In standard deviations, the ITT effect of our intervention on perceptions of government performance is about 0.2 standard deviations, although there is significant missingness on this outcome variable. However, missingness on this outcome is not predicted by treatment status. The other ITT effects are all smaller than 0.05 standard deviations and are not statistically significant.

4.2.3 Prospects for accountability

We also see positive effects of a similar magnitude in the third outcome domain, prospects for accountability, reported in the third panel of Table 4. The estimated treatment effects on the prospects for accountability index are statistically significant only using heteroskedasticity-robust standard errors or when considering LATEs among compliers. The point estimate for the index indicates a positive effect of 0.053 standard deviation units (s.e. 0.045) while the LATE among full compliers indicates a positive effect of 0.170 standard deviation units (s.e. 0.145). Although none of the ITT effects on the constituent outcomes — beliefs about the ability for the MPA to follow through, political efficacy, or the importance of MPA performance in the voting decision — are statistically significant save for one particular case, all ITT estimates are positive and represent increases of about 0.1 standard deviation units. Overall, it appears that the intervention encourages conditions where electoral accountability is more likely to occur.

Although noisy, these results show an overall positive shift in three main outcome domains of incumbent evaluations, government performance, and prospects for accountability. Results indicate that respondents appreciate being asked about their preferences, take the calls as a sign of government competence, and are thinking more than control respondents about performance and policy when deciding whom to vote for in the future. These positive effects could be explained as responses to a signal that the incumbent and the provincial government are expending efforts to obtain feedback from constituents. However, the IVR call could also just be delivering information about recent projects completed by the incumbent and could drive outcomes through that channel. The pilot setup is unable to explicitly differentiate between these two mechanisms, although the second stage of the experiment was aimed at learning about what kind of contact was most valuable.

4.3 Stage Two effects of responsive follow-up calls

It is unclear from the analysis of Stage One results whether the content of the phone call or the mere fact of receiving the phone call drives these effects. In order to parse these possibilities, we study whether the content of a follow-up call matters for evaluations of the incumbent, evaluations of government performance, and prospects for accountability. In the Stage Two analysis, we compare respondents who received a *responsive* follow-up call from the MPA to those who simply got a second call from the MPA, allowing us to control for the number of contacts and focus instead on the level of engagement engendered by the political

communication.

In Table 6, we see that among compliers in Stage One, receiving the *responsive* follow-up generally leads to an increase in most of our outcomes relative to the group that receives the *generic* follow-up, although almost all estimates are imprecisely estimated and the ITT effects on the indices are quite small, between 0.02 and 0.05 standard deviation units. Nonetheless, there is suggestive evidence that responsive linkages are more valuable than mere credit-claiming messages. The first column presents the ITT effects, while columns two and three present LATE effects on partial and full compliers, respectively. We define partial compliers as households that answered the Stage Two *responsive* follow-up call and full compliers as households that listened to the *responsive* follow-up call to completion.

More precisely, every sign is positive except for one outcome in each domain. The only statistically significant result — and the largest standardized point estimate — is the effect on the ranking of MPA performance (ITT of 0.238 on the original scale, 0.140 s.e. or 0.159 SDs; 0.094 s.e.), indicating that responsive linkages may trigger citizens to think more seriously about the promises made by the MPA in the series of phone calls. The only other outcome that is sizably affected is the incumbent party thermometer, where the ITT estimate is around 0.15 standard deviations.

When looking at the marginal effect of the *responsive* follow-up over the *generic* follow up, the LATEs on the three indices for full compliers are all around a 0.1 standard deviation increase. Because all of these units are a subsample of those in the Stage One *contact* treatment, it appears that largest treatment effects of the overall IVR intervention is found among units who receive the *responsive* feedback. This highlights the need not just for credit-claiming and information-seeking, but also closed feedback loops where citizens know that their elected official is aware of their preferences.

