Debates: Voting and Expenditure Responses to Political Communication

Kelly Bidwell | Katherine Casey | Rachel Glennerster

February 2016

Stanford
Center for International
Development

DEBATES:

Voting and Expenditure Responses to Political Communication

Kelly Bidwell	Katherine Casey*	Rachel Glennerster
White House Social and Behavioral Sciences Team	Stanford Graduate School of Business	Abdul Latif Jameel Poverty Action Lab
kellybidwell@gmail.com	kecasey@stanford.edu	rglenner@mit.edu

16 February 2016

Abstract

Candidate debates have a rich history and remain integral to contemporary campaign strategy. There is, however, no evidence that they affect voter behavior. The scarcity of political information in the developing world offers an attractive testing ground. Using experimental variation in Sierra Leone, we find that public debate screenings build political knowledge that changes the way people vote, which triggers a campaign expenditure response by candidates, and fosters accountability pressure that disciplines the subsequent spending of elected officials. We parse the effects of information conveyed about policy versus charisma, and find that both matter. The results show how political communication can trigger a chain of events that begins with voters and ultimately influences policy.

1. Introduction

Debates among candidates for public office have a rich history and offer a unique platform for candidates to communicate. The Lincoln-Douglas senatorial debates of 1858 are a famous early example in the United States. As distinct from other information sources, debates reveal the relative policy positions and competence of rival candidates, cover challengers in an equal fashion to incumbents, and convey comprehensive information ranging from concrete qualifications to more intangible attributes like persuasiveness and charisma. These features have led to some memorable, and highly influential, contests including the first televised presidential debates between Kennedy and Nixon in 1960, and between Mitterand and Giscard in 1974 in France.

Today debates constitute significant campaign events: large numbers of voters watch them; they generate a flurry of media commentary and analysis of candidate performance; and pundits pore over polling data to assess their effects on public opinion (see for example, Shear [2012]). There is, however, no definitive evidence of whether debates have any impact on voter behavior. While the relevant literature is large (see Hellweg, Pfau and Brydon [1992] for review), it relies primarily on cross-sectional analysis of opinion polls with the familiar identification challenges.

In the developing world, debates are less common but arguably no less important. Indeed, the relative scarcity of political information creates scope for the effects of publicizing debates to be more pronounced, persistent, and directly linked to electoral outcomes. Allowing candidates to stand on equal footing and express their views on key policy issues could facilitate the election of more competent individuals. And, by increasing the share of the public that is informed about these pre-election commitments and the resources available to those elected, debates could foster accountability pressure that disciplines the subsequent behavior of the winners in office.

This paper evaluates these claims via a large scale experiment conducted during the 2012 Parliamentary elections in Sierra Leone. The study randomized exposure to debates at three levels—individual, polling center and constituency—to capture the effects of debates on, and the interactions among, voters, candidates and elected officials. We find that debates have strong direct impacts on voters, which trigger indirect effects on candidate campaign expenditure, and ultimately influence the performance of elected politicians.

We first show that debates had substantial impacts on voter behavior. To capture these effects, we worked with an independent media partner to host, film, and disseminate debates in fourteen constituencies. We randomly allocated a "road show" across 224 polling centers that screened

videotapes of the debates in large public gatherings in the five weeks leading up to the election. We find that watching debates led to higher political knowledge, including awareness of candidate qualifications and policy stances; improved alignment between voter policy preferences and those of their selected candidate; and greater voter openness to candidates from all parties. Importantly, the gains in political knowledge translated into changes in votes cast: we document a five percentage point average increase in vote shares for the candidates who performed best during the debates. The effect is significant in both our exit poll data and in the National Electoral Commission's official voting returns. In the context of historical ties between ethnic groups and political parties, candidates who debated well attracted votes from both loyalists and rival ethnic groups, leading to no net impact of debates on the incidence of ethnicity-based voting. Together these results document a high degree of voter responsiveness to information.

