"Press 1 for Roads": Mixed Evidence on Improving Political Communication*

Miriam Golden†

Saad Gulzar‡

Luke Sonnet§

August 16, 2023

Abstract: We report results of a randomized control trial conducted in Pakistan that uses Interactive Voice Response (IVR) technology to augment existing face-to-face communication between politicians and voters. IVR allows politicians to script questions for voters and voters to respond on cell phones. The technology modifies the initiator, scope, content, scale, personalism, and frequency of communication. Both politicians and voters initially exhibit willingness to engage via IVR. However, IVR does not change downstream voter views about politicians or electoral behavior, nor do politicians leverage the opportunity politically. We interpret the null results as well as the reluctance of politicians to engage repeatedly with voters via IVR as functions of the underlying constraints faced by politicians, who lack the means to satisfy voter demands. Descriptive data also suggest that failures of responsiveness may occur for reasons other than clientelism or elite capture. [140 words]

*We are grateful for funding from the Jameel Abdul Latif Poverty Action Lab, the Stanford Global Poverty Center, the UPS Endowment Fund for Transportation, Logistics and Urban Issues at Stanford University, the International Growth Center–Pakistan, and the Empirical Studies of Conflict Project. We also acknowledge essential logistical support from the Consortium for Development Policy Research, the American Institute for Pakistan Studies, and the Center for Economic Research in Pakistan. We thank Mats-Philip Ahrenshop, Jawad Karim, Mamoor Ali Khan, Milad Khattak, Muhammad Parvez, and Jamshed Wazir for research assistance and our enumerators for their work in the field. Seminar participants at NYU, EGAP, the University of Wisconsin-Madison, LUMS, the LSE, Nuffield College, Bocconi, Stanford, Emory, Zurich, ETH, Indiana, CERP, UCSD, Facebook, the Lahore School of Economics, PacDev, DfID-Pakistan, WESSI, the Hertie School, Columbia University, the MPSA, the EPSA, and the APSA made helpful comments on earlier versions of this work, as did Asad Liaqat. The pilot and scale-up reported here were approved by the UCLA IRB (#17-000182). Interviews with Pakistani politicians were approved separately (UCLA IRB #19000719). The pre-analysis plan for the pilot is available at http://egap.org/registration/2476. The pre-analysis plan for the scale-up is available at https://osf.io/vadwn.

†Peter Mair Chair of Comparative Politics, European University Institute (miriam.golden@eui.eu)

‡Assistant Professor of Politics and International Affairs, Princeton University (gulzar@princeton.edu)

§Independent researcher (luke.sonnet@gmail.com)

1 Introduction

Recent work that aims to correct representational problems in less developed democracies draws on informational theories (see Pande (2011) and Dunning et al. (2019)). According to these, political responsiveness is hampered by lack of information among voters about performance metrics for politicians. The lack of information prohibits voters from holding politicians to account. Conversely, politicians lack information about voters' policy preferences, and that reduces the ability of elected officials to enact good policies. Experimental studies therefore provide information to voters — for instance, through performance scorecards (Grossman and Michelitch, 2018) — or to politicians — by providing them citizen polls (Casey, Kamara and Meriggi, 2021; Gulzar, Hai and Paudel, 2021)). In either case, results are mixed or face very substantial take-up issues.

We report results from a large field experiment that draws on the second line of work, in which politicians face information constraints vis-à-vis citizen preferences. We explicitly design the experiment to circumvent some of the challenges identified in previous studies. Nevertheless, we show that progress is harder than anticipated, potentially because of two reasons. First, the information environment may not be as skewed as may be commonly assumed, though there are some margins along which it can be improved. Second, even if the information environment were improved, this may not be sufficient to bring about meaningful changes in governance because of underlying structural obstacles preventing effective information interventions.

Our study is set in Pakistan, an understudied lower-middle income democracy and the world's fifth most populous country. Our intervention augments existing face-to-face interactions between politicians and voters with technologically-based communication in an experiment featuring Interactive Voice Technology (IVR). IVR allows politicians to script and record questions for voters, which we disseminate as robocalls to voters' cell phones. Voters answer the questions using the number keys on their phones. Their responses are then made available in aggregate form to the politician, who can follow up with an additional call that acknowledges and responds to what he learned about voter preferences. The follow-up call closes the communication loop by informing voters that they have been heard and their preferences recorded. We conducted this experiment with the expectation that voters would respond to this new high-frequency, unsolicited, and two-way contacts with their elected representative with greater political support for him and also by moving towards performance-based voting criteria. We expected politicians to respond to the opportunity to communicate more quickly with larger numbers of voters and to receive aggregated

information on their policy preferences by targeting improved service delivery.

We employ mixed methods, collecting multiple types of observational as well as experimental data. Together, these put politician-voter interactions under a microscope from various angles. We present five key findings. *First*, in contrast to previous work using other methods to boost political communication, the use of IVR generates high levels of engagement by citizens. *Second*, improving communication improves public engagement with elected officials on multiple dimensions. It speeds up political interactions, provides better spatial distribution, and reduces scope for particularism. *Third*, however, neither politicians nor citizens follow through in downstream attitudinal or behavioral changes. Politicians do not improve service delivery and voters do not appear more satisfied with their elected representatives. *Fourth*, we use a forecasting exercise to document that both the degree of active engagement and the downstream null results are surprising to, rather than expected by, academic and policy experts. *Fifth*, we provide qualitative evidence that IVR raises expectations among voters while not providing politicians any additional resources to respond. Taken together, the results indicate that boosting communication and information on their own may be inadequate in improving policy responsiveness. In poor countries, even if politicians are well-intentioned and well-informed, they face inherent obstacles to their capacities to respond to voter demands.

Our study contributes to three literatures. The first contribution of our paper is to the specific stream of the accountability agenda that aims to strengthen it with experimental interventions. We contribute to the experimental literature on how politicians interact with citizens during electoral campaigns (Wantchekon, 2003; Gerber, Green and Larimer, 2008; Wantchekon et al., 2015; Kendall, Nannicini and Trebbi, 2015; Cruz et al., 2018; Cantoni and Pons, 2021). We extend this line of inquiry by establishing politician-initiated two-way communication between politicians and voters. Prior work examines political communication initiated by voters (Grossman, Humphreys and Sacramone-Lutz, 2014; Chong et al., 2015; Arias et al., 2019; Buntaine, Nielson and Skaggs, 2019; Dunning et al., 2019; Grossman, Humphreys and Sacramone-Lutz, 2020)[1] or in townhall meetings (Fujiwara and Wantchekon, 2013). In contrast to this line of work, we ask if political communication initiated and led by politicians instead of by voters carries promise. Answering this requires the active cooperation of elected politicians who, as others have noted "are typically unwilling to delegate to researchers (and chance) the decision of which citizens to address and how" (Cantoni and Pons, 2021, p. 381). Our results show that politician-led communication is valued by voters

[1]These studies document limited success, largely because of low and unrepresentative take-up among voters (see Table A.1).

and enlarges communication more effectively than prodding voters to initiate.

Second, we provide new descriptive information on the underpinnings of democratic accountability in the global south. It is commonly believed that accountability fails in such settings because interactions between politicians and citizens are based on clientelistic relations (Wantchekon, 2003; Stokes, 2003; Wantchekon et al., 2015; Larreguy, Marshall and Querubin, 2016; Cruz, Labonne and Querubin, 2017) and that even if they are not, political processes are captured by socially and economically unrepresentative local elites (Bardhan and Mookherjee, 2000). Studies that develop the first argument focus largely on the electoral process, particularly political activities such as vote buying (Vicente, 2014; Arias et al., 2019; Blattman et al., 2019). However, citizens around the world participate in politics in ways other than voting: by reaching out to their political representatives in letters, phone calls, and in person. Recent research has begun to unpack these interactions in the developing world (Bussell, 2019; Kruks-Wisner, 2018; Mohmand, 2019), often corroborating the elite capture thesis. We use a variety of instruments to descriptively characterize these interactions in a setting where one might expect to observe high levels of capture. We present strong evidence that politics is neither as clientelistic nor as skewed towards the interests of an unrepresentative elite as believed. Although we identify several margins along which representation could be improved, our data should generate skepticism among scholars about the assumptions made about the status quo political economy of the global south.

Third, our research contributes to the literature on the role of information and communication technologies on development. Past work studies how data generated by new technologies, such as call detail records, can be used to study development questions (Suri et al., 2021; Blumenstock, Cadamuro and On, 2015; Blumenstock, 2016; Muralidharan, Niehaus and Sukhtankar, 2016; Lewis-Faupel et al., 2016; Callen et al., 2020). We examine the efficacy of a mobile phone technology that is already used extensively by politicians and political parties around the world. We provide causal evidence on the efficacy of its deployment and suggest both avenues where the technology can help and also those where changes might be more muted. Our study shows that remarkable progress can be made in circumventing take-up problems that are persistent (and seemingly intractable) according to previous work.

2 Context and Status Quo Political Engagement

2.1 Local political context

Our research was conducted in Pakistan's third most populous of its four provinces, Khyber Pakhtunkhwa (KP). KP has a population of more than 35 million, mainly rural, inhabitants. Provincial literacy is 57 percent. At the time we conducted the experiment (2017–18), there were 99 directly elected, all-male, Members of the Provincial Assembly, each representing a single-member electoral district. In contrast to the situation in Pakistan's other provinces, KP's provincial legislators face genuine competition in retaining their seats from one election to the next. Sixty percent of MPAs elected in KP in 2013 had never been previously elected to a provincial or the national assembly (compared with 50 in Balochistan and Sindh and only 43 in Punjab).

Only 21 percent of citizen respondents report (in a baseline survey) that they feel they can influence what government does at least a little. However, 84 percent of respondents replied that it was important or very important to them that Pakistan be governed by representatives elected by the people, documenting widely diffused democratic aspirations. Thus, ordinary households feel remote from politics while nonetheless supportive of democratic institutions.

2.2 Status quo interactions between citizens and politicians

In the setting we study, households seek all manner of assistance from their political representatives. MPAs (and politicians generally) are key conduits for assistance for personal and community problems. In our baseline survey, 55 percent of respondents report that the MPA is able to get roads fixed in their communities and 49 precent say that an MPA can help their family members get a job. These results illustrate the extent to which ordinary households rely on political representatives for both development goods and personal help. Politicians are highly visible sources of help in a setting where income is low and unreliable and communities lack much basic infrastructure.

To obtain assistance, citizens make direct, face-to-face requests of their representative. Thus, politicians return to their constituency offices on a weekly basis where, as one ethnographer wrote, "crowds of applicants wait outside to see the politician or a personal assistant in order to get the all-important 'chit' of paper ..." (Wilder, 1999, p. 199). Semi-structured interviews by one of us with more than three dozen current and former Khyber Pakhtunkhwa

MPAs and MNAs found regular complaints about the constant need to visit their constituencies to attend "weddings and funerals" and to spend the day in their *hujra*, the walled-off area outside the home where the male head of household receives other male visitors. The literature reports that in Pakistan, politicians work from their constituency offices (often the *hujra* or equivalent) to directly interact with anywhere from 20 to more than a hundred supplicants over the course of a day (Wilder, 1999, p. 199). These interactions are the modal way for voters to request help of government in the country.