5   Conclusion

The broadly positive outcomes and unusually high take-up rate of our intervention offer promise that IVR can be an effective tool at improving political communication. It may provide actionable information to politicians while also improving respondent evaluations of the incumbent and establishing the conditions for an election based on the incumbent's performance. We find noisy but consistent evidence that receiving calls seeking information about voter preferences as well as credit claiming in the MPA's voice increases perceptions of government competence and improves prospects for accountability. Furthermore, respon-

Table 6: **Marginal Effects of Stage Two Responsive versus Generic Follow-up**

	ITT	Partial Compliers LATE	Full Compliers LATE	N
A. Incumbent evaluation index	0.022 (0.058) [0.062]	0.025 (0.067) [0.071]	0.065 (0.175) [0.188]	669
MPA thermometer (1-10)	0.124 (0.251) [0.162]	0.142 (0.287) [0.186]	0.374 (0.755) [0.494]	669
Inc. party thermometer (1-10)	0.268 (0.261) [0.254]	0.307 (0.299) [0.29]	0.807 (0.788) [0.759]	669
Prefers MPAs party (0/1)	−0.027 (0.038) [0.031]	−0.031 (0.043) [0.035]	−0.081 (0.113) [0.092]	669
B. Government performance index	0.024 (0.051) [0.042]	0.027 (0.058) [0.048]	0.071 (0.154) [0.127]	669
Prov. govt. competent provider (1-4)	0.041 (0.091) [0.051]	0.047 (0.104) [0.058]	0.127 (0.283) [0.156]	528
Elections important (1-4)	0.059 (0.117) [0.06]	0.067 (0.132) [0.068]	0.176 (0.351) [0.18]	612
State looks after (1-5)	0.015 (0.09) [0.033]	0.017 (0.103) [0.038]	0.046 (0.272) [0.103]	663
Party looks after (1-5)	−0.068 (0.12) [0.128]	−0.078 (0.137) [0.147]	−0.204 (0.361) [0.382]	664
MPA looks after (1-5)	0.043 (0.132) [0.093]	0.049 (0.151) [0.106]	0.129 (0.399) [0.282]	664
C. Accountability prospects index	0.047 (0.042) [0.043]	0.053 (0.049) [0.05]	0.143 (0.13) [0.133]	655
MPA likely to follow through (1-4)	0.009 (0.064) [0.057]	0.011 (0.074) [0.066]	0.031 (0.208) [0.186]	563
Political efficacy (1-4)	−0.078 (0.098) [0.11]	−0.09 (0.112) [0.127]	−0.24 (0.3) [0.335]	663
MPA performance important in voting decision (-4.17, 4.17)	0.238 (0.126)* [0.14]*	0.274 (0.145)* [0.16]*	0.728 (0.39)* [0.417]*	574

Notes: *p<0.1; **p<0.05; ***p<0.01. Each cell represents a model fit with either OLS or 2SLS. For the LATEs, *contact* instruments for partial compliance — answering the phone — in the second column and full compliance — answering an IVR question — in the third column. We present two standard errors: (HC2 robust standard errors) and [cluster robust standard errors] at the settlement level. Estimates for domains A and B are specified using the difference-in-differences specification in Equation 2 and estimates in panel C are fit using the simple specification reported in Equation 1. All estimates include settlement FEs. Most models include "supports MPAs party," "owns motorized vehicle," "Knows maj. party not in prov. govt.," and "Knows president name " as covariates. Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "supports MPA party" is the outcome when we exclude the baseline variable of "supports MPA party" as a covariate; for all outcomes in domain C except for "political efficacy," we add baseline values of "political efficacy" as a covariate; when "MPA likely to follow through" is the outcome, we include baseline values of "MPA thermometer" as a covariate. We build the index for each of our outcome domains using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of the constituent outcomes, standardized by the mean and standard deviation of that outcome for the control group in the baseline time period (except for domain C where we just use the control group). All missing values are imputed using the mean of the respective treatment and time group.

sive follow-up calls that directly address the requests of citizens may be most effective at promoting these outcomes. In on-going work with larger samples, we hope to ascertain how different types of IVR contact changes citizen attitudes and electoral behavior.

High take-up of the IVR on the citizens' end potentially signals the high value that citizens attach to this form of political communication. In addition, we observe a large increase in take-up when citizens are asked about development priorities rather than being asked about MPA time-use. This indicates that citizens evaluate communication by elected officials in terms of the salience of the content to their own lives.

We suspect that prior informational studies may have suffered from low take-up when the information disseminated was remote to citizens' everyday concerns. Future work should causally establish wither high take-up is driven by upfront buy-in of politicians or how obvious that buy-in is to citizens. Specifically, does providing a finite set of answers through IVR rather than an open platform drive high take-up or is it driven by delivering the message in the voice of powerful politicians to increase the perceived importance of the communication?

