Electoral Dis-connection: The Limits of Re-Election in Contexts of Weak Accountability

Julie Anne Weaver[1]

Harvard University[2]

Current draft: September 20, 2019

Most recent version available here.

Invited to Revise and Resubmit at the *Journal of Politics*

[1] PhD Candidate, Department of Government, Harvard University, julieanneweaver@fas.harvard.edu.

[2] Thank you to Fran Hagopian, Alisha Holland, Horacio Larreguy, Audrey Latura, Steven Levitsky, Stephanie McNulty, Torben Iversen, Pia Raffler, and participants of the APSA 2019 Annual Meeting, NEWEPS, Behavioral Models of Politics Conference, Harvard-MIT Latin America Discussion Group and Harvard University Comparative Politics Research Workshop for helpful comments, to Chase Harrington for assistance with the survey, and to José Incio for generously sharing data. The research was approved under IRB protocol IRB16-0872 and modifications MOD16-0872-02 and MOD16-0872-03. A pre-analysis plan for the conjoint experiment was registered with Evidence in Governance and Politics (EGAP) prior to the survey being launched (20170811AB). The author gratefully acknowledges funding from Harvard's Foundations of Human Behavior Initiative (FHB), the David Rockefeller Center for Latin American Studies (DRCLAS), and the Institute for Quantitative Social Science (IQSS), and dissertation fieldwork funding from the Inter-American Foundation (IAF) Grassroots Development Fellowship.

Abstract

Holding politicians accountable through re-election has long been a focus of empirical work, yet results are mixed as to whether electoral accountability works in practice. I offer a new theory of voter behavior to explain why electoral accountability may break down. Where voters perceive a greater likelihood of malfeasance in a second term, information about good performance in the first term becomes irrelevant to predicting good performance in the second. Accordingly, voters turn to other accountability institutions for assurance that re-elected incumbents will perform well. I test this argument in the context of mayoral politics in Peru. In a conjoint experiment embedded in an original survey, respondents prefer challengers even when explicitly informed the incumbent performed well, and the effect is strongest among those who doubt good performance will continue if re-elected. Using a regression discontinuity design, I then show that mayors face a significant incumbency disadvantage, and that neither good performance nor voters' access to performance information enables mayors to overcome it. Instead, it is voter trust in accountability institutions that helps attenuate anti-incumbency bias. Together, these results suggest that attempts to improve electoral accountability by expanding voter access to performance information may prove inadequate without strong central-level oversight, and, more broadly, that re-election may fail to generate political accountability when other accountability institutions are weak.

In December 2018, an astounding 85.81% of Peruvian voters in a national referendum opted to ban Congressional re-election. If the possibility of being rewarded with an additional term in office incentivizes politicians' good performance (Ashworth 2012), why might voters reject incumbent re-election, despite the potentially beneficial effects?

For re-elections to generate political accountability, voters have to be both willing to re-elect incumbents and to base their re-election decision at least in part on incumbent performance. An emerging literature suggests, however, that in contrast to findings from the United States and other developed democracies, some developing countries exhibit an incumbency disadvantage, with incumbents systematically less likely than challengers to be elected at the subsequent election (Klašnja 2015; Klašnja and Titiunik 2017; Uppal 2009). In such contexts, we know little about why a bias against incumbents develops nor if good performance in office ameliorates incumbents' electoral handicap, making electoral accountability still possible. If not, what explains the success of the incumbents that are re-elected, despite their electoral disadvantage?

To address these questions, my argument begins from the contention that an incumbency disadvantage can emerge when voters doubt that incumbents will continue to perform well in office if given another chance. This may be particularly likely if second-term mayors are assumed to more successfully engage in corruption than newly elected challengers, given on the job learning about how to avoid detection and the development of networks of collusion (Klašnja 2015; 2016). Weak oversight institutions mean incumbents may be particularly willing to act on this new knowledge because they face little risk of punishment if caught. I argue that in such contexts, voters become skeptical that good incumbent performance in the first term predicts continued good performance in the second. Voters may then perceive that any draw from the challenger pool, who by default will be in her first term, will be less corrupt than a re-elected incumbent. In response, voters will tend toward voting incumbents out of office, rather than distinguishing between good and bad performers to then punish or reward accordingly. However, a perception of strong oversight may attenuate this in-

cumbency disadvantage. Voters may turn to other components of the overall accountability system, such as government oversight institutions and watchdog agencies, to provide stronger assurances that re-elected incumbents will perform well and not engage in corruption. Then the probability of corruption with an incumbent may again be equal or lower to that of a first-term challenger. Overall, to the extent voters doubt whether good types in the first term will behave as good types in the second, performance information becomes an uninformative signal, unless it is coupled with credible enforcement of politicians' behavior in office. Re-election can thereby fail to generate political accountability when other accountability institutions are weak.

I test this argument in the context of mayoral elections in Peru. Using electoral returns data from the three most recent elections, I first use a regression discontinuity design (RDD) to show that Peruvian mayors face an extreme incumbency disadvantage. Incumbents who narrowly win in one election—making their win "as-if" random—are 24 percentage points more likely to lose the subsequent election than the challengers they originally barely beat.

It could be the case, however, that voters tend toward rejecting incumbents because they perform poorly, thus suggesting electoral accountability is actually working well. Or incumbents may be disadvantaged because voters lack accurate information about performance, as a large body of electoral accountability literature argues (see e.g., Dunning et al. 2019). With observational data, simultaneously testing for these two possibilities is problematic due to voters' uneven access to accurate performance metrics in the real-world and because available data may not cover aspects of performance voters value, such as providing personalistic benefits. Instead, to better test for whether performance or information explain voter behavior in re-elections, I embed a candidate choice conjoint experiment within an original household survey. Respondents are presented with two hypothetical candidates, an incumbent and a challenger, and are asked which candidate they would vote for in a mayoral election. They are told explicitly about the incumbent's performance, with five different possible information conditions: no performance information; good (bad) perfor-

mance in implementing public works projects; and good (bad) performance in providing individual financial assistance to constituents. The results demonstrate that respondents on average prefer challengers to incumbents, even when those incumbents were described as performing well. As further support of my proposed mechanism, I show that this effect is concentrated among respondents who, on a later question in the survey, express doubt that a high-performing incumbent will repeat her good performance if re-elected.

Next, to test my argument that confidence in oversight institutions helps attenuate the incumbency disadvantage, I modify the above RDD set-up to test for heterogeneous treatment effects of incumbency. In line with the conjoint experiment's findings, results suggest that high-performing incumbents are no more likely to win re-election than poor performing ones, nor are incumbents more likely to win where voters have access to greater political information. However, when incumbents run in districts where voters report higher trust in oversight institutions, they gain an electoral advantage relative to challengers. A number of robustness checks support the validity of the findings.[3] These results suggest that though voters may tend towards rejecting incumbents overall, if they have confidence that the institutional environment will hold politicians accountable in a subsequent term, they are more willing to reward incumbents with re-election. I also test for a series of alternative explanations, including political competition and candidate traits, yet trust in accountability institutions remains the only factor that improves incumbents' likelihood of winning re-election.

These findings have important implications for electoral accountability and beyond. Part of Mayhew's famous "electoral connection" (1974) is that the incumbent's action in the current period is constrained by her desire for re-election in the next and by her assessment about what criteria voters will use to judge whether she should be re-elected. If incumbents recognize that their chances of being rewarded for their effort are low, their incentive to

[3] I show that it is not just that places with greater trust overall are more likely to re-elect. I test for trust in other government levels, institutions and actors, but it is only trust in oversight institutions that impacts incumbent re-election. Nor is higher trust in accountability institutions a proxy for perceptions of better mayoral performance, since the trust variable and various performance measures are uncorrelated.

perform well in office is reduced. Although more research has traditionally focused on voters punishing poor performers, rewarding good performance is just as important for ensuring responsive and high quality governance (Manin, Przeworski and Stokes 1999). Furthermore, persistent disillusionment with politicians, such as the anti-incumbent attitudes shown here, may generate a self-reinforcing pattern whereby voters increasingly assume the worst about incumbents who then increasingly live up to those low expectations. Over time, this could also lead to disillusionment with democracy itself (Svolik 2013).

This paper contributes, first, to the broader literature on voting behavior in developing countries, and second, to the work on electoral accountability. Standard accountability models (e.g. Fearon 1999; Svolik 2013) assume candidate type remains constant across elections. Instead, and building on work by Klašnja (2015; 2016), my research highlights how incumbents' increasing corruption in subsequent terms indicates candidate type conversion, with important implications for whether these models apply across contexts. Klašnja's research also suggests that an incumbency disadvantage stemming from increasing corruption across terms could be overcome if incumbents can convince voters that the quality of their performance is also increasing over time. Here, I demonstrate a different solution to the electoral accountability problem generated by increasing opportunities for corruption: strong oversight institutions.

This paper departs from the recent incumbency disadvantage literature in developing democracies like Brazil (Klašnja and Titiunik 2017), India (Uppal 2009) and Romania (Klašnja 2015). In these studies, the incumbency disadvantage actually stems from incumbents performing poorly and being voted out of office, reinforcing the existence of electoral accountability itself. In contrast, the results presented here depict a new type of case in which incumbents are disadvantaged merely for being incumbents and, importantly, in which even strong performing incumbents are punished.

In this sense, the paper also challenges the explanation for poor accountability that has been predominant in the electoral accountability literature (e.g. Dunning et al., 2019):

lack of accurate performance information. By solving the information constraint in the conjoint experiment through explicitly informing respondents about incumbent performance, I show how electoral accountability may still fail even with perfect information. The results also build on a recent set of papers that, for the first time, studies cases where voters have access to performance information but punish, rather than reward, good performers (Adida et al 2017; Boas, Hidalgo and Torral 2019; Burstzyn 2016; de Kadt & Lieberman 2017). These findings suggest voters perceive a trade-off in politicians' efforts, such that good performance in one policy domain may be punished if voters actually preferred a different policy (Boas, Hidalgo and Torral 2019), or voters may punish programmatic policy if they prefer personalistic benefits (Adida et al 2017; Burstzyn 2016). However, my results suggest respondents punish both programmatic and personalistic performance, and later survey questions confirm respondents value the performance metrics used. This suggests an entirely new type of case of when good performance is punished.

As I elaborate on further in the conclusion, the theory would reasonably extend to a number of other Latin American countries and beyond. The combination of weak oversight, high corruption and poor performance can set the groundwork for voters' belief that even good types will be corrupted if given the chance. Strong parties may mediate this effect, both by giving politicians an incentive to perform and by increasing the likelihood of incumbent re-election through partisan loyalty and the distribution of clientelist goods. But weak party systems, particularly those without the types of non-party political organization that could facilitate clientelistic exchange, lack the organizational infrastructure to get incumbents re-elected in the face of rising voter doubt about second-term performance.

1 Theory

My theory builds on Fearon's (1999) selection model, in which voters assess candidates in order to identify and elect "good types," meaning politicians that will implement policies

a voter wants. In elections in which the incumbent is running for re-election, voters must choose between the incumbent and one or more challengers. For all candidates, voters have a variety of information at their disposal to assess type, from party affiliation to personal traits to platform promises. For incumbents, voters have an extra, albeit noisy, source of information: the incumbent's performance in office. Voters retrospectively assess this performance to make a prospective decision about the incumbent's type, compared to what voters glean from available information about challenger type.

In this framework, a good type politician is one whose likelihood of performing well in office—meaning performing the way the voter wants—is high. When considering whether to re-elect the incumbent, voters must therefore evaluate both past performance in office and the probability that the performance will be repeated. For incumbents to be rewarded, good incumbent performance must act as a credible signal to voters of the likelihood of continued good performance if elected again.

In most settings, current performance in office *is* a strong predictor of future performance, as whatever component of type—inherent goodness, shared policy preferences, capacity to implement, or external incentive to perform—will continue into the next term. Though standard accountability models (e.g., Fearon 1999; Svolik 2013) assume continuity of type across terms, this may not hold if voters perceive a higher probability of engaging in corruption in a second term, even among incumbents who appeared to perform well in the first. Learning on the job may provide second-term politicians with a greater understanding of how to be corrupt without getting caught by navigating the systems designed to detect malfeasance and identifying potential co-conspirators within government agencies and the business community (Coviello and Gagliarducci 2017; Fisman, Schulz, & Vig 2014; Klašnja 2015). At the same time, weak oversight institutions mean the likelihood of punishment is small, even if detected. This means that if the returns to corruption are high enough, and the likelihood of being caught and punished low enough, even good types in the first term could capitalize on their learning and become corrupted in the second.

Evidence from other contexts of high corruption and weak oversight institutions provides empirical support for the notion that corruption may increase across terms. We can study these dynamics by comparing corruption between incumbents who narrowly won re-election and first-term politicians who narrowly won their first election; to the extent the wins are "as-if" random, the two groups of politicians should be indistinguishable except for their status as first-term or re-elected officials. Research suggests incumbents who narrowly won a second term in office accumulated more wealth during that second term than first-term elected officials who narrowly won, both among mayors in Romania (Klašnja 2015) and state-level politicians in India (Fisman, Schulz, & Vig 2014). Similarly, the procurement practices of narrowly-elected second-term Italian mayors differ from narrowly-elected first-term mayors in ways that suggest the incumbents are engaging in corruption through collusion with contractors (Coviello and Gagliarducci 2017).

I argue that increasing corruption in a second term makes voters tend to prefer challengers even against high-performing incumbents. This is because higher corruption in subsequent terms increases the probability that the draw of a candidate from the challenger pool—who by default will be in her first term—will perform better than the good type who will be corrupted in the second.

