'Press 1 for Roads': Improving Political Communication with New Technology [*]

Miriam Golden[†]

Saad Gulzar[‡]

Luke Sonnet[§]

May 3, 2021

Word count: 10,342 words

Abstract: Politicians face chronic problems collecting accurate information from voters about their policy preferences, a problem especially severe where interactions occur through face-to-face meetings that are potentially dominated by an unrepresentative subgroup of voters. We supplement existing interactions by experimentally providing Interactive Voice Response technology to Pakistani politicians, allowing them to script and record questions for voters and allowing voters to respond on cell phones. The new technology changes the initiator, scope, content, scale, personalism, and frequency of two-way political communication. Our results present mixed evidence on the efficacy of improving political communication when instigated by politicians: even though politicians and voters both exhibit eagerness to engage in this shift, politicians do not follow through with changes in policy-relevant behavior and voters' downstream political attitudes and behavior remain unaltered. We discuss why this might be, emphasizing that face-to-face interactions are not always unrepresentative of voters or inadequate in collecting their preferences. [150 words]

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

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

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

1 Introduction

In representative democracies around the world, politicians struggle to communicate with voters and to collect reliable and accurate information about voters' policy preferences. Elections are blunt instruments that affect who represents voters but do little to convey policy content. Policy-relevant communication between politicians and voters is irregular, infrequent, and unstandardized. Structures for voters to express preferences are weak and there are almost no instruments allowing voters to respond directly to policy choices or to convey policy priorities. Thus, it is perhaps unsurprising that there may be considerable slippage between what voters prefer and the policies that politicians enact. There is, accordingly, a long and distinguished literature studying the hows and whys of elite hijacking of democratic political systems, starting with Mills (1956) and Bachrach and Baratz (1962) and extending more recently to Gilens (2012), Achen and Bartels (2016), Broockman and Skovron (2018), and Gulzar, Hai and Paudel (2021).

This literature identifies at least two distinct theories of slippage in policy representation: differences between characteristics of those who run for office and the electorate[1] and problems that arise in how effectively politicians do their jobs once elected. A part of the second theory, which comprises the focus of this paper, involves communication between politicians and voters. Political elites aim to be responsive to citizens' concerns may lack channels of communication with voters that supply the information necessary to enact responsive policies.[2] We study one possible way that political communication between elected officials and voters can be improved and investigate whether an improvement in communication triggers processes that yield more representative policy. Understanding this process is important because robust political communication forms the foundations upon which effective repre-

---

[1]See Gulzar (2021) for a review of this literature.

[2]There are of course a host of other accounts, such as those that attend to improper incentives to keep elite action in check.

sentation is build: in order for politicians to be responsive, they must know what voters want.

Most prior studies aiming to improve political communication in low-income democracies intervene with the voter to initiate change.[3] These studies provide voters information and some also ask them to express unsolicited opinions to political representatives (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). The hope is that informing and motivating voters catalyzes a bottom-up process of change that will subsequently pressure politicians to become more responsive to voter concerns. However, efforts at improving responsiveness and accountability by asking voters to amplify their political voice have met limited success. Often voters do not use the information provided to change their behavior; when they do, the intervention backfires; or offers of new communication technologies to augment the information flow confront low and unrepresentative take-up (detailed in Table A.1).

In contrast to this line of work, we ask if political communication initiated and led by politicians instead of voters carries greater promise; can it boost the quality of representative institutions, spring-boarding political communication into a virtuous cycle of regular policy feedback? Working directly with Pakistani politicians, we initiate a field experiment that scales up and supplements existing face-to-face interactions between them and their constituents using information technology tools that are already deployed in many comparable contexts but had not been utilized in our field setting. Specifically, we use Interactive Voice Response (IVR) cell phone-based communication. Politicians script and record questions for voters, which are disseminated as robocalls to voters' cell phones. Voters answer the questions using the number keys on their phones. Their responses are then made available

---

[3]Perhaps because scholars have assumed that political communication and responsiveness operate effortlessly and effectively in high-income democracies, we are unaware of many interventions aimed at improving representation in such settings. But see Broockman and Skovron (2018).

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. The expectation is that voters will respond to this new high-frequency, unsolicited, and two-way contact with their elected representative with greater political support for him and also by moving towards performance-based voting criteria. Politicians, on the other hand, will respond to the opportunity to communicate more quickly with larger numbers of voters and to receive aggregated information on their policy preferences by improving service delivery.

Working in Pakistan, an understudied lower-middle income democracy and the fifth most populous country in the world, we employ mixed methods and collect experimental as well as multiple types of observational data. Together, these put politician-voter interactions under a microscope from various angles. Based on our pre-specified analyses, we present four key results. *First*, in contrast to previous work that aims to boost political communication, we see that both politicians and citizens display high levels of engagement when communication is instigated via politicians through IVR. *Second*, improving communication improves public engagement with elected officials on multiple dimensions. It speeds up political interactions, improves their 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. The only positive impact of the intervention consists of suggestive evidence that repeated and responsive communication improves the capacity of voters to use performance-based rather than inherited or particularistic criteria as the basis for electoral choices. *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.

Overall, our results present mixed evidence on the efficacy of improving political communication when instigated by politicians: voters and politicians are willing and even eager to transform how they interact with each other and to adopt more efficient, equitable, and impersonal communication technologies, but whether this improves governance and political representation more broadly remains an open question. In the concluding section of this paper, we speculate about the possible barriers to improved representation that our work suggests may be present in the context we study. There we also highlight the specific theoretical and empirical contributions of our work.

Our paper proceeds as follows. First, we briefly review the literature on how politicians and voters interact and communicate, and explain why we think that IVR could significantly change the nature, scope, and frequency of interactions and the information politicians collect. Second, we present descriptive data we collected that allow us to characterize status quo interactions in our research setting. Third, we present the design of the IVR experiment that we implemented. Fourth, we document the degree of engagement with IVR by politicians and voters. In a fifth section, we analyze downstream attitudes and behavior to show neither party exhibited changes in these domains. We then present results of a forecasting exercise that documents that both engagement and downstream results are surprising to experts. In a final section, we step back to consider the overall importance of our findings, speculate on how to interpret them, and suggest routes for further research.

2 Political Interactions between Voters and Legislators

2.1 Status quo interactions

For politicians to enact representative policies, they must know what voters prefer. A few places institutionalize channels of communication for voters to express preferences via policy referenda. However, referenda mix policy communication and enactment and are not often

used. Instead, in democracies voters typically communicate policy preferences on their own initiative via phone, internet, mail, and in person. In less developed countries, interactions between politicians and voters often overlay the traditional interactions between patrons and clients, or between elites and ordinary citizens. For this and other reasons, including poor mail services, incomplete internet coverage, and the cost of cell phone use, interactions between politicians and voters typically take place face-to-face in less developed democracies.

Face-to-face interactions require the politician visit his constituency on a weekly or more frequent basis in order to meet with voters. In these meetings, the politician receives steams of people making all manner of requests. Some of these are policy-relevant but many others are personalistic — requests for constituency service or for clientelistic goods and services. Face-to-face interactions have been studied in various contexts, giving scholars some information on their frequency, nature, and outcomes (Bussell, 2017; Kruks-Wisner, 2018; Nichter and Peress, 2016). Typical estimates of the numbers of voters involved range from a few dozen to a few hundred per day. With the help of his staff or personal assistants, the politician seeks to resolve as many problems as possible on the spot with phone calls to bureaucrats or others. Other issues may require a second visit or be left unresolved.

The kinds of issues that are brought to the politician in person by voters may be classed into two types: requests for assistance with individual matters (financial assistance, employment, etc.) and requests for improvements of public infrastructure (road paving, water and sanitation, electricity and gas, etc.). The first are typical *clientelistic* disbursements when they involve resources to which the individual lacks entitlement or which are allocated using partisan criteria and *constituency service* when they are allocated using non-partisan and impartial criteria. The second have not been well theorized in contexts of inadequate infrastructure, although in developed countries they are labelled *pork barrel*. We prefer the more neutral term *community* requests.