In an attempt to replicate the pilot findings with more power, test how IVR influences politician behavior and electoral behavior by citizens, and see how take-up changes in larger samples, we are conducting a scale-up of the design across many geographic areas of the province of study. Whereas only one MPA participated in the pilot, 40 MPAs are enrolled in the scale-up, enabling us to study the effects of the intervention on MPA knowledge of constituent preferences, MPA effort in treated areas, and how they respond allocate their time. In addition, the scale-up will test how the positive effects of IVR calls on the domains of government performance and the prospects for accountability identified in the pilot results translate into behavioral outcomes. By measuring the attitudinal outcomes studied in this pilot, as well as how MPA effort and development projects change in the treatment and control polling station areas, we will be able to estimate whether the IVR intervention creates the conditions for accountability that result in greater electoral sanctioning or rewarding of incumbent MPAs.

References

Adida, Claire, Jessica Gottlieb, Eric Kramon and Gwyneth H. McClendon. 2016. "When Good News is Bad ... and When it Isn't: Voter Coordination, Preferences, and Electoral Behavior." Unpublished paper.

Aker, Jenny C., Paul Collier and Pedro C. Vicente. 2016. "Is information power? Using mobile phones and free newspapers during an election in Mozambique." *Review of Economics and Statistics* (0).

Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su. 2010. "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India." Unpublished paper.

Bidwell, Kelly, Katherine Casey and Rachel Glennerster. 2015. "Debates: Voting and Expenditure Responses to Political Communication." Unpublished paper.

Blair, Graeme, Rebecca Littman and Elizabeth Levy Paluck. 2017. "Motivating The Adoption of New Community-Minded Behaviors: An Empirical Test in Nigeria." Unpublished paper.

Bullock, Will, Kosuke Imai and Jacob N Shapiro. 2011. "Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan." *Political Analysis* 19(4):363–384.

Callen, Michael and Ali Hasanain. 2011. "The Punjab Model of Proactive Governance: Empowering Citizens Through Information Communication Technology." Unpublished paper.

Chong, Alberto, Ana L. De La O, Dean Karlan and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice and Participation." *Journal of Politics* 77(1):55–71.

Dale, Allison and Aaron Strauss. 2009. "Don't forget to vote: Text message reminders as a mobilization tool." *American Journal of Political Science* 53(4):787–804.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde and Craig McIntosh. Forthcoming. *Information and Accountability: A New Method of Cumulative Learning.* New York: Cambridge University Press.

Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2016. "Social media and protest participation: Evidence from russia.".

Grimmer, Justin, Solomon Messing and Sean J. Westwood. 2012. "How words and money

cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation." *American Political Science Review* 106(4):703–19.

Grossman, Guy and Kristin Michelitch. 2016. "Transparency and Accountability Initiatives Targeting Politician Performance between Elections: Evidence from a Field Experiment in Subnational Uganda." Unpublished paper.

Grossman, Guy, Kristin Michelitch and Marta Santamaria. Forthcoming. "Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization." *Comparative Political Studies* .

Grossman, Guy, Macartan Humphreys and Gabby Sacramone-Lutz. 2015. "Information Technology and Political Engagement: Mixed Evidence from Uganda." Unpublished paper.

Grossman, Guy, Macartan Humphreys and Gabriella 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.

Kendall, Chad, Tommaso Nannicini and Francesco Trebbi. 2015. "How do voters respond to information? Evidence from a randomized campaign." *The American Economic Review* 105(1):322–353.

Kling, Jeffrey R, Jeffrey B Liebman and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83–119.

Leo, Ben, Robert Morello, Jonathan Mellon, Tiago Peixoto and Stephen Davenport. April 2015. Do Mobile Phone Surveys Work in Poor Countries? Working Paper N. 398 Center for Global Development.

Lieberman, Evan S., Daniel N. Posner and Lily L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60:69–83.

Malhotra, Neil, Melissa R. Michelson, Todd Rogers and Ali Adam Valenzuela. 2011. "Text Messages as Mobilization Tools: The Conditional Effect of Habitual Voting and Election Salience." *American Politics Research* 39(4):664–681.

Tucker, Joshua A, Jonathan Nagler, Megan MacDuffee, Pablo Barbera Metzger, Duncan Penfold-Brown and Richard Bonneau. 2016. "Big Data, Social Media, and Protest." *Computational Social Science* 199.