In settings where voters perceive corruption to be a greater concern among incumbents than first-term officeholders, voters need some assurance of a check on politicians' power in order to believe that an incumbent will perform well if re-elected. This assurance, I argue, can come through trust in the effectiveness of accountability institutions.[4] The judicial system, horizontal accountability institutions, and government organizations given the autonomy to oversee other state agencies—-such as ombudsmen, supreme auditing institutions or special prosecutors (O'Donnell 1998)—all play a key role in overseeing politicians' performance in office. Even in weak institutional environments, monitoring initiatives undertaken by

[4] The intuition is in line with the Cleary and Stokes argument (2006) that the key to a healthy democracy is not trusting individual politicians, but rather skepticism of politicians combined with trust in accountability institutions to keep those politicians in line.

individual oversight agencies can incentivize politicians and bureaucrats to perform well, particularly in terms of reducing corruption (Lagunes 2018; Lagunes 2019; Olken 2007; Di Tella & Schargrodsky 2003). Effective oversight institutions can even work to incentivize mayors to perform in the absence of electoral incentives, such as with term-limited mayors (Ferraz and Finan 2011). Sanctions may be particularly important when norms of good behavior in elected office are weak (Fisman & Miguel 2007).

Overall, to the extent oversight institutions constrain politicians' behavior in office, they increase the likelihood good incumbent performance will be repeated in a second term. Strong oversight can help ensure that good first term performance is a useful informational signal to predict future good performance, thereby increasing the probability high-performing incumbents will perform well in a second term relative to the potential performance of first-term challengers.

Importantly, indicators of actual institutional strength may have an ambiguous effect on voter behavior. On the one hand, if voters see an increase in cases of politicians punished for corruption, they may perceive that the institutions are effective and be more likely to trust an incumbent with a second term. On the other hand, they could interpret the rise in prosecutions as proof that politicians are actually more corrupt than they thought, thereby decreasing their willingness to re-elect. In contrast, voter perceptions of the efficacy of accountability institutions would have an unambiguous impact on behavior.

To assess my theory, I test the following observable predictions.

Prediction 1: Politicians face an incumbency disadvantage, with voters on average preferring challengers to incumbents.

If the incumbency disadvantage is due to voter skepticism about incumbent performance in a second term, then:

Prediction 2a: Voters will on average tend toward preferring challengers even

over incumbents who perform well.

And,

> **Prediction 2b:** Voters will particularly prefer challengers to the extent that they doubt high-performing incumbents will repeat their good performance if re-elected.

Finally, if in such settings, voters turn to oversight institutions for assurance that incumbents will perform well in a second term, then:

> **Prediction 3:** Voters will be more willing to vote for incumbents in re-elections when they have more trust in accountability institutions.

Before turning to the empirical analysis of the Peruvian case, I provide a brief background for the context of the study, including mapping the existing oversight institutions and describing voter perceptions of the likelihood that good incumbent performance will be repeated if re-elected and that corruption is increasing across terms.

2 Accountability and Local Elections in Peru

Peru is divided into 25 regions, 195 provinces and 1647 municipal districts, the latter of which is the focus of this study. A 2002 decentralization law gave significant responsibilities to mayors for managing local affairs and providing public services, from trash clean up and granting business licenses, to building and maintaining roads, irrigation systems, health centers and schools, generating municipal ordinances and overseeing the town's yearly development plan. The law also mandates citizen participation in municipal governance, including through participatory budgeting (McNulty 2011). Districts are largely reliant on central government transfers for their budgets (Loayza, Rigolini and Calvo-Gonzalez 2014).

Elections are held every four years to elect both the district mayor and a group of councilors (*regidores*) who are proportionally allocated seats on the City Council (*Concejo*

Municipal) based on party voteshare. In addition to passing municipal ordinances, councilors are charged with overseeing the mayor's performance and receiving citizen complaints. Voting is mandatory and fines for not voting are enforced, leading to relatively high voting rates (Carpio, Córdova, Larreguy & Weaver 2019). Through the October 2014 election, mayors could be re-elected indefinitely. However, in March 2015, Congress banned immediate re-election for leaders of regions, provinces and districts beginning in the subsequent local elections held in October 2018.

Peru's party system is extremely weak (Levitsky and Cameron 2003; Seawright 2012). The party system breakdown that began in the 1990s ushered in an incredible proliferation of regional, provincial and sometimes even district-level parties, and of quasi-party groupings and alliances. Local political organization is also sufficiently frail that it makes long-term clientelistic exchange difficult (Muñoz 2019).

2.1 Accountability Institutions

I provide a brief overview of the independent agencies, or horizontal accountability institutions, that voters may view as constraining incumbents' behavior. Importantly, these are visible and well-known organizations, based out of the capital, Lima, but with a decentralized network of offices to manage in-person contact and outreach with citizens, including receiving and processing citizen complaints. They receive frequent media attention for their investigative reports and run educational activities like campaigns against vote buying. These institutions may have the formal power to sanction or, like the media, may rely on soft power in the sense of publicizing malfeasance to generate pressure for the formal sanctioning institutions to act (Peruzzotti & Smulovitz 2006).

First, the *Defensoría del Pueblo* (Ombudsmen Office) researches governments' compliance with their various legal requirements. Peru's *Defensoría* is an example of what has become typical of Latin American Ombudsmen offices in terms of going beyond the traditional role of oversight to proactively promote respect for citizens' rights (Pegram 2011).

The *Defensoría* receives and investigates citizen complaints against any level of government or type of government agency, including against mayors. Complaints can be lodged over the phone, online, or in person at one of the 38 offices located throughout the country, including in every region, and in 2014 alone, they received 31,189 complaints (Defensoría del Pueblo 2015). Though they have no formal sanctioning power, they make recommendations to other sanctioning bodies and use the soft power of publicity through disseminating their findings to the media.

Second, Peru's National Elections Board (*Jurado Nacional de Elecciones - JNE)* not only oversees elections, but also takes a proactive role in educating the public about electoral laws and manages the formal complaint system to remove public officials from office.

Third, the *Contraloría General de la República* (Office of the Comptroller General), or the country's independent Supreme Audit Institution, oversees public spending, conducts audits of other government agencies, and can sanction for misconduct as well as send cases for formal prosecution. Like the Ombudsmen, the *Contraloría* receives and investigations citizen complaints about spending irregularities. From 2008 to 2014, the *Contraloría* received 7453 complaints against municipalities, equaling 55% of the total complaints received during that period.[5] Evidence from a field experiment suggests being monitored by the *Contraloría* (albeit in conjunction with a recognized anti-corruption NGO) in small-scale infrastructure projects led municipal governments to spend less per project, a downstream indicator of reduced corruption by eliminating cost inflation (Lagunes 2018).

Finally, in addition to the *Contraloría*, a group of other agencies is tasked with fighting corruption. The Public Prosecutor's Office (*Fiscalía*) has a set of Specialized Prosecutors for Crimes of Corruption of Public Officials and accepts direct complaints from citizens. The Anti-Corruption Prosecutor of the Attorney General (*Procuradaría Pública Especializada en Delitos de Corrupción*) focuses on securing monetary reparations in corruption cases and also manages a citizen complaint system.

[5] Statistics downloaded 25 June 2015, http://www.contraloria.gob.pe/.

2.2 Increasing Corruption in a Second Term

Evidence suggests Peruvians view the opportunity to be corrupt as greater for second-term mayors than for newly elected challengers. In the original survey reported here, respondents were asked if mayors that performed well in the first term would continue to perform well in the second if re-elected. Only 40% said good performance was 'likely' or 'very likely' to be repeated. Those who answered 'unlikely' or 'very unlikely' were then asked 'why'. More than half (53.4%) responded that good performance would not be repeated because re-elected mayors are more likely to be corrupt because they learn how to do so effectively in their first period. Almost a quarter (23.5%) reported that there is not sufficient oversight to ensure re-elected mayors will do a good job in a second term.

In the Congressional debates over the 2015 law prohibiting re-election, lawmakers' justifications for the ban also suggest an assumption of increasing malfeasance in subsequent terms.[6] Every congressperson referenced the fact that the law's aim was to curb corruption. Though the ban was first put forward in 2011, speakers asserted that the reason the bill was finally being debated was the huge corruption scandals involving regional presidents, with the implicated politicians being mentioned by name. Mayors were described as dishonest and corrupt, lacking capacity and a vocation to help their community, and only seeking re-election to continue robbing state coffers. Strong performing mayors were seen as the exception to the rule.

Legislators also referenced the idea that "power corrupts": incumbents may enter politics with good intentions but they quickly see what is possible and realize they want to stay in office indefinitely. There was a prevalent worry about local politicians who wanted to be mayors for life, and the idea that re-election creates *caudillismo* and clientelism. Many explained Peru's high corruption as a result of lack of institutional controls, and their belief that with proper oversight, re-election would not necessarily generate corruption; however,

[6] I read each of the transcripts of the four Congressional debates held over the bill from June 2014 to October 2014. A fifth debate was held May 3, 2015, but it is not available online, and repeated requests to receive it have gone unanswered.

given Peru's generally weak accountability system, curbing corruption required prohibiting re-election.

Peruvians also seem to perceive that re-election has failed to rein in poorly performing mayors, with high public approval for eliminating re-election. One month after the re-election ban passed, polling suggested 67% of Peruvians agreed with the law, and follow-up polls in the years afterward showed these attitudes persisted.[7] Similarly, in the survey presented here, 68.71% of respondents disagreed or strongly disagreed with mayoral re-election. Respondents were asked why they disagreed, and the top answer (46.3%) was that all mayors are bad or corrupt.

In terms of corruption perceptions, in the 2017 LAPOP survey, 77% of Peruvians said more than half of all politicians are corrupt, the 3rd highest rate across the region (Cohen, Lupu & Zechmeister 2017). This perception seems to come from direct experience, as Peru scored 5th out of 27 countries in the percentage of respondents reporting experiencing corruption (29.6%). Though corruption appears rampant, Peruvians are relatively intolerant of it, with only 17.6% of respondents reporting that paying a bribe was justified, slightly lower than the regional average of 20.5%.

3 The Incumbency Disadvantage in Mayoral Politics

To begin, I show that not only are mayors unable to marshal the benefits of office to their advantage, office-holding actually harms their future electoral success (Prediction 1). Following the incumbency advantage literature (Lee 2008), I use a regression discontinuity design to test for the causal effect of incumbency on future electoral outcomes. Simply comparing incumbents' and challengers' voteshare in the next election is not sufficient to isolate the causal effect of being the incumbent, because whatever characteristics ensured their win in the first place would influence their likelihood of winning in the future. Instead, I compare

[7] In both 2016 and 2017, 66% of respondents continued to disagree with mayors having the right to run for re-election. Datum Internacional S.A., *Pulso Perú*, Apri 2015, November 2016, and June 2017.

the electoral success in time t of incumbents who barely won in time *t-1* with the challengers they barely beat. Assuming that their win or loss at time *t-1* is as-if random allows for causal identification of the effect of incumbency on voteshare in the next election.

I use data from Peru's National Elections Board (*Jurado Nacional de Elecciones* - JNE) for the local elections held in 2006, 2010 and 2014.[8] The first local election after the post-Fujimori democratization process took place in 2002, so the 2006 election is the first in which incumbent mayors could have run in the post-2000 era. For each election year (2006, 2010 and 2014), I include the candidates who ran for mayor in the previous election (2002, 2006 and 2010 respectively) and who won or lost within a bandwidth of 9.7 percentage points (the ideal bandwidth following Calonico et al., 2017). In line with critiques of the methodology (Eggers 2017; De Magalhaes 2015), the analysis performed is unconditional on running. Only including incumbents and challengers who chose to run again would bias the results, given both that motivations for running are likely linked with potential outcomes and there is a differential selection into running across challengers and incumbents. To recover unconditional estimates, all candidates from the previous election are included, with those who did not run in the current election being coded as having lost. This yields a dataset of 8948 observations, each of which is a candidate-election year.

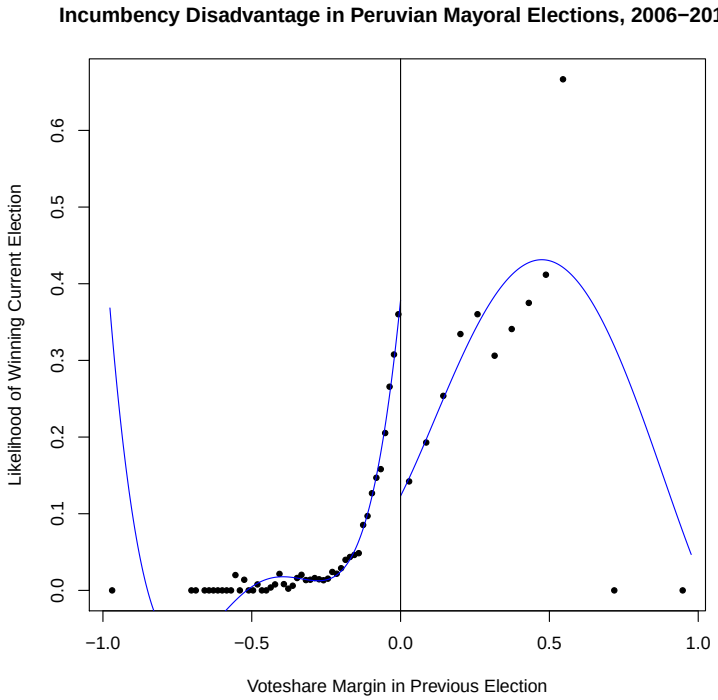
I find a significant incumbency disadvantage: candidates who barely won the previous election (the incumbent) are 24 percentage points less likely to win the subsequent election (and thus, win re-election) compared to candidates who barely lost (the challengers). Figure 1 visually depicts the results. Incumbents, those to the right of 0 in voteshare margin in the previous election, have a lower probability of winning election than the challengers they originally barely beat.