In less developed countries, these interactions come with a mix of burdens and rewards for

both parties. The most obvious burden is that they require large time commitments. The politician must travel to his constituency and devote often personal resources to resolve problems at far remove from the legislative process. He may see this as wasting time away from his principal commitment to policy making. The voter must travel to the politician's constituency office, which may be especially difficult for those who live far from the politician's home. Second, precisely because the interactions are face-to-face, voters who lack the cultural skills to interact effectively with high-status political figures are disadvantaged. Access to the politician may be inequitable, perhaps thereby creating a class of low-level political intermediaries ("brokers," "fixers," and the like) who insert themselves into the process and profit from it, as reported in countries as diverse as India (Manor, 2000) and Argentina (Stokes et al., 2013). On the other hand, both parties may enjoy the direct nature of the interaction. Voters may accrue status from interacting directly with the politician, just as politicians who receive large numbers of visitors may accrue status. In cases that result in on-the-spot problem resolution, the voter knows with certainty that the politician has provided the requested assistance. Finally, face-to-face meetings permit add-on interactions, including the unsavory and illegal — extortion by the politician and bribe payments by the supplicant. Politicians may use meetings to increase their personal income and voters may use them for individual advantage.

Regardless of the mix of burdens and benefits that comes from face-to-face interactions, they operate in a piecemeal fashion and could be made more efficient using information and communications technology. Information technology allows many more voters to communicate with the politician in a shorter period of time. New technology could also bring other benefits: it is well suited to issues that can be dealt with simultaneously for many constituents and it favors communicating solutions that are standardized and rule-observant rather than ad hoc and discretionary. Finally, face-to-face requests allow voters to ask for goods and services to which they are not entitled and that thereby involve political corruption if disbursed to them; and even when that is not the case, they encourage personalistic, partisan-based,

and clientelistic decisions. Information technology has the potential to reduce corruption and clientelism in voter-politician interactions (cf. Muralidharan, Niehaus and Sukhtankar, 2016; Peixoto and Fox, 2017).

Politicians and provincial governments already use information technology to interact with voters in Pakistan (Callen and Hasanain, 2011; Laila et al., 2020). Thus, the idea is already known in the country. More than 80 percent of Pakistanis have cellphone access, making it feasible to use them for widespread political communication (Pakistan Telecom Authority, 2021). Marketing robo-calls are infrequent in the country, and households are not already saturated with unsolicited cell phone calls. This makes it likely that households will answer the phone when it rings. All these factors suggest that our setting is suitable for an intervention based on cell-phone communication.

2.2 How IVR may transform the status quo

As we implemented it, Interactive Voice Response technology changes interactions between voters and representatives on at least seven dimensions: (i) the initiator; (ii) geographic scope; (iii) content; (iv) scale; i.e. the numbers involved; (v) representativeness of voters involved; (vi) personalism; and (vii) frequency. With IVR, politicians solicit responses from voters rather than burdening voters with approaching their representative. Thanks to our sampling techniques, voters who live far from the politician's home have equal access rather than being disadvantaged, as is the case with face-to-face interactions. The content is decided by the politician and by definition involves aggregations of voters; thus, it excludes individualistic, partisan, and clientelistic interactions. Given that we select a random sample of male heads of households, politicians know they are communicating with a broad swath of voters, not only supporters.[4] The scale is larger, allowing more voters to interact with their

---

[4]As in other South Asian countries, such as India, in Pakistan political participation is typically studied at the household level, where men's preferences dominate (Khan, 2017). Our study is not designed to confront the challenge of bringing women into the political realm

representative than in the status quo. If there is socio-economic or partisan bias in the face-to-face interactions characteristic of the status quo, our intervention potentially "flattens" access (Grossman, Humphreys and Sacramone-Lutz, 2014). Interactions are not personally identifiable, potentially freeing the voter to express preferences that he might be reluctant to state in person to his representative. And finally, IVR allows follow-up calls to deepen the interaction.

We study two sets of IVR outcomes: *engagement* and *downstream outcomes*. The first consists of measuring changes effected by IVR over the status quo with regards to the geographic scope, scale, socio-economic representativeness, and frequency of interactions between voters and politicians. We also present descriptive information on the responses of voters and politicians to our requests for involvement with the IVR platform.

We study experimental downstream outcomes to assess whether IVR initiates changes that could improve responsiveness and accountability. In particular, we study whether politicians use their involvement with IVR to improve service delivery to constituencies where voters receive IVR calls and we study whether voters change their views of the politician, participate more in politics, and adopt performance-based voting criteria.

2.3 Prior research offering technology-based political communication

Various prior studies offering voters access to information and communications technology (ICT) to express opinions and preferences to politicians have almost all used short-messaging systems (SMS) rather than IVR. IVR has advantages over SMS because the former removes the burden of initiation from the voter. Unlike SMS, IVR also works well in low-literacy environments. Perhaps because they use SMS, other studies show low take-up by citizens (see

as independent actors, which requires more targeted interventions, such as those reported in Giné and Mansuri (2018); Cheema et al. (2020).

Table A.1), usually in the single digits. Our take-up rates are as high as the maximum rate ever observed in any other study, suggesting the promise of IVR in transforming interactions between voters and politicians in less developed countries.

3 Context and Status Quo Political Engagement

3.1 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 Members of the Provincial Assembly, each representing a single-member electoral district.[5] All directly-elected MPAs were men.

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).[6] Traditional patronage and dynastic networks, although important elsewhere in the country (Chaudhry and Vyborny, 2013; Cheema, Javid and Naseer, 2013; Liaqat et al., 2016), appear less binding in KP, and we report various types of evidence later that show the limited extent of clientelistic interactions in KP between voters and politicians.

The households we study fall outside traditional elite structures: only 21 percent of respon-

---

[5]Additionally, there were 25 seats reserved for women and minority members; these representatives are elected province-wide using proportional representation. We do not admit these MPAs into our sample because they do not have individual geographically-demarcated constituencies and are not named on the ballot.

[6]The average margin of victory for MPAs in KP was 11 percentage points, whereas it was 16 percentage points in Balochistan and Punjab and 30 in Sindh.

dents 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. Finally, access to telecommunications is expanding rapidly; the Pakistan Telecommunication Authority reports that cell phone penetration in the province rose from 42 percent in 2012 to 79 percent in 2020. This combination of marginalization from the political arena, strong democratic aspirations, and newly-won access to modern communications makes KP a good site to investigate whether novel forms of ICT politician-initiated communication improve the scope, efficiency, and equity of constituency service.

3.2   Status quo interactions between citizens and politicians

In the province of Khyber Pakhtunkhwa, households seek all manner of assistance from their political representatives. MPAs (and politicians generally) are key conduits for assistance b th 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.

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). That study found that the most important attribute of candidates for office was that they be "helpful in personal needs and community development" (Wilder, 1999, p. 195). Semi-structured interviews by one of us with more than three dozen current and former Khyber Pakhtunkhwa MPAs and MNAs encountered similar views, as well as regular complaints about the constant need

to visit their constituencies to attend "weddings and funerals"[7] 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.[8] 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, "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).

Such phenomena are standard across South Asia. In the Indian state of Rajasthan, one scholar reports that 76 of surveyed households were engaged in "claim-making" of some kind on politicians (Kruks-Wisner, 2018, p. 15). Another study, drawing on field research from three Indian states, documents that politicians spend 25 percent of their time on face-to-face service-oriented interactions with voters (Bussell, 2019, p. 25).

---

[7]For a discussion of the political importance of weddings and funerals, see Lyon (2004, pp. 194–96).

[8]The *hujra* is a part of traditional male Pashtun culture; status is represented by the number of other men who visit one's *hujra* in the course of an evening (Barth, 1965). In other parts of Pakistan, such as Punjab, anthropologists find that landlords use equivalent sites to receive information and dispense favors in response to requests for help (Elgar, 1960). Lyon (2004, p. 57) reports that in the part of Punjab where he worked, the *hujra* was called the *déra*.

Prior to our research efforts, MPAs had no way to contact voters other than their own loyalists whose phone numbers they had personally collected.

