"Press 1 for Roads": Bridging Communication Gaps in Political Representation

Miriam Golden*

Saad Gulzar†

Luke Sonnet‡

June 1, 2025

Abstract: To investigate whether lack of information is one reason politicians may be unresponsive to voter preferences, we conduct a randomized control trial with senior politicians in Pakistan that collects citizen preferences using Interactive Voice Response (IVR). IVR allows politicians to script questions for citizens and allows the latter to respond on their cell phones. There is strong politician interest in soliciting opinions via IVR; additionally, response rates by citizens are relatively high. Nonetheless, politicians fail to use the new information to modify on-the-ground engagement with voters or service delivery. Nor do we observe improvements in citizen evaluations of politicians or in their electoral support for them. A forecasting exercise shows that experts find these null outcomes unexpected. Our findings reveal that even eager and informed politicians may be unable to respond to citizen preferences due to institutional capacity constraints, highlighting limits of information-centric theories of accountability. [146 words]

Key words: political responsiveness, political representation, legislators, information and communications technology, randomized control trial, service delivery, Pakistan

Word count: 11,300

*Corresponding author. European University Institute (now Stanford University). CDDRL, Encina Hall, 616 Jane Stanford Way, Stanford University CA 94305; tel: +1-310-528-0930; email golden@ucla.edu

†Princeton University (gulzar@princeton.edu)

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

Acknowledgements: We acknowledge essential logistical support from the Consortium for Development Policy Research, the American Institute for Pakistan Studies, and the Center for Economic Research in Pakistan. We thank Mats-Philip Ahrenshop, Jawad Karim, Mamoor Ali Khan, Milad Khattak, Muhammad Parvez, and Jamshed Wazir for research assistance and our enumerators for their work in the field. Seminar participants at NYU, EGAP, the University of Wisconsin-Madison, LUMS, the LSE, Nuffield College, Bocconi, Stanford, Emory, Zurich, ETH, Indiana, CERP, UCSD, Facebook, the Lahore School of Economics, PacDev, DfID-Pakistan, WESSI, the Hertie School, Columbia University, the MPSA, the EPSA, and the APSA made helpful comments on earlier versions of this work, as did Guy Grossman and Asad Liaqat. We are also indebted to Kristóf Gosztonyi for anthropological insights.

Competing Interests: The authors declare none.

Funding: We acknowledge the Jameel Abdul Latif Poverty Action Lab (J-PAL), the Stanford Global Poverty Center, the UPS Endowment Fund for Transportation, Logistics and Urban Issues at Stanford University, the International Growth Center–Pakistan (ICG), and the Empirical Studies of Conflict Project (ESOC).

Ethics and Informed Consent: The pilot and subsequent scale-up research reported here were approved by the Institutional Review Board (IRB) of the University of California at Los Angeles (#17-000182). Interviews with Pakistani politicians were registered separately (UCLA IRB #19000719). The forecasting exercises were approved under expedited review procedures by the Institutional Review Board of the University of California at Los Angeles (#17-000182-AM-00005). Consent to participate was obtained verbally from citizen respondents, in writing from Members of the Provincial Assembly, and in writing from forecasters.

Data Availability Statement: Anonymized data and replication materials for this research will be made available on Dataverse when the article is accepted for publication. The pre-analysis plan for the pilot is available at http://egap.org/registration/2476. The pre-analysis plan for the scale-up is available at https://osf.io/vadwn.

1 Introduction

Across democracies from Michigan townships to Mexican municipalities, citizens complain that their representatives neither listen nor act. One reason, according to a recent stream of research, is that politicians are surprisingly ill informed about voters' preferences (Bafumi and Herron, 2010; Kalla and Porter, 2021; Pereira, 2021; Gulzar, Hai and Paudel, 2021; Walgrave et al., 2023; Walgrave, Soontjens and Sevenans, 2023), even where we might least expect this: settings characterized by well-established representative institutions, freedom of the press, frequent public opinion polling, extensive staff resources for policy makers, and open channels of interaction between them and voters. More accurate information improves reelection chances (Hug et al., 2024), giving politicians a sharp incentive to acquire it. Nevertheless, even where they might easily acquire information, politicians may be resistant to updating (Christensen and Moynihan 2020; but cf. Butler and Nickerson 2011) and exhibit systematic and persistent class and other biases in policymaking that distort voter preferences (Gilens, 2012; Broockman and Skovron, 2018).

In this paper, we study whether giving politicians accurate information about voter preferences aligns policy outcomes with those preferences (Lipset, 1963; Dahl, 1972), and whether voter attitudes towards elected officials become more favorable as a result of this process. We report results of an experimental intervention that recreates the full chain of interactions between politicians and voters using two-sided measurement of both politicians and citizens. We study whether politicians make visible efforts to be more responsive when they receive information directly from citizens about what the latter prefer. We also examine whether establishing new channels of direct communication improves the evaluations by citizens of their political representatives, including by increasing their vote share. To avoid making this a one-off communication activity, which we deemed unlikely to shift opinions or behavior on either side of the interaction, we set up our intervention as repeated conversations between citizens and representatives over an eight-month period, which was concluded by

new elections.

To create an ongoing interactive communication loop between politicians and citizens, one where each side receives feedback from the other, we partnered with 40 elected Pakistani provincial-level politicians, offering half of them opportunities to script and record questions for groups of randomly-selected constituents. We disseminated the questions as robocalls using Interactive Voice Response (IVR) technology, allowing a representative sample of households to respond via their cellphones. We then presented aggregated citizen responses to the provincial representatives to inform them about the distribution of voter preferences. In a second stage, politicians were encouraged to script and record follow-up calls, where they could acknowledge and respond to the first round of citizens' IVR responses. We then evaluated multiple outcomes for politicians and for citizens, including observed changes in citizen respondents' opinions, inferred changes in voting behavior, and shifts in service delivery by the provincial government.

The intervention offered politicians a novel opportunity to gain vastly more information about the preferences of voting-age adults in their constituencies during the run-up to an election than they could do on their own. On the other side, it offered citizen respondents an individualized and easily-accessed channel to communicate directly with their elected representative. As a result of these genuine improvements in the mechanics of existing politician-citizen interactions, both parties were eager to use the new channel of communication that our intervention offered. Politicians exhibited enthusiasm for the exercise: more than twice as many volunteered to participate as we were able to accommodate, documenting that they were willing to engage in costly action to receive new information about citizen preferences. We also find that the citizen population was willing to engage with and communicate their preferences to politicians at rates that were comparatively high. On these measures of implementation, our intervention was successful.

However, the intervention did not change the frequency or type of other on-the-

ground activities that politicians directed to the population or the evaluations of politicians by citizens. On the pre-registered citizen-level outcomes of political support for and evaluation of politicians, the intervention produced no statistically significant changes. Likewise, politicians did not follow up with visible responsive efforts.

These results suggest that information alone is insufficient in improving important aspects of the representative process. Contrary to expectations — our expectations as investigators, the expectations of the politicians who volunteered to participate in the intervention, and the expectations of a large panel of experts we consulted — even apparently well-intentioned politicians confront roadblocks to responsiveness that lie beyond information.

We interpret the results of our intervention as follows: Even politicians who are eager to receive more information about public opinion may be unable to respond effectively without improvements in other aspects of governance. Information without enlarged capacity may dead-end improving the representative process. What we label a *capacity gap* becomes visible when the pre-existing narrow information corridor is enlarged. A barrage of new information may only confront politicians with new citizen demands — demands to which they lack the resources to respond.

The capacity gap that we identify is multifaceted. Is includes fiscal resources, bureaucratic competence, and politician priorities. All of these have been identified in other studies as potentially relevant bottlenecks to political responsiveness in low-capacity settings. Our main contribution lies with solving the informational problem such that we expose the capacity gap, and demonstrate its severity. Many recent investigations have as their focus information gaps between voters and representatives, either because citizens purportedly lack information necessary to evaluate political performance (for instance, Dunning et al. (2019)) or because politicians are not fully informed about what voters want (e.g., Jablonski and Seim (2024)). We highlight both theoretical and empirical limitations of focusing

exclusively or perhaps even principally on information. The capacity gap that we identify between what citizens want — and what politicians would like to provide — is vast in the resource-constrained environment where we work. But studies of the use of direct reporting by citizens to public authorities about the operation of government services via such channels as 311 services suggest that capacity gaps are not uncommon across democratic countries, even very wealthy ones. We probe the nature of this gap in later sections of this paper.

This paper also makes a methodological contribution. There have been numerous experimental interventions that ask voters to communicate with political representatives in efforts aimed at making politicians more responsive (Grossman, Humphreys and Sacramone-Lutz, 2014; Chong et al., 2015; Kruks-Wisner, 2018; Arias et al., 2019; Buntaine, Nielson and Skaggs, 2019; Dunning et al., 2019; Bussell, 2019; Grossman, Humphreys and Sacramone-Lutz, 2020). These typically show very low rates of take-up by citizens, and perhaps for that reason have often not been effective. We establish two-way political communication between politicians and citizens that is *initiated by the politician*. This shifts the burden of instigation to elites, asking citizens only to respond. This shift makes it easier for citizens to express their preferences to politicians, and our take-up rates are correspondingly higher than those observed in other studies. This demonstrates that when communication is instigated by politicians, it is much more successful in establishing ongoing communication. This may hold promise for future interventions.

Our paper proceeds as follows. We first discuss the theory motivating our project in the context of existing literature and briefly present our main hypotheses. We then provide qualitative and quantitative descriptive information about the context of our work and the pre-existing representative process. At that point, we turn to more technical material and detail the experimental design and implementation. A fifth section presents descriptive results regarding how IVR improves on status quo political communication and interaction channels. A sixth section presents the main experimental results of our intervention. Then

we our key interpretation of the results and also consider alternative explanations. A final section interprets and concludes.

2 Theory and Literature

We conceive of democratic representation according to standard accountability theory: voters review the performance of the incumbent and vote to reelect if it has been satisfactory (Ferejohn, 1986). In this formulation, voters evaluate the performance of the incumbent using retrospective sociotropic criteria. Because they seek reelection, elected officials have incentives to deliver policies that most voters prefer. This suggests that politicians will seek information about public opinion — if only out of electoral self-interest — and also that they will use this information as a basis for attempting to align policy outcomes with public preferences.

This broad rational choice framework for understanding democratic representation is nicely complemented by empirical studies from the domain of political behavior. Most of these use data from the United States. They draw on interviews with or surveys of representatives and their constituents or they match characteristics of electoral constituencies with roll call voting and possibly constituency-level policy outcomes. Over many decades of scholarship, this mapping of public preferences onto policy outputs reports considerable correspondence between the two, especially on major domestic policy issues (Miller and Stokes, 1963; Page and Shaprio, 1992). This line of work established that the representative process in a stable democracy such as the United States was relatively well functioning for the median voter over domestic issues.

More recently, however, a new body of research shows that the preferences of elected representatives in the US and elsewhere skew conservative, and that policy systematically favors certain groups, for example, those with higher socioeconomic status (Bartels, 2008; Gilens, 2012; Broockman and Skovron, 2018). There may be many reasons why democracy

in wealthy countries appears to have become less effective in representing the preferences of the median voter, and we only speculate in the broadest of terms here. One factor that has been identified in the literature is that elected officials often have weak and inaccurate perceptions of public opinion (Walgrave et al., 2023; Hug et al., 2024). Additionally, they and their staff typically interact with unrepresentative subgroups from their constituencies, disproportionately including campaign donors (Kalla and Broockman, 2016; Hertel-Fernandez, Mildenberger and Stokes, 2019). A third contributing factor may be that elected officials are themselves usually drawn from higher socioeconomic status groups and may be unfamiliar with or disregard the preferences of other groups in society (Carnes, 2013; Carnes and Lupu, 2023). Finally, elected representatives may display unusual behavioral biases (Sheffer et al., 2017). Overall, the policy and other biases identified in these studies appear to be very general, crossing the divide between developed and less developed democracies.