Following standard practice, Appendix Figure A.1 plots the estimates for a series of bandwidths, showing that the negative and significant effect is robust to bandwidth choice. Appendix Figure A.2 plots results of a McCrary (2008) density test confirming no manipu-

[8] Though the raw data is available publicly, José Luis Incio provided me with a cleaned dataset of all mayors who ran for and won re-election for the 2006, 2010 and 2014 elections.

lation in the running variable around the cutpoint (p-value is 0.6185).

Figure 1: RDD Plot



4 Preferring Challengers Despite Good Performance

The RDD results provide compelling evidence that incumbents are systematically disadvantaged compared to challengers. To understand why, I use a conjoint experiment to identify which candidate characteristics matter for voter behavior in re-elections.[9] The conjoint allows for exploring the micro-foundations of the electoral returns results, providing evidence for the individual-level motivations behind aggregate-level voting patterns. And it facilitates simultaneously testing two prominent explanations from the electoral accountability literature: first, incumbents tend to perform poorly; and second, voters lack accurate performance information with which to judge incumbents.

Conjoint analysis, especially as first applied by Carlson (2015) and developed by Hainmueller, Hopkins and Yamamoto (2014), allows for testing multiple candidate characteristics simultaneously, including estimating the relative importance of each characteristic compared to the others. Choosing between two candidates that vary on multiple dimensions also better reflects real-life voting and reduces the risk of social desirability bias, such as for choosing candidates with a history of poor performance in office or that engage in gift-giving as part of their campaign.

I embedded the conjoint experiment in an original household survey fielded among 1061 respondents in 18 urban, peri-urban and rural municipalities in the region of Cusco, located in the Peruvian Andes (Appendix E provides additional details about the survey and sampling). The surveys were implemented in-person using tablets. The conjoint questions were placed at the beginning of the survey to avoid priming effects. Candidate profiles were randomized at the level of the individual respondent. For each of four question items, two candidate profiles were randomly generated, each of which contained one of the possible values of each of the characteristics. Respondents saw the two candidate profiles side-by-side and were then asked: "Which of the following two candidates would you vote for to be mayor of your district?" The process was repeated four times, so the respondent saw four sets

[9] Pre-Analysis Plan filed prior to launching the survey: `http://egap.org/registration/2762`.

of two randomly-generated candidates, and was asked the same question immediately after viewing each pair. Appendix Figure B.1 shows how the two side-by-side profiles appeared to respondents on the tablet.

I tested eight candidate attributes: incumbency; gender; links with social organizations; being from the respondents' village; personal wealth; indigenous first language; political dynasty; and campaign gift giving.[10] Appendix Table C.1 shows the values each attribute could take and Appendix Table D.1 presents the Spanish language used in the survey. In each pair of profiles, one candidate was always the incumbent and the other was always the challenger. The incumbent could take on one of five performance-related conditions: no performance information; good (bad) performance in terms of public works projects; and good (bad) performance in terms of offering residents individual financial assistance when the need arose.

Electoral accountability research tends to assume a priori knowledge by the researcher of the type of performance that influences voter decision-making. Instead, I use a later survey question to directly test the assumption that voters value the two performance metrics I use. Respondents were asked to rank the importance of six different potential mayoral responsibilities. Qualitative interviews conducted in three regions of Peru prior to fielding the survey suggested it was common for mayors to offer residents individual financial help in cases of emergency. And 80.3% of respondents described providing this kind of aid as a very important or important responsibility of mayors. For public works projects, 95.6% described it as very important or important. In a separate question, respondents were asked to rate the performance of their current mayor, and then to provide the one aspect of their performance that mattered most in their assessment. A full 60% of respondents stated public works projects. And while voters may perceive that public works projects involve corruption, they still perceive implementing them to be a key role that the mayor should fulfill.

[10] As the focus of this paper is incumbency, I do not elaborate here on the selection criteria for including these particular characteristics; rather, they can be thought of as controls that allow for better identifying the impact of incumbent performance on vote choice.

The main outcome of interest is the respondents' choice between two candidates when prompted by the question, "Which of the following two candidates for mayor would you vote for?" The outcome variable corresponding to each candidate is whether the candidate was chosen by the respondent. The dataset includes 7576 unique candidate profiles.

4.1 Results

Figure 2 depicts the main results. The points on the plot show the point estimates of the effect of each candidate attribute on the probability of selection in the two-candidate choice set-up.[11] The interpretation of the estimate is the change in the probability that a profile will be selected when the trait is at the given level as compared to the baseline level. The plot depicts each trait's baseline level as a line with no point estimate. The lines on each point represent 95% confidence intervals, and the dashed vertical line shows zero. Appendix Table F.3 presents point estimates and p-values for each trait.

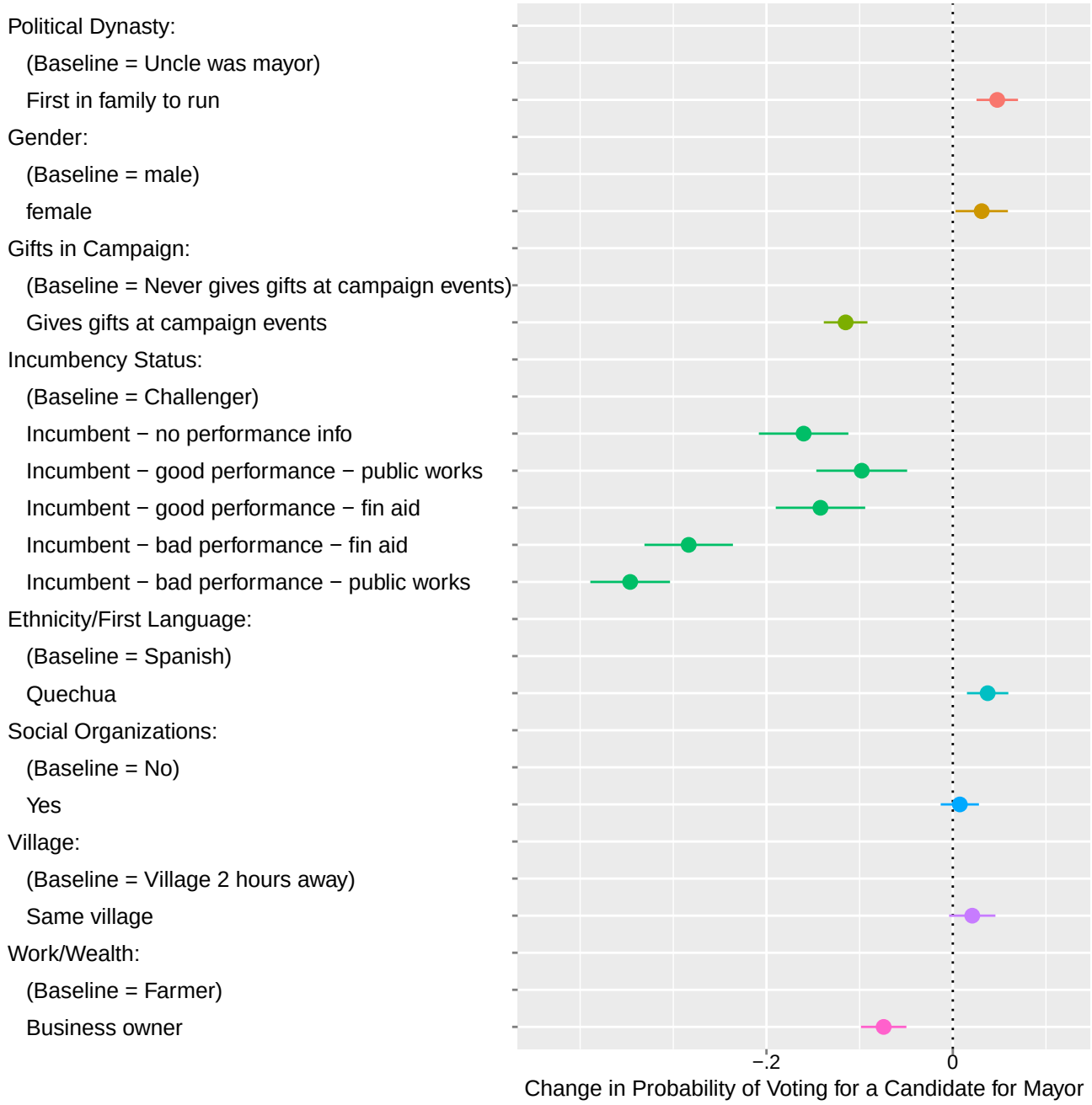
The most noteworthy result is how respondents perceive incumbents. For the incumbency conditions, each point estimate represents the effect of that particular condition compared to being a challenger (which was set as the baseline condition). For example, the effect of the candidate being an incumbent with no performance information compared to being a challenger is that it reduces the likelihood of being selected by .16.

If electoral accountability was working as assumed, respondents would need to both punish poorly performing incumbents *and* reward high-performing incumbents by choosing them over challengers.[12] Instead, the fact that all of the incumbency conditions have negative and significant effects suggests that regardless of the incumbent's performance, challengers are preferred. Most significantly, and in support of Prediction 2a, challengers are preferred even to incumbents that did a *good* job as mayor, and this holds for both of the performance

[11] Following the terminology of Hainmueller et al. (2014), the quantity is the average marginal component effect (AMCE), which represents the average effect of a given candidate trait, in the sense that it is the average of the different effect sizes of that candidate trait given the values of each of the other candidate traits.

[12] This is the expectation recorded in the Pre-Analysis Plan that was registered prior to the survey being fielded.

Figure 2: Effect Sizes of Candidate Traits on Probability of Receiving Respondents' Vote



The dependent variable is whether the candidate profile was selected over the other candidate profile seen in the pair, after being prompted by the question "Which of these candidates would you vote for to be mayor of your municipality?". Clustered standard errors calculated at the level of the individual to account for correlation between the same individual's candidate choices. Bars show 95% confidence intervals. Baseline values have no point estimates or confidence intervals.

measures.

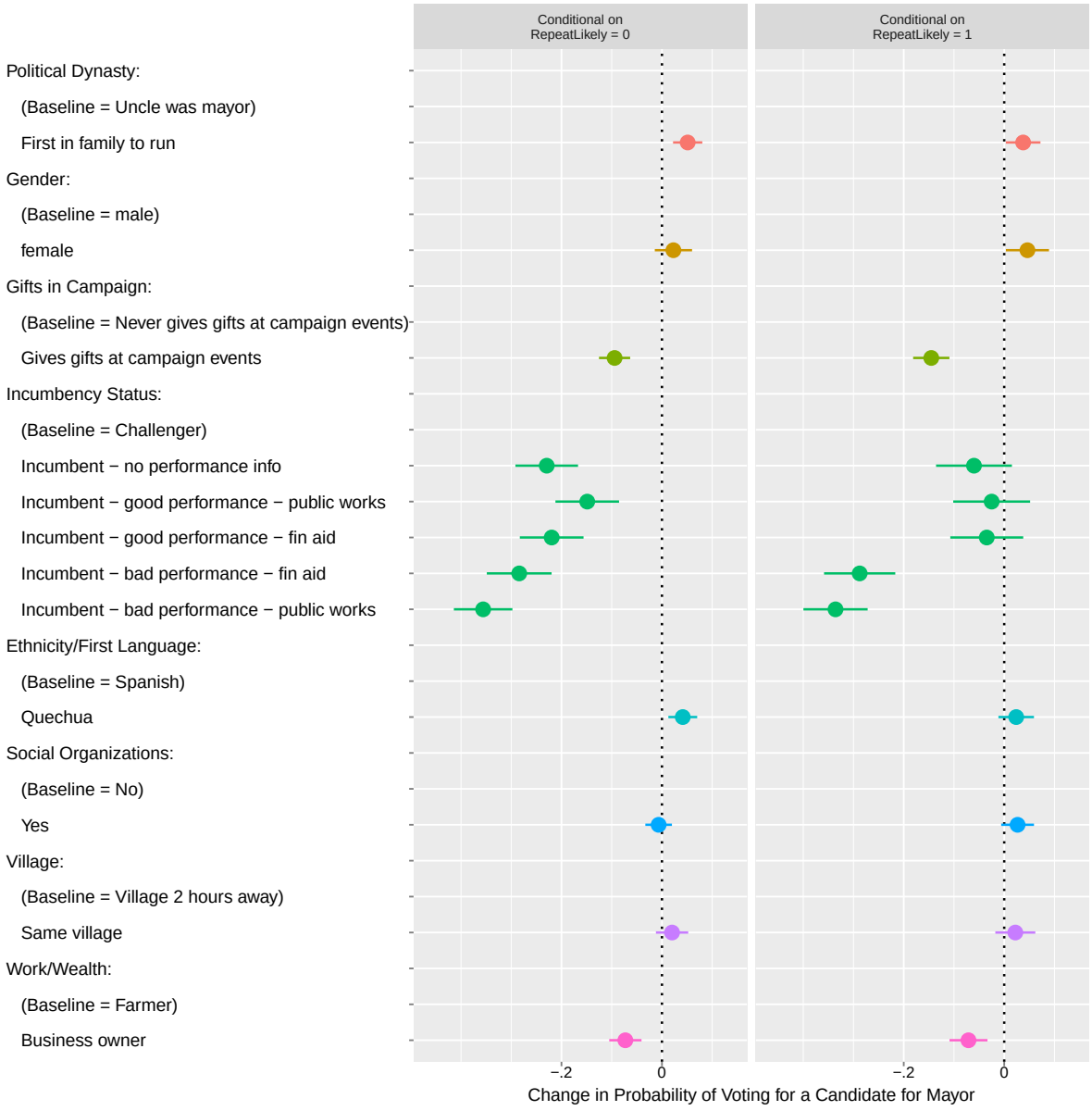
Furthermore, respondents give greater weight to incumbency status than to the other candidate characteristics, as all of the incumbency traits—with the exception of good public works performance compared to giving gifts—have point estimates that are greater in absolute value than any other trait. In other words, not only do voters prefer challengers on average and prefer them even over high-performing incumbents, incumbency status is the most important determinant of vote choice.