A   Endline data collection

This section describes our endline data collection method, demonstrates that is not predicted by treatment, provides evidence of how answers differ across in-person and phone interviews, and shows our main results are qualitatively unchanged when replacing in-person responses with phone responses for the subset of respondents for whom data was collected both ways.

We collect our endline data in two different ways. First, we split the 44 settlements in our study into two groups; households in 22 of the settlements receive the endline as an in-person survey, while households in the other 22 settlements receive the endline through the phone. Several days later, households in 11 settlements that were assigned to have in-person endline surveys, received the exact same endline again, this time over the phone. Households in the other 33 settlements did not receive the endline a second time. The breakdown of settlements and households into different data collection methods is detailed in Table A.1.

Table A.1: **Endline data collection method assignment**

	Phone No follow-up	In-Person No follow-up	In-Person Phone follow-up
# Settlements	22	11	11
# Households	612	304	302

A.1   Collection method not predicted by treatment status

Because our treatments—the stage one *contact* treatment and the stage two *responsive* follow-up—were both randomized within settlements, and the households were cluster randomized into endline data collection method at the settlement level, the endline data collection method is not predicted by treatment. A regression of the endline being administered in person on the stage one *contact* treatment status reveals that *contacted* individuals are 0.02 percentage points less likely to be administered the endline in person (s.e. 3.01), a substantively and statistically insignificant difference. A similar regression of the endline being administered in person on the stage two treatment status reveals that individuals receiving the *responsive* follow-up are 1.59 percentage points less likely to be administered the endline in person (s.e. 3.82), a similarly insignificant difference.

A.2 Answers differ but are correlated

In Table A.2, we display the Pearson correlation coefficient between answers given in-person to answers given on the phone for the 304 households who were surveyed in both ways.

Table A.2: **Correlation between in-person and phone responses**

	Correlation across collection methods
Evaluation Outcomes	
MPA thermometer (1-10)	0.77
Inc. party thermometer (1-10)	0.74
Prefers MPAs party (0/1)	0.83
Performance Outcomes	
Prov. govt. competent provider (1-4)	0.74
Elections important (1-4)	0.86
State looks after (1-5)	0.79
Party looks after (1-5)	0.56
MPA looks after (1-5)	0.80
Accountability Outcomes	
MPA likely to follow through (1-4)	0.84
Political efficacy (1-4)	0.75
MPA performance important in voting decision (1-6)	0.85

Notes: Reported are Pearson correlation coefficients for answers to the same question across in-person and phone endline surveys carried out for 304 households.

A.3 Robustness of results to different data collection method

While the correlations are high, they are not as high as one would have hoped for the same exact questions being asked within a week of one another. Therefore, we replaced the 304 in-person endline answers with the 304 phone endline answers to see whether the results of our main analysis changed when we used the other measurements. In Tables A.3 and A.4, we present replications of Tables 4 and 6 replacing the endline values collecteder than the in-person survey. Therefore, in this robustness check, we keep our initial sample of 1,218, but for 304 of these households we replace their in-person endline survey with the phone endline survey we carried out with them several days later. The endline data for the other 914 households remains as before.

There is some change in statistical significance, with some effects crossing the threshold to statistical significance and results becoming statistically insignificant. However, most of

these changes are on effects that are near thresholds for statistical significance rather than representative of large shifts in point estimates and standard errors. Overall, point estimates and standard errors remain largely unchanged and the summary of the evidence is the same.

Table A.3: **Treatment Effects of Stage One Contact, robustness to data collection method**