Indeed, research in less developed countries describes environments that are even more susceptible to elite skew than what is found in developed democracies. Legislators around the world tend to be relatively well educated, making for a larger gap between them and the average citizen in countries with less educated populations (Carnes et al., 2025). Additionally, as research analyzing data from countries around the world reports, legislator preferences globally are consistently more aligned with the preferences of more affluent groups. This too makes the gap between representatives and the represented larger in poorer than wealthier countries, since average incomes are lower in the former (Lupu and Warner, 2022). Micro-level studies conducted in less developed countries report that politicians gather information about citizen preferences via face-to-face interactions with political intermediaries, known as brokers (Kitschelt and Wilkinson, 2007; Stokes et al., 2013; Bussell, 2019; Auerbach, 2019). These intermediaries are usually better educated than the citizens they speak for (Szwarcberg, 2015), and are often depicted as operating as gatekeepers between voters and higher-level elected officials. All these considerations suggest that less-institutionalized democracies are likely to experience even greater bias in political

representation than wealthy ones.

At the same time, public opinion research shows that voters are often poorly informed about politics and exhibit inconsistent preferences, even in established wealthy democracies (Zaller, 1992). Research specifically on voters in poor democracies argues that they are susceptible to manipulation by higher-level, better-educated political elites; that they offer electoral support when provided patronage and clientelism; and that they fail to evaluate policies on programmatic grounds, responding instead to claims of clan, tribe, and kin (Wantchekon, 2003; Chandra, 2004; Pande, 2011; Dunning et al., 2019). These factors suggest that for political representation to improve, voters would benefit from treatments that helped them engage more meaningfully with the political realm.

Our intervention was designed to interrupt the cycle of inadequate information and elite-skewed communication between politicians and citizens by offering both parties the opportunity to communicate repeatedly and directly using their cell phones. Our intervention was spread over an eight-month pre-electoral period, allowing politicians adequate opportunity to respond with visible policy efforts to the direct expression of policy preferences by citizens.

In designing our study, we built on prior research examining whether citizens can be galvanized into contacting their representatives to report specific problems. This research, as the data reported in Table 1 shows, generally finds very low rates of take-up — usually in the single digits — making it difficult to interpret and generalize its results. We designed our study specifically to improve take-up by shifting the role of communication initiator onto the politician.

Theoretically, doing this offers multiple advantages. First, it aligns with the principles of representative democracy, where voters, having already elected individuals to represent them, are relieved of the additional burden of continuing to initiate communication. Second, by controlling agenda-setting power, politicians may also be more likely to frame communi-

cation in a way that enables them to be more responsive to citizens (Mansuri and Rao, 2013). That is, we expect politicians to ask questions to which they wish to know voter preferences because they have the ability to respond. Third, a limitation of some previous studies has been that they predefine the policy areas of communication, thereby limiting them to a few large topical buckets. For instance, in Blair, Littman and Paluck (2019), communication was limited to reports about corruption (but cf. Grossman, Humphreys and Sacramone-Lutz (2014).) The intervention that we design allows politicians to flexibly choose any domain they want to engage citizens on. Finally, as we discuss in detail below, the content of communication generated by IVR is inherently more likely to focus on broad-based policy rather than particularistic requests by virtue of the fact that questions are asked of large groups of citizens.

In what follows, we report much higher take-up rates than any other study we have been able to locate. As we show when we analyze our results, our most conservative estimate is that more than 17 percent of citizen respondents in our study avail themselves of the opportunity to communicate with their elected representative. This is considerably higher than the single-digit rates reported by other studies.

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

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

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

2.1 Hypotheses

We pre-registered hypotheses about the impact of this new channel of communication on citizen respondents and on the politicians whom we enrolled in IVR. Our main hypotheses regarding the citizen respondents were that use of a new channel of communication with the elected representative would:

1. *Improve an index of evaluations of the incumbent*;
2. *Improve an index of the prospects for electoral accountability*;
3. *Improve an index of political participation.*

Our underlying reasoning was that citizens would respond favorably to the new opportunity to express their views and to communicate preferences directly to their representative. This hypothesis is built on the assumption that voters want political representatives to know their preferences and to produce responsive policy outputs. Thus, voters should evaluate their representative more positively and do so using more policy-oriented criteria when they directly experience responsive communication with the representative.

We specifically expected to observe these improvements where MPAs were more responsive in the IVR interactions. Thus, in our design, we distinguised mere credit claiming from soliciting citizen preferences via a question and we likewise distinguish both of those activities from undertaking a follow-up call that explicitly incorporates the new information about citizen preferences into the text. These permutations of our research design were meant to capture varying doses of responsiveness in the treatment, to which we expected to observe variations in measured attitudinal outcomes among citizen respondents.

On the other side, we expected that the new-found ability to collect information directly from many more citizens than previously would cause changes on the part of the politicians who used IVR. Specifically, we expected them to:

4. *Augment their activities in constituency areas where citizens received IVR calls.*

Our reasoning was that IVR politicians would seek to leverage the new channel of communication with greater visibility among citizen-recipients in order to enhance their vote-getting abilities. We expected to observe this within the timeframe of the research, which was conducted from December 2017 through July 2018, followed immediately by July elections. We tracked activities in selected constituency areas to assess whether IVR politicians made themselves more visible to voters and delivered new or different goods and services.

Finally, we expected that this new form of ongoing and interactive communication between citizens and politicians would allow the IVR politicians to:

5. *Improve their electoral outcomes.*[1]

Our thinking was that the greater visibility that IVR politicians would produce due to the intervention combined with the enhanced political engagement by voters would improve the politicians' electoral outcomes. Thus, improved electoral results would occur via the interaction of two channels: hypothesized positive responses to IVR by voters and hypothesized greater responsiveness to voters by politicians.

Our intervention was thus designed to create a new and fully responsive cycle of engagement between politicians and citizens. It supplements existing face-to-face interactions but also, as we document, provides politicians and voters new opportunities to engage with each other. We expected that both parties would benefit from this new engagement.

3 Context and Status Quo Political Engagement

Our study is set in Pakistan, a large but understudied lower-middle income democracy. As in many other low- and lower-middle income countries, in Pakistan voters have limited

[1]We discuss ethical issues at length later in this paper but note for the moment that because we operated in only 5 percent of polling station areas in a given constituency, our design made it highly unlikely that we could have affected actual election results. Our results confirm that we did not affect electoral outcomes.

ways to express their preferences about public policies or to make claims on government. The equivalent of a 311 government hotline does not exist; many voters are illiterate; the local government system is only partially operative; and opinion polling is infrequent and unreliable. For all these and other reasons, there is every reason to believe that politicians lack accurate and complete information about voter preferences there.

Our intervention augments existing face-to-face interactions between politicians and voters by introducing technologically-based communication through a randomized control trial (RCT) featuring Interactive Voice Technology. IVR allows politicians to script and record questions for voters. We disseminate the questions as robocalls to citizens' cell phones after having enrolled random samples of citizens in the intervention and obtaining their phone numbers. Citizens answer the questions using the number keys on their phones.[2] We then aggregate responses and present them to the politician, who can follow up with an additional call that acknowledges and responds to what he learns about voter preferences. The follow-up call closes the communication loop by informing voters that they have been heard and their preferences recorded.

We work in one of Pakistan's four provinces, Khyber Pakhtunkhwa (KP). There were 99 directly-elected, all male Members of the Provincial Assembly (MPAs) in KP in 2017–18, each representing a single-member district. MPAs assist voters with personal and community problems and control access to the state. Voters are typically aware of who their MPA is, and believe that he can provide community and personal assistance. To verify this, we collected data when beginning the study that shows that 55 percent of citizen respondents

[2]More than 80 percent of Pakistanis have cell phone access, making it feasible to use them for widespread political communication (Pakistan Telecom Authority, 2021). Also, because marketing robocalls are infrequent in Pakistan, households are not already saturated with unsolicited cell phone calls. This makes it likely that households will answer the phone when it rings. Finally, households have no communication channels to reach politicians other than face-to-face interactions, giving them an incentive to use a technologically-based medium. These factors suggest that our setting is suitable for an intervention based on cell-phone communication.

believe that their MPA can get roads fixed and 49 precent say that an MPA can help family members get a job. Voters have confidence in the ability of their provincial representative to provide community and individual assistance, giving them reason to communicate their policy preferences to members of this level of government.

Our intervention includes 40 MPAs out of 47 who volunteered to work with us. We place half in treatment, making IVR available to them, and the other half in control, studying the same outcomes without IVR. (Details are in Section 4.1.) Of the 40 provincial representatives we work with, 27 had been first elected in 2013. In that election, the reelection rate of incumbent KP MPAs was only 11 percent. Thus, the politicians we partner with were extremely insecure in their offices. This would have given them career incentives to improve whatever dimensions of representation they believed could enhance reelection probabilities.[3]

Our citizen respondents were a random sample of (male) heads-of-households; selection and enrollment is described below (see Section 4.1). Eighty-six percent of them report having voted in the prior general election. Although this figure is surely inflated by recall and social desirability biases, we use the terms "citizens," "voters," "households," and "respondents" interchangeably in what follows.[4] In Pakistan, voters have low party identification (Zaman and Mushtaq, 2022), making it difficult for politicians to distinguish supporters from other adult citizens.

In the status quo representative process, delegations of residents approach their provincial representative and make face-to-face requests for assistance. Precisely for this reason, politicians return at least weekly to their constituency offices where, as one ethnographer described it, "crowds of applicants wait outside to see the politician or a personal assistant in order to get the all-important 'chit' of paper . . . " (Wilder, 1999, p. 199). Semi-structured interviews by one of us with more than three dozen MPAs and Members of the National

[3]Of the 40 politicians we worked with, all but seven ran again in 2018, corroborating that most of them wished to retain elected office.

[4]Turnout in the prior 2013 elections is reported to be 53 percent (Gallup Pakistan, 2013).

Assembly (MNAs) elicited complaints about the constant need to attend constituency "weddings and funerals" in order to maintain visibility among and trust of electors. When in their home towns, politicians interact with anywhere from 20 to more than a hundred supplicants over the course of a day (Wilder, 1999, p. 199).[5] This is the main mode of communication between politicians and the people they represent. In the setting where we work, elected officials do not themselves run public opinion polls nor do they have dependable on-the-ground operatives or party machines feeding them information. As a result, politicians have only partial and incomplete information about citizen preferences and problems. Our interviews with politicians revealed that they are aware of this and wish to remedy their informational deficits.

Given the size of the constituencies (averaging 300,000 persons),[6] MPAs can be personally acquainted with only a fraction of their electors. Moreover, the single-member constituencies we study include anywhere from 50 to 294 precincts (in Pakistan these are referred to as polling stations (PS)), many covering equally many separate localized settlements (villages or hamlets). (Urban and peri-urban areas include multiple precincts that are not geographically distinct.) Geographic dispersion makes it difficult for the MPA to visit much of his constituency, especially in the larger ones.

For the MPAs, enrolling in the IVR experiment offered a way to gain potentially valuable information about public opinion that they did not otherwise possess. It also offered a way to communicate at a much larger scale with constituents, and in particular to communicate with constituents who did not approach them in person. In our post-

[5]Although the anthropological research that studies on-the-ground politics has generally been conducted in the province of Punjab and our research was conducted instead in Khyber Pahktunkhta, we have no reason to believe there exist substantial differences in village-level interactions between politicians and voters that would make this description inaccurate for KP.

[6]As a comparative benchmark, this is only 25 percent smaller than the average size of a district that elects a representative to the California state legislature, which is about 400,000 persons.

intervention interviews with MPAs, they explicitly noted the electoral opportunities the intervention offered, explaining that they normally had no way to assess their strength on the ground in a pre-election period.[7] MPAs also reported that they had no way to poll voters independently or to collect the cell phone numbers of those with whom they were not personally acquainted; for instance, they had no way to access something like a phone book. This means that the typical reelection campaign proceeds using incomplete information about where campaign efforts might most productively be made. Moreover, according to the politicians we interviewed, electoral campaigns are expensive and require considerable investment of personal funds. This gives candidates financial incentives to gain information that may be electorally valuable and allow more efficient campaigning.

As we have noted is common in less developed settings, the literature on Pakistan reports that direct contact between voters and politicians are skewed towards higher-income (male) co-partisans (Martin, 2014; Liaqat, Cheema and Mohmand, 2020). Women play little part in public affairs in Pakistan, and on-the-ground interactions between citizens and elected representatives exclude them almost entirely. Face-to-face meetings between politicians and citizens, which usually take place in the politician's home courtyard office, are entirely male. In this political culture, "leaders redirect public resources to benefit kin, friends and clients" (Martin, 2016, p. 67). This is a setting that anthropologists characterize as one of asymmetric factional hierarchies, and is based on status and power (Barth, 1965; Lyon, 2004) in which politicians seek to assemble large clienteles. However, as our interviews repeatedly revealed, limited resources mean politicians are unable to provide assistance to most who request it. (Additional details are in Section 7.) Since politicians thus necessarily 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). The standard assumptions are that access to politicians and policy responsiveness are functions of income, gender, and partisanship.