Consistent with theory, we next find an endogenous response by participating candidates who increased their campaign expenditure in communities where debate screenings were held. While candidates were not informed of which polling centers were assigned to receive screenings, such large public gatherings in rural areas would be fairly easy to track after they occurred. We find evidence that candidates increased their campaign effort, as measured by the number of gifts, monetary value of gifts and number of in-person visits, in communities where the screenings were held. Increased effort and expenditure is consistent with a "swing" voter investment model if the debates made exposed areas appear more competitive, either by making expected vote margins narrower or more uncertain.[1] Theoretically, the effects on competition should be largest where a trailing candidate outperforms the initial frontrunner during the debate, which is exactly what we see in the data. This indirect impact of the experiment on the political market captures the kind of interactions among agents that would contribute to a general equilibrium effect.

Our third, more speculative, set of results traces the effects of debates all the way to policy, where we find some evidence that participation in debates enhanced the subsequent accountability of elected MPs. To assess these effects, we randomly selected 14 constituencies from a pool of 28 to host debates, and then tracked the performance of all 28 winners over their first year in office. We find evidence that debates had positive impacts on constituency engagement and public spending: treated MPs, for example, held twice as many meetings with their constituents and spent

---

[1] See Lindbeck and Weibull (1987), Dixit and Londregan (1996, 1998), and Bardhan and Mookherjee (2010); and also Eifert, Miguel and Posner (2010) and Casey (2015) for applications of political competition to ethnic politics.

2.5 times as much of their discretionary public funds on development projects (as verified by field audits). We find no evidence for effects on participation in Parliamentary sittings or on consistency in promoting the MP's priority sector. While the small sample at this level makes our conclusions more tentative, the finding that debates could enhance accountability, even in areas where direct electoral pressure is limited, is important and particularly so for newer democracies.

To better understand what drives the initial response of voters to debates, we disentangle the influence of information conveyed about policy stance from candidate persona, and find that both matter. A suite of treatment arms randomized at the individual level isolates the "hard facts" content, covering policy and professional qualifications that could easily be delivered in other formats, from the coverage of candidate charisma and persuasiveness that is specific to debates. Some voters watched brief "get to know you" videos of the candidates speaking informally about themselves and their hobbies, which capture persona but exclude policy. Others listened to a radio summary that articulated all the facts about policy positions and professional experience that arose during the debates, but conveyed nothing about persona. Still others watched the full debate. (All three arms were delivered privately to individuals on a tablet device.) We find evidence that voters update their views of candidates in response to information regarding objective facts and personality, but that only debates move them into better policy alignment with candidates and trigger changes in vote choice. This suggests that while both policy preference and persona matter, the combination delivered by debates is more powerful than either factor in isolation.

We close the paper by considering four additional aspects of voter responsiveness to political communication. First, we document survey priming effects on a narrow set of political knowledge outcomes, consistent with Zwane *et al.* (2011). Reassuringly, all our results hold net of priming effects. Second, we detect an immediate dissipation in knowledge gains in the days following debate screenings, but find no further decay over the weeks between the screening and the election. Third, we find little evidence of treatment effect heterogeneity, save that women acquire somewhat less political knowledge from debates than men. And fourth, we find larger effects of debates in the group screening communities versus individual viewing experiments. The divergence is consistent with social mobilization or common knowledge generation reinforcing the impacts of information, or with voters valuing the campaign response that tracked the group screenings. While smaller in magnitude, the private viewing estimates are otherwise similar, which shows that debate exposure has direct effects on voter behavior net of any social mobilization or campaign effects.

In sum, these experiments speak to the central problem in political economy of whether elections effectively discipline candidates and incumbent office holders. Our paper shows how political communication—specifically via interparty debates—can trigger a chain of events that begins with voters, flows through candidates, and ultimately impacts policy.