The results are robust to a series of checks standard for conjoint experiments, as described further in Appendix G. To discount carryover effects, Appendix Figure G.1 shows the estimates are robust across the four different choice tasks, meaning respondent preferences are the same regardless of whether it was the first or subsequent pair of candidate profiles seen. To reject profile order effects, Appendix Figure G.2 shows the stability of estimates regardless of if the candidate with that trait was the first or second in the pair. Finally, I confirm that the randomization across candidate traits (Appendix Table G.1) and across respondent characteristics (Appendix Table G.2) were both successful.

4.2 Mechanisms

I argue that the electoral connection can break down if voters doubt that good performing incumbents will repeat their performance if given a second chance in office. To test this mechanism, I use a later survey question to assess whether beliefs about repeat performance are driving the bias against high-performing incumbents. Respondents were asked to rate the likelihood that an incumbent who performed well in her first term would repeat her performance if re-elected. I divided the sample by beliefs about repeat performance and then re-ran the conjoint analysis. Figure 3 shows the results. Good performance is only punished among those who believe that it is unlikely to be repeated (left panel). Those who perceive a high probability of incumbents' good performance being repeated if re-elected (right panel) are indifferent between challengers and high performers; while they are not

Figure 3: Conjoint Results Conditional on Respondents' Reported Likelihood that Good Performance Will Be Repeated



The left panel shows the analysis conducted only on the sub-sample of respondents who believe that good incumbent performance being repeated if re-elected is 'unlikely' or 'very unlikely', while the right panel shows the analysis for respondents who believe it is 'likely' or 'very likely'. The dependent variable is whether the candidate profile was selected over the other candidate profile seen in the pair, after being prompted by the question "Which of these candidates would you vote for to be mayor of your municipality?". Clustered standard errors calculated at the level of the individual to account for correlation between the same individual's candidate choices. Bars show 95% confidence intervals. Baseline values have no point estimates or confidence intervals.

rewarding good performance, they are not punishing it either. This suggests that while all respondents have some preference for challengers over incumbents, those who question whether good performance will be repeated are driving the results in which strong performers are punished.

As further suggestive evidence, respondents preferred candidates who were described as being the first in their family to run. While intended to be a measure of political dynasty, this result could also be interpreted as voters' preference for candidates who have never held office, which is consistent with a perception that second- or multi-term mayors are more likely to be corrupt.

5 Mediating the Incumbency Bias

Results from the conjoint experiment suggest that neither incumbent poor performance nor voters' lack of performance information explains the incumbency disadvantage. However, some incumbents are getting re-elected, so to understand how, I alter the RDD set-up, using a heterogeneous treatment effects design to test for what "types" of incumbents are able to overcome voters' anti-incumbent bias. In other words, I ask whether certain characteristics of incumbents—either of the districts in which they run, their performance or their individual personal traits—attenuate the incumbency disadvantage.

I combine the RDD dataset with data on voter trust in politicians and accountability institutions, measures of incumbent performance, candidate characteristics, and variables capturing the socioeconomic, political and institutional context at the district level. I perform the following local linear regression, including the triple interaction between the treatment variable (incumbent), the forcing variable (voteshare margin in the previous election)

and the relevant covariate (X), as well as all lower order interactions:

$$y_d = \alpha + \beta_1 incumbent_d + \beta_2 forcing_d + \beta_3 X + \beta_4 incumbent_d \cdot forcing_d + \beta_5 incumbent_d \cdot X + \beta_6 forcing_d \cdot X + \beta_7 incumbent_d \cdot forcing_d \cdot X + \epsilon_d$$

The coefficient of interest is β_5, the coefficient for the interaction term between incumbency and the relevant covariate. If significant, it would suggest heterogeneity in the impact of being an incumbent in terms of that covariate. For example, if the coefficient on the interaction term 'Incumbent X Performance' was significant and positive, it would suggest that while incumbents are disadvantaged on average, those who perform well enjoy a higher probability of getting re-elected than those who perform poorly. All of the interacted covariates have been standardized (mean of zero and standard deviation of 1) to facilitate comparison.

5.1 Does Performance or Information Explain Incumbency Bias?

I begin by using the RDD set-up to provide a further test of the conjoint results that neither good performance nor access to performance information allow incumbents to overcome voter bias and be re-elected. In both cases, results support the conjoint experiment's findings. This section briefly describes these tests, with full details available in Appendix H.

I use four measures of performance to test whether incumbents enjoy an electoral benefit from performing well in office. First, I include whether Peru's conditional cash transfer program (JUNTOS) was added during the mayor's term. Second, I test for performance in terms of execution of the budget for public works projects. Third, I include the change in the district-level Human Development Index (HDI) over the mayor's term. Finally, I use public opinion survey data on how respondents evaluate the mayor's management of municipal governance, available from a yearly national household survey (ENAHO - *Encuesta Nacional de Hogares*) conducted by Peru's National Institute for Statistics and Information.

Table 1 depicts the results. Confirming the conjoint experiment's findings, the four measures of mayoral performance are not significant, suggesting that even incumbents who performed well in terms of these metrics still face an incumbency disadvantage, with the challengers they barely beat being preferred. Of course, this relationship holds only for these particular performance measures, and it could be the case that voters do reward for other aspects of performance. Still, the results broadly support the voter preferences expressed in the conjoint experiment, and thus provide a real-world test of the conjoint's validity.

Table 1: RDD and Incumbent Re-election: Impact of Incumbent Performance

	DV: Candidate Won Election			
	(1)	(2)	(3)	(4)
Incumbent	−0.278***	−0.288***	−0.248***	−0.357***
	(0.022)	(0.024)	(0.019)	(0.049)
Voteshare Margin Previous	3.113***	2.984***	2.898***	3.380***
	(0.235)	(0.261)	(0.196)	(0.576)
Incumbent X Voteshare Margin Previous	−2.591***	−2.448***	−2.196***	−2.593***
	(0.313)	(0.337)	(0.270)	(0.655)
Incumbent X CCT Added	−0.033			
	(0.021)			
Incumbent X Publ Works Budget Spent		−0.018		
		(0.025)		
Incumbent X Change HDI			−0.022	
			(0.020)	
Incumbent X Rate Municipal Performance				0.026
				(0.052)
Constant	0.386***	0.407***	0.371***	0.462***
	(0.015)	(0.016)	(0.013)	(0.035)
Observations	6422	5548	8915	1371

Note: p<0.1; **p<0.05; ***p<0.01. District-level clustered standard errors reported. Non-interactions and triple interactions 'Incumbent X Margin Previous X Variable' estimated but not reported.

Next, to test the hypothesis that voters would re-elect incumbents if they had access to accurate performance information, I use district-level data from the 2007 census, creating an average across three measures of owning information-related technology (cell phones, television and internet). I also include two ENAHO survey questions, asking how often respondents inform themselves about politics and how much interest they have in politics.

As Table 2 shows, none of the variables associated with having greater access to po-

litical information were significant, suggesting that an information constraint—not knowing how the incumbent performed—does not explain why voters prefer challengers. Given that respondents had perfect information in the conjoint and still on average preferred challengers, these results again provide strong support for the validity of the conjoint experiment's findings.

Table 2: RDD and Incumbent Re-election: Impact of Political Information

	DV: Candidate Won Election		
	(1)	(2)	(3)
Incumbent	−0.247***	−0.266***	−0.265***
	(0.019)	(0.029)	(0.029)
Voteshare Margin Previous	2.907***	2.865***	2.873***
	(0.196)	(0.309)	(0.308)
Incumbent X Voteshare Margin Previous	−2.262***	−1.780***	−1.810***
	(0.264)	(0.420)	(0.419)
Incumbent X Communication Technology Index	0.019		
	(0.020)		
Incumbent X Politically Informed		0.034	
		(0.030)	
Incumbent X Political Interest			−0.011
			(0.027)
Constant	0.372***	0.378***	0.379***
	(0.013)	(0.020)	(0.020)
Observations	8904	3892	3894

Note: p<0.1; **p<0.05; ***p<0.01. District-level clustered standard errors reported. Non-interactions and triple interactions 'Incumbent X Margin Previous X Variable' estimated but not reported.

5.2 The Role of Strong Accountability Institutions

To review, I argue that for electoral accountability to work in practice, voters have to be confident incumbents will continue to perform well in office if given another chance. If voters perceive increasing corruption in a second term, then even seemingly good types in the first term may become bad types in the second. In such contexts, before they re-elect, voters need an extra assurance that incumbent behavior will be constrained despite the increased

draw of corruption, and this assurance can come through belief in the effectiveness of other accountability institutions.

To test whether voters' trust in oversight institutions impacts their willingness to reward incumbents (Prediction 3), I use responses from the ENAHO survey described earlier for a series of questions about trust in different government agencies and social institutions.[13] The survey is implemented on a monthly basis, so I include responses from January-October for 2006 and January-September for 2010 and 2014 (the mayoral election was held in November in 2006 and October in 2010 and 2014). I use only the most immediate time frame prior to the election because I hypothesize that this period would be the most likely to influence voter behavior; earlier lack of trust may already be resolved or have been forgotten by the time the election comes around.

I create a district-level measure of trust in oversight institutions by averaging responses to the trust question for all of the accountability-related institutions that are included in the ENAHO survey. In 2006, 2010 and 2014, the institutions are the National Elections Board; judicial branch; Ombudsmen; Municipal Government; written press; and radio and television. The 2014 survey also asks about trust in four government agencies designed to fight corruption, so I include an average of these responses as well. The institutions are the Anti-Corruption Prosecutor of the Attorney General; Office of the Comptroller General; Public Prosecutor's Office; and the High-level Anti-Corruption Commission.

The results, presented in Column 1 of Table 3, provide strong support for Prediction 3, that when voters report more trust in oversight institutions, they are more likely to vote for incumbents. The trust measures go from lower to higher trust, so a positive coefficient can be interpreted as higher levels of trust being correlated with a greater likelihood of the incumbent being re-elected. The positive and significant coefficient on the interaction between incumbency and trust in oversight institutions suggests that incumbents who run

[13] The question reads 'Currently, do you trust institutions such as ...', listing the institution then giving five options: none, a little, sufficient, a lot, I don't know. The latter answer I coded as missing. The questions are asked identically across years.

Table 3: RDD and Incumbent Re-election: Voter Trust in Institutions

	DV: Candidate Won Election		
	(1)	(2)	(3)
Incumbent	−0.276***	−0.283***	−0.283***
	(0.029)	(0.028)	(0.028)
Voteshare Margin Previous	3.127***	3.126***	3.125***
	(0.305)	(0.297)	(0.297)
Incumbent X Voteshare Margin Previous	−2.205***	−2.057***	−2.062***
	(0.410)	(0.399)	(0.400)
Incumbent X Trust Accountability Institutions	0.063**		
	(0.030)		
Incumbent X Trust Other Politicians		0.011	
		(0.028)	
Incumbent X Trust Police Army			0.027
			(0.029)
Constant	0.390***	0.393***	0.393***
	(0.020)	(0.019)	(0.019)
Observations	3934	4139	4139

Note: p<0.1; **p<0.05; ***p<0.01. District-level clustered standard errors reported. Non-interactions and triple interactions 'Incumbent X Margin Previous X Variable' estimated but not reported.

in districts in which respondents reported higher levels of trust in accountability institutions receive an electoral boost relative to challengers.

I conduct a series of checks on the validity of the trust measure and the results. First, I test two measures of trust in institutions that are unrelated to accountability. One is trust in other politicians and levels of government, which is a district average of reported trust in the regional government, provincial government, Congress and political parties. The other is trust in the army and police. As Columns 2 and 3 of Table 3 show, neither of these measures is significant, suggesting that it is not the case that places with greater trust overall are more likely to re-elect; rather, it is only trust in the specific institutions of oversight that matters for incumbent re-election.

Second, it is also not the case that higher trust in accountability institutions is merely a proxy for perceptions of better mayoral performance. The trust in oversight institutions variable is not correlated with any of the four performance measures.[14] In addition, recall that respondents were specifically asked about their satisfaction with the municipal government's public management, which unlike trust in municipal or other institutions, *is* a direct measure of the incumbent mayor's performance. However, as was shown in Column 4 of Table 1, this variable is insignificant, while trust in oversight institutions is significant and positive. This suggests both that the trust question is distinct from voters' assessment of performance, and that what impacts incumbents' electoral fortunes is not performance, but trust in oversight institutions.