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

As we later detail (see Section 5), aspects of the conventional understanding are in fact inaccurate for our setting. Politics is neither as clientelistic nor as elite-skewed as believed. Although there is substantial demand for individual benefits such as employment, which may be interpreted as clientelistic, most status quo citizen-initiated requests are for community improvements. In addition these requests come from a representative sample of households. However, we note persistent deficiencies that include an absence of independent political engagement by women.

4 Experimental Design and Implementation

We now provide details of the experimental protocol, as well as our data collection strategy. These technical details are important for interested readers to evaluate the design of the study. We describe the multiple steps in sample selection; two *stages* in the rollout of IVR (an initial contact/question and then a followup call) as well as the randomization (of MPAs and also of citizen respondents); and multiple types and levels of data collected over an approximately two-year period.

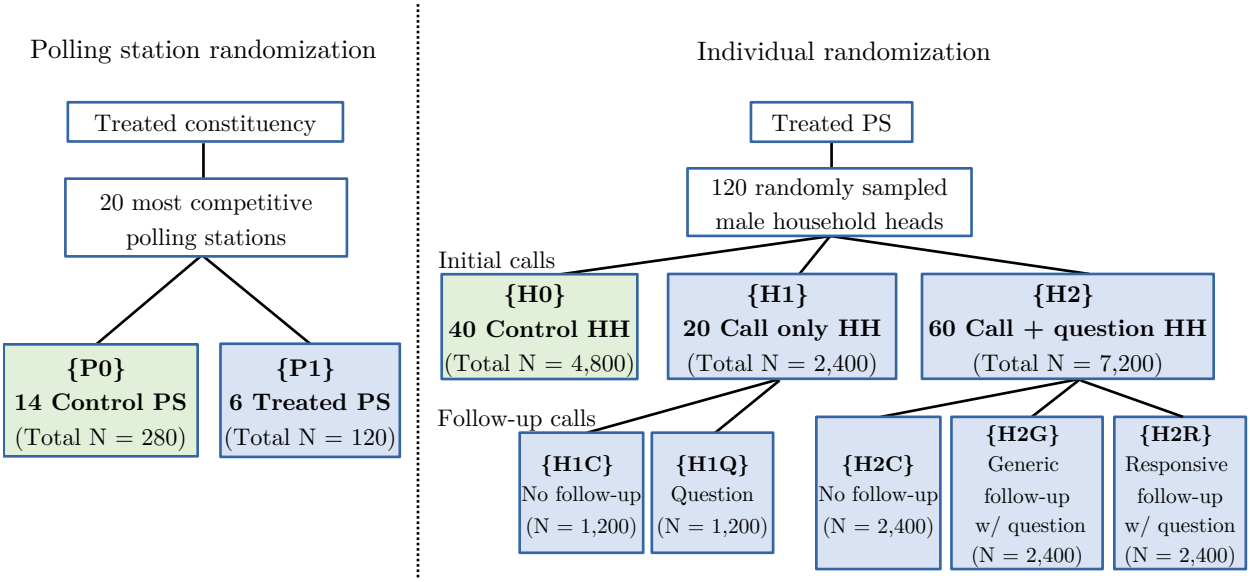
4.1 Selecting the Random Samples

How were MPAs and citizen respondents selected for the research? The process can be broken down into three steps, which are depicted in Figure 1.

In the *first step*, we select specific MPAs to enroll in the research.[8] Of the 99 directly-elected MPAs then serving in the KP assembly, our implementing partner identified an initial 47 who expressed interest in using IVR to interact with citizens. We randomly select 40 of these, who generally come from slightly more competitive and urban constituencies than other MPAs. Blocking on political party, we then randomly select 20 of the 40 MPAs to enroll into treatment. Treatment gives the MPA access to the IVR technology through our

[8]Appendix B discusses the ethics of working with incumbents alone.

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



research team. The other 20 are put into control.[9]

The *second step* selects the areas inside each constituency where we work with citizens. Within each of the 20 treated constituencies — that is, those held by the 20 MPAs selected for IVR — we select the 20 polling stations with the smallest absolute margin of victory of the incumbent MPA. The left panel of Figure 1 provides a diagram of the randomization process at the polling station level. We randomize six of the 20 polling station areas into treatment {P1} and 14 into control {P0} in a two-step process as follows: (1) We first randomly sample 12 polling stations out of the 20 and then, for those 12 polling stations (2) create matched pairs using a Mahalanobis distance score that incorporates the total number of registered voters and raw vote totals for large parties.[10] We assign one of each pair to treatment and the other to control.

In the *third step* we select the specific households we work with. Starting at a central

[9]None of the MPAs in the province had the resources or skill to use IVR on their own, nor did they have any way to obtain the cell phone numbers of constituents.

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

location in the PS catchment area, a random walk sampled 120 households within each treated PS area. A random walk procedure was used because of the absence of a census-based sampling frame.[11] The goal was to enroll a random sample of households within each PS area. The right panel of Figure 1 shows household-level randomization. We enroll male heads-of-household only, owing to insurmountable obstacles collecting phone numbers of women.[12]

The total sample comprises 14,400 male heads of household.[13]

4.2 Treatment Details

We conducted IVR calls in two stages/rounds, with each based on separate randomization procedures from among the enrolled respondents. We describe each in turn.

4.2.1 First Stage of Randomization: Initial Calls

The objective of the *first stage* is for the politician to contact voters and ask a question, labelled "initial calls" in Figure 1.

The call begins with the MPA introducing himself and then credit-claiming about some recent activities. Then the MPA asks a question to which he seeks feedback from

[11] At the time of our study, the last census had been conducted in 1998.

[12] We made explict efforts during pilot work to enroll female respondents. But even when female enumerators approached female citizens, an overwhelming share either refused to provide their phone numbers or reported having no regular access to a phone to begin with. This is to be expected as, in South Asia, phone ownership among women is very low (Roy, 2012). Consequently, the IVR experiment can be interpreted as expanding contact across households by politicians but retaining the male nature of contact that characterizes the political status quo. We attempt to circumvent some of these limitations in other work on enabling women's political participation in Pakistan.

[13] The descriptive statistics that characterize it (see Figure 9) seem reasonable given the province where we conduct our study. The citizen respondents we enroll in the study are generally young (modal age between 30 and 39), of modest income, somewhat educated (modal 10-13 years of education), and reveal low levels of political knowledge (nearly half of them fail to accurately identify the then-President of the country).

respondents. The modal script recorded by the IVR politicians is depicted in Figure 2.[14] The recordings bundle political advertising and credit-claiming with preference elicitation. This bundling parallels pre-existing face-to-face interactions in the setting where we work, and in constituency outreach more broadly. Allowing the MPAs to credit-claim at the start of the calls makes the citizen contact more natural.

Figure 2: **Modal Stage One IVR Script**

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

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

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

Thank you for taking your valuable time. Good bye.

Blocking on copartisanship with the MPA, 40 of the 120 households in each PS are placed into a control condition and receive no contact other than the baseline and endline surveys {H0}, 20 receive a version of the round one IVR call with a credit-claiming message but no MPA question(s) {H1}, and 60 receive a credit-claiming message as well as an IVR question(s) {H2}. This design allows us to to study not only the effect of IVR contact in the initial calls but also to separate the impact of credit claiming from the effect of asking for a response to a question, as discussed in Section 2.1. The sample sizes in this round are reflected by the N values reported in each box in Figure 1.

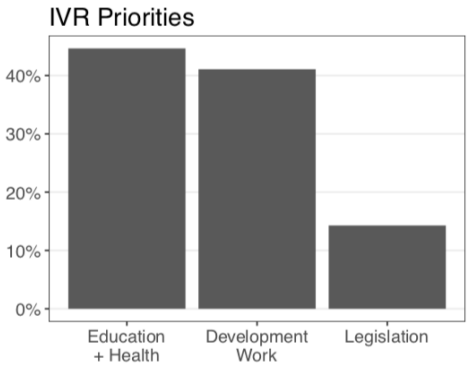
[14]We discuss why politicians record this script and analyze its strengths and weaknesses in Section 6.3.

Once the calls were delivered and citizens' responses were received, the research team collated the responses and provided the MPA aggregate demographics from the baseline survey as well as an aggregated version of the IVR responses. An excerpt from a sample report appears in Figure 3. The report provides the MPA the names of polling station locations where the information is collected as a way of signaling the spatial coverage of IVR. Finally, there is also information about how respondents evaluate the MPA at the polling station level and the respondents' preferred party by polling station level. This aggregated information allows an MPA to assess his popularity at a fine-grained level, pointing out where he might need to campaign to gain reelection. In the status quo low-information setting without political polling, this information is new to the MPA; prior to viewing it, he had no way to know the then-current proportion of citizens whose support he enjoyed. This concludes the first stage of IVR calls.

Figure 3: **MPA Report Excerpt**

Participating Settlements

Polling Station	Settlement
GGPS Gheba	Gheba
GGPS Gheba	Gheba Maira
GGPS Pandak	Pandak
GHS Sarai Saleh	Feroz Khan
GHS Sarai Saleh	Machis Factory
GHS Sarai Saleh	Maki Masjid
GHS Sarai Saleh	Masjid Bilal
GHS Sarai Saleh	Moh Eid gah
GHS Sarai Saleh	Moh Jadeed
GHS Sarai Saleh	Railway Station
GHS-Pharalla	Pharalla
GHS-Thanda Choha	Thanda Choha
GPS Kehka	Kehka



4.2.2 Second Stage of Randomization: Follow-up Calls

Once the MPA receives aggregate information about responses to the first stage of the IVR, we proceed to the *second stage*. The objective of this stage of IVR, described as "follow-up calls" in Figure 1, is to study if repeated contact, as well as deeper contact, affects voter evaluations. The MPA is offered the ability to record another round of IVR as a follow-up

call to the first round. In the second round, the MPA reports to respondents what he learned from responses to first stage calls and how he intends to proceed. The MPA also records a second question, to which voters again are requested to respond. Figure 4 shows an example of the kinds of stage two questions that were asked. Because of temporal proximity to the elections, many MPAs chose to ask questions related to electioneering.

Figure 4: **Example of Stage Two IVR Call**

Introduction and Credit Claiming: Assalam U Alaikum. I am [MPA NAME], your elected MPA from [CONSTITUENCY NAME]. I am calling to let you know about my programs for the coming months and my last year in office. Being an opposition member, I have worked on education, health, employment, roads, streets, electricity.

Generic addition: You got a call on my behalf a while ago in which I asked for your opinion on what development works I should focus on in the constituency. Thank you for your response.

Responsive addition: You got a call on my behalf a while ago in which I asked for your opinion on what development works I should focus on in the constituency. I have reviewed the feedback and the majority of you wanted more work to be done on roads. If you trust me and vote for me again, I will work hard on improving roads.

Additional Question: I want to ask you another question. How should I spend my time before the upcoming election? If you want me to focus on more rallies press 1, if you want more open courts press 2.

Thank you.

The objective of randomization in this round was to study if deeper and more meaningful communication by the MPA improved outcomes: specifically, we wanted to test if a) repeated contact mattered by itself; or b) if an acknowledgment of the first round responses was important; or c) if responsiveness to the first round, beyond mere acknowledgement, was needed. To explore these possibilities, this round splits those who had received first-round calls into three groups: those who receive no follow-up call and those who receive a follow-up call containing one of two types of components, labelled "generic" or "responsive" in Figure 4. Respondents in {H1C} and {H2C} (see Figure 1) receive no follow-up call, while all other

respondents receive a follow-up call in which the MPA asks a new question {H1Q, H2G, H2R}. Respondents in {H1Q} receive only the new IVR question, while respondents in {H2G} and {H2R} also receive a *generic* or a *responsive* message. In the generic response, the MPA acknowledges the first round IVR question and thanks respondents. In the responsive message, the MPA also details what he will do based on the information collected in the initial round of robocalls that asked for constituent input. The main goal of these distinctions is to estimate the total effect of the most interactive and deepest IVR communication we could generate {H2R} as well as marginal effects of call components.