	ITT	Partial Compliers LATE	Full Compliers LATE	N
A. Incumbent evaluation index	0.054 (0.064) [0.052]	0.063 (0.074) [0.06]	0.177 (0.211) [0.169]	1185
MPA thermometer (1-10)	0.091 (0.182) [0.167]	0.105 (0.211) [0.193]	0.299 (0.597) [0.544]	1184
Inc. party thermometer (1-10)	0.341 (0.212) [0.218]	0.393 (0.244) [0.251]	1.115 (0.693) [0.695]	1184
Prefers MPAs party (0/1)	0.006 (0.028) [0.018]	0.007 (0.033) [0.02]	0.019 (0.093) [0.057]	1184
B. Government performance index	0.062 (0.041) [0.034]*	0.072 (0.047) [0.039]*	0.204 (0.134) [0.111]*	1184
Prov. govt. competent provider (1-4)	0.148 (0.078)* [0.074]**	0.17 (0.09)* [0.085]**	0.458 (0.244)* [0.227]**	929
Elections important (1-4)	0.011 (0.09) [0.046]	0.013 (0.103) [0.053]	0.036 (0.288) [0.149]	1043
State looks after (1-5)	0.005 (0.073) [0.057]	0.006 (0.085) [0.066]	0.018 (0.24) [0.186]	1175
Party looks after (1-5)	0.025 (0.09) [0.094]	0.029 (0.104) [0.109]	0.081 (0.295) [0.306]	1173
MPA looks after (1-5)	0.002 (0.096) [0.057]	0.002 (0.11) [0.066]	0.005 (0.313) [0.186]	1179
C. Accountability prospects index	0.055 (0.036) [0.042]	0.064 (0.042) [0.049]	0.179 (0.117) [0.138]	1156
MPA likely to follow through (1-4)	0.027 (0.058) [0.073]	0.031 (0.067) [0.084]	0.09 (0.193) [0.242]	965
Political efficacy (1-4)	0.025 (0.088) [0.088]	0.029 (0.101) [0.102]	0.083 (0.286) [0.288]	1174
MPA performance important in voting decision (-4.17, 4.17)	0.116 (0.105) [0.088]	0.135 (0.122) [0.101]	0.389 (0.351) [0.293]	961

Notes: *p<0.1; **p<0.05; ***p<0.01. This table uses phone endline survey data instead of in-person survey data for 304 households. Each cell represents a model fit with either OLS or 2SLS. For the LATEs, *contact* instruments for partial compliance — answering the phone — in the second column and full compliance — answering an IVR question — in the third column. We present two standard errors: (HC2 robust standard errors) and [cluster robust standard errors] at the settlement level. Estimates for domains A and B are specified using the difference-in-differences specification in Equation 2 while estimates in panel C are fit using the simple specification reported in Equation 1. All estimates include settlement FEs. Most models include "supports MPA party," "owns motorized vehicle," "Knows maj. party not in prov. govt.," and "Knows president name " as covariates. Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "supports MPA party" is the outcome, we exclude the baseline variable "supports MPA party" as a covariate; for all outcomes in domain C except for "political efficacy," we add baseline values of "political efficacy" as a covariate; when "MPA likely to follow through" is the outcome, we include baseline values of "MPA thermometer" as a covariate. We build the index for each of our outcome domains using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of the constituent outcomes, standardized by the mean and standard deviation of that outcome for the control group in the baseline time period (except for domain C where we just use the control group). All missing values are imputed using the mean of the respective treatment and time group.

Table A.4: **Marginal Effects of Stage 2 Responsive versus Generic Follow-up, robustness to data collection method**

	ITT	Partial Compliers LATE	Full Compliers LATE	N
A. Incumbent evaluation index	0.003 (0.06) [0.063]	0.004 (0.069) [0.072]	0.01 (0.182) [0.19]	666
MPA thermometer (1-10)	0.084 (0.254) [0.178]	0.096 (0.29) [0.204]	0.254 (0.765) [0.54]	666
Inc. party thermometer (1-10)	0.214 (0.267) [0.251]	0.245 (0.306) [0.287]	0.646 (0.807) [0.753]	666
Prefers MPAs party (0/1)	−0.036 (0.038) [0.031]	−0.041 (0.044) [0.035]	−0.108 (0.115) [0.092]	666
B. Government performance index	0.035 (0.052) [0.044]	0.04 (0.059) [0.05]	0.105 (0.156) [0.133]	666
Prov. govt. competent provider (1-4)	0.02 (0.096) [0.059]	0.023 (0.111) [0.068]	0.063 (0.307) [0.186]	513
Elections important (1-4)	0.076 (0.117) [0.062]	0.086 (0.133) [0.071]	0.233 (0.359) [0.194]	584
State looks after (1-5)	0.021 (0.092) [0.04]	0.024 (0.105) [0.046]	0.064 (0.278) [0.122]	662
Party looks after (1-5)	−0.057 (0.122) [0.119]	−0.065 (0.139) [0.137]	−0.171 (0.367) [0.357]	661
MPA looks after (1-5)	0.094 (0.132) [0.097]	0.108 (0.151) [0.111]	0.286 (0.401) [0.3]	661
C. Accountability prospects index	0.077 (0.044)* [0.044]*	0.089 (0.05)* [0.051]*	0.237 (0.134)* [0.137]*	652
MPA likely to follow through (1-4)	0.049 (0.068) [0.057]	0.057 (0.079) [0.066]	0.159 (0.218) [0.185]	535
Political efficacy (1-4)	−0.014 (0.1) [0.123]	−0.016 (0.115) [0.142]	−0.043 (0.307) [0.379]	660
MPA performance important in voting decision (-4.17, 4.17)	0.205 (0.128) [0.132]	0.233 (0.146) [0.15]	0.627 (0.393) [0.406]	532