The literature on debates in American politics is large, however few studies credibly identify causal effects (Prior [2012]). Those that do produce mixed results: one experiment finds that debates impact voter assessment of candidates (Fridkin *et al.* [2007]); others find no meaningful effects on political attitudes (Wald and Lupfer [1978]) or opinions (Mullainathan, Washington and Azari [2010]); and two explore how the medium of delivery—via television versus radio—affects voter evaluation of candidates (McKinnon, Tedesco and Kaid [1993], Druckman [2003]). Our private viewing experiments extend this literature by evaluating the impacts of debates in an information poor environment, unpacking voter responses to the distinct types of information delivered, and documenting effects on actual votes cast. The scale and intensity of the group screenings offers a new contribution. Interestingly, the screenings generate effects that are similar in magnitude, and yet much more persistent, than those found for one-sided campaign advertising in wealthier countries (see Gerber *et al.* [2011] in the U.S. and Kendall, Nannicini and Trebbi [2015] in Italy). Potential dilution of effect—via diminishing marginal returns to information or drowning out by the deluge of political commentary—is less likely in low information environments like Sierra Leone, where debate effects persisted over several weeks.

Publicizing debates is typically the purview of mass media outlets. Standard models in the media effects literature show how access to politically informative news enhances voter responsiveness to politician quality and effort, which in turn generates stronger incentives for politicians to perform in office (Stromberg 2015). Under this view, the effects of publicizing debates can be interpreted as consequences of increasing the share of the electorate that is politically informed. As such, our results on voter responsiveness are consistent with evidence that media coverage of politics affects party choice (Enikolopov, Petrova and Zhuravskaya [2011]) and voting based on candidate quality (Ferraz and Finan [2008]).[2] Our findings that debate exposure attracts campaign expenditure and induces better MP performance resonate with evidence that more informed electorates elicit greater politician effort (Snyder and Stromberg 2010), attract

---

[2] See also Gentzkow (2006) and Gentzkow, Shapiro and Sinkinson (2011) on turnout, and DellaVigna and Kaplan (2007) on biased coverage and voting.

public funding (Stromberg 2004), suffer less leakage (Reinikka and Svensson 2005), and see their needs better met (Besley and Burgess 2002). In our empirical context, the reach of mass media is particularly constrained (television ownership is below 10% and adult illiteracy near 70%). Radio, however, has emerged as an important source of political information and our implementing partner is an active contributor. The context affords an unusual degree of control over media exposure, and our results preview the role a more developed media might play in poor countries (see also Paluck and Green 2009).

In development economics, testing the efficacy of debates contributes to the literature on information provision in elections (see Pande [2011] for review). Studies like Banerjee *et al.* (2011), among others, show that providing specific information about candidate qualifications can impact voting. Debates are distinctive in that they provide more comprehensive information about candidates, including information about persuasion and charisma, which are plausibly productivity-enhancing traits in legislatures. Moreover, if no sufficient statistic of political competence is available, the generality of debates could be important for three reasons. From a theoretical perspective, comprehensiveness eases concerns that increasing transparency along one dimension will simply reallocate politician effort towards those more observable actions, regardless of their impact on welfare (e.g. Liessem and Gersbach [2003] on multi-tasking, or Cranes-Wrone, Herron and Shotts [2001] and Prat [2005] on pandering). Pragmatically, it makes it harder for politicians to unravel the impact of the intervention: for example, it is easier for them to discredit a scorecard-style information campaign (Humphreys and Weinstein [2012]) than a video of their own statements. And, by covering a range of issues and allowing candidates to make a positive case, debates may be less likely to backfire than single issue interventions, which have been found in some cases to depress turnout (Chong *et al.* [2015]) and increase vote buying (Cruz, Keefer and LaBonne [2015]).