4   Experimental Design and Implementation

4.1   Sources of data

Our study uses information from eight data collection instruments (see Table E). We began in 2018 with a (i) *baseline* survey to enroll $14,400$ voters into the study; in 2019 we conducted an (ii) *endline* survey to collect information on experimental outcomes. We merged these data with 2019 polling station level (iii) *election* information that we collected. In 2019, we also conducted face-to-face semi-structured (iv) *interviews* with approximately three dozen MPAs and MNAs from KP. During the course of the intervention, we surveyed 240 (v) *key informants* across the 20 treated constituencies. In 2020, we conducted a (vi) *descriptive* telephone survey of $3,600$ respondents subsampled from the initial $14,400$ to collect information on pre-existing (status quo) political communication.[9] In addition, we surveyed $2,000$ students, academics, and policy implementors in advanced countries and in Pakistan in 2019 to collect information for a (vii) *forecasting* exercise. Finally, the IVR experiment generated (viii) *intervention* data on content, take-up, and response rates. We use these eight labels to identify each type of data in our analysis.

4.2   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 using IVR to interact with voters. The 47 MPAs generally come from slightly more competitive and urban constituencies than

---

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

their uninterested counterparts. We randomly select 20 of the 47 MPAs — blocking on political party — 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.

Figure 1: Experimental design at the polling station and household levels
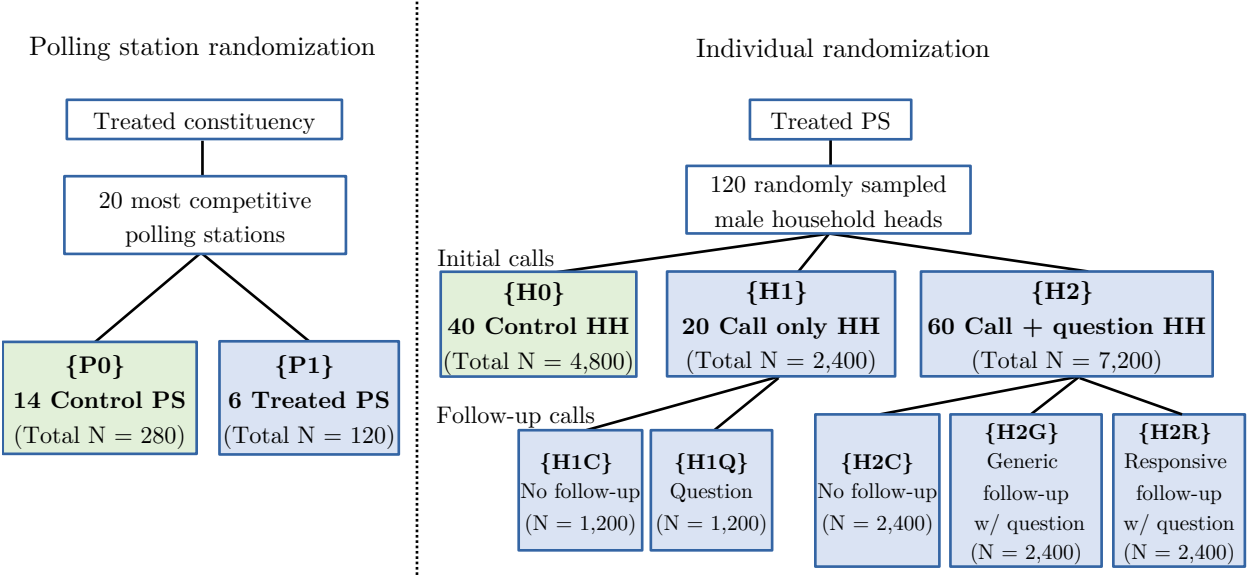
The MPAs in our sample are elected from constituencies with between 50 and 294 polling stations (PS). The median number of registered voters per KP polling station is $1,333$; each MPA represents an average of 300,000 people. Within each treated constituency, we select the 20 polling stations with the smallest margin of victory of the incumbent MPA.[10] 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.[11] We

---

[10]Margin of victory is the PS-level difference in vote shares between the incumbent MPA and the best challenger in the most recent election (either the 2013 general elections or a subsequent bye-election). We exclude female-only polling stations.

[11]A large party is 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.

assign one of each pair to treatment and the other to control.[12]

The right panel of Figure 1 shows show 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.[13] We refer interchangeably to respondents as heads-of-household, households, individuals, and respondents. 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} 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

---

[12]For analyses of electoral outcomes, we compare the six treated PSs with all 14 control PSs, but owing to cost constraints we can only compare the matched PSs when analyzing MPA effort using key informant interviews.

[13]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.

{H2R} as well as marginal effects of call components.

4.3 Content of the IVR calls

Each first-stage call contains two sections. First, the MPA introduces himself and credit-claims for recent activites. 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. A sample call illustrating the credit-claiming message and modal question is:

Stage One Scripts

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.

We use automated recall patterns to call multiple times staggered throughout the day until the respondent answers. Respondents in the IVR question treatment condition who answer the phone but not the question are contacted again to give them other opportunities to respond to the question. After calls are completed, we collate responses and report aggregate demographics from our baseline survey as well as aggregate IVR responses to the MPA. Part of one such report is shown in Figure 2.

After receiving first-stage responses, MPAs record follow-up calls and second IVR questions.

Figure 2: MPA sample report
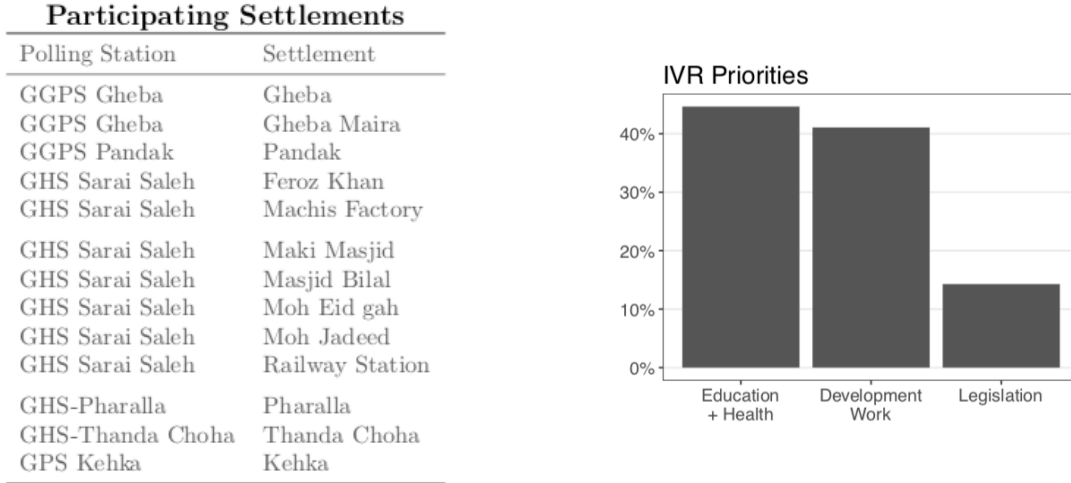
These go out to respondents using a similar strategy of recontacting those who do not initially answer the phone. Details on the intervention timeline appear in Appendix D.

5 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. Status quo political communication consists of face-to-face citizen-instigated meetings with political representatives; use of other forms of communication (e.g. phone calls, email, etc.) by ordinary voters is almost completely unknown in our context. Status quo communication is thus inherently limited to individuals willing and able to access the MPA in person, usually in his constituency office but occasionally also at public occasions (weddings and funerals).

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

We first examine if community characteristics correlate with who interacts with an MPA. If interactions are geographically concentrated under the status quo, some geographic areas will be potentially underserved.

To study this, we examine data from the *descriptive* survey. In Figure 3, 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 by 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. Results of 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 the constituency level differences instead of polling station level differences.[14] Thus, IVR contact is more evenly spread across geography than in-person contact by our random sample.

Table 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

---

[14]Note that in order to make the comparison between modes of communication equivalent, we estimate the variance and the distribution of contact rates on the same follow-up subsample. If we estimate the variance in PS-level IVR contact rates from the full baseline sample, then it will have mechanically lower variance.