4.3 Sources of data and levels of analysis

We assemble data from a variety of sources to carry out two separate levels of analysis. Given the large sample size of the data we use for individual citizens, we are able to conduct quantitative analyses of responses to questions that citizens were asked. We can do this separately for treatment and control groups both before and after the intervention. With only 20 MPAs enrolled in the IVR experiment, we conduct qualitative analyses of their involvement in and reactions to the intervention, both by interviewing them and via data collected from key informants in their home constituencies. Key informants were educated, non-partisan adult residents who agreed to a schedule of meetings to update our research team about any observed on-the-ground politician activities that might have occurred following the IVR calls.

Analyzing data on more than 14,000 citizen respondents as well as data on the 40 MPAs allows our study to use mixed methods. Overall, we analyze information that was collected via nine separate instruments, enumerated in Table 2. (The table also reports the source of data for each figure and table that appears in this paper.) In early 2017 we undertook a (a) *pilot* survey with a single MPA to demonstrate the project's operational feasibility; the take-up rate for the pilot was extremely high, with 31 percent of respondents answering at least one IVR question, although results were statistically insignificant

for changes in respondent opinions [redacted]. In late 2017, we distributed a (b) *baseline* survey to enroll 14,399 voters in the study. That survey also administered informed consent and collected cell phone numbers. We undertook the IVR experiment itself in 2018 and it generated (c) *intervention* data on content, take-up, and response rates. In 2019, we conducted an (d) *endline* survey of citizens to collect information on attitudinal outcomes. In 2020, we conducted a (e) *descriptive* telephone survey of 3,600 respondents, subsampled from the initial 14,400, from whom we collected information on preexisting (status quo) political communication and interactions with MPAs.[15] During the intervention, we surveyed 240 (f) *key informants* across all 20 treated constituencies to collect information about politician behavior during and after IVR. We also merged household-level data with 2019 polling station level data to assemble (g) *election* data. In 2019, we conducted lengthy face-to-face semi-structured (h) *interviews* with approximately three dozen provincial and national parliamentarians from KP. Finally, in 2019 we surveyed 400 students, academics, and policy implementors in advanced economic countries and in Pakistan in a (i) *forecasting* exercise that asked respondents to tell us whether they thought the intervention would be successful, showing them data from the 2017 pilot.

5 Communication Transformations Offered by IVR

Before we discuss the impacts of the IVR experiment itself, we analyze the specific improvements in patterns of communication that the intervention offered. This section presents these descriptive results. Given that IVR is now widely used for political, as well as non-political public sector communication, these results provide information about the reach and scope of this technology that may be relevant across multiple domains.

In reporting descriptive results, we compare patterns of what we call *status quo* com-

[15]We had hoped to observe face-to-face interactions between MPAs and citizens but the Covid-19 pandemic made that infeasible except for a single day-long pilot observation. We used that to inform the questions in the descriptive survey instrument.

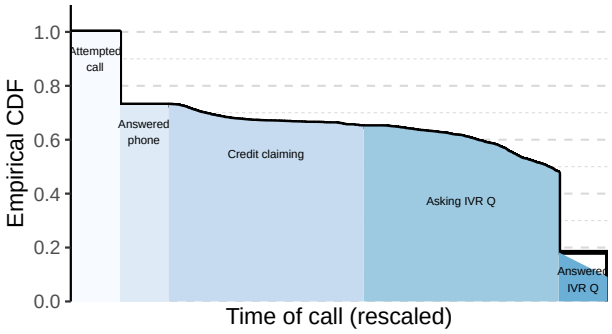
Table 2: **Datasets collected during research**

Dataset	Sample size	Notes	Tables and figures
Respondent level			
Pilot survey (2017)	1,218 HHs	Conducted with one MPA.	Data used in forecasting exercise presented in Figure L.1
Baseline survey (2017–2018)	14,399 HHs	Every respondent baselined (except for one duplicate phone number). Some tables use only a subset (e.g. {H2}) when appropriate. Often IVR compliance statistics (e.g. answering the IVR question) come from this data as it is available regardless of whether we endline.	Figures 6 and 9; Table I.2
Intervention data (2018)	1,247 HH	Respondents who answered the phone and those that answered a question.	Figures 5, 6, 7, 8, and 9; Table D.2
Endline survey (2018)	13,988 HHs	Every respondent we could reach for the endline. Note that some tables that use this data use only a subset (e.g. {H2}) when appropriate and when noted in the table/table notes.	Tables 3, G.1, G.2, G.3, I.1, I.3, and I.4; Figure H.1
Descriptive survey (2020)	2,863 HHs	Those we could reach via phone out of 3,600 HHs randomly subsampled from the experimental group assigned to {H2}.	Figures 6, 7, 8, and 9.
Polling station area level			
Key informant survey (2018)	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 J.1
Electoral data (2019)	300 PS areas	The max set of polling stations at which electoral data available (larger than key informant data due to cost of interviewing key informants).	Table K.1
Politician/constituency level			
Interviews (2019)	35 KP MPAs and MNAs	Hour-long face-to-face semi-structured open-ended interviews.	Paper text
Other			
Expert *forecasts* (2019)	400 respondents	Solicited in-person and over email. Includes undergraduate students, research/development practitioners, and academics around the world. Information provided for forecasts drew on results of pilot study.	Figure L.1

munication with those elicited by IVR technology. In the status quo — the pre-existing system of communication and representation — citizens meet face-to-face with their provincial representatives, often traveling in small delegations from their home villages to do so. We compare five aspects of these interactions to what happens under the new communication channel that we established using cell phone calls and IVR. These are uptake, geographic scope, scale, content, and descriptive representativeness. We discuss each in turn.

Uptake Under IVR, 73 percent of respondents answer the phone, about half are still on the call while the MPA asks a question, and 17.3 percent of respondents answer at least one question posed by the MPA. The biggest drop off occurs after the question is asked but before respondents answer, as is shown in Figure 5. This is consistent with citizen respondents being unfamiliar with the technology at the time we did our experiment, and we expect these numbers would improve with more familiarity. For instance, respondents may have been confused by the instructions about pressing a number to respond. Compared to other ICT interactions and compared to our *descriptive* data on rates of face-to-face interactions between politicians and voters, even this rate is high, as we have documented in Table 1. Thus, IVR has good uptake; as we show momentarily, IVR allows more than double the number of citizens to communicate directly with their representative then status quo modalities.

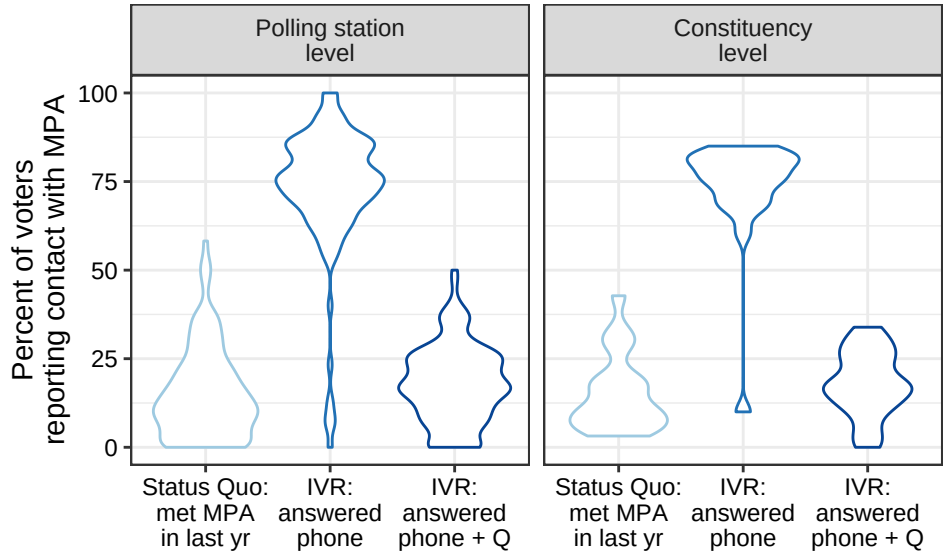
Figure 5: **Dropoff during IVR call**



Notes: Proportion of respondents who remain on the IVR call at various stages of the call. The x-axis is rescaled for illustrative purposes to account for differing call lengths across MPAs.

Scope In Figure 6, we compare the distribution of status quo in-person interactions with what IVR calls achieve across polling stations (left) and constituencies (right). In substantial numbers of polling stations, no one has met his MPA in person in the last year. This number goes to almost zero for IVR contact, defined as answering the phone; a single round of IVR, spanning days instead of a whole year, thus provides complete geographic penetration, including to remote areas whose residents are typically denied direct political access to their MPA. Results are similar if we restrict attention only to IVR respondents who answer a question or if we consider constituency-level differences instead of polling station-level differences. IVR clearly enlarges the scope of contact between citizens and MPAs over status quo in-person interactions. Finally, we also document that citizens who live further from the MPA's constituency office are disadvantaged by the need to travel to face-to-face meetings (see Table E.1). We do not observe the same in the case of IVR contact, because it equalizes access regardless of location. IVR thus democratizes access across constituencies that have large geographic spread.

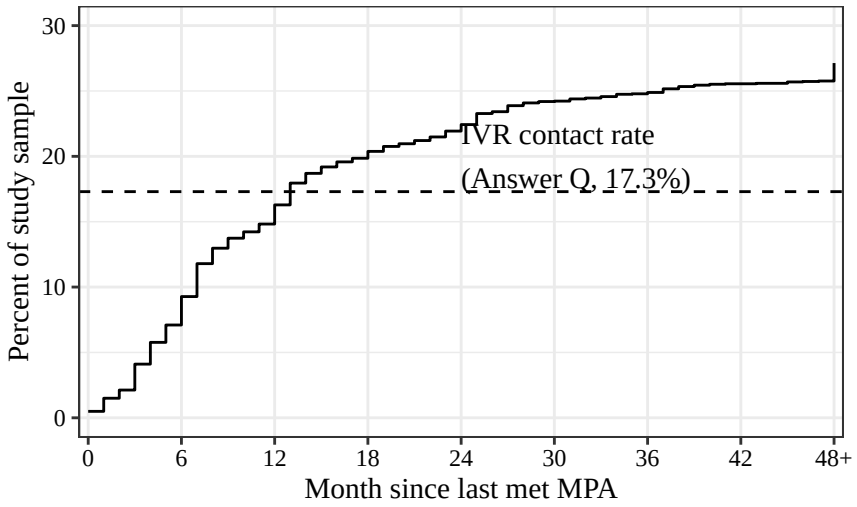
Figure 6: **Distribution of contact rates across polling stations (left) and constituencies (right)**



Notes: Distribution of contact 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.

Scale Next, we analyze the total number of constituents who interact with their elected representative in any given period. Only about a third of all households report *ever* having met their MPA in person; we take this as the upper-limit on face-to-face interactions over *any* period of time. IVR is not only able to more than double the rate of interaction in just a single round of calls, it also establishes as much two-way interaction in one round of calls as occurs face-to-face over a full year, as we show in Figure 7. Thus, IVR vastly improves the scale of contact between citizens and representatives.

Figure 7: **Descriptive evidence on scale of communication transformations under IVR**

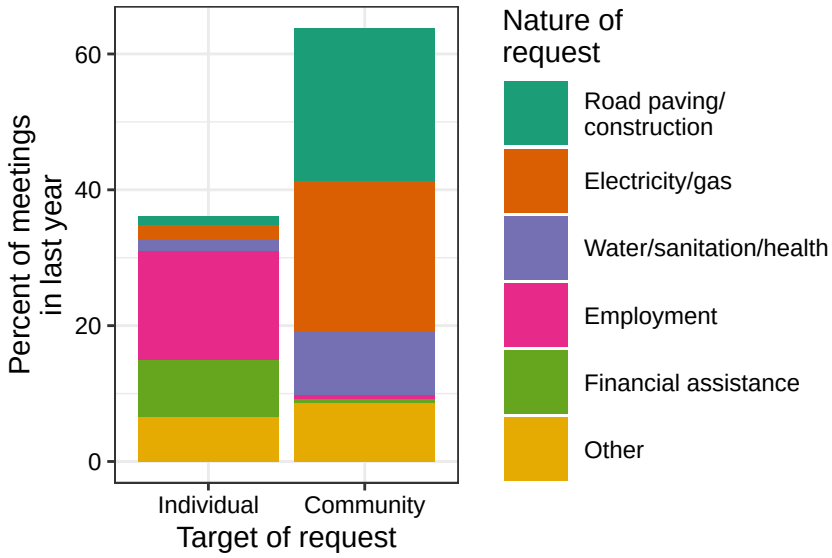


Notes: The solid (increasing) line plots the cumulative distribution of time since survey respondents last met their MPA in person. The dashed (flat) line is the average rate of contact under IVR (where contact is defined as having answered the question).