Our approach of working with political candidates in the course of their actual campaigns follows in the tradition of Wantchekon and co-authors.[3] They find that public deliberation between one party's representative and constituents decreases the prevalence of clientelism and increases electoral support for the participating party in Benin and the Philippines. We instead focus on the interaction between rival parties, where the head-to-head debates were designed to reveal information about the relative quality and policy differences between candidates. Information on

---

[3] See Wantchekon (2003), Fujiwara and Wantchekon (2013), and Wantchekon *et al.* (2015).

the complete choice set straightforwardly helps voters identify the candidate associated with the highest utility level (in the tradition of Hotelling [1929]); and matters more if voting exhibits context dependence, where relative comparisons are also relevant (Callander and Wilson [2006]). Our study further relates to papers that identify the indirect effects of voter interventions on politician behavior. Our empirical results are consistent with theoretical predictions in Casey (2015), who argues that information provision increases voter responsiveness to candidate characteristics, which in turn creates uncertainty around party vote shares and attracts additional campaign expenditure. The finding that the effects of debates flow through voter behavior to ultimately impact policy echoes Fujiwara (2015), who links an improvement in voting technology, via changes in the *de facto* composition of the electorate, to more redistributive policy in Brazil.

The rest of the paper is structured as follows. Section 2 explains the institutional context and experimental design. Section 3 details the research hypotheses, econometric specifications and pre-analysis plans. Section 4 discusses evidence for treatment effects on voters, candidates, and elected officials. Section 5 explores mechanisms. Section 6 concludes with policy considerations.

2. Institutional Context and Experimental Design

2.1. Context

Sierra Leone has 112 Parliamentary constituencies, which are single member jurisdictions elected by first-past-the-post plurality. The winning MP represents the local area, containing approximately 40,000 residents, in the national legislature. In these elections, the ethnic composition of voters in a given constituency predicts the corresponding party vote shares with remarkable accuracy. These correlations arise from historical ties between the All People's Congress (APC) party and the ethnic groups in the North, most prominently the Temne; and between the Sierra Leone People's Party (SLPP) and groups in the South, most prominently the Mende (see Kandeh [1992]). As an example of the contemporary strength of these loyalties, 89 percent of respondents in the control group of this study reported voting for the MP candidate from the party that is historically associated with their ethnic group.

How unusual are debates in this context? At the regional level, interest in debates has increased markedly over the past decade with at least thirteen African countries holding candidate debates.[4] In Sierra Leone, debates remain rare but not unheard-of: Presidential debates were held before the

---

[4] Source: www.debatesinternational.org/countries, accessed 2 February 2016.

2007 and 2012 elections, however in no debate did both major party candidates participate. The dissemination vehicle studied in this experiment, via mobile cinema, was certainly novel. Before our intervention, respondents reported that they rely primarily on the radio (43%), followed by their friends and family (33%), for information about politics. To explore the impact of debates in this context, our research design involves randomization on three levels: constituency, polling center, and individual. We discuss the implementation details for each level below.

2.2. Candidate Debates and Constituency-level Randomization

Before the 2012 Parliamentary candidates were officially announced, we selected what we estimated would be the 28 most competitive races for inclusion in our constituency sample. While we used a variety of metrics to do this—including the 2007 vote margin, the ethnic-partisan bias favoring one party over the other, and whether the seat changed parties in the previous election—*ex post* these races ended up being neither the most nor least competitive in 2012 (see Appendix Figure A.1). The vote margins within our sample thus represent a broad subset (ranging from 0.14 to 0.75) of the national distribution (which ranges from 0.01 to 0.91). We then randomly selected 14 constituencies from this set, stratifying on the degree of ethnic-party bias, to host debates. All randomizations were done on a computer. Appendix Table A.1 compares characteristics of constituencies, candidates and winning MPs across treatment assignment, and shows that the randomization achieved reasonable balance.

Our implementing partner and host of the debates, Search for Common Ground (SFCG), is a well-respected nonpartisan civil society organization with a strong presence in local media. SFCG provides a range of programs, including regular radio shows relayed by stations throughout the country, focused on their mission of promoting peace and reconciliation. SFCG invited candidates from the three largest parties—the APC, SLPP, and the latter's splinter party, the People's Movement for Democratic Change (PMDC)—who were contesting a given seat to participate in a debate. No other parties won seats in the previous election, and these parties respectively held 59, 39 and 9 percent of the seats in Parliament at the time.