Notes: *p<0.1; **p<0.05; ***p<0.01. This table uses phone endline survey data instead of in-person survey data for 304 households. Each cell represents a model fit with either OLS or 2SLS. For the LATEs, *contact* instruments for partial compliance — answering the phone — in the second column and full compliance — answering an IVR question — in the third column. We present two standard errors: (HC2 robust standard errors) and [cluster robust standard errors] at the settlement level. Estimates for domains A and B are specified using the difference-in-differences specification in Equation 2 and estimates in panel C are fit using the simple specification reported in Equation 1. All estimates include settlement FEs. Most models include "supports MPAs party," "owns motorized vehicle," "Knows maj. party not in prov. govt.," and "Knows president name " as covariates. Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "supports MPA party" is the outcome when we exclude the baseline variable of "supports MPA party" as a covariate; for all outcomes in domain C except for "political efficacy," we add baseline values of "political efficacy" as a covariate; when "MPA likely to follow through" is the outcome, we include baseline values of "MPA thermometer" as a covariate. We build the index for each of our outcome domains using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of the constituent outcomes, standardized by the mean and standard deviation of that outcome for the control group in the baseline time period (except for domain C where we just use the control group). All missing values are imputed using the mean of the respective treatment and time group.

B   Analysis with survey weights

Here we replicate our main analyses using survey weights. Because we sample 11 of the 25 Village Council areas in our study district with simple random sampling. Furthermore, within sampled Village Council areas, we sample four settlements using simple random sampling and have to adjust for the size of those settlements. Thus for each unit i, we apply the following weights in all analyses:

$$w_{isv} = \frac{n}{N} \frac{m_v}{M_v} \frac{o_s}{O_s} = \frac{11}{25} \frac{4}{M_v} \frac{25}{O_s},$$

where n is sampled Village Council areas (11), N is the total number of Village Council areas (25), m_v is the number of sampled settlements in Village Council v (4), M_v is the number of settlements in Village Council v (varies), o_s is the number of sampled households in settlement s (25), and O_s is the number of households in settlement s (varies).

Table B.1 replicates the main analysis table but with survey weights added. The effects are qualitatively similar.

Table B.1: **Treatment Effects of Stage One Contact Treatment, with sampling weights**

	ITT	Partial Compliers LATE	Full Compliers LATE	N
A. Incumbent evaluation index	0.035 (0.074) [0.044]	0.041 (0.086) [0.051]	0.113 (0.237) [0.139]	1192
MPA thermometer (1-10)	0.03 (0.201) [0.156]	0.035 (0.234) [0.182]	0.096 (0.647) [0.5]	1191
Inc. party thermometer (1-10)	0.274 (0.241) [0.172]	0.319 (0.28) [0.199]	0.883 (0.774) [0.535]*	1191
Prefers MPAs party (0/1)	0.001 (0.034) [0.016]	0.001 (0.039) [0.019]	0.004 (0.109) [0.052]	1191
B. Government performance index	0.113 (0.047)** [0.04]***	0.132 (0.055)** [0.048]***	0.365 (0.154)** [0.132]***	1191
Prov. govt. competent provider (1-4)	0.203 (0.085)** [0.076]***	0.236 (0.098)** [0.089]***	0.622 (0.263)** [0.234]***	959
Elections important (1-4)	0.032 (0.121) [0.044]	0.038 (0.139) [0.051]	0.102 (0.379) [0.138]	1102
State looks after (1-5)	0.006 (0.072) [0.05]	0.007 (0.084) [0.059]	0.018 (0.233) [0.162]	1182
Party looks after (1-5)	0.118 (0.109) [0.097]	0.137 (0.127) [0.114]	0.378 (0.352) [0.314]	1180
MPA looks after (1-5)	0.149 (0.124) [0.088]*	0.174 (0.144) [0.105]*	0.479 (0.397) [0.282]*	1186
C. Accountability prospects index	0.04 (0.046) [0.045]	0.046 (0.053) [0.052]	0.127 (0.147) [0.146]	1163
MPA likely to follow through (1-4)	−0.01 (0.063) [0.068]	−0.011 (0.072) [0.078]	−0.033 (0.207) [0.225]	1002
Political efficacy (1-4)	0.052 (0.105) [0.089]	0.061 (0.121) [0.104]	0.169 (0.337) [0.287]	1181
MPA performance important in voting decision (-4.17, 4.17)	0.105 (0.167) [0.232]	0.122 (0.193) [0.268]	0.348 (0.551) [0.772]	1029