Content Figure 8 examines the kinds of requests individuals make when meeting their MPA face-to-face. We bin requests into two "targets:" requests that relate to the individual or their household (labeled "individual"), and those that relate to the broader "community." Two-thirds of requests seek community improvements. That is, the data show that even in face-to-face interactions, voters more frequently request basic public infrastructure than individual clientelistic goods. However, face-to-face interactions almost never involve discussions of public policy and legislation. This is in contrast to IVR; Table D.2 shows the distribution

of answers of those responding to the first-stage IVR question. A quarter of first-stage IVR respondents would like their representative to prioritize legislation over other activities. Thus, IVR alters the content of discussion between MPAs and citizens, allowing citizens to voice programmatic concerns that are almost entirely absent from face-to-face interactions.

Figure 8: **Target and nature of requests made to MPA in face-to-face meetings**



Notes: Data are subset to respondents who report having met their MPA in the last year. Targets of requests are coded by whether they pertain to the individual and his household (labeled "individual") or whether they pertain to the community.

Representativeness We also consider which voters MPAs are able to reach using IVR across three groups, displayed in Figure 9: (a) people who report having met their MPA in person in the last year; (b) respondents enrolled in IVR who answer a question; and (c) all randomly sampled respondents enrolled in IVR. These data show that both modes of communication reach constituents who are representative along various dimensions, including demographic (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).[16] Face-to-face interactions are not elite-biased, a finding we highlight

[16]Both methods also are unsuccessful at reaching women as independent voters in the household (Cheema et al., 2023).

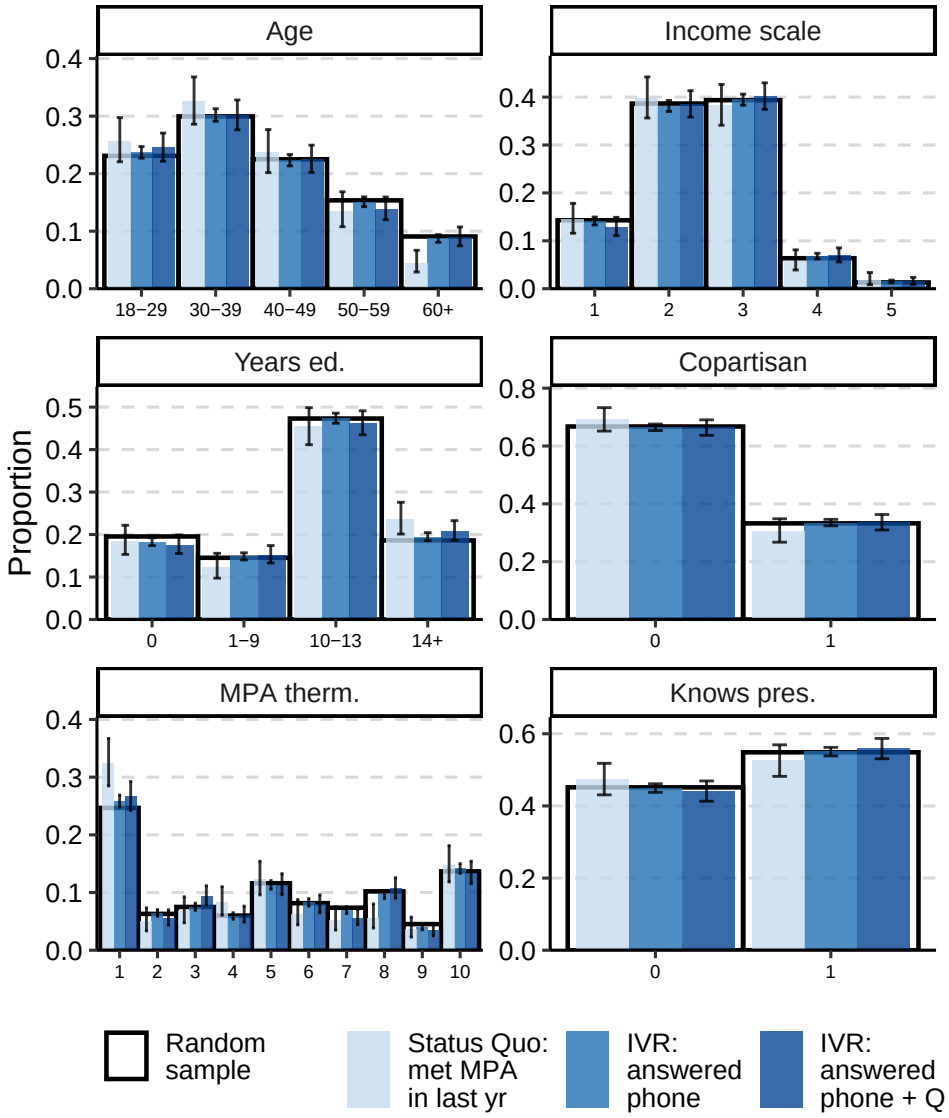
below. Moreover, IVR replicates the pattern of reaching a representative sample on observed covariates. In particular, both modalities reach equal proportions of respondents as regards income and education, and politicians interact with more respondents who are not their supporters than with copartisans.

We consider this a particularly important descriptive result of our data collection. It documents that status quo communication and interaction between elected officials and citizens are not necessarily accurately depicted in much of the literature. Instead, the comparison of IVR with status quo face-to-face interactions between MPAs and citizens shows that the latter exhibit good representativeness as regards income, education, political knowledge, and co-partisanship. MPAs meet with citizens who, by our reading, are not particularly elite on these dimensions. MPAs do not filter out supporters of other political parties nor do they meet with only the wealthiest and most educated strata. The delegations of citizens who approach them mainly ask for club goods, suggesting that communities have organized to send representatives to engage in claim-making of their elected official. This is vastly different than the picture that depicted in much existing literature (e.g. Kitschelt and Wilkinson, 2007) but resonates with an emerging view that communities in less developed countries may be politically assertive, organized, and regularly seek improvements to public welfare (Auerbach, 2019; Bussell, 2019; Kruks-Wisner, 2018).

6 Downstream Results of the IVR Experiment

Results reported in the previous section show that IVR improves four specific dimensions of interaction between politicians and constituents: uptake, geographic scope, content, and scale. Now we examine whether IVR affects voters' attitudes and self-reported political behavior as well as politician service delivery. The results we report in this section stem from the experimental aspects of our research. Except as noted below and detailed in Appendix A, the outcome measures, equation specifications, and treatment effects reported are all

Figure 9: **Constituent characteristics by mode of contact**



Notes: Box plots include 95% confidence intervals.

pre-registered. The order of our presentation mirrors that of the order of the hypotheses introduced in Section 2.1.

6.1 Effects on voter attitudes and behavior

Outcomes: We have 13,988 individuals in our *endline* dataset which, combined with the *baseline* dataset, constitute the data analyzed in this section.[17] At the individual level, we focus on three main outcome indices from survey responses: *evaluations of the incumbent*, *political participation*, and *prospects for electoral accountability*.[18]

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

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

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

Individual level effects: Table 3 reports results at the individual level. The first column presents the standardized control mean, which is zero by construction. The second set of columns presents the treatment effect of any IVR call in standard deviation units as

[17] We were unable to recontact 411 individuals at endline. Results are similar with inverse probability of attrition weights in Table G.1.

[18] Details regarding indices construction appear in Appendix F

well as the corresponding sample size. 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 captured by any index, any of their subcomponents (see Appendix I), or when analyzing effects among compliers (see Table G.3). In the third set of columns, we consider the effect of the fully responsive IVR intervention, whose respondents received an initial call with a question and a subsequent follow-up acknowledging first-stage responses. Although we did not pre-register this comparison, we examine it because it is the deepest possible use of IVR in our research. We have the greatest expectation of positive effects for this high-dosage treatment. Even here, we do not observe any treatment effect. Other results analyzing differences between various treatment arms are reported in Appendix I.

Table 3: **Effects of any IVR call and effects of full IVR treatment on individual citizen outcomes**

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

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

6.2 Evidence of voter engagement

Despite null effects of treatment on voter attitudes and behavior, some components of the intervention successfully affected *voter engagement*. The data shows that when citizens receive a call from their MPA that includes a question — as opposed to an exclusively credit-claiming call — they are more likely to answer a subsequent call from the same politician (see Table I.2). We infer from this that citizens appreciate being included in a policy discussion with

their political representative and would like to continue the conversation. Citizens appear eager for new opportunities to express their views to elected officials. Mere communication, however, is insufficient to change voting or attitudinal outcomes (see Table I.3), in line with findings reported in studies such as Mansuri and Rao (2013) and Lieberman, Posner and Tsai (2014).

6.3 Effects on politician behavior

Next, we evaluate whether IVR involvement affects the behavior of politicians. We examine the kinds of messages they craft as well as whether they direct more attention, spending, or public services to treated localities. We analyze information that comes from interviews with key informants in treated localities. As noted earlier, key informants are educated local residents who agreed to a series of meetings with us to update us on local activities by the MPA. For details, see Appendix J.

We allowed partner MPAs to craft their own IVR messages. We anticipated that this would permit them to tailor communications to treated localities. However, politicians did not craft specific messages. The modal question (see Section 4.2) was, by our reading, general and imprecise. Indeed, 16 of 20 partner MPAs asked identical questions, recycling a prompt provided by our staff. That does not mean that politicians did not invest in the experiment, however. Instead, we interpret their behavior as consistent with the idea that they are experimenting with a new technology to see for themselves how useful it might be for them. Despite limited time and resources, they were willing to engage with their constituents using a new mode of communication. However, perhaps because of lack of experience with anything like robocalls as well as their busy schedules, the politicians put little forethought into the specific messages they recorded.

The politicians enrolled in our study interacted repeatedly with us in the eight months preceding an election. This might have encouraged them to improve service delivery in the

polling station areas targeted for calls or to visit these locations more often to augment the political visibility offered by the calls. Given the importance that MPAs believe that voters put on visits to constituency areas, we use a simple and easily measured metric of effort; namely, the number of visits to a polling station area. A quantitative evaluation of key informant data shows no effect on this measure of MPA effort (see Appendix J.)

6.4 Effects on aggregate electoral outcomes

Our design additionally allows identifying the effects of saturating one polling station area with IVR calls while leaving other areas untouched. This enables us to estimate whether increased communication changes the turnout rate or the MPA's vote share at the polling station level. Again, we do not observe any treatment effects (see Appendix K).[19]

6.5 What did we expect? A forecasting exercise

Were the results we observe obvious ex ante? It is easy to dismiss null results as obvious ex post. To probe this, we conduct forecasting exercises that poll potential policy consumers of this research in Pakistan and academics in the United States and Europe, asking them about their beliefs about the experimental impacts without seeing the results. The results document that other academics and policy experts expected the IVR intervention to succeed (see in Appendix L). Our null findings are thus surprising and therefore, puzzling.

[19]We note for the record that even in polling station areas with the highest saturation of calls, a sample size of 120 respondents would have required massive spillover to the more than 1,100 other voters to produce electoral effects visible at the polling station level, let alone across the 50 to 294 polling station areas that comprise a constituency. The intervention was not designed at a scale to potentially affect actual election outcomes.

7 Discussion: The *Capacity Gap* in Representation

In this section, we lay out our interpretation of the results and also consider possible alternative explanations for the null downstream findings, despite initial promise and engagement with improved communication. We are interested in explaining why politicians did not take advantage of the intervention to visit treated areas more often and why citizen respondents did not change their evaluations of their representatives after involvement with IVR. To understand the surprising outcomes we encounter, we turn to qualitative research methods, mainly drawing on lengthy interviews we conducted with 16 of the 20 treated MPAs and another 20 MPAs and national-level equivalents, Members of the National Assembly (MNAs). In these interviews, we asked open-ended questions about the IVR intervention and about political representation more broadly.