Each of the fourteen debates followed a standardized format. The SFCG moderator opened the debates by introducing the candidates and explaining the basic roles and responsibilities of office. A casual "get to know you" section followed, where the candidates spoke informally about where they were from, their family and hobbies. Then five national policy questions were posed and each candidate was given two to three minutes to respond to each question. The first policy question

concerned the candidate's top priority for additional government spending. The second covered plans for spending the constituency facilitation fund (CFF), which is an untied 43.8 million Leones (approximately US$ 11K) grant given annually to each MP. It is intended to support development projects in, and the MP's own travel to and from, their constituency. The third asked for the candidate's strategy to uplift the youth, where "youth" is defined by the government as 18 to 35 year olds. This demographic segment faces high unemployment and their historic disenfranchisement and frustration were seen by many as a contributing factor to the country's civil war (1991 to 2002). Fourth was whether the candidate, if elected, would vote in favor of the Gender Equity Bill (GEB), a 30% quota for women's representation in government that was introduced but never voted on by the previous Parliament. The last national policy question asked for the candidate's assessment of the implementation of free healthcare (FHC), a major initiative by the incumbent government to provide free care to children under five and pregnant or nursing women. Each debate closed with two local policy questions, tailored to prominent issues in the host constituency. All debates were conducted in Krio, Sierra Leone's *lingua franca*.

2.3. Group Screenings and Polling Center-level Randomization

Within the fourteen constituencies selected to host debates, we first selected polling centers for the group screenings. All citizens had to register anew for this election, and the polling centers—typically a primary school or community center—are where they registered and later voted. This sample drew in 224 polling centers that had fewer total registered voters (471 on average) and were located further away from their nearest neighboring center (2.4 miles) than the population in general. SFCG took videotapes of the debates on a "road show" to 112 of these centers, selected randomly. In the eight constituencies where there were a sufficient number of polling centers left over, we randomly selected 40 of the remaining larger and closer together centers to host the individual-level private viewing treatments (discussed in Section 2.4). Note that the individual-level experiment was thus administered in a completely separate set of communities than the public screenings. A few months before the interventions, we conducted a household listing of registered voters in all 264 polling centers to develop the sampling frame for survey respondents.

The "road show" or mobile cinema treatment at the polling center level consisted of an evening showing of the video of the relevant debate projected at a convenient public place, usually on the side of the polling center itself, in the weeks leading up to the election. Typical protocols for these screenings were as follows: host polling center and satellite communities were notified in advance

and invited to attend the screening; 25 randomly selected residents (using data from the earlier listing exercise) were provided a small incentive (10 cooking spice cubes) to attend the screenings; the video was played once in a pause and play format that inserted translation into the relevant local language after each question; and the video was played a second time without translation. Secondary screenings were held earlier in the day in the largest accessible satellite communities (85 in total). Overall, the mobile cinema visited one quarter of the polling centers in these fourteen constituencies. As the centers were smaller than average, and not everyone in the catchment area attended, we estimate (very roughly) that 6% of all registered voters were directly exposed.

At the time of screening in treated polling centers, 16 of the 25 respondents who received attendance incentives were surveyed. We later conducted exit polls on Election Day and the days immediately after in all 224 treated and control polling centers. To avoid any differential attrition or selection across treatment assignment, the 5,600 exit poll respondents were drawn from the original household listing in both treatment and control polling centers and surveyed at their residence. In what follows, we will be estimating intention to treat effects, where 82% of exit poll respondents indicated that they attended a debate screening, as did 4% of those in the control group. The comparison of means across treatment assignment for voter characteristics in Appendix Table A.1 validates the polling center-level randomization.

2.4. Private Viewing Treatments and Individual-level Randomization

Within each of the 40 polling centers assigned to administer the private viewing experiment, we randomized individuals to six distinct treatment arms. To do so, we first divided households into bins based on the gender composition of registered voters, assigned treatment arms at the household level within each bin, and then selected one respondent per household to participate.