5.3 Other Alternative Explanations

In addition to incumbent performance and voters' access to performance information, I also test for and reject a series of other alternative explanations. I briefly describe the tests and results here, and present full details in Appendix H. I begin by examining whether particular

[14] The correlations are, respectively: Average spending of public works budget (.073); CCT added to district (-.006); Change in HDI (.095); and rating of municipal performance in 2014 (.116).

political features of the municipality might impact incumbents' electoral success. First, since more competitive districts may be simply harder for incumbents to win, I include the effective number of parties from the previous election. Second, given that elections are won with a plurality and no run-off, I use precinct-level returns to gauge whether a candidate had wide support across villages or relied on securing high support in a small set of villages. Third, Peruvian law allows voters to petition to put the standing mayor up for a recall. Following Holland and Incio (2018), I include a binary measure for whether a petition was made to begin gathering signatures and whether enough signatures were gathered to hold a recall in both the current and the previous term.[15] As Appendix Table H.1 shows, none of these variables are significant, suggesting they do not impact incumbents' likelihood of winning re-election.

Finally, particular characteristics of the incumbent, say age or past experience, could impact whether voters are willing to re-elect. I test the seven candidate characteristics included in the information candidates must submit to the National Elections Board, available for the 2014 electoral outcomes. As Appendix Table H.2 shows, none of the candidate characteristics is significant, suggesting that the incumbency disadvantage is not mediated by particular traits of the incumbent herself.

6 Conclusion

In the broadest sense, this paper aimed to assess the conditions under which re-elections succeed in generating political accountability. It begins from the premise that, at a minimum, electoral accountability requires that voters are willing to re-elect high-performing incumbents, thereby incentivizing politicians' good behavior in office. However, leveraging various empirical strategies, I find that voters are consistently biased against incumbents and, importantly, good performance seems unable to help incumbents overcome their disadvantage and get re-elected.

[15] Again, the data was generously provided by José Incio.

I present evidence in support of my theoretical argument for why. In contexts in which second-term politicians are assumed to be more corrupt than first, voters question whether good incumbent performance in one period is an accurate predictor of good performance in the second. But where non-electoral accountability institutions are perceived to be strong, they provide an assurance of continued good performance, making voters willing to re-elect.

These results have important implications for the study of political accountability and democracy more broadly. If we assume that re-election has some role to play in generating accountability, then these findings should cause alarm. They also highlight the importance of understanding how different types of accountability operate not in isolation, but how they interact. In particular, re-election may fail to generate political accountability when other complementary mechanisms of accountability are weak. This points to the need to have a minimum set of accountability institutions in place for any one individual institution to work well.

Furthermore, the results found in the Peru case may be expected to travel elsewhere. Weak states and poor incumbent performance can generate disdain for politicians (Mainwaring 2006), that when combined with a perception of widespread corruption, lays the groundwork for the view that even good types can be corrupted in subsequent terms. With no parties or strong clientelistic machines to deliver votes to incumbents, a perception of increasing malfeasance across terms can translate into an incumbency disadvantage.

Taking just the case of Latin American countries, Guatemala, Ecuador and Panama join Peru in having both weak states and frail party systems. These countries also have relatively low re-election rates compared to other countries in the region. In Peru, incumbent re-election was about 33% across the 2006, 2010 and 2014 elections, and in Panama in 2019, it was 32% (Jiménez 2019). In Ecuador, mayors were re-elected 32% of the time in 2009, but the rate fell even further to only 22% in 2014 (Mejia Acosta & Meneses 2019). In Guatemala, incumbent re-election is the highest in the group at 41.2%, but evidence using a regression discontinuity design suggests an incumbency disadvantage (Morales 2014) similar

in magnitude to that of Peru.

Overall, then, other developing democracies with similarly weak institutions and frail party systems may fall into an electoral disconnection disequilibrium in which voters reject the very premise of re-election and ignore politicians' good performance when they make voting decisions. In response, politicians have little incentive to excel in office, potentially setting in motion a race to the bottom, with poor candidates increasingly running and even worse performance in office reinforcing voters' initial hunch to throw them all out.

7 References

Mejia Acosta, Andres, and Karla Meneses. 2019. "Who Benefits? Intergovernmental Transfers, Subnational Politics and Local Spending in Ecuador." *Regional & Federal Studies* 29(2): 219-247.

Adida, Claire, Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2017. "Breaking the Clientelistic Voting Equilibrium: The Joint Importance of Salience and Coordination." AidDataWorking Paper 48.

Ashworth, Scott. 2012. "Electoral accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15: 183-201.

Boas, Taylor, C., F.Daniel Hidalgo, and Guillermo Torral. 2019. "Accountability Backlash: Negative Electoral Responses to Public Service Provision in Brazil". Unpublished manuscript.

Burstzyn, Leonardo. 2016. "Poverty and the Political Economy of Public Education Spending: Evidence from Brazil." *Journal of the European Economic Association* 14(5): 1101-1128.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "rdrobust: Software for Regression-Discontinuity Designs." The Stata Journal 17(2): 372-404.
Carlson, Elizabeth. 2015. "Ethnic Voting and Accountability in Africa: A Choice Experiment in Uganda." *World Politics* 67(2): 353-385.

Carpio, Miguel Ángel, Beatriz Córdova, Horacio Larreguy, and Julie Anne Weaver. 2019. "Understanding the General Equilibrium Effects of Compulsory Voting on Policy: Evidence from Peru." Unpublished manuscript.

Cleary, Mathew R., and Susan C. Stokes. 2006. *Democracy and the Culture of Skepticism: Political Trust in Argentina and Mexico.* New York: Russell Sage Foundation.

Cohen, Mollie J., Noam Lupu, and Elizabeth J. Zechmeister. (eds). 2017. *The Political Culture of Democracy in the Americas, 2016/17: A Comparative Study of Democracy and Governance.* LAPOP, Vanderbilt University.

Coviello, Decio, and Stefano Gagliarducci. 2017. "Tenure in Office and Public Procurement." *American Economic Journal: Economic Policy* 9(3): 59-105.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis, eds. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I.* Cambridge University Press.

Eggers, Andrew C. 2017. "Quality-based Explanations of Incumbency Effects." *The Journal*

of Politics 79(4): 1315-1328.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types Versus Sanctioning Poor Performance." In Adam Przeworski, Susan C. Stokes and Bernard Manin, eds., *Democracy, Accountability, and Representation.* (Vol. 2). Cambridge University Press.

Ferraz, C., & Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review* 101(4): 1274-1311.

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets." *Journal of Political Economy* 115(6): 1020-1048.

Fisman, Raymond, Florian Schulz and Vikrant Vig. 2014. "The Private Returns to Public Office." *Journal of Political Economy* 122(4): 806-862.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22:01, 1-30.

Holland, Alisha C. and José Incio. 2018. "Imperfect Recall: The Politics of Subnational Office Removals." *Comparative Political Studies* 1-29.

De Janvry, A., Frederico Finan and Elisabeth Sadoulet. 2012. "Local Electoral Incentives and Decentralized Program Performance." *Review of Economics and Statistics* 94(3): 672-685.

Jiménez, Thaylin. 2019. "La Mayoría de Alcaldes Electos Son Nuevos; 11 Lograron la Reelección." *Metro Libre*. `https://metrolibre.com/actualidad-y-pol%C3%ADtica/nacionales/167621-la-mayor%C3%ADa-de-alcaldes-electos-son-nuevos-11-lograron-la-reeleccin.html` (accessed June 19, 2019).

Johannessen, Peter G. 2018. "Visibility and Local Electoral Accountability." Unpublished manuscript.

de Kadt, Daniel, and Evan S. Lieberman. 2017. "Nuanced Accountability: Voter Responses to Service Delivery in Southern Africa." *British Journal of Political Science* 1-31.

Klašnja, Marko, and Rocio Titiunik. 2017. "The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability." *American Political Science Review* 111(1): 129-148.

Klašnja, Marko. 2015. "Corruption and the Incumbency Disadvantage: Theory and Evidence." *The Journal of Politics* 77(4): 928-942.

Klašnja, Marko. 2016. "Increasing Rents and Incumbency Disadvantage." *Journal of Theoretical Politics* 28(2): 225-265.

Lagunes, Paul. 2018. *Guardians of Accountability: A Field Experiment on Corruption and Inefficiency in Local Public Works.* Unpublished manuscript.

Lagunes, Paul. 2019. *The Eye and the Whip.* Book manuscript.

Lee, David S. 2008. Randomized Experiments from Non-random Selection in U.S. House Elections. *Journal of Econometrics* 142(2): 675-697.

Levitsky, Steven, and Maxwell A. Cameron. 2003. "Democracy Without Parties? Political Parties and Regime Change in Fujimori's Peru." *Latin American Politics and Society* 45(3): 1-33.

Linos, Elizabeth. 2013. "Do Conditional Cash Transfer Programs Shift Votes? Evidence From the Honduran PRAF." *Electoral studies* 32(4): 864-874.

Loayza, Norman V., Jamele Rigolini, and Oscar Calvo-González. 2014. "More Than You Can Handle: Decentralization and Spending Ability of Peruvian Municipalities." *Economics & Politics* 26(1): 56-78.

Mainwaring, Scott. 2006. "The Crisis of Representation in the Andes." *Journal of Democracy* 17(3): 13-27.

Manin, Bernard, Adam Przeworski, and Susan C. Stokes. 1999. "Elections and Representation." In Adam Przeworski, Susan C. Stokes and Bernard Manin, eds., *Democracy, Accountability, and Representation.* (Vol. 2). Cambridge University Press.

Mayhew, David R. 1974. *Congress: The Electoral Connection.* Yale University Press.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Re-gression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2): 698-714.

McNulty, Stephanie. 2011. *Voice and Vote: Decentralization and Participation in Post-Fujimori Peru.* Stanford University Press.

De Magalhaes, Leandro. 2015. "Incumbency Effects in a Comparative Perspective: Evidence from Brazilian Mayoral Elections." *Political Analysis* 23(1): 113-126.

Morales Carrera, Ivan. 2014. "Efecto Incumbente en Elecciones Municipales: Un Análisis de Regresión Discontinua para Guatemala." *Revista de Análisis Económico* 29(2): 113-150.

Muñoz, Paula. 2019. *Buying Audiences: Clientelism and Electoral Campaigns When Parties are Weak.* Cambridge: Cambridge University Press.

de la O, Ana L. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science* 57(1): 1-14.

O'Donnell, Guillermo A. 1998. "Horizontal Accountability in New Democracies." *Journal of Democracy* 9(3): 112-126.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy* 115(2).

Pegram, Thomas. 2011. "Weak Institutions, Rights Claims and Pathways to Compliance: The Transformative Role of the Peruvian Human Rights Ombudsman." *Oxford Development Studies* 39(02): 229-251.

Peruzzotti, Enrique, and Carolina Smulovitz. 2006. "Social Accountability: An Introduction." In Enrique Peruzzotti & Carolina Smulovitz (Eds.), *Enforcing the Rule of Law: Social Accountability in the New Latin American Democracies.* Pittsburgh, PA: University of Pittsburgh Press.

Seawright, Jason. 2012. *Party-system Collapse: The Roots of Crisis in Peru and Venezuela.* Stanford: Stanford University Press.

Svolik, Milan W. 2013. "Learning to Love Democracy: Electoral Accountability and the Success of Democracy." *American Journal of Political Science* 57(3): 685-702.

Di Tella, R., & Schargrodsky, E. (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. The Journal of Law and Economics, 46(1), 269-292.

Uppal, Yogesh. 2009. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Public Choice* 138(1): 9-27.

Appendices

Appendix A RDD Robustness Checks

Figure A.1: RDD Bandwidth Sensitivity Check

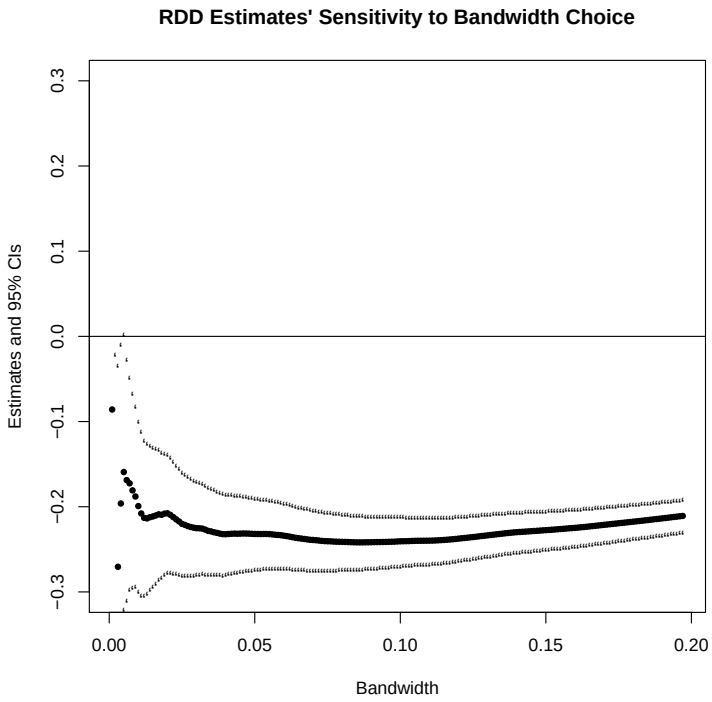
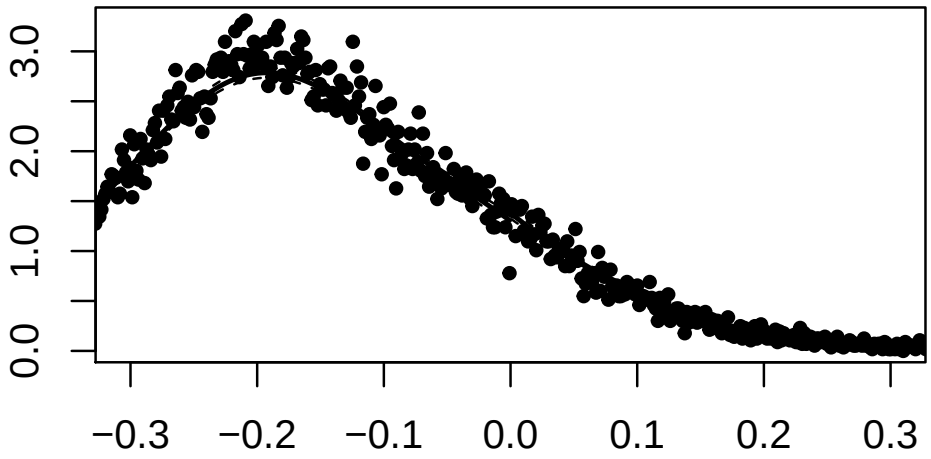