Our interpretation invokes the concept of a *capacity gap* between what politicians seek to do and their abilities to deliver responsive policies. This gap has long been noted in the literature about developing countries (Huntington, 1968). The initial postwar research agenda for scholars of developing nations was largely an effort to understand when the frustrations and disappointments that resulted from such a capacity gap would spill over into mass support of anti-democratic (i.e. communist) insurgencies. However, this older line of research has been largely obscured in more recent political science, whose focus has been the incentive and selection discrepancies that may underlie poor political responsiveness in developing countries. Some research has pointed to the ongoing importance of bureaucratic capacity (Williams, 2017; Dasgupta and Kapur, 2020) and fiscal constraints (Besley and Persson, 2011), but these appear to constitute minor themes in the political science literature. Nonetheless, it is this line of thought that we believe useful to understand the unexpected results of our research.

Before we delve into the concept of a capacity gap, we unpack more fully why the null results on pre-specified outcomes were so surprising. This lays the ground for under-

standing why changing information seems inadequate to improving responsiveness. There are three reasons we expected before the fact that our intervention would succeed: initial MPA enthusiasm; the concrete informational benefits the intervention provided MPAs; and the forecasting intervention.

MPAs themselves were very enthusiastic about working with us. The experiment was embraced by senior politicians who were Members of the Khyber Pakhtunkhwa Provincial Assembly in Pakistan. Most of these individuals had long political experience in their home communities, even if many were newly elected to the provincial assembly. They were accustomed to the demands of vote-getting and knowledgeable about the communities they represented. And more MPAs volunteered to partner with us than we could accommodate, suggesting that they saw potential benefits to IVR.

Second, MPAs clearly benefitted from aspects of the intervention. MPAs reported that the information they gained when we delivered aggregate feedback from voters was politically helpful because it revealed specific polling station areas where they had more or less electoral support.[20] That is, MPAs considered the information useful for electoral goals. In addition, many claimed that they lacked existing methods to collect this information.[21] Thus, the information the IVR experiment offered was new. Finally, MPAs reported that information that they subsequently collected to understand more fully the citizen feedback they had received brought to their attention reports of specific public works failures (e.g. non-operational water pipes). Thus, the information provided via IVR helped them to identify problems that they could solve in the short-term.

The third reason we expected our intervention to succeed comes from the results of our forecasts. These show that local as well as globally-located academics and also policy practitioners expected the intervention to generate engagement and also to change voter

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

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

behavior. Thus, knowledgable experts believed the IVR intervention would produce results.

Yet MPAs failed to redirect efforts towards areas where voters received IVR communication or to make other observable efforts to leverage the intervention for their own stated electoral goals. The data we collected from key informants shows no change in MPA behavior on these indicators. And the evaluations that citizen respondents made of their representatives did not improve. What happened?

A first hint of why our work produced null results comes from the failure of MPAs to make repeated IVR recordings in the period of the intervention, which covered the eight months preceding joint provincial and national elections. The research was designed to leverage the political aspirations of the enrolled MPAs, more than 80 percent of whom ran for reelection in 2018 just after the intervention ended. Despite their electoral ambitions, most MPAs recorded only a single question and a single follow-up response. Our staff approached them repeatedly during the months prior to the elections with offers to make additional recordings but were rebuffed. We were surprised by this because it seemed that MPAs were missing out on opportunities for self-promotion, credit-claiming, enhanced name recognition, and visibility among voters in the critical pre-election period. Communicating with voters via prerecorded messages was also a relatively painless and quick activity, certainly compared with traveling from the capital city back to home constituencies and interacting directly with constituents. Why did MPAs almost uniformly decide not to pursue continued opportunities for IVR interactions with citizens?

In interviews, the MPAs reported that voters tended to misinterpret the IVR questions as *commitments* to provide new infrastructure.[22] When they went back to their village offices after the IVR calls, some MPAs confronted angry voters wanting to know when they would make good on their promises.[23] As a result, despite their initial enthusiasm, MPAs

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

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

disengaged from the experiment and did not record additional messages. Asking citizens what they wanted only raised public expectations without providing additional fiscal resources or bureaucratic capacity with which to satisfy them — and this in an environment where MPAs are unable to satisfy many of the existing requests that they receive in person.[24] In interviews, MPAs explained that much of their job regularly consisted of dealing with dissatisfied constituents because inadequate fiscal resources meant the MPAs had to pick and choose which communities would get which limited and inadequate — and perhaps never completely operational — public goods. They were constantly confronting the temptation to make promises they could not keep. On top of this, MPAs expressed considerable frustration with the nature of voter requests, which they deemed often "inappropriate," occasionally illegal, and frequently misaligned with their own responsibilities. For instance, MPAs reported that voters regularly sought them out for jobs because of poor labor market or agricultural conditions, although provincial representatives had no way legally to provide job opportunities. MPAs contended that voters failed to appreciate the limited sphere of authority of an MPA and sought assistance with problems that lay beyond the MPA's jurisdiction or competence. These kinds of misunderstandings, many said, were ongoing but were sharpened by the IVR intervention.

The experiment thus generated misaligned expectations. Citizen respondents who answered a question deemed their representative unreliable and unresponsive due to his inability to follow up on their demands. MPAs, on their side, now had more information about voter preferences but were no more able to use it effectively. They were thus additionally frustrated by the limited capacities of their offices. It is perhaps unsurprising, given this, that some citizen respondents who engaged with their representative via IVR may even have reduced their evaluations of him (see Table 3), in line with Kruks-Wisner (2021) and Cheng

[24] A chronic inability to respond to citizen requests was noted by respondents 2, 3, 4, 5, 6, and 7, group interview 22 April 2019; respondent 10, 24 April 2019; respondents 12 and 13, group interview, 24 April 2019; and respondent 15, 25 April 2019.

and Liu (2018). And it is perhaps unsurprising as well that the MPAs we partnered with disengaged from the intervention, and in interviews with us expressed a mixture of frustration and embarrassment.

The IVR intervention was designed precisely to raise voters' expectations of politicians and to encourage them to use performance-based criteria to evaluate elected officials. Ironically, to the extent that our intervention succeeded in doing this, it frustrated elected representatives and dead-ended the process of change.

This unbalanced interaction also suggests the dynamics that might potentially trigger a new, virtuous cycle of electoral accountability and improved performance. Since IVR communication raised expectations among voters, using it could potentially prod politicians to reconsider their use of existing resources. That is, *if MPAs had been aware in advance of how voters might respond to IVR*, they might have prepared to respond flexibly and quickly. This could have included setting aside resources in advance, preparing civil service officers, and counseling staff. But using political communication to break out of the current equilibrium characterized by low expectations and poor responsiveness might require extensive preparation of MPAs and likely of bureaucratic officers as well.

We make one additional point before proceeding to consider alternative explanations. The Pakistani provincial authorities we worked with evinced surprise at the results of disseminating messages and questions to voters via IVR. They, too, expected the intervention to succeed — that is why they agreed to work with us in the first place. Their disappointment mimics the results of research conducted in the United States that finds that political elites are not very good at knowing what messages will be convincing to the public (Broockman et al., 2024). Even in high-information environments, political communication is a learning process. When provided a new communication modality, politicians will have to experiment with it repeatedly over time to calibrate it to electoral goals.

7.1 Alternative explanations

Our attention to the capacity gap is not meant to single out this concept as the exclusive cause of why politicians disengaged from the IVR intervention. Politicians face complex social and political environments, and the concept of the capacity gap is inherently multifaceted and in addition incomplete as an explanation. In this section, we consider some alternative or additional interpretations of our results.

Time horizon of the study: Would our study have generated different results if we had conducted it over a longer period of time, allowing MPAs to continue to call respondents and ask new, perhaps more finely tailored, questions? While plausible, it seems unlikely to us that MPAs would have engaged more deeply with IVR had it been offered outside of the context of an upcoming election — which would have been the case if the intervention had continued for longer or started earlier. We selected to work in the eight months prior to an election precisely to leverage the opportunity to engage MPAs when they were most sensitive to public opinion. But even then, MPAs were not willing to continue to asking questions.

Modality of IVR: Another possibility is that voters place intrinsic value on face-to-face interactions with their representatives and would not consider IVR calls genuinely responsive regardless of their depth, number, or frequency. There is some evidence in favor of this interpretation. In our interviews with MPAs, they repeatedly claimed that voters expect regular on-the-ground interactions. Many MPAs we talked to noted that they spend at least one full day each week attending events in their communities in order to interact in person with voters.[25] Our interpretation of these in-person interactions is that voters want to meet their MPA and have their names known to him in order to forge a direct connection with a politically powerful person in the event of need, however. We are agnostic about whether voters place intrinsic value on these interactions; we can only state with some confidence

[25]Respondents 2, 3, 4, 5, 6, and 7, group interview 22 April 2019; respondents 10 and 11, 24 April 2019.

that they appear to use them for risk reduction (cf. Scott, 1977).

If voters do place high intrinsic value on face-to-face interactions with politicians, it would be surprising for them to answer the IVR questions in the first place and to continue answering their phone in the second round (see Table I.2). Given the relatively high response rates to the intervention, we question whether many voters have such firm and pronounced preferences for in-person interactions that they would not supplement them with electronic modes of communication. Finally, it is worth noting that the IVR intervention supplements face-to-face communications — that is, voters retain the ability to seek an in-person meeting with their MPA and our intervention did not ask them to choose between modalities of contact with their representative.

Party constraints on politician activities: It is also possible that MPAs in Pakistan lack autonomy because political party leaders exercise strict control over the delivery of club goods to localities. For instance, there is some evidence that the governing party directs constituency spending to the areas that are most electorally helpful to it, and in doing so limits the autonomy and discretion of representatives of lower levels of government (Malik, 2019). However, while party leaders in Pakistan exercise some control over the activities of local politicians, this control is far from perfect, as exhibited by the very frequent party switching by politicians that occurs. When central control is exercised, it is imperfect and incomplete, still leaving considerable discretion for how resources and influence are spent locally.

Power considerations: One final concern with the null results is that the experiment might have been statistically underpowered and therefore unable to detect effects even if they were to exist. Although this may be the case at the polling station level, it is unlikely to be true at the household level, where we have a large sample. We can formalize this by computing equivalence confidence intervals, as proposed in Hartman and Hidalgo (2018). These equivalence confidence intervals provide the range outside of which true effects are

unlikely to lie given the data and our significance level of 0.05. As shown in Appendix Figure H.1, 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.

8 Conclusions

We undertook a large-scale intervention, partnering with 40 Pakistani elected members of a provincial assembly, to facilitate the use of Interactive Voice Response communication with nearly 15,000 citizens. The technology allowed politicians to ask questions and citizens to respond using their cell phone keypads. We expected this to improve representation, but instead we found that politicians did not improve service delivery along the lines requested and that voters did not improve their evaluations of politicians.

Our interpretation of these results is that elected politicians in our setting operate under hidden constraints — which we call a capacity gap — that neither they nor we anticipated when we began the IVR treatment. Politicians were enthusiastic partners at the outset of our intervention. But enhancing contact and communication between them and thousands of persons they represented allowed the latter to express their preferences while not augmenting the resources controlled by the former. Politicians may have already known what voters wanted from government — our interviews suggested they certainly thought they did — but their ongoing failure to respond was baked into the low-capacity setting in which they operate. Only additional research can disentangle the various competing potential explanations and interpretations of this capacity gap. Identifying it explicitly speaks to the need for additional systematic on-the-ground research into the operation of distributive politics in Pakistan and other developing countries.

The dynamics we uncover seem likely to characterize many democracies, which struggle to stay abreast of citizen complaints (through 311 systems, for instance) and to provide universal public goods. Capacity gaps of the magnitude we document may also characterize many local and state governments, for instance, and our null results may recur whenever subnational budgets are outstripped by bottom-up demands.