Figure 3: Distribution of contact rates across polling stations (left) and constituencies (right)
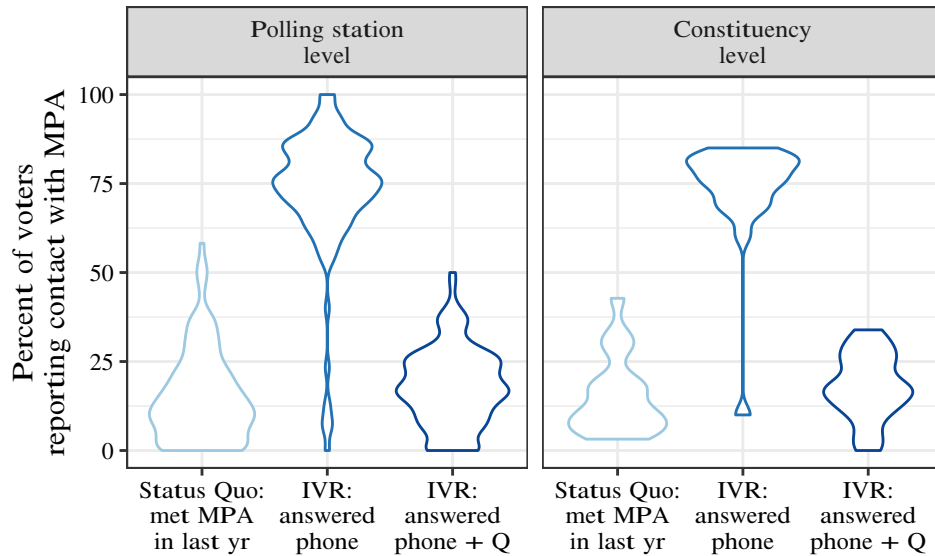
Notes: Each violin plot shows a distribution of contact rates between politicians and voters. For example, in the left panel, polling stations with no reported contact lie at zero and those where everyone reports contact with their MPA lie at 100.

constituency office.[15] 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

---

[15]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.

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.

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

| | Outcome | | | |
	Status quo % who met MPA in last yr		IVR % 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.

5.2   Content: What gets communicated to the MPA?

In Figure 4, 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) or instead for the community. Two-thirds of requests seek community improvements rather than individual clientelistic goods. These data show that even in face-to-face interactions, voters more frequently request basic public infrastructure

than clientelistic goods.[16] This contrasts sharply with a large literature that depicts political interactions in the global south as oriented mainly towards clientelism and patronage (Chandra, 2004; Kitschelt and Wilkinson, 2007; Nichter, 2018; Bussell, 2019). Like this larger literature, our data confirm that requests targeted to individual claimants mainly involve jobs or financial assistance. But most voters who meet with their MPA in our setting do so to ask for improvements in roads, electricity, or water for their communities.

Figure 4: Target and nature of requests made to MPA in status quo meetings
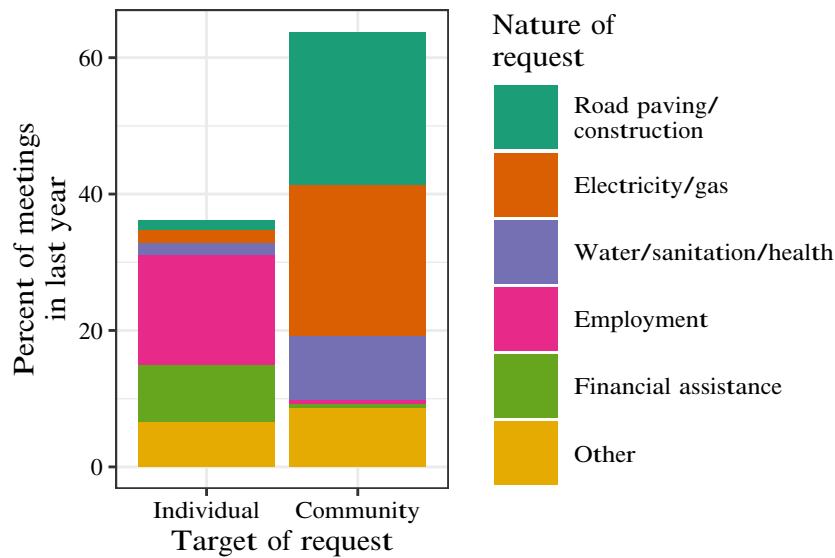
Notes: Data are subset to respondents who report meeting their MPA in the last year.

As the modal IVR call content documented (reported in Section 4.3), IVR content involves broad policy orientations and priorities; it could potentially be adapted to poll communities on infrastructure conditions, for instance.

---

[16]This figure is considerably higher than that reported for reasons for contacting local government offices in Pakistan. A study based on 2004 data reports that half of such contacts involve a "problem with local service (water, sewerage, health etc.)" and another 3 pecent "property and land issues" (Cockcroft et al., 2005, p. 24, table 3). The remaining contacts involved personal issues.

5.3 Scale: How many people communicate with their MPA?

The third dimension we investigate is the total number of constituents who interact with their elected representative in any given period. In Table 2, we show contact rates using data from the *descriptive* survey: only nine percent of respondents report meeting their MPA in-person over the last six months and only 15 percent report meeting him over the last year. Thus, status quo interactions involve small proportions of voters.

IVR data show vast improvement in the numbers of voters who interact with their representative. Fully 73 percent of the population in our random IVR sample of 14,400 answer the phone to accept a call from their MPA. About 17 percent of respondents stay on the phone long enough to hear and answer a question, signaling substantial engagement with the communication system and the MPA. Phone calls initiated by a politician to communicate with voters are thus highly effective in establishing contact and eliciting feedback.

Table 2: Status quo and IVR interaction rates

Interaction type	Interaction percentage	Source
Status quo face-to-face (1 yr)	14.8%	Descriptive survey
Status quo face-to-face (6 mos)	9.0%	Descriptive survey
IVR: answer phone	73.1%	Experimental sample {H1, H2}
IVR: answer phone and question	17.3%	Experimental sample {H2}

Status quo interactions by time. Status quo rates of contact vary significantly over time. In data gathered in the *descriptive* survey of the subsample of respondents in the {H2} condition, citizens report when they most recently met their MPA. We plot the data in Figure 5. 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. Second, IVR is not only able to more than double the rate of interaction with just a single round of calls, it is also establishes as much two-way interaction in one round of calls as occurs face-to-face over a full year. This is shown by the intersection between the dotted

and solid lines. 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 per call, this represents the enormous gains in scale that can occur if IVR is used to boost interactions between an MPA and constituents.

Figure 5: Status quo by contact window
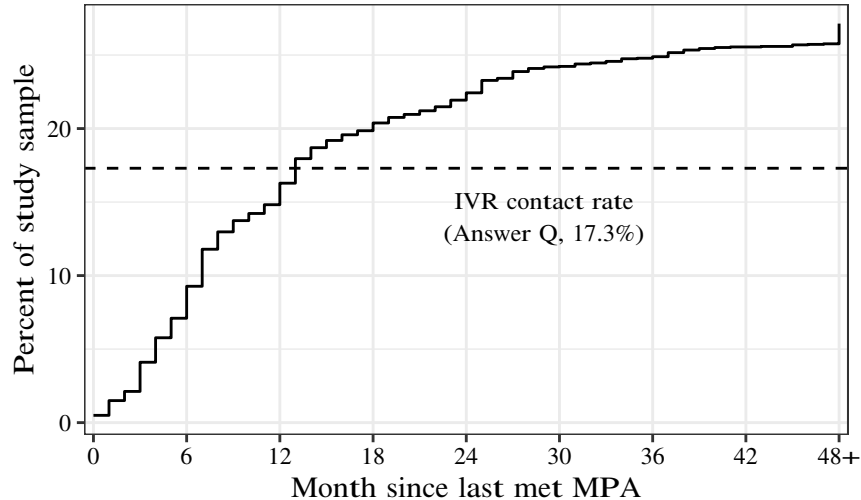
Notes: The solid line plots the cumulative distribution of time since survey respondents last met their MPA in person. The dashed line is the average rate of contact under IVR (answered question).