Direct contacts between Pakistani voters and politicians are reported to be skewed towards higher-income male co-partisans (Martin, 2014; Liaqat, Cheema and Mohmand, 2020). This is part of a political culture in which "leaders redirect public resources to benefit kin, friends and clients" (Martin, 2016, p. 67). In a setting that anthropologists characterize as one of asymmetric factional hierarchies based on status and power (Barth, 1965; Lyon, 2004), politicians seek to assemble large clienteles but, as our semi-structured interviews in KP repeatedly revealed, limited resources mean they are unable to provide assistance to most who request it. Since politicians thus must make discretionary allocative decisions, other scholars report that "those who are richer and part of the village elite have greater bargaining power vis-à-vis leaders than poorer, non-elite members ..." (Mohmand, 2019, p. 24).

3 Experimental Design and Implementation

3.1 Sources of data

Our study uses information from eight data collection instruments (see Table F.1). We began in late 2017 with a (i) *baseline* survey to enroll 14,400 voters into the study. We undertook the IVR experiment itself in 2018 and it generated (ii) *intervention* data on content, take-up, and response rates. In 2019, we conducted an (iii) *endline* survey to collect information on experimental outcomes. In 2020, we conducted a (iv) *descriptive* telephone survey of 3,600 respondents subsampled from the initial 14,400 to collect information on pre-existing (status quo) political communication.[2] During the course of the intervention, we surveyed 240 (v) *key informants* across the 20 treated constituencies to collect auxiliary information about politician behavior. We also merged the household-level data with 2019 polling station level (vi) *election* information that we collected. In 2019, we also conducted lengthy face-to-face semi-structured (vii) *interviews* with approximately three dozen MPAs and MNAs from KP.

[2]Because of COVID-19, face-to-face surveying was not possible at that time.

In addition, we surveyed 400 students, academics, and policy implementors in advanced countries and in Pakistan in 2019 to collect information for a (viii) *forecasting* exercise. We use these eight labels to identify each type of data in our analysis.

3.2 Randomization

Of the 99 directly-elected MPAs in the KP assembly, our implementing partner identified an initial 47 who expressed interest in working with us and in using IVR to interact with voters. The 47 MPAs generally come from slightly more competitive and urban constituencies than their uninterested counterparts. Blocking on political party, we randomly select 20 of the 47 MPAs to enroll into treatment and thus give access to the IVR technology through our research team. Figure 1 provides a diagram of the randomization process at both the polling station and household levels.

Within each treated constituency, we select the 20 polling stations with the smallest absolute margin of victory of the incumbent MPA.Among these 20 polling stations, we randomize six into treatment {P1} and 14 into control {P0} in a two-step process. We first randomly sample 12 polling stations out of the 20 and then, for those 12 polling stations, create matched pairs using a Mahalanobis distance score that incorporates the total number of registered voters and raw vote totals for large parties.[3] We assign one of each pair to treatment and the other to control.

The right panel of Figure 1 shows household-level randomization. A random walk sampled 120 households within each treated PS area, starting at a central location in the PS catchment area. We enroll male heads-of-household only, owing to obstacles collecting phone numbers of women.[4] Our total sample is 14,400 male heads of household, whom we survey once before (baseline) and once after (endline) the intervention. Blocking on co-partisanship with the MPA, 40 of the 120 PS-level households are placed into a control condition and receive no contact other than the surveys {H0}, 20 receive an IVR call with a credit-claiming message but no question(s) {H1}, and 60 receive a credit-claiming message as well as an IVR question(s) {H2}. Total sample sizes are reflected by the N values reported in each box.

A second stage splits those who receive first-stage calls to either receive no follow-up call or receive a follow-up call containing specific components. Respondents in {H1C} and {H2C}

[3]We define a large party as any party that received more than 100 votes in any of the 12 PSs or that received an average of 20 votes across all 12 PSs in the 2013 elections.

[4]Using female enumerators, ninety percent of the women we asked refused to provide their phone numbers or did not have regular access to a phone. The exclusion of women may raise ethical issues. These are discussed in Appendix C.

receive no follow-up call, while all other respondents receive a follow-up call in which the MPA asks a new question via IVR {H1Q, H2G, H2R}. Respondents in {H1Q} receive only the new IVR question, while respondents in {H2G} and {H2R} also receive a *generic* or *responsive* message, respectively, where the MPA acknowledges the first stage IVR question. The *responsive* message details how the MPA will act based on the information collected in the initial robocall in which he asked for constituent input while the *generic* message simply thanks respondents for their input. The main goal of these randomizations is to estimate the total effect of the most interactive and deepest IVR communication we could generate {H2R} as well as marginal effects of call components.

3.3 Content of the IVR calls

Each first-stage call contains two sections. First, the MPA introduces himself and credit-claims for recent activities. Second, he asks a question seeking voter feedback. The average initial recording lasted two minutes, which with the MPA's permission we edited down to an average of 85 seconds for clarity and to retain voter interest. See Appendix E for a sample script.

After calls are completed, we collate responses and report aggregate demographics from our baseline survey as well as aggregate IVR responses to the MPA. After receiving first-stage responses, MPAs record follow-up calls and second IVR questions. Details on the intervention timeline appear in Appendix D.

4 Communication Transformations Offered by IVR

In this section, we report descriptive results comparing what we call status quo communication with the IVR technology we deployed, where by status quo, we mean the communication patterns between voters and MPAs prior to the IVR intervention.

4.1 Geographic scope: Where are the locations where voters communicate with the MPA?

We first examine if community characteristics correlate with who interacts with an MPA in the *descriptive* survey. If interactions are geographically concentrated under the status quo, some geographic areas may be potentially underserved. In Figure 2a, we depict the

distribution of in-person interactions and IVR calls across polling stations (left) and constituencies (right). There are substantial numbers of polling stations where no one reports having met his MPA in person in the last year. This number goes to almost zero for IVR contact, defined as answering the phone; IVR thus reaches all locations, including those excluded from direct political access. If we restrict attention to only IVR respondents who answer a question, contact still improves under IVR. One-sided F-tests demonstrate that the variance in PS status quo contact is greater than the variance in polling station IVR question response rates ($p = 0.036$) but not IVR phone response rates ($p = 1$). Results are similar if we consider constituency level differences instead of polling station level differences. Finally, we also examine contact by distance to the MPA's constituency office and find that voters who live further from the MPA's constituency office are disadvantaged by the need to travel to face-to-face meetings (see Appendix Table G.1).

4.2 Scale: How many people communicate with their MPA?

The second dimension we investigate, using the *descriptive* survey, is the total number of constituents who interact with their elected representative in any given period. In the subsample of respondents in the {H2} condition, citizens report when they most recently met their MPA. Only about a third of all households report *ever* having met their MPA in person; we take this as the upper-limit on face-to-face interactions over any period of time. IVR is not only able to more than double the rate of interaction with just a single round of calls, it also establishes as much two-way interaction in one round of calls as occurs face-to-face over a full year. (See Figure G.1a for details.) Since IVR calls are automated, require about a week to roll out via the technology provider, and cost at most a few U.S. cents each, this represents the enormous gains in scale that occur if IVR is used to boost interactions between an MPA and constituents.

4.3 Content: What gets communicated to the MPA?

In Figure 2b, we present data from the *descriptive* survey that shows the target and nature of requests made by individuals who report having met face-to-face with their MPA in the last year. We separate requests into those where respondents ask for goods or services for the individual (or household) and where they instead make requests on behalf of their community. With the caveat that these are self-reported data, we find that two-thirds of requests seek community improvements rather than individual goods. These data thus show

that even in face-to-face interactions, voters more frequently request basic public infrastructure than clientelistic goods. However, face-to-face interactions seem almost never to involve discussions of public policy and legislation. This is in contrast to IVR; as the modal IVR call content documents (reported in Section 3.3), IVR offers voters the opportunity to urge that their elected representative prioritize broad policy orientations and legislation. In the context we study, legislation might for instance involve prohibiting child marriages, enforcing greater gender equality, and protection of human rights. By construction, IVR suppresses purely individualistic requests.

In Table E.1, we report the distribution of answers of those answering the first-stage IVR question. Only a quarter of respondents suggest their representative prioritize legislation. Three-quarters convey preferences for development goods (roads, electricity, piped water) or for improvements in health and education; the latter is especially important for respondents residing in rural areas, where basic facilities are often absent or inadequate (e.g. schools without lavatories).

Overall, IVR and status quo face-to-face interactions are broadly similar in that in both cases, a large majority — ranging from two-thirds to three-quarters — of voters ask politicians for improvements in basic public infrastructure. If we take the (normative) position that voters should not ask political representatives for jobs or individual financial assistant, then IVR improves face-to-face interactions by naturally suppressing such communication.

Analyzing drop-off rates during IVR calls, we find that of the respondents who were called, 73 percent answer the phone. By the time the MPA finishes asking a question, about half the respondents are still on the call. The biggest drop-off occurs when respondents are asked to answer a question, which is completed by 17.3 percent of those called.[5] Whether this is a large or small proportion depends on one's baseline; compared to other ICT interactions (reported in Table A.1) and compared to our *descriptive* data on rates of face-to-face interactions, it is excellent. See Figure G.1b for details.

4.4 Socio-economic and partisan representativeness: Who communicates with his MPA?

A fourth dimension we study concerns *who* MPAs are able to reach using IVR. Figure 2c compares the characteristics of three groups: (i) people who report in our *descriptive* sur-

[5]Because IVR is still not in general use in Pakistan, even in the private sector, this may signal that citizens are still familiarizing themselves with how it works.

vey having met their MPA in person at any time in the last year; (ii) randomly sampled respondents enrolled in IVR who answer a question (based on *intervention* data); and (iii) randomly sampled respondents enrolled in IVR (using data from the *baseline* survey). These data show that *both* modes of communication do a good job reaching constituents who are representative along various dimensions we measure, including demographics (age), socio-economic status (income and education), partisanship (co-partisan and MPA thermometer scale), and political knowledge (whether the respondent accurately identifies Pakistan's president).[6] Face-to-face interactions are not elite-biased and IVR replicates the pattern of reaching a representative sample of those who pick up the phone as well as those who answer a question.

5 Downstream Results of IVR Communication

Taken together, the results in the previous section show that IVR offers clear improvements on two dimensions of interaction between politicians and their constituents: scope and scale. IVR may also improve the content of interactions, assuming we wish to exclude individual requests for assistance. Now we ask whether IVR shows effects on individual attitudes, self-reported political behavior, and on politician service delivery. Except as noted below and detailed in Appendix B, the outcome measures, equation specifications, and treatment effects reported in this section were all pre-registered. Appendix J discusses power considerations in the interpretation of these results.

5.1 Effects on individual attitudes and behavior

Data: Our main study sample comprises 14,399 male heads-of-household.[7] Of these, we were unable to recontact 411 (2.9 percent) at endline.[8] As a result, we have 13,988 individuals in our *endline* dataset which, combined with the *baseline* dataset, constitute the data analyzed in this section.[9]

[6] Both methods also do an equally bad job reaching women as independent voters in the household. Neither the status quo nor IVR are designed to counter Pakistan's massive gender imbalances in political communication (Khan, 2017).

[7] We baselined 14,400 respondents; one was dropped after randomization due to a duplicate phone number.