Recent literature encourages disseminating null results of research in order to reduce publication bias (Franco, Malhotra and Simonovits, 2014). We agree, and add that providing plausible — even if speculative — interpretations of null results adds to the case. Previous research on spring-boarding communication between politicians and voters reports low take-up by voters. Our research shows that a politician-led approach resolves the take-up challenge but introduces other, perhaps more fundamental and previously unobserved complications — complications that neither we nor our politician partners anticipated. Even if politicians wish to respond to voter preferences, in many developing countries they may lack the capacity to do so. A more robust communication infrastructure between politicians and voters may need a concurrent boost in the arrangements and resources that allows politicians to make promises they can keep. Politicians may back off from using new communication technologies if these expose them to voter disappointment. This response is reasonable, and suggests that many things must change simultaneously to improve political responsiveness where it is low. Future work should randomize *capacity boosts* — discretionary funds or staff — alongside information flows.

References

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

Auerbach, Adam Michael. 2019. *Demanding development: The politics of public goods provision in India's urban slums*. Cambridge University Press.

Bafumi, Joseph and Michael C. Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(3):519–42.

Bartels, Larry M. 2008. *Unequal Democracy: The Political Econom of the New Gilded Age*. Princeton: Princeton University Press.

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

Besley, Timothy and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton: Princeton University Press.

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).

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

Broockman, David E., Joshua L. Kalla, Christian Caballero and Easton Matthew. 2024. "Political practitioners poorly predict which messages persuade the public." *Proceedings of the National Academy of Sciences* 121(45):e2400076121.

Buntaine, Mark T., Daniel L. Nielson and Jacob T. Skaggs. 2019. "Escaping the Disen-

gagement Dilemma: Two Field Experiments on Motivating Citizens to Report on Public Services." *British Journal of Political Science* 51(2):685–705.

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* 31(1):108–27.

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

Butler, Daniel M. and David W. Nickerson. 2011. "Can learning constituency opinion affect how legislators vote? Results from a field experiment." *Quarterly Journal of Political Science* 6(1):55–83.

Carnes, Nicholas. 2013. *White-Collar Government: The Hidden Role of Class in Economic Policy Making*. Chicago: University of Chicago Press.

Carnes, Nicholas, Joshua Ferrer, Miriam Golden, Esme Lillywhite, Noam Lupu and Eugenia Nazrullaeva. 2025. "The Global Legislators Database: Characteristics of National Legislators in the World's Democracies." *British Journal of Political Science* 55.

Carnes, Nicholas and Noam Lupu. 2023. "The Economic Backgrounds of Politicians." *Annual Review of Political Science* 26:253–70.

Casey, Katherine, Rachel Glennester, Edwards Miguel and Maarten Voors. 2023. "Skill versus Voice in Local Development." *Review of Economics and Statistics* 105(2):311–26.

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

Cheema, Ali, Sarah Khan, Asad Liaqat and Shandana Khan Mohmand. 2023. "Canvassing the gatekeepers: A field experiment to increase women voters' turnout in Pakistan." *American Political Science Review* 117(1):1–21.

Cheng, Tony and Shelley Liu. 2018. "Service cynicism: How civic disengagement develops." *Politics & Society* 46(1):101–29.

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.

Christensen, Julian and Donald P. Moynihan. 2020. "Motivated reasoning and policy information: politicians are more resistant to debiasing interventions than the general public." *Behavioural Public Policy* 8(1):47–68.

Dahl, Robert A. 1972. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.

Dasgupta, Aditya and Devesh Kapur. 2020. "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India." *American Political Science Review* 114(4):1316–34.

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.

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.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1–3):5–25.

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

Franco, Annie, Neil Malhotra and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer Problem." *Science* 345(6203):1502–05.

Gallup Pakistan. 2013. "10th General Elections in Pakistan 2013.". Available at: https://gallup.com.pk/bb_old_site/election/10GIER2013.pdf.

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

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

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

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

Gulzar, Saad, Zuhad Hai and Binod Kumar Paudel. 2021. "Information, candidate selection, and the quality of representation: Evidence from Nepal." *The Journal of Politics* 83(4):1511–1528.

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

Hertel-Fernandez, Alexander, Matto Mildenberger and Leah C. Stokes. 2019. "Legislative Staff and Representation in Congress." *American Political Science Review* 113(1):1–18.

Hug, Simon, Frédéric Varone, Luzia Helfer, Stefaan Walgrave, Karolin Soontjens and Lior Sheffer. 2024. "The role of politicians' perceptual accuracy of voter opinions in their electoral career." *Legislative Studies Quarterly* 50(1):107–22.

Huntington, Samuel P. 1968. *Political Order in Changing Societies*. New Haven: Yale University Press.

Jablonski, Ryan and Brigitte Seim. 2024. "What Politicians Do Not Know Can Hurt You: The Effects of Information on Politicians' Spending Decisions." *American Political Science Review* 118(3):1497–1517.

Kalla, Joshua L. and David E. Broockman. 2016. "Campaign Contributions Faciliate Access to Congressional Officials: A Randomized Field Experiment." *American Journal of Political Science* 60(3):545–58.

Kalla, Joshua L. and Ethan Porter. 2021. "Correcting Bias in Perceptions of Public Opinion Among American Elected Officials: Results from Two Field Experiments." *British Journal of Political Science* 51:1792–1800.

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

Kruks-Wisner, Gabrielle. 2021. "Great expectations, Great Grievances: the politics of citizens' complaints in India." *Comparative Politics* 54(1):27–49.

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

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.

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.

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

Lipset, Seymour Martin. 1963. *Political Man: The Social Bases of Politics*. Expanded edition ed. Baltimore: Johns Hopkins University Press.

Lupu, Noam and Zach Warner. 2022. "Affluence and Congruence: Unequal Representation around the World." *Journal of Politics* 84(1):276–90.

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.

Malik, Rabia. 2019. "(A)Political Constituency Development Funds: Evidence from Pakistan." *British Journal of Political Science* 51(3):963–80.

Mansuri, Ghazala and Vijayendra Rao. 2013. "Can participation be induced? Some evidence from developing countries." *Critical Review of International Social and Political Philosophy* 16(2):284–304.

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.

Miller, Warren E. and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1):45–56.

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

Page, Benjamin I. and Robert Y. Shaprio. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.

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

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

Pereira, Miguel M. 2021. "Understanding and Reducing Biases in Elite Beliefs About the Electorate." *American Political Science Review* 115(4):1308–24.

Roy, Nilanjana S. 2012. "Mobile Phones Offer Indian Women a Better Life." *The New York Times* . Accessed: 2025-04-08.
URL: *https://www.nytimes.com/2012/05/23/world/asia/23iht-letter23.html*

Scott, James C. 1977. *The Moral Economy Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*. New Haven: Yale University Press.

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

Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave and Tamir Sheafer.

2017. "Nonrepresentative representatives: an experimental study of the decision making of elected politicians." *The American Political Science Review* 112(2):302–321.

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.

Szwarcberg, Mariela. 2015. *Mobilizing Poor Voters: Machine Politics, Clientelism, and Social Networks in Argentina*. New York: Cambridge University Press.

Walgrave, Stefaan, Arno Jansen, Julie Sevenans, Karolin Soontjens, Jean-Benoit Pilet, Nathalie Brack, Frédéric Varone, Luzia Helfer, Rens Vliegenthart, Toni van der Meer, Christian Breunig, Stefanie Bailer, Lior Sheffer and Peter John Loewen. 2023. "Inaccurate Politicians: Elected Representatives' Estimations of Public Opinion in Four Countries." *Journal of Politics* 85(1):209–22.

Walgrave, Stefaan, Karolin Soontjens and Julie Sevenans. 2023. *Politicians' Reading of Public Opinion and its Biases*. Oxford: Oxford University Press.

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

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

Williams, Martin J. 2017. "The political economy of unfinished development projects: Corruption, clientelism, or collective choice?" *American Political Science Review* 111(4):705–723.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zaman, Adil and Muhammad Mushtaq. 2022. "Public Perspective About Party Identification in Peshawar Valley, Pakistan." *Pakistan Journal of Social Research* 4(2):500–513.

Appendices

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

Key informant interviews: 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 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.

B Ethics of the intervention

Prior to any field activities involving human subjects, we sought permission from Institutional Review Board (IRB) at [redacted] Interviews with politicians conducted in 2019 were approved in an IRB amendment.

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

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

  There were practical reasons for working exclusively with sitting MPAs. Our experiment was rolled out in the nine months prior to our best guess of when the next election would be held. Because of the lead time required for the experiment, it was not feasible to undertake it immediately prior to the election. At the time the experiment was conducted, the identity of assembly candidates in the upcoming 2018 elections was unknown. (Indeed, seven of the 40 incumbents we work with did not run again.) In addition, our funding agency [redacted] 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.

- Voters' expectations: A second concern is that communication with voters by politicians might raise the expectations of the former which, if unfufilled, might create citizen disappointment. We consider that offering politicians new opportunities to speak to their constituents allows politicians to shape voter expectations. That is, a goal of our project was precisely to encourage politicians to use technology to increase voters' expectatations — with an eye to responding in a timely manner. We do not consider that allowing politicians to raise the expectations of citizens is inherently ethically problematic; instead, we deem politicians the natural judges in a democracy of how, when, and whether to do that.

- Political interference: A third ethical concern is that our intervention might have affected political outcomes. While we personally do not consider it unambiguously unethical for academic interventions to interfere with political processes, our sample sizes are nevertheless too small for this to have been possible. For evidence, see the discussion in Section 6.4.

- Gender: A final ethical concern is that we work only with men and exclude women. As we note however (see fn. 4.1), even when we sent female staff to approach women, they were unwilling to provide their phone numbers and participate in the study. Although 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 heads-of-households 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 respondents, 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.

C 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 activities — baseline survey, two stages of phone calls, and endline survey — is reported in Table C.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

Table C.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.

meetings and because some MPAs recorded their calls later when they found 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.

D Balance Table

Table D.1 presents evidence that randomization was successful . Only one hypothesis out of the 24 that we test is significant at the 95% level, which is lower than what statistical chance would yield.

Table D.1: Balance Table

	Tier 1 - Full Sample				**Tier 2 - by H1**			**Tier 3 - by H2**			
Variable	H0	H1	H2	P-value	H1C	H1Q	P-value	H2C	H2G	H2R	P-value
Household Characteristics											
MPA Thermometer	2.635	2.57	2.62	0.651	1.974	1.926	0.712	2.978	2.8	2.822	0.104
	(0.257)	(0.26)	(0.257)		(0.508)	(0.513)		(0.405)	(0.402)	(0.401)	
MPA Party Thermometer	5.531	5.547	5.547	0.957	5.892	6.008	0.427	5.687	5.536	5.627	0.331
	(0.367)	(0.371)	(0.367)		(0.969)	(0.969)		(0.539)	(0.539)	(0.54)	
Education Years	7.786	7.95	8.009	0.048	9.228	9.172	0.788	8.358	8.403	8.489	0.67
	(0.435)	(0.438)	(0.432)		(0.858)	(0.852)		(0.602)	(0.599)	(0.602)	
Income Scale	2.53	2.512	2.542	0.253	2.401	2.399	0.961	2.53	2.544	2.526	0.749
	(0.076)	(0.077)	(0.076)		(0.169)	(0.171)		(0.111)	(0.111)	(0.111)	
Copartisan	0.367	0.364	0.367	0.957	0.344	0.356	0.549	0.371	0.364	0.366	0.884
	(0.044)	(0.045)	(0.044)		(0.11)	(0.11)		(0.063)	(0.063)	(0.063)	
Legislation Importance	3.176	3.176	3.177	0.999	3.055	3.045	0.751	3.241	3.235	3.236	0.956
	(0.042)	(0.043)	(0.041)		(0.053)	(0.053)		(0.067)	(0.067)	(0.067)	
Development Importance	3.165	3.189	3.183	0.329	3.079	3.121	0.21	3.191	3.213	3.219	0.47
	(0.062)	(0.062)	(0.061)		(0.101)	(0.1)		(0.089)	(0.088)	(0.089)	
Efficacy	1.754	1.798	1.781	0.167	1.663	1.637	0.559	1.768	1.788	1.794	0.705
	(0.088)	(0.089)	(0.088)		(0.168)	(0.168)		(0.125)	(0.125)	(0.125)	

Table D.2: Responses to first-stage IVR questions: descriptive data

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

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

E Additional Descriptive Results

E.1 How contact varies by distance

Table E.1 reports results from two sets of polling station level regressions, distinguished according to whether the respondent reports having met his MPA in person in the last year or whether he answered an IVR question. The regressions study the relationship between distance to an MPA's constituency office and each type of contact with the MPA. We measure distance as the geodesic distance between the respondent's polling station and the MPA's constituency office.[26] 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 meet in person with the MPA; moving from the 25th percentile (4km) to the 75th (15km) percentile on distance corresponds to a 2.5 percentage point decrease in the percent of respondents who report meeting face-to-face with the MPA. Thus, voters who live further from the MPA's constituency office are disadvantaged by the need to

[26]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.