The specific treatment arms were as follows: (i) debate treatment, where individuals were shown the exact same debate screened in polling centers but now on a tablet; (ii) "get to know you" treatment, where individuals were shown a short video of the candidates speaking informally about their hobbies and interests; (iii) "radio report" treatment, where individuals listened to a journalistic summary of the main policy positions articulated by the candidates during the debates; (iv) surveyed control, where individuals were given the same survey as the one that accompanied treatments i to iii, but were not shown any media; and (v) pure control, where individuals were not surveyed until Election Day, and whose only contact with the research team at time of treatment implementation was to record basic demographics. A sixth arm participated in a lab-in-the-field

experiment (analyzed in related work) that exposed voters only to photos and 20 second video clips of candidates to assess, for example, whether voters could infer candidate ethnicity from physical appearance. No other political information was conveyed and this arm is thus grouped with the controls.

We assigned 400 individuals per treatment arm and 600 to the surveyed control group. Unlike for the polling center-level intervention, the exact same respondents who participated in the individual treatment arms were resurveyed in the exit polls. As we had perfect compliance and minimal attrition (6 percent overall), average treatment effect estimates for the individual treatment arms are comparable to treatment on the treated effects. Appendix Table A.2 presents voter characteristics, including attrition, across treatment arms, and validates that the individual randomization created reasonably balanced groups.

3. Hypotheses and Econometric Framework

The three levels of random assignment were designed to test separate but related questions. The group screening experiment, randomized at the level of polling center, evaluates whether debates affect voter behavior and whether candidates respond to debate publicity by changing the allocation of campaign expenditure. The series of treatment arms randomized at the individual level isolates how voters respond to the distinct types of information conveyed by debates. The constituency-level experiment, which randomized which races held debates, tests whether participating in a debate as a candidate changes the future behavior of the elected MP. Below we discuss the specific hypotheses and econometric specifications for each experiment separately.

3.1. Group Screening Experiment

As the group screening experiment captures the response of multiple actors to the debate road show, for clarity we refer to questions about voters as "domain A," questions about participating MP candidates as "domain B," and questions about other party officials as "domain C." Starting with voters, we test the following specific hypotheses:

A1. Exposure to debates increases political knowledge and leads to more informed voting, including (i) general political knowledge; (ii) knowledge of individual candidate attributes; and (iii) candidate policy stances

A2. Exposure to debates increases policy alignment

A3. Exposure to debates increases vote shares for the candidate who performed the best in the debate

10

A4. Exposure to debates increases the willingness to vote across party lines

A5. Exposure to debates enhances voter openness to other parties

We also explored some secondary hypotheses, namely that exposure to debates: (i) mobilizes the public and leads to greater turnout; (ii) increases the perceived legitimacy of elections; and (iii) increases interest in politics.

Analysis of treatment effects for domain A takes the form:

$$Y_{ipc} = \beta_0 + \beta_1 T_{pc} + X_{ipc}'\gamma + Z_{pc}'\delta + W_{ipc}'\theta + \mu_c + \varepsilon_{ipc} \qquad (1)$$

where outcome Y (e.g. vote choice) is measured for individual i registered in polling center p within Parliamentary constituency c; T is an indicator variable equal to one if the polling center received the debate group screening treatment; X is a vector of indicator variables that denote the stratification bin from which exit poll respondents were drawn (based on age and gender); Z is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (based on the number of registered voters and distance to nearest neighboring center); W is a vector of individual controls selected by hypothesis from the set {gender, age, years of schooling, polygamous marriage, farming occupation and radio ownership}[5] based on their predictive power in the control group data; c is a set of constituency-specific fixed effects (the level of debate); and

ε is an idiosyncratic error term clustered at the polling center level. The coefficient of interest is β_1, which captures intention to treat effects. We also test for heterogeneous effects by: (i) competitiveness of the constituency; (ii) candidate performance; and (iii) subgroups of voters by gender, age and fluency in Krio.