[8] The p-value for an F-test of the joint significance of a regression of attrition on the two first stage treatment dummies is 0.77, indicating that attrition is not predicted by treatment.

[9] As a robustness check, we replicate our main analysis (reported in Table 1) using inverse probability of attrition weights (see Table I.1). Results are nearly identical due to the large sample size and relatively low rate of attrition.

Outcomes: At the individual level we focus on three main outcome indices. First, we study effects of the intervention on *evaluations of the incumbent*. The index comprises four outcomes: (1) a feeling thermometer for the MPA himself (1–10); (2) a feeling thermometer for the MPA's party (1–10); (3) a binary indicator for whether the respondent reports having voted for the MPA (or the incumbent MPA's party if the incumbent did not run again); and (4) the inverse of the ranking of the MPA among his top four challengers in the 2018 elections. Second, we study effects of the intervention on *political participation*. This index consists of three outcomes: (1) a binary indicator for whether or not the respondent voted in the 2018 election; (2) a binary indicator for whether or not the respondent attended a rally in the period leading up to the 2018 elections; and (3) a binary indicator for whether or not the respondent attended a political meeting before the 2018 elections. Third, we study effects of the intervention on *prospects for electoral accountability*. This index is made up of three outcomes: (1) a measure of self-stated political efficacy (1–5); (2) how important incumbent performance is in an individual's vote choice (1–6); (3) the number of conversations the respondent had about politics in the two weeks before the endline survey. The first item is included because we think voters are more likely to attempt to evaluate the performance of the incumbent when they have higher levels of political efficacy. We include the second item to measure whether the voter thinks that performance criteria should be used when deciding for whom to vote. The final item is included because we contend that voters cannot enforce political accountability without some political engagement that includes discussing issues.[10]

Estimation: We estimate effects on the indices and constituent outcomes with:

$$Y_{hpm(t=1)} = \tau D_h + \alpha Y_{hpm(t=0)} + \beta^{'} \mathbf{H}_{hpm(t=0)} + \lambda_p + \epsilon_h,$$

where $Y_{hpm(t=1)}$ is the outcome Y for household h in polling station p at endline ($t = 1$), D_h is a binary indicator for treatment status, $Y_{hpm(t=0)}$ is the pre-treatment outcome Y collected in the baseline (if available), and λ_p is a polling station fixed effect. The vector of household variables, $\mathbf{H}_{hpm(t=0)}$, is a set of pre-treatment variables selected from {age by decade, an income scale, education bins, an index of political knowledge, a set of indicators for the party the respondent supports, a binary measure of reported turnout in 2013, and a binary measure of support for the MPA's party} if they predict the outcome in the control group.[11]

[10] We detail how we build these indices in Appendix H. We also show that adding this outcome to the political participation index does not change any of the conclusions reported throughout the paper (see Table I.2).

[11] We admit pre-treatment household characteristics into the specification if the F-statistic of a regression of the outcome on that pre-treatment variable is significant at the 0.05 level. For variables like age, where

We use heteroskedasticity-consistent standard errors (HC2), since treatments of interest are assigned at the household level.

Individual level effects: We present treatment effects for receiving either part of the IVR treatment (groups {H1} and {H2}) as well as receiving the deepest, or "full" responsive IVR intervention ({H2R}) in Table 1. In the first column, we present the control mean; because this is the reference group to which we scaled all indices, the mean is 0 and the standard deviation is 1 by design. The second set of columns presents the treatment effect of any call, along with total sample size used in estimating that treatment effect. The absolute values of all treatment effects are smaller than 0.02 standard deviations, and all are statistically indistinguishable from zero. We find no effects of receiving any kind of IVR call on individual attitudes for any of the indices. Even when we split out the three indices into their component measures (see Appendix K), there are no treatment effects on any outcome. Local average treatment effects among compliers are also substantively small and statistically insignificant (see Table I.3).

In the third set of columns, we consider the effect of the fully responsive IVR intervention, whose respondents received an initial call with an IVR question and a subsequent follow-up call acknowledging first-stage responses. Although we did not pre-register this comparison, we include it here because we consider this to be the deepest use of the IVR tool deployed. We have the greatest expectation for positive treatment effects for this group of respondents. Nonetheless, we again see small treatment effects, with the largest treatment effect of +0.25 standard deviations on the prospects for accountability index. Again, no treatment effects are statistically significant. Other results analyzing differences between various treatment arms are reported in Appendix K.

5.2 Effects on aggregate electoral outcomes

We now turn to effects on aggregate electoral outcomes. One of the features of our experimental design was the ability to identify the effects of saturating one PS area with IVR calls while leaving other PS areas untouched. This allows us to estimate whether voters reward or punish their MPA for communicating via IVR with dozens of households in a PS area, as well as whether increased communication changes the turnout rate at the polling station level. Unsurprisingly, given the lack of individual-level effects, we find no evidence of any

we have a set of dummies for the decade of the respondent, we admit all of the age dummies if the p-value on the joint test of their significance is less than 0.05

intervention effect on vote shares for our partner MPA (or his party) or for turnout. We present the full specification and results in Appendix M.

5.3 Effects on politician behavior

We evaluate whether IVR had effects on the behavior of politicians by examining the kinds of messages they crafted during the *intervention* as well as whether they invested more effort in service delivery in treated localities. Our information comes from interviews with *key informants* recruited in treated localities.

For the IVR calls, we allowed partner MPAs to craft messages of their choosing. We anticipated that this would permit politicians to tailor their communication to treated localities. However, politicians did not invest in crafting messages that were specific to the geographic area. The modal question (see Section 3.3) was, by our reading, general and somewhat imprecise. Indeed, 16 of our 19 partner MPAs who recorded a question asked identical questions, recycling one of two prompts our staff provided to assist MPAs in crafting questions.

The politicians who enrolled in our study interacted repeatedly with us in the nine months preceding an election. This might have encouraged them to improve service delivery in the polling station areas where voters received IVR calls, perhaps with a view towards augmenting the enhanced political visibility offered by IVR calls. This additional effort could have created a second channel for voters to receive new information about their representative. In a quantitative evaluation of the key informant interview data, we see no evidence of any effect on MPA effort. (See Appendix L for details.)

Another marker of low politician engagement with the opportunity to communicate with a large and representative sample of households using automation is the failure of our research to repeat on schedule. We designed the study to obtain repeated IVR recordings — as many as half a dozen per MPA — in the seven months preceding provincial and national elections. Our design was meant to leverage reelection aspirations of enrolled MPAs.[12] In the end, we recorded only a single question and a single follow-up response with each MPA because of across-the-board resistance on their part to repeated meetings to record additional rounds. We discuss this more in the final section of this paper.

[12] Appendix C discusses the ethics of working with incumbents alone.

5.4 Evidence of voter engagement

Despite null effects of treatment on voter attitudes and behavior, there is evidence that some components of the intervention affect voter engagement. We evaluate whether citizens are interested in engaging with politicians using IVR. The data shows that when citizens receive a call from their MPA that includes a question — as opposed to an exclusively credit-claiming call — they are more likely to answer a subsequent call from the same politician (see Table K.2). This suggests that citizens appreciate being included in policy discussions with politicians and would like to continue the conversation. Mere communication, however, is insufficient to change downstream voter outcomes (see Table K.3).

5.5 What did we expect? A forecasting exercise.

Were the weak results we observe obvious ex ante? To find out, we conduct forecasting exercises that poll potential consumers of this research in Pakistan and academics in the United States and Europe, asking them about their beliefs about the impacts of the experiment before seeing the results. We find that our 400 forecasters significantly and substantively underestimate how many people answer the phone and overestimate the treatment effects across our three downstream outcomes. (See Appendix N for details.) Our interpretation is that forecasters expect the intervention to produce relatively large results. Thus, the actual null results are surprising to both academics and policy experts.

6 Interpretation and Conclusions

The experiment we conducted was enthusiastically embraced by the Members of the Provincial Assembly we partnered with, and academics as well as policy practitioners expected it to succeed. Nonetheless, MPAs disengaged from the experiment, refusing to continue recording messages for voters. They also failed to redirect efforts towards the areas where voters received IVR communication. On the other side, the evaluations voters made of their representatives did not improve. What happened?

In open-ended follow-up interviews with 16 of the 20 treated MPAs, we asked them this very question. MPAs reported that the information they gained when we delivered aggregate feedback from voters was politically very helpful to them, showing them the areas of their

constituency with more or less voter support.[13] Many claimed that the information we provided was new and that they had no existing methods to collect similar information.[14] Yet, they also reported that voters tended to misinterpret the IVR questions as commitments by the MPA to provide new infrastructure.[15] MPAs stated that when they went back to their hjura after dissemination of the IVR calls, they were confronted with angry voters wanting to know when they would make good on their IVR promises.[16] As a result, MPAs uniformly disengaged from the experiment and refused to continue recording messages. Asking voters what they wanted only raised expectations without providing additional resources to politicians with which to satisfy them. For the MPAs, the experiment thus backfired; voters who answered a question deemed their representative "a failure" due to his inability to follow up with new infrastructure.

Previous research on spring-boarding communication between politicians and voters faced problems of take-up from voters. This article shows that a politician led approach resolves the key take-up challenge, but it comes with additional, perhaps more fundamental, complications. Even if politicians wish to respond to voter preferences, in many developing countries they lack the means to do so. A more robust communication infrastructure between politicians and voters may therefore need a concurrent boost in the institutional arrangements that make it possible for politicians to make promises they can keep.

[13]For instance, respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

[14]Reported for instance by respondent 27, interviewed 26 April 2019; respondent 26, interviewed 26 April 2019; respondent 11, interviewed 24 April 2019.

[15]Respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

[16]Respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

References

Arias, Eric, Pablo Balán, Horacio Larreguy, John Marshall and Pablo Querubín. 2019. "Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks." *American Political Science Review* 113:475–98.

Bardhan, Pranab and Dilip Mookherjee. 2000. "Capture and Governance at Local and National Levels." *American Economic Review* 90(2):135–39.

Barth, Fredrik. 1965. *Political Leadership among Swat Pathans*. London School of Economics Monographs on Social Anthropology (Book 19) London: Berg Publishers.

Blattman, Christopher, Horracio Larreguy, Benjamin Marx and Otis R. Reid. 2019. Eat Widely, Vote Wisely? Lessons from a Campaign Against Vote Buying in Uganda. Working Paper, 26293 National Bureau of Economic Research (NBER).

Blumenstock, Joshua Evan. 2016. "Fighting poverty with data." *Science* 353(6301):753–754.

Blumenstock, Joshua, Gabriel Cadamuro and Robert On. 2015. "Predicting poverty and wealth from mobile phone metadata." *Science* 350(6264):1073–1076.

Buntaine, Mark T., Daniel L. Nielson and Jacob T. Skaggs. 2019. "Escaping the Disengagement Dilemma: Two Field Experiments on Motivating Citizens to Report on Public Services." *British Journal of Political Science* pp. 1–21.

Bussell, Jennifer. 2019. *Clients and Constituents: Political Reponsiveness in Patronage Democracies*. New York: Oxford University Press.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan and Arman Rezaee. 2020. "Data and policy decisions: Experimental evidence from Pakistan." *Journal of Development Economics* 146:102523.