Notes: *p<0.1; **p<0.05; ***p<0.01. Each cell represents a model fit with either OLS or 2SLS. For the LATEs, *contact* instruments for partial compliance — answering the phone — in the second column and full compliance — answering an IVR question — in the third column. We present two standard errors: (HC2 robust standard errors) and [cluster robust standard errors] at the settlement level. Estimates for domains A and B are specified using the difference-in-differences specification in Equation 2 and estimates in panel C are fit using the simple specification reported in Equation 1. All estimates include settlement FEs. Most models include "supports MPAs party," "owns motorized vehicle," "Knows maj. party not in prov. govt.," and "Knows president name " as covariates. Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "supports MPA party" is the outcome when we exclude the baseline variable of "supports MPA party" as a covariate; for all outcomes in domain C except for "political efficacy," we add baseline values of "political efficacy" as a covariate; when "MPA likely to follow through" is the outcome, we include baseline values of "MPA thermometer" as a covariate. We build the index for each of our outcome domains using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of the constituent outcomes, standardized by the mean and standard deviation of that outcome for the control group in the baseline time period (except for domain C where we just use the control group). All missing values are imputed using the mean of the respective treatment and time group. All regressions are estimated with sampling weights to account for cluster sampling.

C Spillover effects

Because we set aside one Village Council area as a pure control area, we can compare control households in Village Council areas with treated units to these pure control households. Mostly due to power and the inability to include settlement fixed effects, we are unable to show clear spillover effects. If anything, they are mostly positive and substantively small.

Table C.1: **Spillover Effects of Stage 1 Contact Treatment**

	Control - Pure Control)	N
A. Incumbent evaluation index	0.022 (0.128) [0.073]	420
MPA thermometer (1-10)	0.059 (0.307) [0.3]	419
Inc. party thermometer (1-10)	0.126 (0.355) [0.317]	419
Prefers MPAs party (0/1)	0 (0.057) [0.019]	419
B. Government performance index	0.101 (0.072) [0.077]	419
Prov. govt. competent provider (1-4)	0.312 (0.154)** [0.201]	348
Elections important (1-4)	0.12 (0.151) [0.084]	395
State looks after (1-5)	0.067 (0.129) [0.053]	416
Party looks after (1-5)	0.04 (0.139) [0.199]	413
MPA looks after (1-5)	−0.072 (0.154) [0.069]	419
C. Accountability prospects index	−0.01 (0.066) [0.085]	409
MPA likely to follow through (1-4)	0.032 (0.096) [0.139]	351
Political efficacy (1-4)	−0.153 (0.159) [0.112]	416
MPA performance important in voting decision (-4.17, 4.17)	0.1 (0.21) [0.222]	361

Notes: *p<0.1; **p<0.05; ***p<0.01. Each cell represents results of a different OLS model. Estimates in panels A and B are fit using the difference-in-differences specification in Equation 2 while Estimates in panel C is fit using the simple specification in Equation 1. Most models include the following covariates: co-partisanship with the incumbent ("Supports MPAs party"), baseline assets ("Owns motorized vehicle"), and baseline political knowledge ("Knows maj. party not in prov. govt." and "Knows president name "). Exceptions as specified in our pre-analysis plan are made for the following outcomes: when "Supports MPAs party" is the outcome when we exclude the baseline variable of "Supports MPAs party" as a covariate; for all outcomes in Panel C except for "Political efficacy", we add baseline values of "Political efficacy" as a covariate; lastly, when "MPA likely to follow through" is the outcome we include baseline values of "MPA thermometer" as a covariate. For each of our three outcome domains, the index is built from the constituent outcomes using the method described in Kling, Liebman and Katz (2007). This method creates an index as a sum of standardized variables, where we construct each standardized variable by subtracting the mean of the control group and dividing by the standard deviation within the control group. All missing values are imputed using the mean of the treatment group for the missing data. For the two indices where we have all of the baseline values, we construct the indices by standardizing measures to the post-treatment control mean and standard deviation and imputing missing data by mean of the treatment group *and* time period. We present two standard errors: standard robust HC2 standard errors and cluster robust standard errors at the settlement level.