IVR interactions over call-time.   In Figure 6, we present information on the proportion of constituents IVR retains at all stages of the phone call. We see that of the 100 percent of respondents who agreed to participate and were called, nearly 80 percent answer the phone. There is then a gradual reduction in respondents still on the phone as the MPA first makes a credit-claiming statement and then asked his question. By the time the MPA finishes asking a question, about half the respondents are still on the call. This is often more than a minute and a half into call time. The biggest drop-off occurs when respondents are asked to answer a question which, as we have already noted, is completed by only 17.3 percent

of those called.[17] 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.

Figure 6: IVR contact by length of call
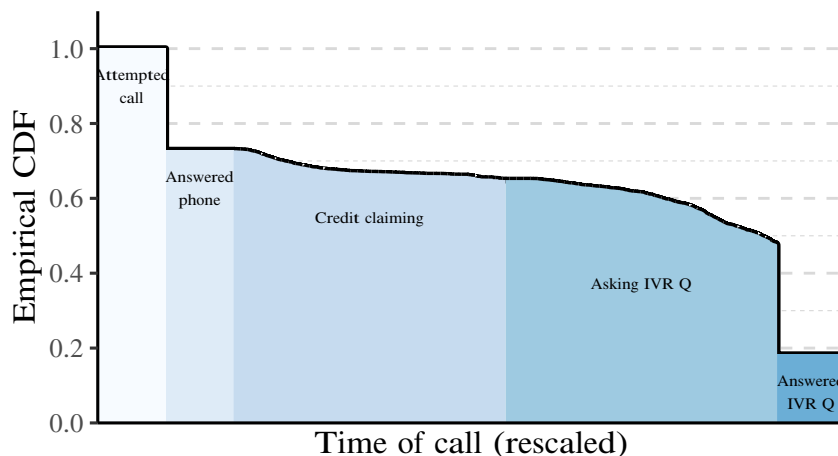
Notes: The figure shows the proportion of respondents who remain on the phone at various stages of the IVR call. The x-axis is rescaled for illustrative purposes to account for differing call lengths across MPAs.

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

A fourth dimension we study concerns *whom* MPAs are able to reach using IVR. Figure 7 compares the characteristics of three groups: (i) people who report in our *descriptive* survey 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). Perhaps surprisingly, we find that *all* modes of communication do a good job reaching constituents who are

---

[17]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; thus, there remains a large number whose voices could potentially be heard once they become more familiar with the mechanics of answering questions using IVR.

representative along 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).[18] 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.

5.5   Discussion

Overall, we find that in the setting we study, status quo interactions involve a small number of constituents but these individuals are representative along standard dimensions of the broader male population. Perhaps because they were sited in areas of the country that are known to have deep clientelistic networks, other studies conducted in Pakistan (Cheema et al., 2020) — like those conducted in Africa (Grossman, Humphreys and Sacramone-Lutz, 2020) — report socio-economic bias for face-to-face interactions that we do not find. However, status quo methods are relatively slow in reaching constituents: in-person contacts under the status quo takes about a year to reach as many people as one wave of IVR calls reaches in a week. In other words, IVR can be powerful in spring-boarding regular and timely interactions between the MPA and constituents, enlarging the *scale* of interactions. In addition, IVR also can flatten access across space, thereby improving the *scope* of interactions, which is important if interest articulation is a necessary condition for gaining public infrastructure from the state. Finally, IVR modifies the *content* of interactions because it is inhospitable to clientelistic and personalistic transactions. Thus, IVR offers improvements on three dimensions of interaction: scope, scale, and content.

---

[18]All 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).

Figure 7: Constituent characteristics by mode of contact (with 95 percent confidence intervals)
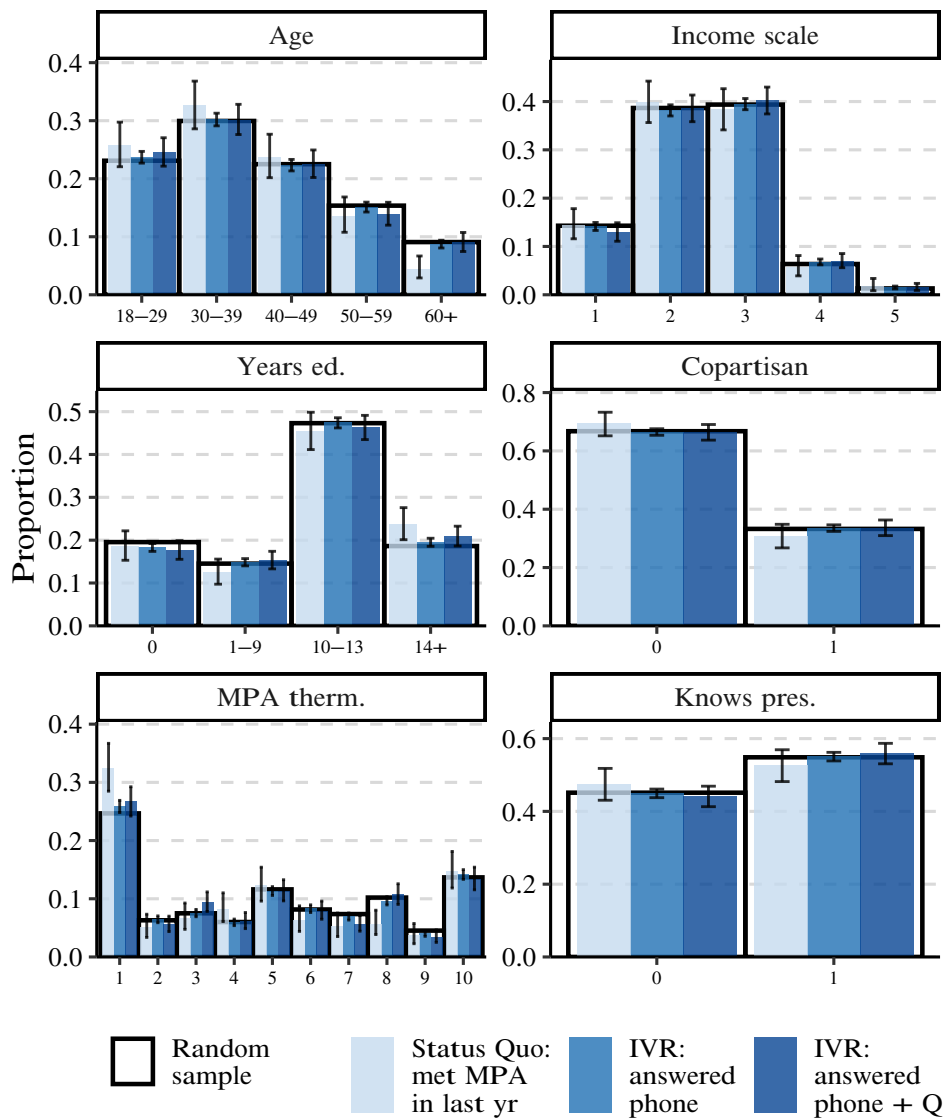
6 Downstream Results of IVR Communication

In this section, we look beyond IVR usage and ask whether IVR shows effects on individual attitudes and self-reported political behavior and on politician service delivery. Except as noted below and detailed in Appendix B, outcome measures, equation specifications, and treatment effects were all pre-registered.

6.1 Effects on individual attitudes and behavior

Data: Our main study sample comprises 14,399 male heads-of-household.[19] Of the 14,399 individuals for whom we have baseline data, we were unable to recontact 411 (2.9 percent) at endline.[20] 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.[21]

Outcomes: At the individual level — male heads-of-household — 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 claims to have voted for the MPA (or the incumbent MPA's party if he 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 to an individual's vote choice (1–6); (3) the number of conversations the respondent had about politics in the two weeks before

---

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

[20]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 condition.

[21]As a robustness check, we replicate our main analysis (reported in Table 3) using inverse probability of attrition weights (see Table G.1). Results are nearly identical due to the large sample size and relatively low rate of attrition.

the endline survey. The third index is conceptually the most complex. 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.[22]

Estimation: We estimate effects on the indices and constituent outcomes using an OLS specification:

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

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

---

[22]We report how we build these indices in Appendix F. We also examine whether adding this outcome to the political participation index changes any of the conclusions throughout the paper. It does not, and we report these alternative index results in Table G.2.

[23]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 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

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 3. 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 I), there are no treatment effects on any outcome. Furthermore, local average treatment effects among compliers are also substantively small and remain statistically insignificant (see Table G.3).