Cantoni, Enrico and Vincent Pons. 2021. "Do interactions with candidates increase voter support and participation? Experimental evidence from Italy." *Economics & Politics* 33(2):379–402.

Casey, Katherine, Abou Bakarr Kamara and Niccoló F Meriggi. 2021. "An experiment in candidate selection." *American Economic Review* 111(5):1575–1612.

Casey, Katherine, Rachel Glennester, Edwards Miguel and Maarten Voors. 2018. Skill versus Voice in Local Development. National Bureau of Economic Research (NBER) Working Paper 25022.

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.

Christensen, Garret, Jeremy Freese and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How To Do Open Science*. Oakland: University of California Press.

Cruz, Cesi, Julien Labonne and Pablo Querubin. 2017. "Politician family networks and electoral outcomes: Evidence from the Philippines." *American Economic Review* 107(10):3006–37.

Cruz, Cesi, Philip Keefer, Julien Labonne and Francesco Trebbi. 2018. Making policies matter: Voter responses to campaign promises. Technical report National Bureau of Economic Research.

DellaVigna, Stefano and Devin Pope. 2018. "Predicting experimental results: Who knows what?" *Journal of Political Economy* 126(6):2410–56.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde and Craig McIntosh, eds. 2019. *Information and Accountability: A New Method of Cumulative Learning*. New York: Cambridge University Press.

Fujiwara, Thomas and Leonard Wantchekon. 2013. "Can informed public deliberation overcome clientelism? Experimental evidence from Benin." *American Economic Journal: Applied Economics* 5(4):241–255.

Gerber, Alan, Donald P. Green and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102(1):33–48.

Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda.".

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.

Grossman, Guy, Macartan Humphreys and Gabriella Sacramone-Lutz. 2020. "Information Technology and Political Engagement: Mixed Evidence from Uganda." *Journal of Politics* 82(4):1321–36.

Gulzar, Saad, Zuhad Hai and Binod Kuman Paudel. 2021. "Information, Candidate Selection, and the Quality of Representation: Evidence from Nepal." *Journal of Politics* .

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.

Khan, Sarah. 2017. "Personal is political: Prospects for women's substantive representation in Pakistan." *Unpublished manuscript* .

Kruks-Wisner, Garielle. 2018. *Claiming the State: Active Citizenship and Social Welfare in Rural India*. Cambridge: Cambridge University Press.

Larreguy, Horacio, John Marshall and Pablo Querubin. 2016. "Parties, brokers, and voter mobilization: How turnout buying depends upon the party's capacity to monitor brokers." *American Political Science Review* 110(1):160–79.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A Olken and Rohini Pande. 2016. "Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia." *American Economic Journal: Economic Policy* 8(3):258–83.

Liaqat, Asad, Ali Cheema and Shandana Khan Mohmand. 2020. Who Do Politicians Talk To? Political Contact in Urban Punjab. In *Pakistan's Political Parties*, ed. Mariam Mufti, Sahar Shafqat and Niloufer Siddiqui. Washington, D.C.: Georgetown University Press.

Lyon, Stephen M. 2004. *An Anthropological Analysis of Local Politics and Patronage in a Pakistani Village*. Mellen Series in Anthropology (Volume 12) Lewiston: Edwin Mellen Press.

Martin, Nicolas. 2014. The Dark Side of Patronage in the Pakistani Punjab. In *Patronage as Politics in South Asia*, ed. Anastasia Piliavsky. Delhi: Cambridge University Press.

Martin, Nicolas. 2016. *Politics, Landlords and Islam in Pakistan*. London: Routledge.

Mohmand, Shandana Khan. 2019. *Crafty Oligarchs, Savvy Voters: Democracy under Inequality in Rural Pakistan*. Cambridge: Cambridge University Press.

Muralidharan, Karthik, Paul Niehaus and Sandip Sukhtankar. 2016. "Building State Capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106(10):2895–929.

Pande, Rohini. 2011. "Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies." *Annual Review of Economics* 3:215–37.

Stokes, Susan. 2003. "Perverse Accountability: Monitoring Voters and Buying Votes." Paper prepared for delivery at the Annual Meeting of the American Political Science Association, Philadelphia.

Suri, Tavneet, Jenny Aker, Catia Batista, Michael Callen, Tarek Ghani, William Jack, Leora Klapper, Emma Riley, Simone Schaner and Sandip Sukhtankar. 2021. "Mobile money." *50 Years of Central Banking in Kenya* p. 146.

Vicente, Pedro C. 2014. "Is Vote Buying Effective? Evidence from a Field Experiment in West Africa." *The Economic Journal* 124(574):F356–F387.

Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55(03):399–422.

Wantchekon, Leonard, Gabriel López-Moctezuma, Thomas Fujiwara, Cecilia Pe Lero and Daniel Rubenson. 2015. "Policy Deliberation and Voter Persuasion: Experimental Evidence from an Election in the Philippines." Unpublished paper.

Wilder, Andrew R. 1999. *The Pakistani Voter: Electoral Politics and Voting Behaviour in the Punjab*. Oxford: Oxford University Press.

7 Figures and Tables

Figure 1: **Experimental design at the polling station and household levels**

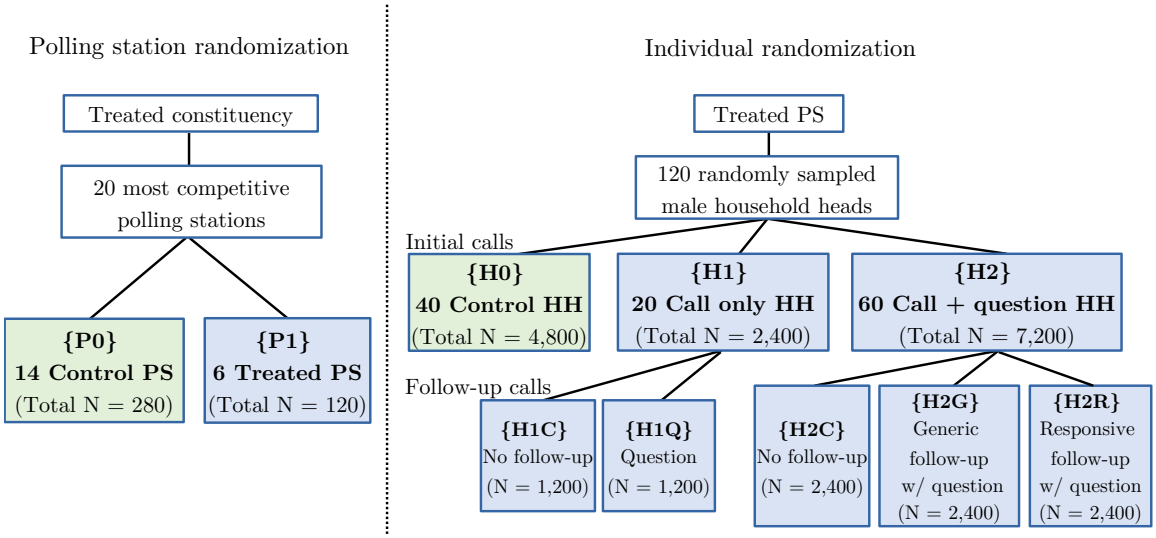
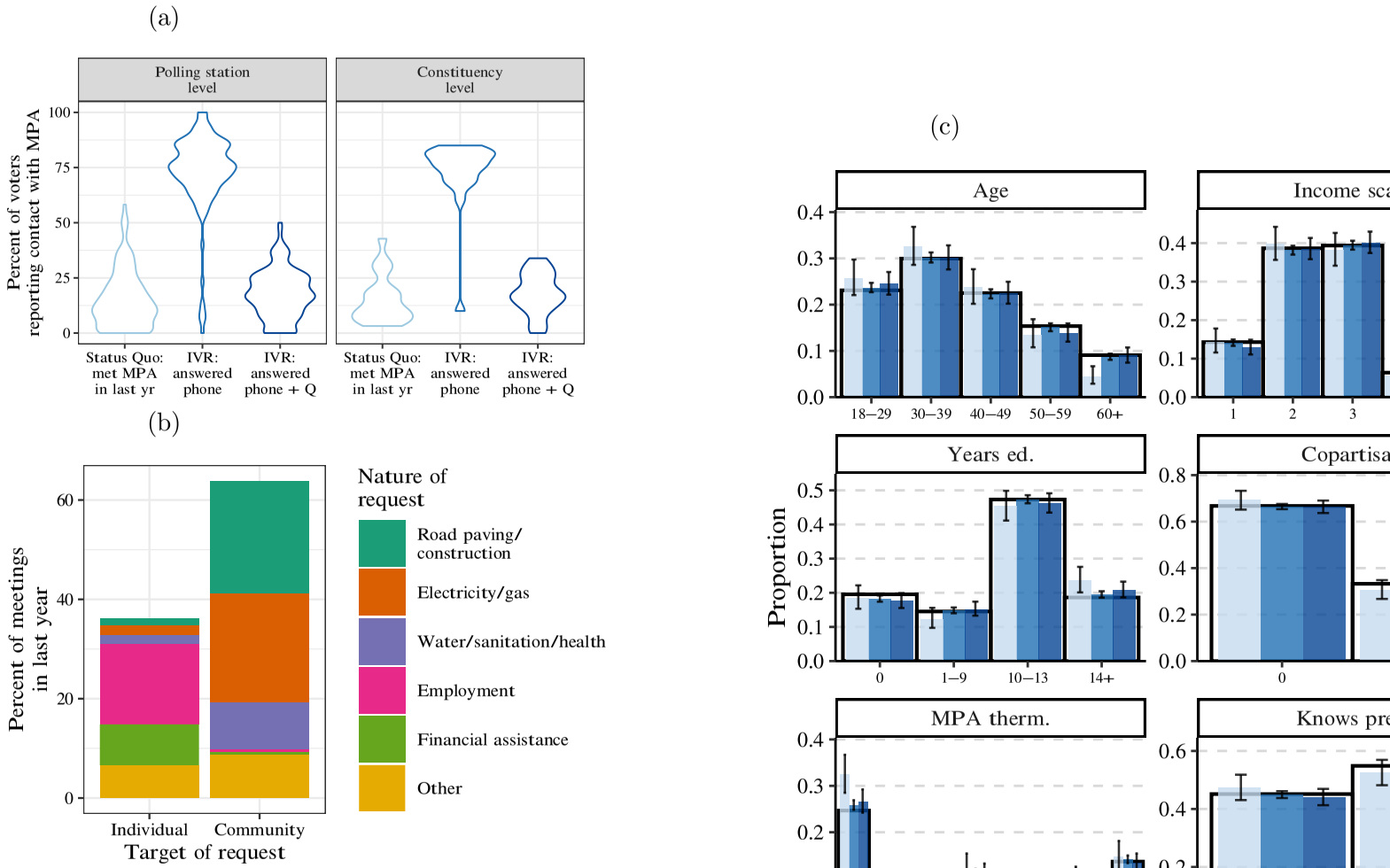


Figure 2: **Descriptive Evidence on Communication Transformations Under IVR**



Notes: (a) Distribution of contact across polling stations (left) and constituencies (right) between politicians and voters. For example, in the left panel, polling stations with no reported contact lie at zero and those where everyone reports contact with their MPA lie at 100. (b) Status quo by contact window. Data are subset to respondents who report meeting their MPA in the last year. (c) Constituent characteristics by mode of contact (with 95 percent confidence intervals). 95% confidence intervals displayed.