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

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

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

travel for face-to-face meetings.

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, no outcome is imputed and the 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 control households, when we make comparisons between treatment arms, the standard deviation in the comparison group is not always equal to one.

We first study effects of the intervention on *evaluations of the incumbent*. The index comprises four outcomes: (1) a feeling thermometer for the MPA himself (1–10); (2) a feeling thermometer for the MPA's party (1–10); (3) a binary indicator for whether the respondent reports having voted for the MPA (or the incumbent MPA's party if the incumbent did not run again); and (4) the inverse of the ranking of the MPA among his top four challengers in the 2018 elections. Second, we study effects of the intervention on *political participation*. This index consists of three outcomes: (1) a binary indicator for whether or not the respon-

[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).

dent voted in the 2018 election; (2) a binary indicator for whether or not the respondent attended a rally in the period leading up to the 2018 elections; and (3) a binary indicator for whether or not the respondent attended a political meeting before the 2018 elections. Third, we study effects of the intervention on *prospects for electoral accountability*. This index is made up of three outcomes: (1) a measure of self-stated political efficacy (1–5); (2) how important incumbent performance is in an individual's vote choice (1–6); (3) the number of conversations the respondent had about politics in the two weeks before the endline survey. The first item is included because we think voters are more likely to attempt to evaluate the performance of the incumbent when they have higher levels of political efficacy. We include the second item to measure whether the voter thinks that performance criteria should be used when deciding for whom to vote. The final item is included because we contend that voters cannot enforce political accountability without some political engagement that includes discussing issues.[28]

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 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.

[28]We also show that adding this outcome to the political participation index does not change any of the conclusions reported throughout the paper (see Table G.2).

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

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

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses. Pre-treatment control variables not displayed; see Section 6.1 for details. Control means, treatment effects, and all estimates 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

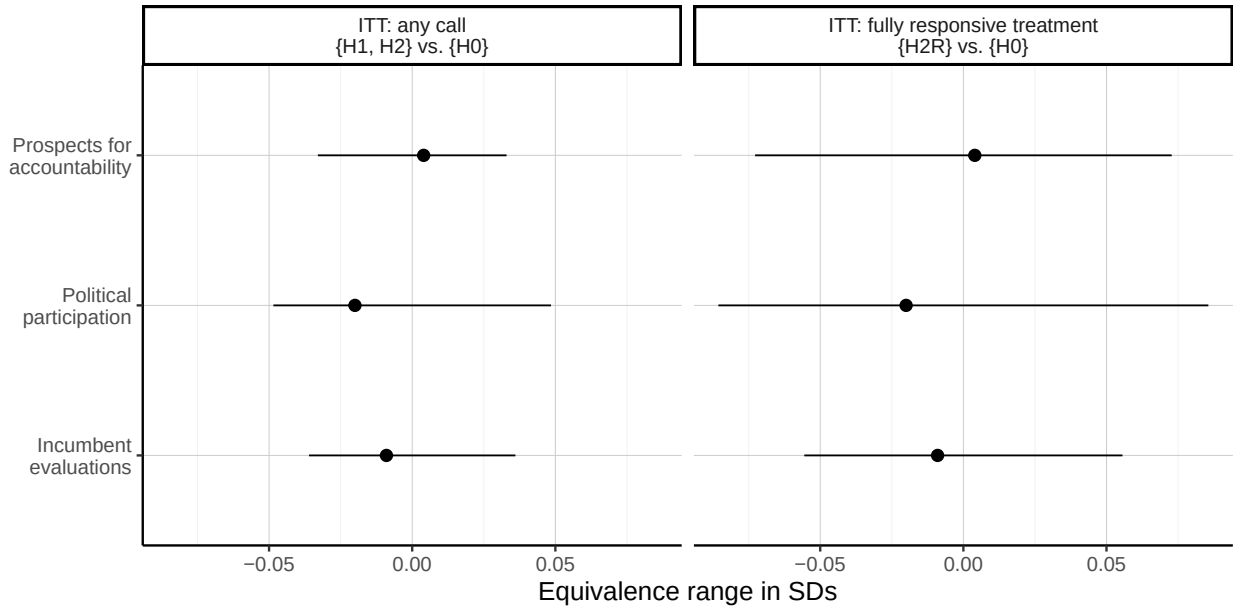
	Control mean: no call {H0}	ITT: any call {H1, H2} vs. {H0}		ITT: full responsive treatment {H2R} vs. {H0}	
Outcome	μ	τ	N	τ	N
Incumbent evaluations index	0.000	−0.009	13757	−0.016	6539
	(1.000)	(0.009)		(0.013)	
MPA feeling thermometer (1-10)	4.864	−0.056	13753	−0.087	6536
	(3.340)	(0.038)		(0.058)	
MPA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

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 6.1 for details.

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 receiving 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.

H Robustness of null findings in downstream results

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.

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

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

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

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.

Below we expand tables reported in the main paper to include index components, and we include additional treatment group comparisons.

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

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

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 6.1 for details.

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

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

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

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

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

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 6.1 for details.

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

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

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 6.1 for details.

J Effects on politician behavior

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

We ask key informants whether politicians visit the polling station area: the goal is to evaluate whether they visit areas more where we direct their IVR phone calls. We also ask key informants whether politicians exert effort in delivering public services, including schools, roads, health facilities, employment conditions, electricity provision, gas provision, water provision, rubbish collection, and general security.

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

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

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

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

K 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 communicate with households in the six treated polling stations.

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

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

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

K.0.1 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.

K.0.2 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 redelimitation and because some of our partner MPAs did not seek office again, not all partner MPAs were 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.

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

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

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

K.0.3 Effects within treated constituencies: Table K.1 presents intervention effects from treated polling stations compared to control polling stations within treated MPA constituencies.

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

L Forecasting Details and Results

We conduct forecasting exercises that poll potential policy consumers of this research in Pakistan and academics in the United States and Europe, asking them about their beliefs about the experimental impacts without seeing the results. To do this, we provide them information about out pilot study — including the fact that the pilot did not result in statistically significant changes in respondents' evaluations of their MPAs. Nonetheless, the 400 forecasters significantly and substantively overestimate treatment effects across our three downstream respondent-level outcomes (the incumbent evaluation index, the political participation index, and the prospects for accountability index)[redacted].[29]

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

We analyze results from two sets of forecasts: the first with 283 (out of 1,584 solicited) academic experts in or associated with the United States[30] and the second with 117 (out of 172 solicited) university students, faculty, and policy practitioners in Pakistan, for an overall total of 400 responses. The two waves allow examination of differences between local and global expertise (Casey et al., 2023). After briefly describing the context and the experiment, we ask respondents to forecast the take-up rates for the IVR calls as well as impacts on downstream results on the three key voter indices analyzed above. Details are available in [redacted]. We used these four simple outcomes in order to keep the forecasting tasks relatively simple and thereby more reliable.

Results: Panel A in Figure L.1 shows that, on average, nearly 40 percent more people answer the phone than forecasters predict. However, conditional on whether the IVR call was answered, forecasters were more likely to think that respondents would answer an IVR question than the number that actually did. That is, forecasters underestimate whether respondents would answer the phone but overestimate whether respondents would answer an IVR question.

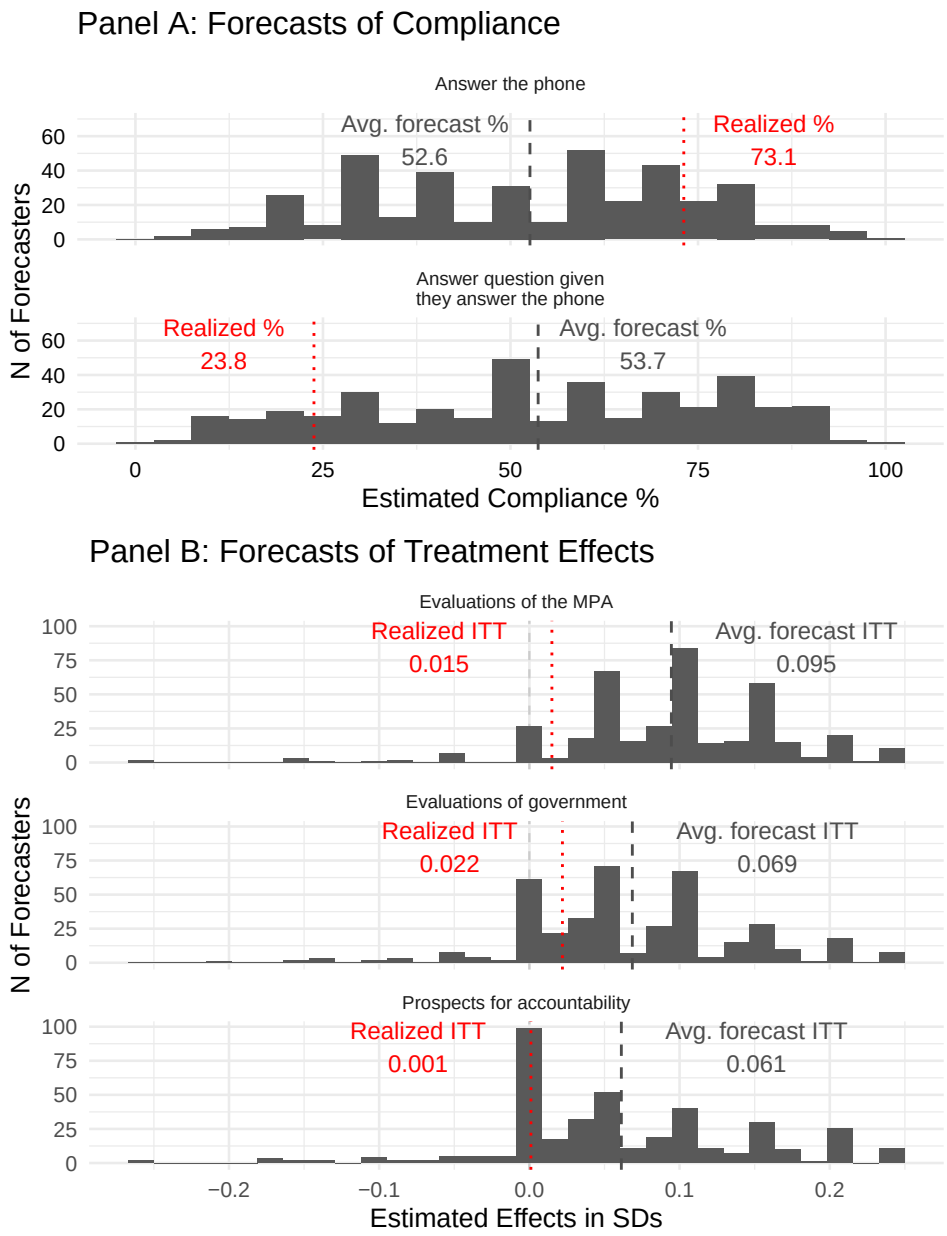
Panel B contrasts the forecasts against realized results on the three downstream outcome indices. Forecasters were asked to predict the intent-to-treat effect for each index. Forecasters overestimate the size of all effects: in every case, average forecasts were substantially (and statistically significantly)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

Of course, forecasts of field experiments might generally be overly optimistic. One

[29] Forecasts were incentivized: respondents whose responses were closest to the truth received small Amazon giftcards.

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

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



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

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). A second reason is that interested academics and policy experts might have a kind of intellectual vested interest in seeing interventions produce results [redacted].

Whether for both or for other reasons, the results of the forecasting research shows

that the results of the intervention 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 that the intervention produced unanticipated results for knowledgeable and interested persons in the country of research and elsewhere. The key interpretation that we stress is that there were high prior expectations among research consumers of the intervention in advance of knowing the results. While after-the-fact reasoning may claim the null results were to be expected, the forecasting results show that they were not. Indeed, an important advantage of forecasting research in advance of knowing the results is precisely that forecasting can reveal genuine surprise of null (or other) results.

Appendix References

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

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

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