Table 3: Effects of any IVR call and effects of full IVR treatment on household head outcomes

Outcome indices	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
	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 below do not represent the full 13, 988 individuals for whom we collect both baseline and endline data.

The letters in braces refer to the experimental groups described in Figure 1.

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 also a follow-up call that acknowledges 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 absolute standard deviation 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 I. The IVR intervention appears to have little effect on individual attitudes or self-reported political behaviorsi of citizens receiving calls or questions.

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

6.3 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 *key informant* interviews.

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 on average did not invest in crafting messages that were likely to res-

onate with voters or that were specific to the geographic area. The modal question (see Section 4.3) was, by our reading not very specific and was consequently of low potential interest interest to voters. 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. So although politicians engaged eagerly with the intervention — we had more than double the number of MPA volunteers than we could work with — they shied away from using the opportunity to deepen their interactions with voters more than superficially.

In addition, politicians do not use the new IVR channel as a complement to other interactions with voters. The politicians who enrolled in our study interacted repeatedly with us in the nine months preceding an election. This could have encouraged them to exert more effort and improve service delivery in the polling station areas where voters received IVR calls, perhaps with a view towards augmenting the enhanced political visibility brought by IVR calls. This additional effort could have created a second channel for voters to receive new information about their representative. But MPAs did not exert additional effort.

To measure this, 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 4 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.

Another marker of low politician compliance and failure to engage with the opportunity to communicate with a large and representative sample of households is the failure of our research to repeat on schedule. We designed the study to obtain repeated MPA IVR recordings — as many as half a dozen per MPA — in the seven months preceeding provincial and national elections. Our design was meant to leverage reelection aspirations of enrolled MPAs.[24]

---

[24]On the ethics of working with incumbents and not including challengers, see Appendix

Table 4: ITT effects of IVR calls on MPA effort in polling station areas

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

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.

In the end, we recorded only a single question and a single follow-up response with each MPA because of resistance to repeated meetings to record additional rounds. Thus, although we intended to engage voters repeatedly, we ended up with a light-touch intervention.

Finally, other (non-experimental) evidence suggests that some politicians in our sample may have deliberately disengaged from the experiment, despite having volunteered to participate in it. In the first stage of the intervention, one MPA refused to record an IVR question, claiming he already knew everything there was to know about his constituency. In the second stage, three MPAs were either unwilling to record the different components of the phone call as per the experimental design or were consistently unavailable despite repeated requests to meet. In Table 5 we code these four MPAs as "partial" compliers and compare them to the rest of the directly-elected MPAs in the KP Assembly on a host of covariates. Although the sample is obviously small, the data show that partial compliers are in safer seats, have longer parliamentary tenure, and are less likely to be affiliated with the Pakistan Tehreek-C.

e-Insaf (PTI), a newer reform-oriented party that took power in KP in 2013. In addition, non-compliers exhibit lower attendance in parliament. One plausible interpretation of these patterns is that politicians facing less electoral competition see little need to use new or augmented forms of communication with voters. More surprising, however, even politicians who volunteered for IVR treatment and who faced more serious reelection problems did not use it to leverage their standing, status, and visibility with voters by visiting treated areas and improving service delivery.

Table 5: Observable differences among MPAs by degree of experimental compliance

| MPA type | N | Average | | | | |
		Vote share	Victory margin	Tenure	Legislative attendance	Member PTI
In sample						
Partial compliance	4	0.44	0.22	1.50	42.5	25.00
Full compliance	16	0.27	0.05	1.31	66.9	56.25
Out of sample	79	0.36	0.12	1.35	62.5	36.71

Notes: "In sample" MPAs are our 20 treated partners and "out of sample" are the remaining 79 directly-elected KP MPAs. PTI is Pakistan Tehreek-e-Insaf, the then-governing party in KP.

6.4 Evidence of voter engagement

Despite the null effects of treatment on voter attitudes and behavior, there is evidence that some components of the intervention affect voter engagement. In Table 6, we present data to 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. This suggests that citizens appreciate being included in policy discussions with politicians and would like to continue the conversation. Although this improvement is not large enough to change downstream voter outcomes (see Table I.2),

it suggests that on-going IVR interactions could gradually affect political attitudes and behavior. Potential changes were not sufficient to be visible in the timeframe of our study, however.

Table 6: ITT effects of initial call type on follow-up pickup rates

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

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.

To further explore this, we study downstream outcomes when respondents receive a single call from their MPA compared to also receiving an additional follow-up call. The results reported in Table 7 show suggestive evidence that repeated contact boosts citizens' prospects for accountability. When combined with the evidence just reported that voters like to engage in more substantive interactions that include a question, these results document nascent willingness among voters to engage on policy with politicians. But politicians apparently have to repeat their efforts to communicate with and hear from voters, and they have to demonstrate that they are responsive to feedback.

6.5 Robustness of null effects: power calculations

One concern with the null results reported in this section 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

Table 7: Marginal effects of any follow-up call versus only initial call on household head outcomes

Outcome indices	Control mean: only stage 1 call {H1C, H2C}	ITT: marg effect of follow-up call {H1Q, H2G, H2R} vs. {H1C, H2C}	
	μ	τ	N
Incumbent evaluations index	0.004	−0.001	7777
	(0.987)	(0.011)	
Political participation index	0.004	0.008	7792
	(0.985)	(0.021)	
Prospects for accountability index	−0.066	0.042†	7779
	(1.006)	(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.

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). Equivalence confidence intervals represent the range of hypothetical treatment effects consistent with the data; treatment effects outside these intervals are rejected as too large given the data. Equivalence confidence intervals for the main effects of the intervention at the household level (reported in Table 3) all have upper bounds at or below 0.05 standard deviations (see Appendix H). In other words, if the true treatment effect were 0.05 standard deviations or larger, we would only get our observed treatment effects less than 5 percent of the time. This shows that large treatment effects are highly implausible, making it unlikely that the null results on downstream outcomes result from lack of power.

7 What Did We Expect? A Forecasting Exercise

Were the mixed results we observe obvious ex ante? To find out, we conduct forecasting exercises that poll potential users of this research in Pakistan and from academia in the United States and Europe, asking them about their beliefs about the impacts of the experiment before seeing the results. Our aim 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.[25] 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[26] 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 provide forecasts on 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 8 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 question than the number who actually did. That is, forecasters underestimated whether respondents would answer the phone but overestimated whether respondents would answer an IVR question.

---

[25]The exercise follows DellaVigna and Pope (2018), which documents the utility of using expert forecasts to understand what we can learn from experiments. Recent work in political science has followed a similar approach (Dunning et al., 2019; Humphreys, Sanchez de la Sierra and van der Windt, 2019).