Table 1: **Effects of any IVR call and effects of full IVR treatment on household head outcomes**

	Control mean: no call	ITT: any call		ITT: full responsive treatment	
	{H0}	{H1, H2} vs. {H0}		{H2R} vs. {H0}	
Outcome indices	μ	τ	N	τ	N
Incumbent evaluations index	0.000	-0.009	13757	-0.016	6539
	(1.000)	(0.009)		(0.013)	
Political participation index	0.000	-0.020	13780	0.004	6551
	(1.000)	(0.016)		(0.025)	
Prospects for accountability index	0.000	0.004	13759	0.025	6539
	(1.000)	(0.017)		(0.026)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see text for details. Because our preferred specification includes pre-treatment covariates and the baseline measure of the outcome may have some missingness and because there is some missingness on the outcomes themselves, the sample sizes in the tables do not represent the full $13,988$ individuals from whom we collect both baseline and endline data.
The letters in braces refer to the experimental groups described in Figure 1.

Appendices

A Take-up Rates Reported by Other Technology-Based Political Communication Interventions

Table A.1: Take-up Rates and Modes of Communication Across Studies

Study Authors	Country	Mode of Communication	Take-up Rate (%)
Grossman, Humphreys and Sacramone-Lutz (2014)	Uganda	SMS	5.8
Leo et al. (April 2015)	Various	IVR	4.3
Ferrali et al. (2016)	Uganda	SMS	2–7
Grossman, Michelitch and Santamaria (2017)	Uganda	SMS	4.8
Erlich et al. (2018)	South Africa	SMS+	2–14
Blair, Littman and Paluck (2019)	Nigeria	SMS	0.1
Buntaine, Hunnicutt and Komakech (2020)	Uganda	SMS	10
Grossman, Humphreys and Sacramone-Lutz (2020)	Uganda	SMS	<0.1
Shaul-Cohen and Lev-On (2020)	Israel	SMS	4–18

Notes: SMS refers to Short Messaging Service. SMS+ refers to a 5-channel study.

B Registry of changes from the pre-analysis plan (PAP)

Key informant interviews used in Table L.1: We originally intended to conduct multiple waves of key informant interviews but were prevented from doing so when field activities were interrupted by government security services. As a result, we deviate from the PAP in two ways. We use heteroskedasticity-consistent standard errors (HC2) rather than clustered standard errors, since the the treatments of interest are assigned at the polling station level, the same level as outcomes. We also remove wave fixed effects since we have none.

Measuring PS level effects: We deviate from our pre-specified analysis, where we had a post-treatment variable (whether the MPA ran again) on the right hand side. We removed that variable in the analysis.

C Ethics of the intervention

Prior to any field activities involving human subjects, we sought permission from Institutional Review Board (IRB) at the University of California at Los Angeles (UCLA). Stanford University's IRB had agreed to subordinate to UCLA.

Our intervention raises at least three major ethical concerns: partisanship, possible interference in the democratic process, and the gender imbalance of our study.

- Partisanship: A first ethical concern is that we partner with incumbent MPAs and do not offer IVR communication to challengers. This might bias the political process in favor of incumbents or their parties.

There were practical reasons for working exclusively with sitting MPAs. Our experiment was rolled out in the nine months prior to our best guess of when the next election would be held. Because of the lead time required for the experiment, it was not feasible to undertake it immediately prior to the election. At the time the experiment was conducted, the identity of assembly candidates in the upcoming 2018 elections was unknown. (Indeed, 25 percent of the incumbents we work with did not run again.) In addition, our funding agency (J-PAL) prohibited us from conducting research that could have been construed as interfering with or participating in the election campaign. For both of these reasons, it was not possible to work in the month just prior the election when campaigning was underway and when the candidates had been selected and their identities known.

These logistical considerations explain the timing of our research, but do not necessarily speak directly to the ethics of working with sitting MPAs. Some might worry that our work would shore up their political power. We were relatively unconcerned about this, however. The reason is that the MPAs we work with did not constitute an entrenched political elite. Indeed, 60 percent of those enrolled in the IVR treatment had never served previously in the provincial or national assembly, and 50 percent of them did not get reelected (either because they were not on the ballot or because they did not win the seat). Similarly, elections in KP are often closely fought; the average margin of victory in 2013 was 11 percent, considerably lower than in Pakistan's other three provinces. Indeed, we considered that providing a new way for MPAs to communicate with voters might encourage a largely inexperienced group of politicians to learn to do their jobs more effectively and thus might improve political representation in KP.

- Political interference: A second ethical concern is that our intervention might have affected political outcomes. However, our sample sizes are too small for this to have been possible. For evidence, see the discussion in Section 5.2.
- Gender: A final ethical concern is that we work only with men and exclude women. As we note however (see fn. 4), even when we sent female staff to approach women voters, they were unwilling to provide their phone numbers and participate in the study. Even though our study thus excluded women, we do not think that our activities harmed them. The reason is that we do not believe that providing a representative sample of male voters new ways to express political and policy opinions is necessarily detrimental to female voters. We would have considered the intervention to carry more problematic ethical implications if it had given voice to an unrepresentative sample of wealthier, older, or more conservative male voters, for instance. That said, we remain troubled by the failure to enroll female respondents in the study and hope to examine how these barriers can be overcome in future work.

D Timeline of the intervention

Owing to weather and security constraints, we implemented the intervention in five phases. Our implementing partner put four of the 20 sampled constituencies into treatment in each

Table D.1: Phased implementation of intervention and surveys

Phase	Baseline survey	Initial calls	Follow-up calls	Endline survey
1	December 2017 - February	February	May - July	August - September
2	January - February	March	May - June	September - October
3	February	April - May	June - July	August - October
4	February - April	May	July	September - October
5	March - April	May	July	August - October

Notes: All activities took place in 2018 unless noted otherwise. Constituency numbers (e.g. "PK-50") were assigned after redelimitation for the 2002 elections. These numbers were changed following a 2018 redelimitation.

stage, and then rolled out the baseline and recordings of messages with MPAs on a staggered basis. The timing of all activities — baseline survey, two stages of phone calls, and endline survey — is reported in Table D.1. In each constituency, households were recruited, provided informed consent, and baselined over a two week period. Simultaneously, meetings with MPAs were ongoing to enroll them in the project, introduce the technology to them, and work with them to script and record their initial calls to constituents. Recordings were made either with field staff on site or later on the MPA's own time. Because of the nature of the meetings and because some MPAs recorded their calls later when they found more time, many recordings were made on mobile phones, although we edited the files for clarity afterwards. Follow-up calls were made in similar fashion. The endline survey was rolled out after the general elections were held on July 25, 2018, also in phases due to the weather as well as security concerns.

E Stage One Script

Introduction and Credit Claiming: Assalam U Alaikum. I am [MPA NAME], your elected MPA. In the past nine years, girls' degree college, boys' degree college, and hundreds of primary, middle, high and higher secondary schools have been built. I have constructed a link road. In addition to this, gas lines to each house have been or will be completed. Furthermore, I have worked hard to speak for the people and their rights on the floor of the assembly.

Question: What do you think I should focus on going forward?

- Education and health, press 1
- Development works, press 2
- Legislation, press 3
- If you want to listen again, press 9

Thank you for taking your valuable time. Good bye.

F Datasets used for tables and figures

Table F.1 enumerates the datasets used in the paper. The main sources of data are: (i) a series of three surveys conducted at the individual respondent level; (ii) key informant interviews conducted at the polling station level; (iii) electoral data to measure outcomes at the polling station area level; (iv) forecasting data collected from external experts; and (v) data generated by the intervention itself. The information in the table clarifies which dataset, and which subsets, are used in each table and figure in the paper.

Table E.1: Responses to first-stage IVR questions: descriptive data

Response	Count	Percent
Education and Health	432	35
Development Works	552	44
Legislation	250	20
Total	1,247	99

Notes: Excludes 13 respondents who pressed 4 in response to "If you are satisfied with my services, Press 4," recorded by one MPA. Total reported includes these subjects and thus does not add up to 100%. Reports numbers only in relation to number of persons who answered the IVR call and question.

Table F.1: Datasets used for each table and figure in the paper

Dataset	Sample size	Notes	Tables & Figures
Respondent level			
Baseline survey	14,399 HHs	Every respondent baselined (except for one duplicate phone number). Some tables use only a subset (e.g. {H2}) when appropriate. Often IVR compliance statistics (e.g. answering the IVR question) come from this data as it is available regardless of whether we endline.	Table K.2; Figures G.1b and 2c (random sample and IVR communicators)
Intervention data	1,247 HH	Respondents who answered the phone and answered a question	Table E.1
Endline survey	13,988 HHs	Every respondent we could reach for the endline. Note that some tables that use this data use only a subset (e.g. {H2}) when appropriate and when noted in the table/table notes.	Tables 1, I.1, I.2, I.3, K.1, K.3, and K.4; Figure J.1
Descriptive survey	2,863 HHs	Those we could reach via phone survey out of 3,600 HHs randomly subsampled from the experimental group assigned to {H2}.	Figures G.1a, 2c (status quo communicators), 2b, and 2a
Polling station area level			
Key informant survey	240 PS areas	PS level data largely from mean responses of two key informants in each of the 6 matched pairs of treatment and control polling station areas per MPA.	Table L.1
Electoral data	300 PS areas	The max set of polling stations for which electoral data could be available (broader than key informant data due to cost of interviewing key informants)	Table M.1
Politician/constituency level			
Interviews	35 KP MPAs and MNAs	Hour-long face-to-face semi-structured open-ended interviews	Reported in text of paper
Other			
Expert *forecasts*	400 respondents	Solicited in-person and over email. Includes undergraduate students, research/development practitioners, and academics around the world.	Figure N.1

Table G.1: Distance to MPA's constituency office and average interaction rate, by PS area

	Outcome			
	Status quo		IVR	
	% who met MPA in last yr		% who answer IVR question	
Intercept	16.699***		17.750***	
	(2.854)		(2.461)	
Middle 2 PS	−1.441		−0.250	
	(1.252)		(1.646)	
Furthest 2 PS	−3.026†		0.750	
	(1.699)		(1.709)	
Distance (km)		−0.225*		0.004
		(0.104)		(0.102)
Constituency FEs		Yes		Yes
Num. Obs.	120	120	120	120
R2	0.009	0.704	0.002	0.710

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. Standard errors clustered by constituency in parentheses. The first and third models have indicators for polling station distance by tercile within constituency. The omitted category is the first tercile, which includes the two closest polling stations.