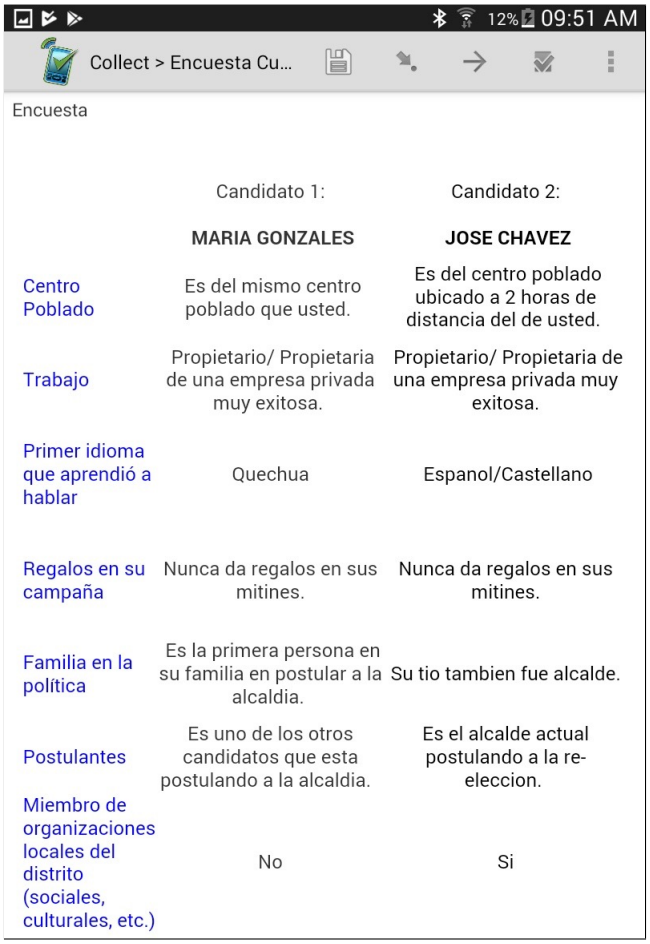
Figure A.2: McCrary Test



Plot of results of the McCrary (2008) density test of discontinuities confirming there is no manipulation in the running variable around the cutpoint. The discontinuity is 0.0163 and the p-value is 0.6185. Note that since many candidates run in each municipal election, there are more losers (observations with voteshare margin below zero) than there are winners (observations with voteshare margin above zero).

Appendix B Conjoint As It Appeared To Respondents

Figure B.1: Tablet Image of Two Candidate Comparison



Appendix C Conjoint Experiment: Candidate Attributes

Table C.1: The Eight Candidate Attributes Used in the Conjoint Experiment

Attribute	**Values**	**Language Used in the Profile**
Incumbency Status	Incumbent: no performance information	He/she is the current mayor running for reelection.
	Incumbent: good programmatic performance	He/she is the current mayor running for reelection. As mayor, he/she implemented a lot of public works projects, like building roads and irrigation canals.
	Incumbent: bad programmatic performance	He/she is the current mayor running for reelection. As mayor, he/she did not implement hardly any public works projects, like building roads and irrigation canals.
	Incumbent: good personalistic performance	He/she is the current mayor running for reelection. As mayor, he/she gave individual financial help to residents if they asked for it.
	Incumbent: bad personalistic performance	He/she is the current mayor running for reelection. As mayor, he/she did not give individual financial help to residents if they asked for it.
	Challenger: no performance information	He/she is one of the other candidates running for mayor.
Community ties (1)	Strong links with social organizations	He/she is a member of local social and cultural organizations in the municipality.
	Weak links with social organizations	He/she is not a member of local social and cultural organizations in the municipality.
Community ties (2)	From respondent's *centro poblado* (village)	He/she is from your centro poblado.
	From other *centro poblado* (village)	He/she is from the centro poblado that is 2 hours away from yours.
Personal wealth	Wealthy/access to own resources	He/she is the owner of a very successful private company.
	Modest economic background	He/she works as a farmer.
Ethnicity	Indigenous	The first language that he/she learned at home was Quechua.
	Not indigenous	The first language that he/she learned at home was Spanish.
Political dynasty	Member of a political dynasty	His/her uncle was also previously mayor.
	First in family to run for office	He/she is the first in his/her family to run for mayor.
Campaign gift giving	Gives out gifts	He/she gives out gifts at campaign rallies.
	Does not give out gifts	He/she never gives out gifts at campaign rallies.
Gender	Male	Male name
	Female	Female name

Appendix D Spanish Language Used in Conjoint Experiment

Table D.1: Spanish Language Used in Conjoint Experiment

Candidate Characteristic	**Language Used in the Profile**
Postulantes	Es el alcalde actual postulando a la reelección.
	Es el alcalde actual postulando a la reelección. Como alcalde, ha hecho muchas obras públicas, como por ejemplo calles y canales de riego.
	Es el alcalde actual postulando a la reelección. Como alcalde, no ha hecho casi ninguna obra pública, por ejemplo ninguna calle ni ningn canal de riego.
	Es el alcalde actual postulando a la reelección. Como alcalde, daba ayuda financiera a la población cuando se lo pedían.
	Es el alcalde actual postulando a la reelección. Como alcalde, no daba ayuda financiera a la población cuando se lo pedían.
	Es uno de los candidatos que está postulando contra el alcalde.
Miembro de organizaciones locales del distrito (sociales, culturales, etc.)	Sí
	No
Centro Poblado	Es del centro poblado ubicado a 2 horas de distancia del de usted.
	Es del mismo centro poblado que usted.
Trabajo	Propietario/Propietaria de una empresa privada muy exitosa.
	Trabaja como agricultor.
Primer idioma que aprendió a hablar	Español/Castellano
	Quechua
Familia en la política	Es la primera persona en su familia en postular a la alcaldía.
	Su tío también fue alcalde.
Regalos en su campaña	Da regalos en sus mítines.
	Nunca da regalos en sus mítines.

Appendix E Additional Information on the Survey and Sampling

The survey was fielded August 12-23, 2017. I designed and managed all aspects of the survey and its implementation. I contracted out with the Peru country office of Innovations for Poverty Action to program the survey on the tablets, rental of and technical trouble shooting with the tablets, and use of IPA servers for daily uploading of the data.

Respondents were required to be between 18 and 70 years old, speak Spanish and be residents of the municipality. A daily, per-enumerator gender quota was employed to ensure equal numbers of male and female respondents.

I selected the region of Cusco for the survey because of the wide variety of districts across the different electoral accountability types I wanted to include in the stratification described below ('never run', 'always lose' and recent incumbent re-election). The region is also home to districts that rank both high and low on social accountability engagement, including on indicators of social conflict over the mayor's performance, lodging complaints against the mayor with the Peruvian Ombudsmen, and self-reported participation in accountability-related civil society organizations in a yearly survey from Peru's National Statistics Institute. At the same time, implementing the survey within one region allows me to control for socio-political, historical and geographic factors that likely vary across regions. Finally, I had already conducted qualitative fieldwork in two municipalities in the region, during which I confirmed the presence of the accountability-related phenomenon to be studied in the survey.

The 18 municipalities were randomly sampled, then individual residents were randomly sampled from within them. I reduced the possible sampling frame of municipalities in two ways. First, as in the broader study, I excluded the 13 municipalities that are both provincial and district capitals, because these districts have only one mayor whose responsibilities cover managing both the province and the district in which the provincial capital is located. In these cases, the line of accountability for district-level performance is confused

with that of provincial-level performance. Second, for budgetary reasons, I had to ensure enumerators could visit the selected municipalities and return within the same day, so I reduced the set of municipalities to include only those within an approximately 3-hour bus drive from Cusco city (the regional capital). The 18 municipalities were drawn from this final potential sample of 59 of Cusco's 95 districts.

To ensure variation in respondents across municipalities with distinct variants of electoral accountability, and because these variants are not uniform across the population, I stratified the sample on municipal type, according to four electoral accountability outcomes. The first two are what I term 'never run' and 'always lose' districts, meaning those in which across the three most recent mayoral elections, three different incumbents either always chose not to run, or always ran but always lost; in these districts, the electoral connection has almost certainly broken down. In the third, on the other hand, where incumbents won re-election in the most recent election (2014), it is at least possible electoral accountability is functioning well, with incumbents presumably being rewarded for good performance. The fourth strata of municipalities, referred to below as "other", is composed of all those districts that are not included in the first three types. It thereby represents districts with electoral accountability outcomes that fall somewhere in between the two extremes: incumbents often, though not necessarily always, run for office, and sometimes, though not always, win re-election.

Within each of the municipal types, municipalities were drawn probability proportional to size (PPS), then a uniform number of respondents was sampled within each municipality (about 60 in each of the 18 municipalities). In each of the two districts sampled in urban Cusco, enumerators randomly sampled respondents from two different neighborhoods. In the 16 remaining rural and peri-urban municipalities, in 11 cases, respondents were sampled from the main population center where the municipal government is located, and in 5 cases, enumerators were able to sample from smaller villages in addition to the main town.

Appendix F Additional Information on the Conjoint Experiment

Two additional questions were included as part of the conjoint: 1) "If this candidate were elected, how likely is it that he/she would do a good job as mayor?"; 2) "If this candidate were elected, how likely is it that he/she would respond well to complaints or suggestions from people like you?". To avoid priming effects on the main outcome of interest (the forced choice between two candidates), these two rating questions come after the four choice questions, and thus were only asked for the two candidates from the last choice set. The results are largely identical to the main conjoint question, and therefore are not presented here. The only difference is that the performance condition for good public works is not distinguishable from zero in both of the questions, and the good performance in financial assistance condition is not distinguishable from zero in the responsiveness question (though it is negative and significant in the 'doing a good job as mayor' question). The sample size in both questions is much smaller (1967 observations), making precision difficult.

I implemented two constraints to the randomization of the individual candidate characteristics. First, for the incumbency status characteristic, one of the two candidate profiles was always the challenger, since it would be impossible to have two incumbents running against each other. Second, two female candidates were never presented within the same pair of profiles. Because there are so few female mayoral candidates in Peru, it is unrealistic for two women to run against each other (outside of Lima, at least). Thus there are more male candidates than female (2/3 to 1/3), because there were pairs of candidates who were both male, but no pairs of candidates who were both female.

All other combinations of attributes were possible, meaning all values had an equal probability of being assigned to each candidate. This means that for all other candidate characteristics, the two profiles could contain the same values (for example, they could both be native Quechua speakers or both give gifts at rallies).

Due to limitations in the survey software used and because it was implemented in-person rather than online, it was not possible to completely randomize the presentation order of the candidate characteristics. Instead, each respondent was randomly selected to see the candidate characteristics listed in one of three possible orders.[16] For ease of presentation, a given respondent saw all sets of candidate profiles presented in the same order.

One concern with the conjoint set-up is if respondents are asked to choose between hypothetical candidates with traits they they might not actually see in the political context in which they live. For example, if corruption is viewed as widespread, an incumbent mayor described as not being corrupt might be unimaginable. To address this concern, I included questions in the broader survey to assess the prevalence of two of the candidate traits for which this might be a problem.

First, a respondent may have never seen an incumbent mayor who they believed did a good job in the two performance aspects I test (public works and individual financial assistance). I therefore included the questions "How many of the past mayors in this municipality have implemented public works projects?" and "How many of the past mayors in this municipality have provided residents with individual financial help when they have asked for it for a family or health emergency?"[17]

Table F.1 shows responses to the two questions across the full sample. We would be worried about the hypothetical candidates being unrealistic if all respondents reported that either 'all' or 'none' of the mayors implemented the two performance indicators. For public works, the responses are well distributed, with few in the 'all' or 'none' categories. For financial assistance, few respondents answered 'all' though 19% answered 'none'. Still, these results suggest on the whole that an incumbent who either did or did not provide

[16] To randomly select the three orders, I first generated all possible order permutations of the seven candidate characteristics in which the presentation order could vary (names are always presented at the top, next to candidate number). To ensure the three orders all began with a different candidate characteristic, I first divided the orders into seven groups such that in each group, all orders began with the same characteristic. I randomly sampled one order from each of the seven groups, then randomly sampled three of those orders to arrive at the final set of three possible presentation orders.

[17] The two questions appeared together but the order was randomized, such that about half of respondents saw the public works question first, while the other half saw the financial assistance question first.

public works or financial assistance while in office is a reasonably realistic candidate for most respondents.