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

Figure 8: Forecast and realized compliance and treatment effects

Panel A: Forecasts of Compliance
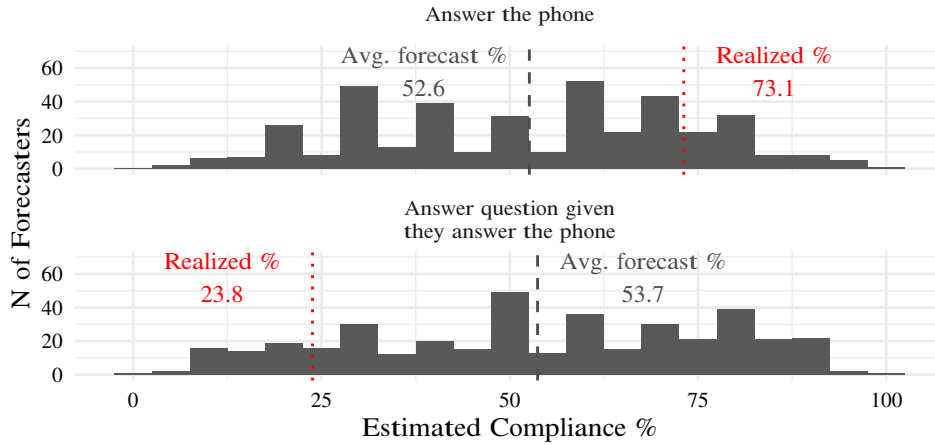
Panel B: Forecasts of Treatment Effects
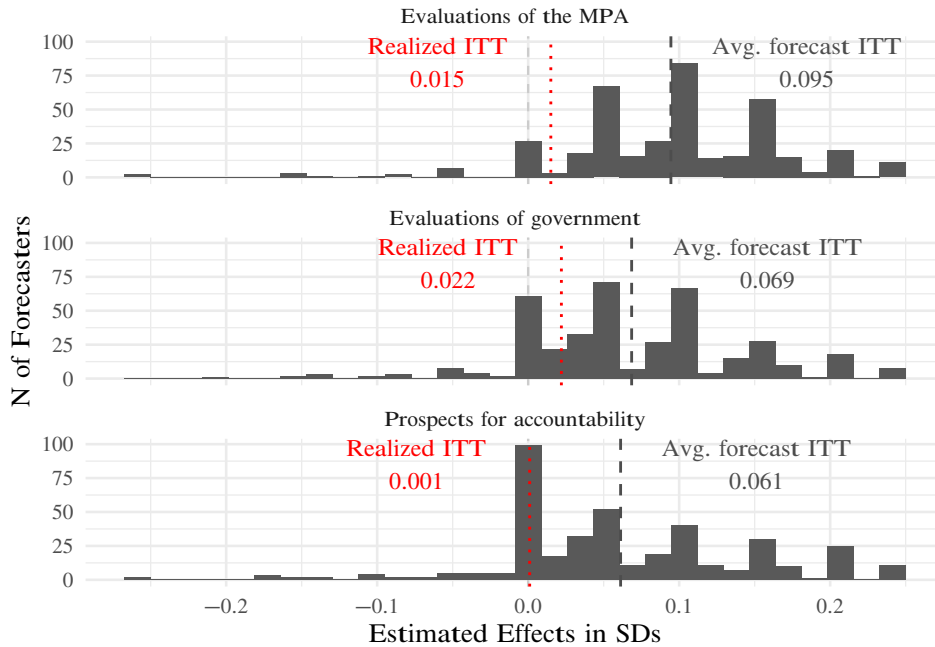
Notes: All differences between the average forecast and realized estimates are statistically significant with p<0.001.

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 generally 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.

8 Conclusions

Politicians in developing countries spend a lot of time meeting with voters and performing constituency service. Yet modes of communication have largely remained stagnant over the last few decades, relying primarily on traditional face-to-face communication at the politician's constituency office. Modern communications tools that have transformed other fields such as marketing are slowly being deployed by enterprising politicians across South Asia and beyond. This paper is a first attempt to examine the efficacy of such tools in the large, important, and understudied context of Pakistan.

We report three major findings. For policy practitioners in particular, we highlight that our study proved unusually effective in eliciting citizen feedback. We show that politician-initiated IVR has greater potential than the ICT methods that have been previously used in successfully engaging citizens in less developed democracies. When politicians call voters, voters pick up the phone and most of them go on to listen to what their representative has to say. A sizable fraction of respondents even use IVR to answer a question and a sizable fraction pick up the phone when the politician calls again. Take-up rates by voters

are higher in our study than those reported in almost all studies that place the burden of initiating communication on voters and that offer them other modes of technology-based communication. Our research design does not allow us to untangle how much of the high take-up rate is driven by shifting initiation to the politician and how much by adopting IVR in place of SMS. Future research may wish to pursue this specific issue.

Second, for scholars of comparative politics, we highlight that data we collected and analyzed show that governance patterns may be substantially different than current dominant theories suggest. We find that pre-existing non-IVR interactions between politicians and voters are not inequitable along standard measures such as social class or partisanship. The only bias we uncover is that voters living further from the politician's home are less likely to interact with him in person, presumably because of travel costs. IVR is effective in enlarging the geographic scope of interactions but otherwise does not improve further the ability of politicians to interact with a representative sample of households. Indeed, *they already interact with a representative sample of households.* Moreover, households that approach their provincial legislator for face-to-face meetings generally do so on behalf of larger groups in the community, as documented by the fact that they generally make requests for basic public infrastructure. Overall, then, our study finds that politics in Pakistan differs from standard accounts along two dimensions: it is more programmatic than standard clientelistic depictions and more representative than standard elite-centered accounts.

The third and perhaps most interesting finding that we highlight speaks especially to field experimenalists. An intriguing and puzzling aspect of our research is the lack of compliance on the part of the politicians who volunteered to participate in the intervention. Some refused to meet with us at all, and all were reluctant to meet with us according to the originally-scheduled monthly frequency. The average MPA who participated in our study failed to leverage it to improve service delivery to the parts of his constituency where IVR called were delivered, which could have amplified the intervention and improved his political reputation

and visibility. The intervention had been designed to study repeated IVR and follow-up calls but was transformed into a single-shot activity by the difficulties we experienced in meeting with the MPAs. Thus, it is perhaps not surprising that downstream attitudes and behaviors by voters did not respond to what had become a light-touch intervention. But lack of politician engagement on various dimensions is surprising.

Why did politicians fail to comply with the partnership to which they had agreed? Ironically, one possible reason lies with the frequency and representativeness of status quo face-to-face interactions. These interactions already deliver considerable information to MPAs about community needs. Perhaps MPAs saw IVR as a time-consuming addition that would provide little new information. Moreover, given that MPAs receive regular requests for public service improvements, they may already be using all the financial and bureaucratic resources available to them to respond.

Our experience on the ground in conducting the experiment we report suggests that working with politicians to encourage more responsive representation needs to take the general institutional environment into account. In a setting characterized by severe fiscal constraints, more government funding may not be available on top of those already disbursed. Thus, interactions between politicians and voters may be locked into an inefficient equilibrium where poor fiscal resources mean politicians cannot benefit politically from more information about voter preferences because they have little more to deliver. In this case, the main constraints on improved responsiveness lies with the level of development and fiscal extraction, not with any inherent lack of desire on the part of elected officials. Low MPA compliance in our intervention may thus be a reasonable posture given the political economy of Pakistan today. IVR could speed up and enlarge the geographic scope of communication but without improvements in the capacity of government to respond to citizen needs, it has the potential to frustrate both voters and politicians. To change voting behavior, we may have to change politician behavior; but to do this, larger institutional transformation may be necessary. Our

work asks experimentalists to step back and reconsider more fully the institutional incentives and constraints under which politicians operate.

References

Achen, Christopher H. and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government.* Princeton: Princeton University Press.

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.

Bachrach, Peter and Morton S. Baratz. 1962. "Two Faces of Power." *American Political Science Review* 56(4):947–52.

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

Broockman, David E. and Christopher Skovron. 2018. "Bias in Perceptions of Public Opinion among Political Elites." *American Political Science Review* 112(3):542–63.

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. 2017. "Serving Clients and Constituents: Experimental Evidence on Political Responsiveness in India." Unpublished paper.

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

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

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.

Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India.* Cambridge: Cambridge University Press.

Chaudhry, Azam and Kate Vyborny. 2013. "Patronage in Rural Punjab: Evidence from a New Household Survey Dataset." *Lahore Journal of Economics* 18:183–209.

Cheema, Ali, Hassan Javid and Muhammad Farooq Naseer. 2013. Dynastic Politics in Punjab: Facts, Myths and their Implications. Working Paper, No. 01-13. Institute of Development and Economic Alternatives (IDEAS).

Cheema, Ali, Sarah Khan, Shandana Khan-Mohmand and Asad Liaqat. 2020. "Canvassing the Gatekeepers: A Field Experiment to Increase Women?s Electoral Turnout in Pakistan." *Unpublished manuscript* .

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.

Cockcroft, Anne, Neil Andersson, Khalid Omer, Noor Ansari, Amir Khan, Ubaid Ullah Chaudhry and Sohail Saeed. 2005. *Social Audit of Governance and Delivery of Public Services.* Islamabad: Community Information, Empowerment and Transparency (CIET).

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.

Elgar, Zekiye. 1960. *A Punjabi Village in Pakistan*. New York: Columbia University Press.

Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton: Princeton University Press.

Giné, Xavier and Ghazala Mansuri. 2018. "Together we will: Experimental evidence on female voting behavior in Pakistan." *American Economic Journal: Applied Economics* 10(1):207–35.