G Additional Descriptive Results

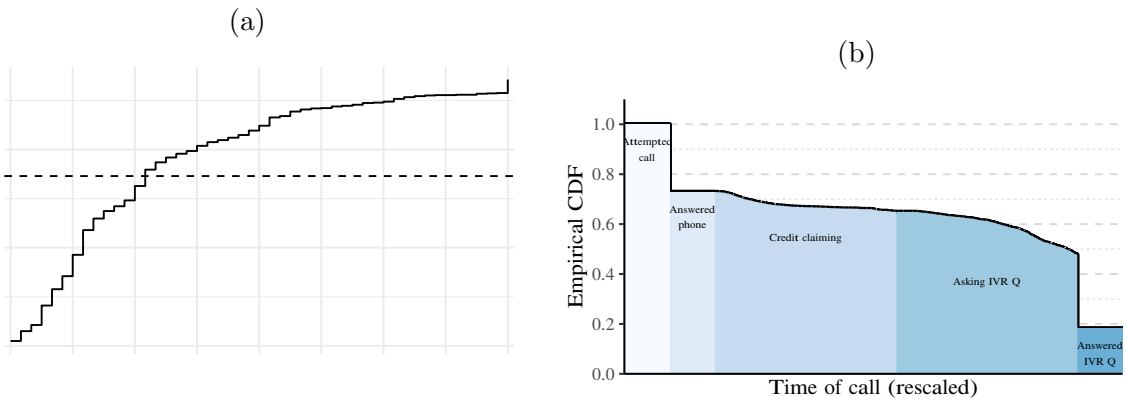
G.1 How contact varies by distance

Table G.1 follows up with results from two sets of polling station level regressions, distinguished according to whether the respondent reports having met his MPA in person in the last year or whether he answered an IVR question. The regressions study the relationship between distance to an MPA's constituency office and each type of contact with the MPA. We measure distance as the geodesic distance between the respondent's polling station and the MPA's constituency office.[17] We specify distance in two different ways, controlling for constituency level effects in two different ways to deal with the variation in constituency size in our sample: in the first column for each outcome we regress the percent of respondents reporting contact with the MPA on distance with constituency fixed effects; in the second column we instead aggregate respondents into terciles of the distance of the respondent's polling station within each constituency. In both cases, there is evidence that distance to the MPA more strongly predicts status quo contact rates than IVR, and that IVR thus reaches a more geographically diverse set of constituents. In the first column, the coefficient on distance shows that for each kilometer further from the MPA's constituency office, 0.2 percentage point fewer respondents met in person with the MPA; moving from the 25th percentile (4km) to the 75th (15km) percentile on distance corresponds to a 2.5 percentage point decrease in the percent of respondents who report meeting face-to-face with the MPA. Thus, voters who live further from the MPA's constituency office are disadvantaged by the need to travel for face-to-face meetings.

[17]For IVR communication, the MPA is located at his 2013 constituency office whereas for status quo communication, he is located at his 2018 constituency office. Using the 2018 MPA's constituency office for the distance calculations for IVR produces substantively similar results.

G.2 The scale of contact under the status-quo and IVR

Figure G.1: Descriptive Evidence on Scale of Communication Transformations Under IVR



Notes: (a) Target and nature of requests made to MPA in status quo meetings. The solid line plots the cumulative distribution of time since survey respondents last met their MPA in person. The dashed line is the average rate of contact under IVR (answered question). (b) Proportion of respondents who remain on the IVR call at various stages of the call. The x-axis is rescaled for illustrative purposes to account for differing call lengths across MPAs.

H Household level index construction

At the household level we build each index following Kling, Liebman and Katz (2007). We first standardize all of the component outcomes by the mean and standard deviation of the outcome in the group that received none of the IVR intervention calls {H0}. We then impute all missing component outcome means to the average of that component outcome in the stage one household treatment group. Note that if there is missingness on all component outcomes for an index for a particular individual, none of the outcomes is imputed and that individual is dropped. We restandardize the indices with respect to the control group (always defined as {H0}), so that effects are interpretable in standard deviation units of the index.[18] Furthermore, because we restandardize only once and with respect to the control households, when we make comparisons between treatment arms, the standard deviation in the comparison group is not always equal to one.

I Alternative specifications for downstream household results

In this section we present the robustness of the null experimental results presented in Table 1 to (i) attrition, (ii) alternative specifications of two indices, and (iii) considering compliance rates by estimating local average treatment effects among compliers.

[18]The original indices were interpretable as an average of standard deviation unit treatment effects on the component measures, rather than as a standard deviation treatment effect on the index itself (Kling, Liebman and Katz, 2007).

First, we present robustness of the main results to attrition. The results in Table I.1 replicate the main, downstream household level results in Table 1 but account for attrition using inverse probability of attrition weights. Using the full experimental sample we first estimate the probability a respondent attrited and then use these estimated probabilities to weight the non-attritors to overrepresent the respondents who have similar characteristics to those respondents who attrited. If the model estimating the probability of attrition is well-specified, then these weights will unbiasedly estimate the treatment effect among the full sample, including attriters. While we do not expect our model to be perfect, this is a common approach to dealing with attrition, especially when treatment status does not predict attrition and when it is not severe.

The weights for non-attriters used in the analysis below are

$$w_i = \frac{1}{\hat{p}_i}$$

where $\hat{p}_i$ is predicted probability of non-attrition from a regression of non-attrition on co-partisanship, age bins, income scale groups, education bins, political knowledge, and MPA feeling thermometer bins.

Second, political conversations could instead be considered political participation rather than a precursor to accountable electoral politics. As such, we rebuild the indices with political conversations moved to the political participation index and present the results in Table I.2.

Third, compliance with the full IVR treatments was around 17 percentage points. As such, local average treatment effects among compliers will be larger than intent to treat effects. We present local average treatment effects among compliers in Table I.3 where we define compliance with getting any call {H1, H2} as answering the first stage phone call and we define compliance with the full, responsive treatment {H2R} as answering the first stage IVR question and answering the second stage phone call. For all analyses, we use the same specifications as the main results and instrument for the binary indicator of compliance with the treatment assignment. Even with these fairly restrictive definitions of compliance, the treatment effects remain substantively small and, unsurprisingly given the nature of the estimating local treatment effects among compliers, they remain statistically insignificant.

Table I.1: Effect of any IVR call and effect of full IVR treatment on household head outcomes - including index component measures and weighted for attrition

	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
Outcome	μ	τ	N	τ	N
Incumbent evaluations index	0.000	−0.009	13757	−0.016	6539
	(1.000)	(0.009)		(0.013)	
MPA feeling thermometer (1-10)	4.864	−0.056	13753	−0.087	6536
	(3.340)	(0.038)		(0.058)	
MPA party feeling thermometer (1-10)	4.536	−0.018	13758	−0.019	6538
	(3.501)	(0.035)		(0.056)	
Voted for MPA (0/1)	0.337	−0.004	13753	−0.010	6538
	(0.473)	(0.004)		(0.007)	
Inverse rank of MPA (1-5)	2.661	−0.001	13309	−0.012	6307
	(1.478)	(0.018)		(0.027)	
Political participation index	0.000	−0.021	13780	0.004	6551
	(1.000)	(0.016)		(0.025)	
Voted (0/1)	0.985	−0.001	13260	0.002	6282
	(0.122)	(0.002)		(0.003)	
Attended rally (0/1)	0.239	−0.008	13760	−0.001	6539
	(0.427)	(0.007)		(0.011)	
Attended political meeting (0/1)	0.180	−0.007	13780	−0.002	6551
	(0.385)	(0.006)		(0.010)	
Prospects for accountability index	0.000	0.004	13759	0.025	6539
	(1.000)	(0.017)		(0.026)	
Political efficacy (1-5)	3.781	0.003	13930	0.058†	6618
	(1.163)	(0.020)		(0.030)	
Vote choice based on performance (1-6)	4.267	−0.007	13703	−0.020	6514
	(1.684)	(0.025)		(0.039)	
N political conversations	3.739	0.030	13978	0.029	6642
	(2.466)	(0.040)		(0.062)	
Global index	0.000	−0.015	13950	0.005	6629
	(1.000)	(0.013)		(0.020)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses. Pre-treatment control variables not displayed; see Section 5.1 for details. Control means, treatment effects, and all estiamtes of uncertainty are weighted using inverse probability of attrition weights. These weights were generated using a linear model incorporating income, education, age, political knowledge, MPA feeling thermometers, and copartisanship with the MPA.

Table I.2: Effect of any IVR call and effect of full IVR treatment on household head outcomes - moving conversations to participation index

	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
Outcome	μ	τ	N	τ	N
Incumbent evaluations index	0.000	−0.009	13757	−0.016	6539
	(1.000)	(0.009)		(0.013)	
MPA feeling thermometer (1-10)	4.864	−0.056	13753	−0.087	6536
	(3.340)	(0.038)		(0.058)	
MPA party feeling thermometer (1-10)	4.536	−0.018	13758	−0.019	6538
	(3.501)	(0.035)		(0.056)	
Voted for MPA (0/1)	0.337	−0.004	13753	−0.010	6538
	(0.473)	(0.004)		(0.007)	
Inverse rank of MPA (1-5)	2.661	−0.001	13309	−0.012	6307
	(1.478)	(0.018)		(0.027)	
Political participation index (w/ convs.)	0.000	−0.012	13780	0.009	6551
	(1.000)	(0.016)		(0.024)	
Voted (0/1)	0.985	−0.001	13260	0.002	6282
	(0.122)	(0.002)		(0.003)	
Attended rally (0/1)	0.239	−0.008	13760	−0.001	6539
	(0.427)	(0.007)		(0.011)	
Attended political meeting (0/1)	0.180	−0.007	13780	−0.002	6551
	(0.385)	(0.006)		(0.010)	
N political conversations	3.739	0.030	13978	0.028	6642
	(2.466)	(0.040)		(0.062)	
Prospects for accountability index (no convs.)	0.000	−0.004	13759	0.023	6539
	(1.000)	(0.016)		(0.025)	
Political efficacy (1-5)	3.781	0.003	13930	0.059†	6618
	(1.163)	(0.020)		(0.030)	
Vote choice based on performance (1-6)	4.267	−0.007	13703	−0.020	6514
	(1.684)	(0.025)		(0.039)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Table I.3: Effect of any IVR call and effect of full IVR treatment on household head outcomes — local average treatment effects among compliers