D   SMS encouragement experiment

We also conducted an SMS encouragement experiment before the first round of calls in order to boost compliance. We build on the evidence gathered in SMS mobilization campaigns encouraging electoral turnout that indicates positive effects of mobilization messages in developed contexts in both high and low salience elections (Dale and Strauss, 2009; Malhotra et al., 2011) as well as in developing countries (Aker, Collier and Vicente, 2016).[13] We craft SMS messages to study whether some types of messages are more likely to encourage citizens to spend a small amount of time responding to their MPA's question. The full texts of the SMS treatments is in Table D.5.

The SMS message was sent one hour before the IVR calls were placed. We assigned households that were going to receive an IVR call to one of six SMS treatment conditions. The various treatment groups can be found in Table D.1. The SMS experiment took the shape of a two by three experiment where: (1) respondents either received a generic participation encouragement (*placebo*), an encouragement focusing on the *community* benefits to participating, or an encouragement focusing on the *personal* benefits to participating; and (2) respondents either had no mention of peer participation (*no peer*) or were reminded at the end of the encouragement message that their *peers* were also participating.

Table D.1: Number of Individuals by SMS Treatment Arm

	Placebo	Community	Personal
No Peer	120	120	120
Peer	120	156	156

The results suggest that, if anything, simple encouragements are more effective than SMS encouragements that attempt to induce peer pressure or focus on particularistic rather than communal benefits. First, in Table D.3 we display the raw proportions and standard errors of households who did or did not answer the IVR call by SMS encouragement condition. None of the comparisons between these proportions are statistically significant except for the comparison between the *peer* message and non-peer message among those told to think about the encouragement as personally beneficial. This may indicate that peer effects are largest when people consider the personal stakes of political participation, although we caution against over interpreting this effect in a small sample.

---

[13]However, there is also evidence that SMS messages may be insufficient to prompt citizens in developing countries to report public service deficiencies or corruption (Blair, Littman and Paluck, 2017; Grossman, Michelitch and Santamaria, Forthcoming).

Table D.2: SMS Treatment Conditions

	Placebo	Community	Personal
No Peer	Hello (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how to do a good job for everyone in his constituency. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how best to help you and your family. We hope you can take the call.
Peer	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how to do a good job for everyone in his constituency. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how best to help you and your family. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.

Second, in Table D.4 we display the raw proportions and standard errors of whether individuals answered a question in the IVR call. This represents a higher level of engagement with the call. As we can see, the stand-out group both in size and statistical significance, is the pure placebo group, indicating that longer messages may have hurt participation in this intervention.

Table D.3: Proportion of Respondents Answering the Phone by SMS Treatment

	Placebo	Community	Personal
No Peer	0.875	0.85	0.825
	(0.03)	(0.033)	(0.035)
Peer	0.875	0.859	0.897
	(0.03)	(0.028)	(0.024)

Table D.4: Proportion of Respondents Answering Any Question by SMS Treatment

	Placebo	Community	Personal
No Peer	0.425	0.3	0.292
	(0.045)	(0.042)	(0.042)
Peer	0.258	0.263	0.321
	(0.04)	(0.035)	(0.037)

Table D.5: SMS Treatment Conditions

	Placebo	**Community**	**Personal**
No Peer	Hello (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how to do a good job for everyone in his constituency. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how best to help you and your family. We hope you can take the call.
Peer	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how to do a good job for everyone in his constituency. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.	Hi (name from survey). In a few hours you will receive a call from your [MPA NAME] from this number so that he can talk to you about how best to help you and your family. [MPA NAME] has also called a lot of other people and most of them have responded. We hope you can take the call.

The International Growth Centre
(IGC) aims to promote sustainable
growth in developing countries
by providing demand-led policy
advice based on frontier research.

IGC

**International
Growth Centre**