The second candidate trait that may be unrealistic is not giving out gifts at campaign rallies, especially considering that results from prior qualitative interviews suggested the practice was widespread. To assess this possibility, I included in the broader survey the following question: "In the last mayoral election in October 2014, how many of the candidates gave out gifts at their campaign events, such as at rallies or meetings?" Again, the distribution of responses suggests that candidates that both offer and do not offer gifts are realistic, as Table F.2 shows.

Table F.1: How Many Mayors in this Municipality...

	All	Majority	Half	Some	None
... implemented public works projects?	7.9%	20.3%	16.7%	46.9%	1.4%
... provided individual financial help?	6.6%	10%	10.2%	39.3%	19%

Note: Rounding to tenth of a percent. Percentages for 'No response' and 'Don't Know' not shown.

Table F.2: In the Last Mayoral Election in October 2014, How Many Candidates...

	All	Majority	Half	Some	None
... gave gifts as part of their campaign?	26.9%	27.9%	9%	19.5%	5.8%

Note: Rounding to tenth of a percent. Percentages for 'No response' and 'Don't Know' not shown.

Table F.3 presents the point estimates and p-values for each of the candidate attributes. The same results are depicted visually in Figure 2 in the main text.

Table F.3: Average Marginal Component Effects (AMCE)

Attribute	**Level**	**AMCE Estimate**
Political Dynasty	*Baseline: Uncle was also previously mayor*	
	First in family to run	0.048 ***
		(0.0114)
Gender	*Baseline: Male*	
	Female	0.031 *
		(0.0143)
Gifts in Campaigns	*Baseline: Does not give gifts*	
	Gives gifts	-0.115 ***
		(0.012)
Incumbent	*Baseline: Challenger*	
	Incumbent: No performance information	-0.16 ***
		(0.0245)
	Incumbent: Good performance - public works	-0.098 ***
		(0.0245)
	Incumbent: Good performance - fin aid	-0.142 ***
		(0.0245)
	Incumbent: Bad performance - fin aid	-0.283 ***
		(0.0242)
	Incumbent: Bad performance - public works	-0.346 ***
		(0.0218)
First Language	*Baseline: Spanish*	
	Quechua	0.037 ***
		(0.0113)
Social Organizations	*Baseline: No*	
	Yes	0.008
		(0.0105)
Village	*Baseline: Village 2 hours away*	
	Same village	0.021
		(0.0127)
Work/Wealth	*Baseline: Farmer*	
	Business owner	-0.074 ***
		(0.0124)
Observations		7576

Notes:*p<0.1; **p<0.05; ***p<0.01

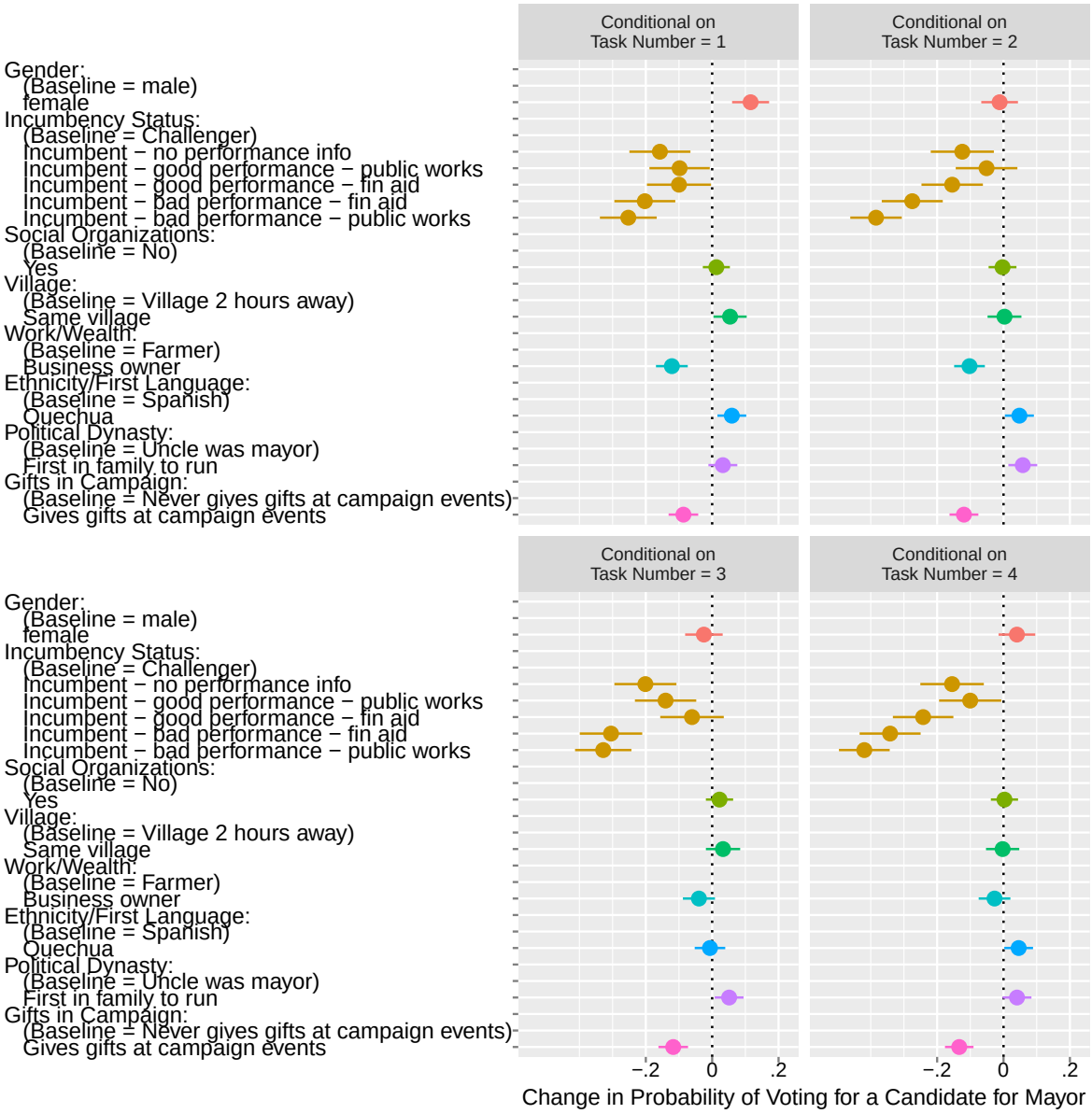
Appendix G Conjoint Experiment Robustness Checks

I conduct a series of robustness checks which are standard practice for conjoint experiments. First, respondents were shown four different pairs of candidates and asked to select their preferred candidate for mayor after each pair (or task). One concern is therefore whether there are carryover effects, meaning respondents' preferences for candidate attributes change across rounds of questions. Figure G.1 shows how the results are robust across the four different tasks. In other words, respondents' bias against incumbents holds regardless of whether it is the first or subsequent pair of candidates they saw.

Second, it might also be the case that respondents systematically prefer the first (or second) candidate profile they see in the pair. Figure G.2 discounts this possibility by showing that the estimates are the same regardless of if the candidate with that particular trait was the first or second candidate profile shown (in other words, if the candidate was on the left or right of the tablet screen).

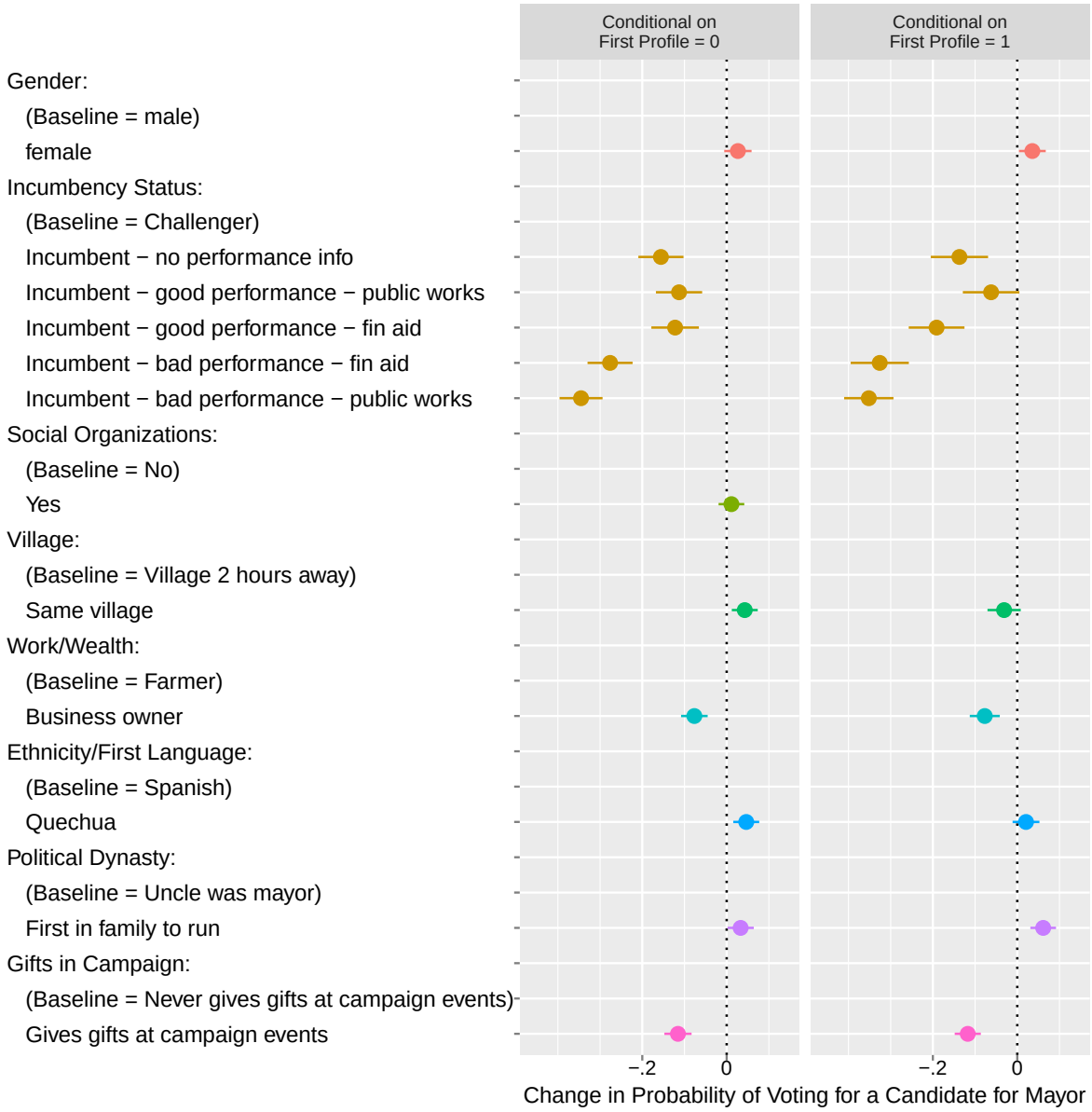
Finally, I conduct two checks to confirm that the randomization was successful. First, Table G.1 shows the frequency distributions of each of the candidate traits across the sample, which match the randomization constraints described in Appendix F. Second, I test whether the candidate traits were randomly viewed across respondents. Results may be skewed if, for example, women were more likely to see profiles of incumbents with good public works performance than were male respondents. Columns 1-5 in Table G.2 show separate regressions of each candidate trait on the five respondent traits. An omnibus F-test is insignificant, suggesting that the candidate traits are jointly insignificant in each of the five respondent trait models.

Figure G.1: Conjoint Results Conditional on Task Number



Each panel shows the analysis conducted only on the sub-sample of candidates shown in the corresponding task number. The dependent variable is whether the candidate profile was selected over the other candidate profile seen in the pair, after being prompted by the question "Which of these candidates would you vote for to be mayor of your municipality?". Clustered standard errors calculated at the level of the individual to account for correlation between the same individual's candidate choices. Bars show 95% confidence intervals. Baseline values have no point estimates or confidence intervals.

Figure G.2: Conjoint Results Conditional on Profile Order



Each panel shows the analysis conducted only on the sub-sample of candidates shown in the first or second profile respectively. The dependent variable is whether the candidate profile was selected over the other candidate profile seen in the pair, after being prompted by the question "Which of these candidates would you vote for to be mayor of your municipality?". Clustered standard errors calculated at the level of the individual to account for correlation between the same individual's candidate choices. Bars show 95% confidence intervals. Baseline values have no point estimates or confidence intervals.