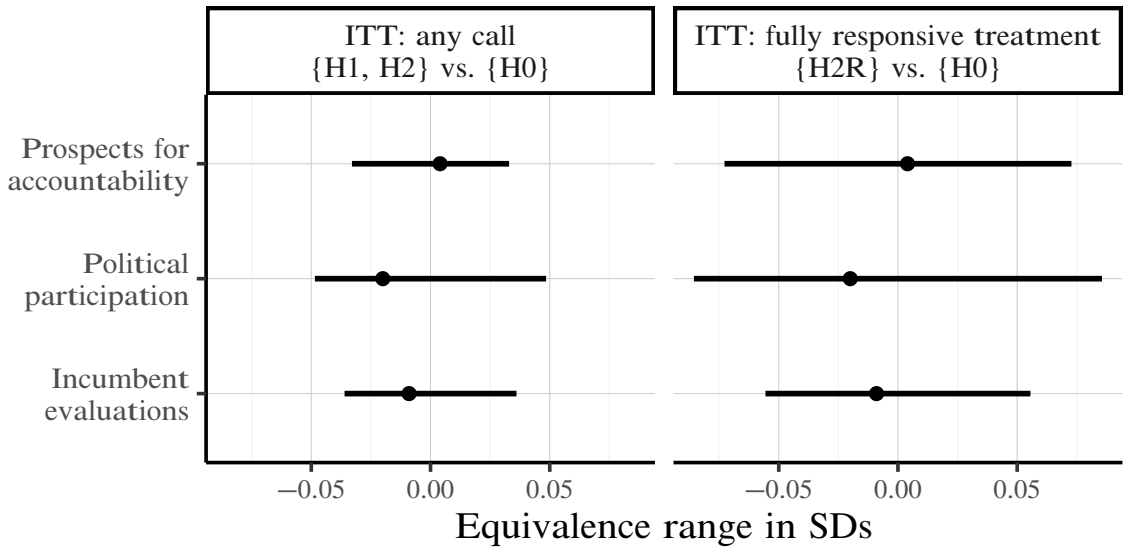
	Control mean:	LATE: any call (answered phone)		LATE: full responsive treatment (answered first q and second call)	
	no call {H0}	{H1, H2} vs. {H0}		{H2R} vs. {H0}	
Outcome	μ	τ	N	τ	N
Incumbent evaluations index	0.000	−0.012	13757	−0.108	6539
	(1.000)	(0.012)		(0.091)	
MPA feeling thermometer (1-10)	4.864	−0.077	13753	−0.596	6536
	(3.340)	(0.052)		(0.399)	
MPA party feeling thermometer (1-10)	4.536	−0.024	13758	−0.128	6538
	(3.501)	(0.048)		(0.385)	
Voted for MPA (0/1)	0.337	−0.006	13753	−0.066	6538
	(0.473)	(0.006)		(0.047)	
Inverse rank of MPA (1-5)	2.661	−0.001	13309	−0.084	6307
	(1.478)	(0.024)		(0.187)	
Political participation index	0.000	−0.028	13780	0.027	6551
	(1.000)	(0.022)		(0.171)	
Voted (0/1)	0.985	−0.001	13260	0.011	6282
	(0.122)	(0.003)		(0.023)	
Attended rally (0/1)	0.239	−0.011	13760	−0.004	6539
	(0.427)	(0.009)		(0.073)	
Attended political meeting (0/1)	0.180	−0.009	13780	−0.014	6551
	(0.385)	(0.009)		(0.067)	
Prospects for accountability index	0.000	0.006	13759	0.174	6539
	(1.000)	(0.023)		(0.177)	
Political efficacy (1-5)	3.781	0.004	13930	0.401†	6618
	(1.163)	(0.027)		(0.209)	
Vote choice based on performance (1-6)	4.267	−0.010	13703	−0.137	6514
	(1.684)	(0.034)		(0.267)	
N political conversations	3.739	0.041	13978	0.196	6642
	(2.466)	(0.055)		(0.428)	
Global index	0.000	−0.020	13950	0.037	6629
	(1.000)	(0.018)		(0.135)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.

J Robustness of null findings in downstream results

One concern with the null results is that the experiment might have been statistically underpowered and therefore unable to detect effects even if they exist. Although this may be the case at the polling station level, it is unlikely to be true at the household level, where we have a large sample. We can formalize this by computing equivalence confidence intervals, as proposed in Hartman and Hidalgo (2018). We estimate equivalence confidence intervals that contain treatment effects that are small enough that we cannot reject the null that they are too large. In other words, the values within the equivalence confidence intervals that we report in Figure J.1 are small enough to be consistent with the data; larger treatment effects can be rejected as too large given the data at our given significance level (here, 0.05). Therefore, instead of relying on failing to reject the null to establish a null effect, these confidence intervals allow us to find the largest treatment effects — those on the ends of the equivalence confidence intervals — that we cannot reject as too large. For the effect of receiving any call in the left panel, all treatment effects on individual attitudes and self-reported behavior larger than |0.05| sds can be rejected as too large, while effects larger than |0.08| sds are inconsistent with the observed effect of the full treatment (as seen in the right panel). These "largest possible effects" are quite small and reflect the substantial power of our design to detect meaningful individual level effects.

Figure J.1: Equivalence confidence intervals for main household level treatment effects



This figure contains the realized treatment effects and equivalence confidence intervals (Hartman and Hidalgo, 2018) for the two main analyses we report in Table 1. The points are the realized treatment effects and the equivalence confidence intervals are built at the 0.05 level. Our three main outcome indices are on the y-axis and the treatment effects in the original units of the outcomes, standard deviations, are on the x-axis. The equivalence confidence intervals represent the range of hypothetical treatment effects that are consistent with our data and estimated treatment effects. Any hypothetical treatment effect outside these intervals can be rejected by an equivalence test as too large at the 0.05 level.

K Additional downstream results

We preregistered other analyses between various treatment arms: the marginal effect of receiving an initial call with questions (H2) versus receiving the initial call with no questions (H1); the marginal effect of receiving a responsive follow-up call (H2R) versus a generic follow-up call (H2G); and the marginal effect of receiving *any* follow-up call (H1G + H2G + H2R) versus no follow-up call (H1C + H2C). The first two analyses report similar treatment effects , where we find no large substantive effects with no statistically significant treatment effects. The only treatment effect that is statistically significant at even the 0.1 level comes when considering the marginal effect of receiving any follow-up call.

This section expands tables in the main paper to include index components, and it includes additional treatment group comparisons.

Table K.1: **Effect of any IVR call and full IVR treatment on household head outcomes including index components**

	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
Outcome	μ	τ	N	τ	N
Incumbent evaluations index	0.000	−0.009	13757	−0.016	6539
	(1.000)	(0.009)		(0.013)	
MPA feeling thermometer (1-10)	4.864	−0.056	13753	−0.087	6536
	(3.340)	(0.038)		(0.058)	
MPA party feeling thermometer (1-10)	4.536	−0.018	13758	−0.019	6538
	(3.501)	(0.035)		(0.056)	
Voted for MPA (0/1)	0.337	−0.004	13753	−0.010	6538
	(0.473)	(0.004)		(0.007)	
Inverse rank of MPA (1-5)	2.661	−0.001	13309	−0.012	6307
	(1.478)	(0.018)		(0.027)	
Political participation index	0.000	−0.020	13780	0.004	6551
	(1.000)	(0.016)		(0.025)	
Voted (0/1)	0.985	−0.001	13260	0.002	6282
	(0.122)	(0.002)		(0.003)	
Attended rally (0/1)	0.239	−0.008	13760	−0.001	6539
	(0.427)	(0.007)		(0.011)	
Attended political meeting (0/1)	0.180	−0.007	13780	−0.002	6551
	(0.385)	(0.006)		(0.010)	
Prospects for accountability index	0.000	0.004	13759	0.025	6539
	(1.000)	(0.017)		(0.026)	
Political efficacy (1-5)	3.781	0.003	13930	0.059†	6618
	(1.163)	(0.020)		(0.030)	
Vote choice based on performance (1-6)	4.267	−0.007	13703	−0.020	6514
	(1.684)	(0.025)		(0.039)	
N political conversations	3.739	0.030	13978	0.028	6642
	(2.466)	(0.040)		(0.062)	
Global index	0.000	−0.014	13950	0.005	6629
	(1.000)	(0.013)		(0.020)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Table K.2: **ITT effects of initial call type on follow-up pickup rates**

	Control mean: call only {H1}	ATE: effect of getting asked IVR question vs. call only {H2} vs. {H1}	
Outcome	μ	τ	N
Answered follow-up phone call (0/1)	0.787	0.036*	3718
	(0.410)	(0.015)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001.
Notes: Heteroskedasticity-consistent (HC2) standard errors in parentheses. Because nothing about a respondent's treatment condition is revealed before picking up the phone, initial call treatment status (e.g. {H1} or {H2}) only affects the respondent once he answers the initial call. Therefore, we subset the analysis to respondents who answer the first call.

Table K.3: **Marginal effect of IVR question in initial call on household head outcomes including index component measures**

	Control mean: credit claiming call only {H1}	ITT: marg effect of IVR q {H2} vs. {H1}	
Outcome	μ	τ	N
Incumbent evaluations index	−0.018	0.004	9164
	(1.007)	(0.012)	
MPA feeling thermometer (1-10)	4.778	0.014	9162
	(3.377)	(0.051)	
MPA party feeling thermometer (1-10)	4.480	0.031	9166
	(3.486)	(0.046)	
Voted for MPA (0/1)	0.331	−0.002	9160
	(0.471)	(0.005)	
Inverse rank of MPA (1-5)	2.646	0.009	8866
	(1.473)	(0.023)	
Political participation index	−0.014	−0.009	9179
	(0.987)	(0.021)	
Voted (0/1)	0.985	−0.001	8834
	(0.123)	(0.003)	
Attended rally (0/1)	0.234	−0.001	9167
	(0.423)	(0.009)	
Attended political meeting (0/1)	0.176	−0.003	9179
	(0.381)	(0.008)	
Prospects for accountability index	−0.003	0.016	9166
	(0.995)	(0.022)	
Political efficacy (1-5)	3.789	−0.003	9283
	(1.146)	(0.026)	
Vote choice based on performance (1-6)	4.257	0.010	9129
	(1.681)	(0.033)	
N political conversations	3.723	0.054	9312
	(2.357)	(0.052)	
Global index	−0.022	0.006	9293
	(1.009)	(0.017)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Table K.4: **Marginal effect of responsive vs. generic follow-up call on household head outcomes including index components**

	Mean: Generic Follow-up {H3G}	ITT: Marg. Effect of Responsive Follow-up {H3R} vs. {H3G}	
Outcome	μ	τ	N
Incumbent Evaluations Index	0.000	−0.016	6539
	(1.000)	(0.013)	
MPA Feeling Thermometer (1-10)	4.864	−0.087	6536
	(3.340)	(0.058)	
MPA Party Feeling Thermometer (1-10)	4.536	−0.019	6538
	(3.501)	(0.056)	
Voted for MPA (0/1)	0.337	−0.010	6538
	(0.473)	(0.007)	
Inverse Rank of MPA (1-5)	2.661	−0.012	6307
	(1.478)	(0.027)	
Political Participation Index	0.000	0.004	6551
	(1.000)	(0.025)	
Voted (0/1)	0.985	0.002	6282
	(0.122)	(0.003)	
Attended rally (0/1)	0.239	−0.001	6539
	(0.427)	(0.011)	
Attended political meeting (0/1)	0.180	−0.002	6551
	(0.385)	(0.010)	
Prospects for Accountability Index	0.000	0.025	6539
	(1.000)	(0.026)	
Political efficacy (1-5)	3.781	0.059†	6618
	(1.163)	(0.030)	
Vote choice based on performance (1-6)	4.267	−0.020	6514
	(1.684)	(0.039)	
N political conversations	3.739	0.028	6642
	(2.466)	(0.062)	
Global Index	0.000	0.005	6629
	(1.000)	(0.020)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

L Effects on politician behavior

To measure politician behavior, we administer structured *key informant* interviews in the six treated polling station areas in each constituency where voters receive IVR and in six of the 14 control polling station areas where they do not. We conduct key informant interviews before and after the intervention. Key informants were generally salaried individuals (school teachers, for instance) who had no personal or professional relationship with the MPA and who were not involved in campaigning for any political party. Enumerators selected two key informants in each polling station area, both of whom were deemed likely to be available for repeated interviews.