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. 2021. "Who Enters Politics and Why?" *Annual Review of Political Science* 24.

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

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

Humphreys, Macartan, Raul Sanchez de la Sierra and Peter van der Windt. 2019. "Exporting democratic practices: Evidence from a village governance intervention in Eastern Congo." *Journal of Development Economics* 140:279–301.

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

Kitschelt, Herbert and Steven I. Wilkinson, eds. 2007. *Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition.* Cambridge: Cambridge University Press.

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

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

Laila, Ume, Najma Sadiq, Tahir Mehmood and Muhammad Farhan Fiaz. 2020. "E-governance as a roadmap to good governance: A digital Punjab perspective." *Journal of Accounting and Finance in Emerging Economies* 6(3):905–15.

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.

Liaqat, Asad, Michael Callen, Ali Cheema, Adnan Khan, Farooq Naseer and Jacob N. Shapiro. 2016. "Candidate Connections vs. Party Performance: Evidence on Vote Choice in a Low Information Environment." Unpublished paper.

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.

Manor, James. 2000. "Small-Time Political Fixers in India's States: "Towel over Armpit"." *Asian Survey* 40(5):816–835.

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.

Mills, C. Wright. 1956. *The Power Elite.* New edition. ed. New York: Oxford University Press.

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.

Nichter, Simeon. 2018. *Votes for Survival: Relational Clientelism in Latin America.* New York: Cambridge University Press.

Nichter, Simeon and Michael Peress. 2016. "Request Fulfilling: When Citizens Demand Clientelist Benefits." *Comparative Political Studies* 50(8):1086–117.

Pakistan Telecom Authority. 2021. "Telecom Indicators.".
URL: *https://www.pta.gov.pk/en/telecom-indicators*

Peixoto, Tiago and Jonathan A. Fox. 2017. When Does ICT-Enabled Citizen Voice Lead to Government Responsiveness? In *Civic Tech in the Global South: Assessing Technology for the Public Good*, ed. Tiago Peixoto and Micah L. Sifry. Washington, D.C.: The World Bank and Personal Democracy Press.

Stokes, Susan C., Thad Dunning, Marcelo Nazareno and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics.* New York: Cambridge University Press.

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

Appendices

A Take-up rates of other technology-based political communication interventions

Table A.1: Take-up rates and modes of communication across studies

Study authors	Country location	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 to 7%
Grossman, Michelitch and Santamaria (2017)	Uganda	SMS	4.8%
Erlich et al. (2018)	South Africa	SMS+	2 to 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 to 18%

Notes: SMS refers to short-messaging systems. SMS+ refers to a 5-channel study.

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

Key informant interviews used in Table 4: 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. The first two are of general relevance to the political science discipline. The American Political Science Association's has a Guide to Professional Ethics regarding human subjects protection (APSA Ad Hoc Committee on Human Subjects, 2019). That Guide raises issues of partisanship and possible interference in the democratic process that are potentially relevant to our study. A third issue concerns 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. This issue is explicitly raised by APSA Ad Hoc Committee on Human Subjects (2019). However, our sample sizes are too small for this to have been possible. For evidence, see the discussion in Section 6.2.

- Gender: A final ethical concern is that we work only with men and exclude women. As we note however (see fn. 13), 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 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.

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.

E   Analysis datasets and relevant tables and figures

Table E.1 details 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: Analysis datasets 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 can endline them.	Tables 2 (rows 3 and 4), 6; Figures 6, 7 (random sample and IVR communicators)
Endline survey	13,988 HHs	Every respondent we could reach for the endline. Note that some tables that use this data only use a subset (e.g. {H2}) when appropriate and is noted in the table/table notes.	Tables 3, 7, G.1, G.2, G.3, I.1, I.2, I.3; Figure H.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}.	Table 2 (rows 1 and 2); Figures 5, 7 (status quo communicators), 4, 3
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 4
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 J.1
Politician/constituency level			
All 99 directly-elected KP Members of the Provincial Assembly	99 MPAs		Table 5
Interviews	35 KP MPAs and MNAs		NA
Other			
Expert *forecasts*	400 respondents	Solicited in-person and over email. Includes undergraduate students, research/development practitioners, and academics around the world.	Figure 8

F   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.[27] 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.

G   Alternative specifications for downstream household results

In this section we present the robustness of the null experimental results presented in Table 3 to (i) attrition, (ii) alternative specifications of two indices, and (iii) considering compliance rates by estimating local average treatment effects among compliers.

First, we present robustness of the main results to attrition. The results in Table G.1 replicate the main, downstream household level results in Table 3 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

---

[27]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).

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

Outcome	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
	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 6.1 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 G.2: Effect of any IVR call and effect of full IVR treatment on household head outcomes - moving conversations to participation index

Outcome	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
	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 6.1 for details.

9

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

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

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.

H   Robustness of null findings in downstream results

In this section, we report equivalence tests demonstrating the power of our design to detect null downstream household level findings. We follow Hartman and Hidalgo (2018) and 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 H.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.

I   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.

Figure H.1: Equivalence confidence intervals for main household level treatment effects
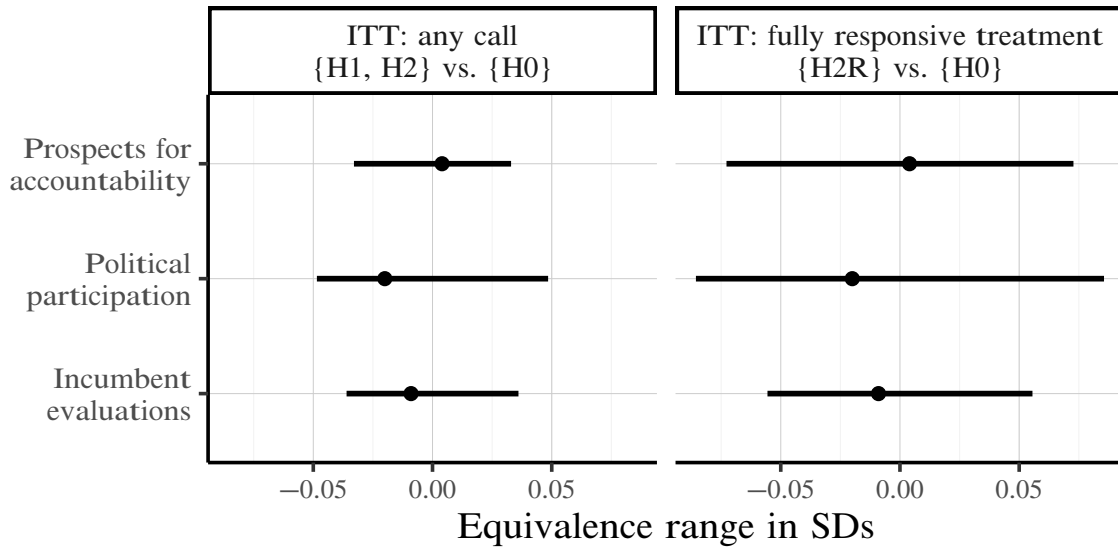
This figure contains the realized treatment effects and equivalence confidence intervals (Hartman and Hidalgo, 2018) for the two main analyses we report in Table 3. The points are the realized treatment effects and the equivalence confidence intervals are built at the 0.05 level. Our three main outcome 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.

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

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

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

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 6.1 for details.

13

Table I.2: Marginal effect of IVR question in initial call on household head outcomes including index component measures

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

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 6.1 for details.

Table I.3: Marginal effect of responsive vs. generic follow-up call on household head outcomes including index component measures

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

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 6.1 for details.

J   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 J.1 presents intervention effects from treated polling stations compared to control polling stations within treated MPA constituencies.

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

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

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.

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, Nazrulleava and Wolton, 2018). Elections in KP often have more than 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.

Appendix References

APSA Ad Hoc Committee on Human Subjects. 2019. Report of the APSA Ad-Hoc Committee on Human Subjects Research. Technical report American Political Science Association.

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., Eugenia Nazrulleava and Stephane Wolton. 2018. "Reelection, Political Selection, and Development." Unpublished paper.

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.

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.