Table G.1: Randomization Check: Frequency of Candidate Attributes

Attribute	**Values**	**Frequency Across Sample**	**Randomization Constraint**
Incumbency Status	Incumbent: no performance information	10.1%	10%
	Incumbent: good programmatic performance	10.2%	10%
	Incumbent: bad programmatic performance	10.1%	10%
	Incumbent: good personalistic performance	9.7%	10%
	Incumbent: bad personalistic performance	9.8%	10%
	Challenger: no performance information	50%	50%
Community ties (1)	Strong links with social organizations	50.1%	50%
	Weak links with social organizations	49.9%	50%
Community ties (2)	From respondent's *centro poblado* (village)	50.6%	50%
	From other *centro poblado* (village)	49.4%	50%
Personal wealth	Wealthy/access to own resources	50.3%	50%
	Modest economic background	49.7%	50%
Ethnicity	Indigenous	49.4%	50%
	Not indigenous	50.6%	50%
Political dynasty	Member of a political dynasty	50.3%	50%
	First in family to run for office	49.7%	50%
Campaign gift giving	Gives out gifts	49.4%	50%
	Does not give out gifts	50.6%	50%
Gender	Male	64.1%	$66.\overline{6}\%$
	Female	35.9%	$33.\overline{3}\%$

Table G.2: Randomization Check: Candidate Attributes and Respondent Characteristics

	Dependent variable: Respondent Characteristic				
	Age	Education	Gender	Income	Maternal Language
	(1)	(2)	(3)	(4)	(5)
Political Dynasty (Uncle was mayor)	0.016	−0.017	0.024**	−0.008	−0.009
	(0.019)	(0.044)	(0.012)	(0.026)	(0.010)
Gender (Male)	−0.008	0.031	−0.003	0.011	−0.003
	(0.020)	(0.046)	(0.012)	(0.027)	(0.010)
Gifts in Campaign (Never)	0.046**	0.047	−0.003	0.031	0.014
	(0.019)	(0.044)	(0.012)	(0.026)	(0.010)
Incumbent: good performance - public works	−0.038	0.026	−0.014	0.057	0.010
	(0.043)	(0.099)	(0.026)	(0.058)	(0.023)
Incumbent: good performance - fin aid	−0.057	−0.083	−0.012	0.015	0.013
	(0.044)	(0.102)	(0.027)	(0.060)	(0.023)
Incumbent: bad performance - fin aid	−0.023	0.069	0.021	0.015	−0.016
	(0.045)	(0.105)	(0.027)	(0.062)	(0.024)
Incumbent: bad performance - public works	−0.089**	−0.024	0.011	−0.016	−0.006
	(0.043)	(0.100)	(0.026)	(0.059)	(0.023)
Incumbent: Challenger	−0.022	−0.023	−0.006	0.020	0.011
	(0.035)	(0.082)	(0.021)	(0.048)	(0.019)
First Language (Quechua)	0.003	0.071	−0.002	0.026	−0.001
	(0.019)	(0.044)	(0.012)	(0.026)	(0.010)
Social Organizations (Yes)	0.040*	−0.044	−0.014	0.010	0.023**
	(0.022)	(0.051)	(0.013)	(0.030)	(0.012)
Village (Same)	0.016	−0.004	0.003	−0.019	−0.022**
	(0.020)	(0.047)	(0.012)	(0.027)	(0.011)
Work (Farmer)	0.019	−0.024	0.006	−0.058**	0.010
	(0.020)	(0.046)	(0.012)	(0.027)	(0.010)
Constant	1.862***	4.909***	0.510***	2.611***	0.737***
	(0.043)	(0.100)	(0.026)	(0.059)	(0.023)
Observations	7,576	7,544	7,576	7,296	7,544
F Statistic	1.418 (df = 12; 7563)	0.656 (df = 12; 7531)	0.707 (df = 12; 7563)	0.782 (df = 12; 7283)	1.121 (df = 12; 7531)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix H Additional Information about Tests for Alternative Explanations

This section provides additional detail about the tests discussed in the main paper to discount the alternative hypotheses of incumbent performance and access to information. Then it describes tests to reject additional alternative hypotheses related to the the political context and candidate characteristics.

I use four different measures to test whether incumbent performance explains the incumbency disadvantage. First, I include whether Peru's conditional cash transfer program (JUNTOS) was added during the mayor's term.[18] Unlike corruption, which is often hidden, this outcome is quite visible; put simply, voters know if they or their neighbor became beneficiaries of JUNTOS. Furthermore, it is highly likely voters value this aspect of performance because of its direct link to their immediate well-being. In contrast, a mayor's success in reducing bureaucratic red tape may not be important enough to impact voting decisions. Finally, it is likely local mayors would receive credit for the program's expansion. As with other CCT programs, JUNTOS' beneficiaries are meant to be selected using objective poverty indicators, and the program is managed from the central, not the local level. But given the possibility of local-level manipulation, or at least the perception of it, the likelihood mayors are trying to claim credit for the program, and the fact that local residents may not know or care what level of government implements it, it seems reasonable to hypothesize that voters would reward the mayor with re-election if the program was expanded in their district. Research from Brazil, Honduras and Mexico, which have similarly structured CCT programs, suggests that incumbent mayors are in fact rewarded electorally for CCT roll-out in their municipalities (de Janvry, Finan and Sadoulet 2012; Linos 2013; de la O 2013).

[18] For each election, I code as 1 every district that was added during the mayor's term prior to the election, and code all other JUNTOS districts as 0. There were 292, 307, and 430 new districts added during the mayoral administrations before the 2006, 2010 and 2014 elections respectively. I treat as the full universe of cases the 1176 districts that had the JUNTOS program in 2016 in order to compare districts that, because of their poverty and eligibility status, *could* have received JUNTOS with those that actually did. All other districts are marked as missing. Data received from the JUNTOS program in March 2016.

Second, I test for performance in terms of execution of the budget for public works projects. Variation in spending of the budget has been linked with mayoral capacity in Peru (Loayza, Rigolini and Calvo-Gonzalez 2014), and once a recall election was initiated, Peruvian incumbents with lower budget execution were more likely to be successfully removed from office through the recall (Holland and Incio 2018). In other contexts, voters have been shown to reward incumbents for spending on visible public works projects (Johannessen 2018). Furthermore, results from the survey presented in this paper show how 95.6% percent of respondents describe providing public works projects as a very important or important responsibility of the mayor. Respondents were also asked to rate the performance of their current mayor, and then to explain what was the most important aspect of their performance that impacted their assessment. A full 62.4% said public works projects.

Third, I include the change in the district-level Human Development Index (HDI) over the mayor's term, with the logic that if voters see their socio-economic well-being improved, they are likely to reward the mayor. The HDI is a composite indicator that combines three key measures of development: life expectancy at birth, years of schooling, and GNI per capita. Based on data from Peru's National Institute for Statistics and Information,[19] the HDI for every district is calculated for 2003, 2007, 2010, 2011 and 2012.[20] The HDI years do not match up perfectly with election years: for the 2006 election, I used the difference between a district's 2003 and 2007 HDI; for 2010, I used the difference between 2007 and 2010; and for the 2014 election, I used the difference between HDI in 2010 and 2012.

Finally, I use public opinion survey data on how respondents evaluate the mayor's management of municipal governance. The data comes from a yearly national household survey (ENAHO - *Encuesta Nacional de Hogares*) that is a random sample of districts and of households within those districts, conducted by Peru's National Institute for Statistics and Information. The survey includes a series of questions asking respondents to rate the

[19] *Instituto Nacional de Estadística e Informática* (INEI), https://www.inei.gob.pe/

[20] http://www.pe.undp.org/content/peru/es/home/library/poverty/Informesobredesarrollohumano2013/IDHPeru2013.html (Downloaded 10 March 2016)

performance of different levels of government (national, regional, provincial and municipal).[21] I create a district average of responses to the question about the municipal level.[22]

Table 1 in the main text depicts the results. None of the four measures of mayoral performance are significant, suggesting that performing well in these four measures does not give incumbents any electoral benefit in terms of overcoming the incumbency disadvantage.

I also test the hypothesis that lack of access to political information prevents voters from effectively sanctioning or rewarding incumbents. I use district-level data from the 2007 census, creating an average across three measures of owning information-related technology (cell phones, television and internet). I also include two ENAHO survey questions, asking how often respondents inform themselves about politics and how much interest they have in politics. As Table 2 in the main text shows, none of the variables associated with having greater access to political information were significant, suggesting that an information constraint—not knowing how the incumbent is performing—is not the explanation for why voters prefer challengers.

Next, I test two explanations related to the political context. Since districts that are more competitive electorally may be simply harder for incumbents to win, I include the effective number of parties from the previous election (time *t-1*).[23] I also gauge whether a candidate's support was widespread across villages or more concentrated, as perhaps successful incumbents are those who cultivate sufficiently strong support in a few key villages. Given that elections are won with with a plurality and no run-off, candidates need not earn support from across the municipality to win. Using precinct level returns, I calculate the variance of the precinct voteshare across all precincts in the district for each candidate from the previous election (time *t-1*).[24] A lower variance would indicate the candidate had re-

[21] The answer options are: Very good; Good; Bad; Very bad; Don't know/Don't answer.

[22] The question is only available beginning in 2011, meaning it can only be tested with data from the 2014 election.

[23] I include data from time *t-1* because candidate-level data is missing for candidates who choose not to run in time *t*. Recall the analysis is unconditional on running, so all candidates from time *t-1* are included, even if they do not run in time *t*.

[24] A special thank you to José Incio who provided me with the effective number of parties measure and the precinct-level elections returns.

ceived consistent levels of voter support across villages. As Appendix Table H.1 shows, none of these variables are significant, suggesting they do not explain when incumbents are able to overcome the incumbency disadvantage.

Third, a unique and potentially relevant feature of Peruvian local politics is the prevalence of recalls, a constitutionally-mandated policy through which citizens can petition to put the standing mayor up for a recall vote. Recalls have become surprisingly common in Peruvian municipalities, though a 2015 law requiring that recalls be held much later in a mayor's term and allowing for the removed mayor to pick her successor has dampened their use significantly (Holland and Incio 2018). To see if districts with high recalls behave differently, and to evaluate if the incumbency bias is driven by municipalities who tend toward holding recalls, following Holland and Incio (2018), I included a binary measure for whether a petition was made to begin gathering signatures and whether enough signatures were gathered to hold a recall in both the current and the previous term.[25] Again, none of these variables is significant, as Appendix Table H.1 shows.

Finally, particular characteristics of the incumbent could impact whether voters are willing to give that incumbent a second chance in office despite an overall preference for challengers. All candidates running for mayor are required to submit an *hoja de vida* to the National Elections Board (*Jurado Nacional de Elecciones* - JNE) which includes information like age, gender, education and past experience. Beginning in the 2010 elections, candidates had to submit their *hoja de vida* in an electronic format that was then uploaded to a JNE portal.[26] Though the candidate data is available for both the 2010 and 2014 elections, the 2010 RDD data only includes candidates who ran in 2006, so I can only use the 2014 RDD data, which includes candidates from 2010. The available characteristics include age, education, whether the candidate has civil or penal charges pending, years residing

[25] The data was generously provided by José Incio.

[26] Data from the *hojas de vida* was originally obtained through webscraping of the *Jurado Nacional de Elecciones*' website, Infogob.com.pe, through assistance from the Harvard Institute for Quantitative Social Science. Additional data was provided directly by the JNE. The datasets were obtained through joint work with Horacio Larreguy and Miguel Ángel Carpio.

in the municipality, whether they have previously held any elected office, and if they have held a leadership position within their party. As Appendix Table H.2 shows, none of these candidate characteristics is significant, suggesting that these particular traits do not impact incumbents' electoral disadvantage.

Table H.1: RDD and Incumbent Re-election: Political Competition and Recalls

	DV: Candidate Won Election				
	(1)	(2)	(3)	(4)	(5)
Incumbent	−0.293***	−0.246***	−0.247***	−0.248***	−0.248***
	(0.024)	(0.019)	(0.019)	(0.020)	(0.019)
Voteshare Margin Previous	2.953***	2.663***	2.890***	2.903***	2.890***
	(0.267)	(0.200)	(0.197)	(0.197)	(0.196)
Incumbent X Voteshare Margin Previous	−2.352***	−1.957***	−2.216***	−2.223***	−2.185***
	(0.344)	(0.289)	(0.270)	(0.270)	(0.271)
Incumbent X Voteshare Variance Previous	−0.003				
	(0.025)				
Incumbent X ENP		0.028			
		(0.020)			
Incumbent X Recall Requested			0.002		
			(0.019)		
Incumbent X Recall Held				0.010	
				(0.017)	
Incumbent X Recall Held Previous					0.020
					(0.019)
Constant	0.407***	0.367***	0.371***	0.372***	0.371***
	(0.016)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	5,441	8,913	8,913	8,913	8,913

Note: p<0.1; **p<0.05; ***p<0.01. District-level clustered standard errors reported. Non-interactions and triple interactions 'Incumbent X Margin Previous X Variable' estimated but not reported.

Table H.2: RDD and Incumbent Re-election: Candidate Characteristics

	DV: Candidate Won Election						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incumbent	−0.248***	−0.301***	−0.296***	−0.311***	−0.299***	−0.301***	−0.297***
	(0.019)	(0.035)	(0.036)	(0.035)	(0.036)	(0.036)	(0.036)
Voteshare Margin Previous	2.885***	2.734***	2.713***	2.924***	2.712***	2.771***	2.749***
	(0.197)	(0.412)	(0.420)	(0.404)	(0.426)	(0.417)	(0.421)
Incumbent X Voteshare Margin Previous	−2.184***	−2.170***	−2.172***	−2.433***	−2.124***	−2.161***	−2.245***
	(0.272)	(0.489)	(0.495)	(0.472)	(0.501)	(0.499)	(0.499)
Incumbent X Female	−0.011						
	(0.014)						
Incumbent X Age		0.049					
		(0.031)					
Incumbent X Post 2ndary Educ			−0.013				
			(0.032)				
Incumbent X Penal Civil Charges				0.025			
				(0.030)			
Incumbent X Years Residing District					−0.024		
					(0.032)		
Incumbent X Previously Elected						0.035	
						(0.031)	
Incumbent X Party Position							0.041
							(0.034)
Constant	0.371***	0.409***	0.407***	0.424***	0.409***	0.409***	0.409***
	(0.013)	(0.025)	(0.025)	(0.024)	(0.026)	(0.025)	(0.025)
Observations	8,919	2,394	2,389	2,643	2,372	2,394	2,394

Note: p<0.1; **p<0.05; ***p<0.01. District-level clustered standard errors reported. Non-interactions and triple interactions 'Incumbent X Margin Previous X Variable' estimated but not reported.