We ask key informants whether politicians visit the polling station area: the goal is to evaluate whether they visit areas more where we direct their IVR phone calls. We also

ask key informants whether politicians exert effort in delivering public services, including schools, roads, health facilities, employment conditions, electricity provision, gas provision, water provision, rubbish collection, and general security.

We estimate effects on these outcomes with OLS using the following specification:

$$Y_{pm(t=1)} = \tau D_p + \alpha Y_{pm(t=0)} + \lambda_m + \epsilon_{pm},$$

where $Y_{pm(t=1)}$ is outcome Y at polling station p at endline ($t = 1$), D_p is a binary indicator for treatment status, $Y_{pm(t=0)}$ is the pre-treatment outcome Y collected in the baseline (if available), and λ_m is a constituency fixed effect. We use heteroskedasticity-consistent standard errors (HC2) since the the treatments of interest here are assigned at the polling station level, the same level as the outcomes.

Table L.1 shows that there is little evidence of effects on politician effort in places they administer IVR, although estimates are not very precisely estimated due to the relatively small number of observations. Key informants report that, in control areas, MPAs made some effort in only 0.48 of the nine public goods domains. In addition, on average, politicians are reported as making some kind of effort in just over a quarter of areas (the control mean is 0.27), and there is almost no effect of treatment on this. Very few areas (only 0.05) saw their MPA visit in June, again virtually unaffected by treatment. Overall, we observe little effect on politician behavior from being enrolled in treatment: their low levels of effort remain as before.

Table L.1: **ITT effects of IVR calls on MPA effort in polling station areas**

	Control mean: control PS {P0}	ITT: treated PS {P1} vs. {P0}	
Outcome	μ	τ	N
N of domains where MPA made effort	0.483	0.017	240
	(1.004)	(0.090)	
Any MPA effort (0/1)	0.267	0.050	240
	(0.444)	(0.047)	
Any MPA visit in June (0/1)	0.053	0.035	240
	(0.153)	(0.022)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses. Pre-treatment control variables not displayed; see text for details.

M Polling station level treatment effects

Here we analyze downstream results at an aggregate level: polling station level electoral returns. As these results are realized after both voters and politicians have acted in response to treatment, results represent short-term equilibrium experimental outcomes. We compare outcomes for the six treated polling stations to those for the 14 polling stations that we did not treat within the set of 20 most competitive polling stations for each MPA. Thus, all polling stations are within a partner MPA's constituency although the MPA only used IVR to households in the six treated polling stations.

Data: Since we randomize across 20 polling stations in 20 MPA areas, we should have outcome data from 400 polling stations. We successfully collected official election data from only 341 constituencies. The remaining data are missing because: (i) some .psf format returns released by the Election Commission are illegible; (ii) in some instances we were unable to match our polling stations with polling stations resulting from a subsequent redelimitation; and (iii) initial results for 20 polling stations in one constituency were annulled because of low female turnout and the later results have not been made publicity available by the Election Commission.

Estimation: We estimate effects using OLS and the following specification:

$$Y_{pm(t=2018)} = \tau D_p + \alpha Y_{pm(t=2013)} + \lambda_m + \epsilon_{pm},$$

where $Y_{pm(t=2018)}$ is the outcome Y for polling station p in MPA constituency m in the 2018 election, D_p is a binary indicator for treatment status, $Y_{pm(t=2013)}$ is the pre-treatment outcome Y in the 2013 election, and λ_m is an MPA constituency fixed effect. As before, we use HC2 standard errors, since the treatment assignment is at the polling-station level.

We estimate effects on two outcomes at the polling station level: the vote share for the incumbent (partner) MPA and the turnout rate. Because of re-delimitation and because some of our partner MPAs did not seek office again, our partner MPAs were not candidates in 2018 in every polling station in our sample. In cases where the partner MPA was not a candidate for any party, we code for the candidate from the party with which our partner MPA was last associated.

Effects within treated constituencies: Table M.1 presents intervention effects from treated polling stations compared to control polling stations within treated MPA constituencies.

On average, evaluation of incumbents and turnout in elections in control areas remain low; about 33 percent and 47 percent respectively. The low vote shares received by incumbent MPAs are consistent with the generally low reelection rates of incumbents across the developing world (Golden and Nazrullaeva, 2023). Elections in KP often have more than

Table M.1: ITT effects of IVR calls on polling station voting outcomes

	Control mean: control PS {P0}	ITT: treated PS {P1} vs. {P0}	
Outcome	μ	τ	N
Incumbent MPA vote share	0.332	0.002	341
	(0.165)	(0.016)	
Turnout share	0.477	0.006	288
	(0.109)	(0.012)	

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001.
Notes: Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Notes: Results presented here are ITT effects estimated using OLS.

two competitive candidates, meaning the local political environment is unstable and highly competitive, which feeds into low reelection rates.

In general, we do not find evidence to suggest that treatment affected election results, either in terms of stated incumbent vote share — whose point estimate is very close to zero — or voter turnout. Due to high attrition in the sample of polling stations, we are unable to state that we have estimated a precise null effect of the intervention on voting behavior, however.

N What Did We Expect? A Forecasting Exercise

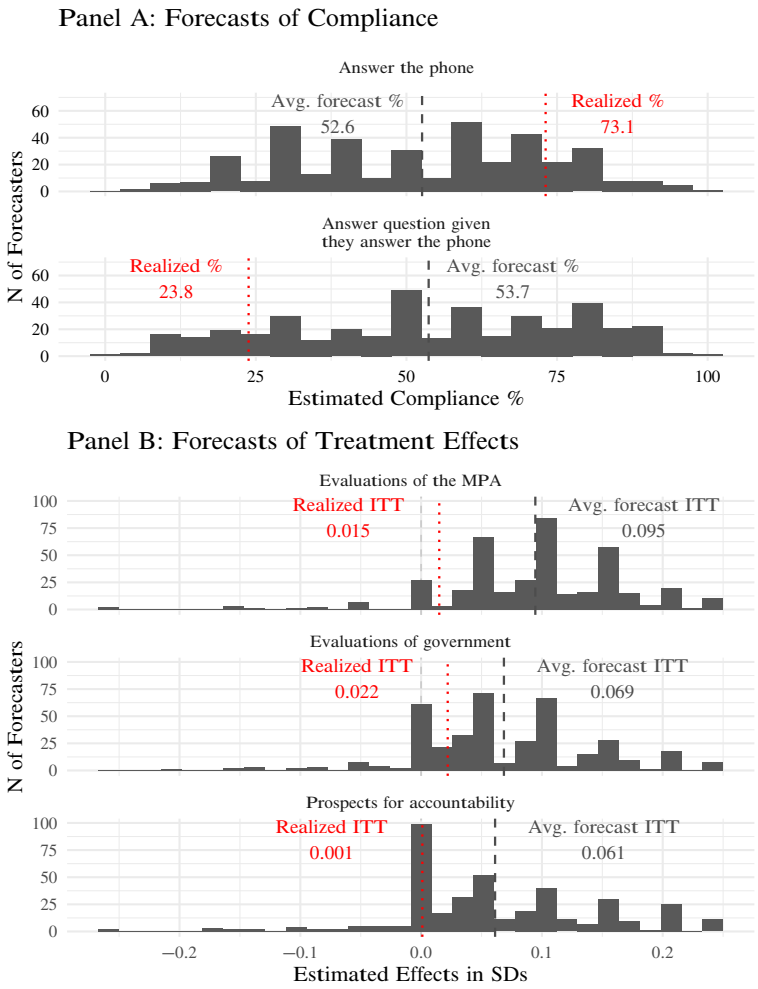
The aim of the forecasting exercise is not to get a representative set of forecasters but rather to capture the ex ante beliefs of those who express an interest in the research by volunteering their time. The exercise follows DellaVigna and Pope (2018), which documents the utility of using expert forecasts to understand what we can learn from experiments. We assume interest and relative expertise among such volunteers.

We analyze results from two sets of forecasts: a first with 283 (out of 1,584 solicited) academic experts in or associated with the United States[19] and a second with 117 (out of 172 solicited) university students, faculty, and policy practitioners in Pakistan, for a total of 400 responses. The two waves allow examination of differences between local and global expertise (Casey et al., 2018). After briefly describing the context and the experiment, we ask respondents to forecast the take-up rates for the IVR calls as well as impacts on downstream results on the three key voter indices analyzed above.

Results: Panel A in Figure N.1 shows that, on average, nearly 40 percent more people answer the phone than forecasters predict. However, conditional on whether the IVR call was answered, forecasters were more likely to think that respondents would answer an IVR

[19]Academics "associated" with the United States were members of the American Political Science Association's Organized Section in Comparative Politics.

Figure N.1: Forecast and realized compliance and treatment effects



Notes: All differences between the average forecast and realized estimates are statistically significant with p<0.001.

question than the number who actually did. That is, forecasters underestimate whether respondents would answer the phone but overestimate whether respondents would answer an IVR question.

Panel B contrasts the forecasts against realized results on the three downstream outcome indices. Forecasters were asked to predict the intent-to-treat effect for each index. Forecasters overestimate the size of all effects: in every case, average forecasts were substantially (and statistically significantly) above realized average treatment effects. In other words, forecasters expect the intervention to produce relatively large results.

Of course, forecasts of field experiments might generally be overly optimistic. One reason lies with publication bias: almost all field experiments that get published report successful interventions. Those that produce null results are typically difficult or even impossible to publish (Christensen, Freese and Miguel, 2019). Perhaps in part for this reason, forecasting

shows that the results of the intervention we report were not predicted by interested experts. Experts expect take-up to be less than was the case and they expect downstream results to be larger than those obtained. These differences show the intervention produced unanticipated results.

Appendix References

Blair, Graeme, Rebecca Littman and Elizabeth Levy Paluck. 2019. "Motivating The Adoption of New Community-Minded Behaviors: An Empirical Test in Nigeria." *Science Advances* 5(3).

Buntaine, Mark T., Patrick Hunnicutt and Polycarp Komakech. 2020. "The Challenges of Using Citizen Reporting to Improve Public Services: A Field Experiment on Solid Waste Services in Uganda." *Journal of Public Administration Research and Theory* .

Erlich, Aaron, Danielle F. Jung, James D. Long and Craig McIntosh. 2018. "The double-edged sword of mobilizing citizens via mobile phone in developing countries." *Development Engineering* 3:34–46.

Ferrali, Romain, Guy Grossman, Melina Platas Izama and Jonathan Rodden. 2016. "Explaining Uptake of the U-Bridge Program: Network Analysis Final Report." Unpublished paper.

Golden, Miriam A. and Eugenia Nazrullaeva. 2023. *The Puzzle of Clientelism: Elections and Political Discretion Around the World*. Elements in Political Economy New York: Cambridge University Press.

Grossman, Guy, Kristin Michelitch and Marta Santamaria. 2017. "Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization." *Comparative Political Studies* 50(10):1325–57.

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.

Grossman, Guy, Macartan Humphreys and Gabriella Sacramone-Lutz. 2020. "Information Technology and Political Engagement: Mixed Evidence from Uganda." *Journal of Politics* 82(4):1321–36.

Hartman, Erin and F. Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." *American Journal of Political Science* 62:1000–13.

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.

Shaul-Cohen, Sara and Azi Lev-On. 2020. "Smartphones, Text Messages, and Political Participation." *Mobile Media and Communication* 8(1):62–82.