We registered a pre-analysis plan (PAP) to govern this analysis before the exit poll (our primary source of data) was completed (Appendix A).[6] The plan includes the research hypotheses, the list of specific outcome measures under each hypothesis, and the econometric specifications. Pre-specification has two main benefits.[7] First, it provides a structure for multiple inference corrections: we estimate effects for both individual outcomes and hypothesis-level indices, and adjust standard errors to account for the number of tests we run within and across hypotheses.

---

[5] Interest in politics was removed from the pre-specified set as it is endogenous to treatment. While radio acquisition could conceivably follow from heightened interest, we find no evidence for treatment effects on ownership (see Appendix Table A.1).

[6] We registered the first and main PAP governing this analysis with the Abdul Latif Jameel Poverty Action Lab on November 20, 2012. We later migrated the PAP when the American Economic Association's trial registry opened, where our entry can be found (https://www.socialscienceregistry.org/trials/26).

[7] See Casey, Glennerster and Miguel (2012) and Olken (2015) for discussion of PAPs.

Specifically, we follow Kling, Liebman and Katz (2007) to construct a mean effects index that orients each individual outcome so that larger numbers imply better outcomes, translates them into standard deviation units with reference to the control group mean and standard deviation, and computes the equally weighted average of all transformed outcomes under a given hypothesis. Following Anderson (2008), we apply family wise error rate (FWER) corrections at the hypothesis level, which strongly control the probability of making any Type-I error; and apply false discovery rate (FDR) adjustments at the individual outcome level, which control the expected proportion of rejections that are Type-I errors.[8] He notes that the former are quite conservative, as may be appropriate for assessing overall effectiveness and making policy decisions about scaling up implementation. We also report the per comparison, or "naïve," *p*-value for all estimates, which are appropriate for those with an *a priori* interest in the specific outcome or hypothesis presented.

The second benefit of pre-specification is that it leverages statistical power by bolstering the credibility of one-sided tests where theory only supports one direction of effect. Unless otherwise stated, results in our discussion are based on one-sided tests in the direction indicated in hypothesis statement. We also show results for more conventional two-sided tests, under conservative specifications that further exclude control variables, in Appendix Tables A.3 and A.4. We flag in the text estimates that fall from the 95% to 90% confidence level under these adjustments.[9]

Turning to candidates, the group screening experiment tests for effects of debate publicity on participating MP candidates (domain B) and on other party officials (domain C). The hypothesis for the two sets of actors is the same:

> B1(C1). Candidate (party) allocation of campaign effort and expenditure is responsive to debate publicity

We are interested in whether campaign investment complements or substitutes for debate publicity, and thus conduct two-sided tests. As mentioned earlier, if debates make exposed areas appear more competitive, then a swing voter model predicts that resources will flow toward communities where the screenings were held. Recall that while we did not inform the candidates of which polling centers were assigned to treatment, the screenings were large public events whose locations would

---

[8] We do not adjust across research domains (e.g. across voters in the group versus individual treatments), as each domain concerns a distinct sample, covering different agents, datasets and/or randomizations.

[9] Overall, we estimate treatment effects for 55 individual outcomes concerning voters, candidates and politicians in Tables 2, 3 and 5. Thirty estimates have *p*-values less than 0.050 under our preferred specification. Of these, five estimates fall below the 95% confidence level when we remove controls and conduct two-sided tests, where the highest resulting *p*-value is 0.105.

not have been difficult to track after they occurred. These hypotheses capture an endogenous response of candidates to the polling center-level treatment assignment. Treatment effects on voters in domain A thus reflect the combination of exposure to debate and the campaign response. By contrast, comparisons across the individual-level treatment arms under domain D below isolate a "pure" debates effect, net of any campaign, survey priming, or social mobilization effects.

To test hypothesis B1 concerning participating MP candidates, we use the same econometric specification as in (1), save the outcomes are linked to individual candidates: e.g., an outcome Y (such as receiving a gift) is measured for individual i in relation to candidate m where the individual is registered in polling center p within Parliamentary constituency c. This analysis thus leverages detailed campaign data on individual voter-candidate pairs.

Hypothesis C1 considers the campaign response of political parties more generally. Data for this domain were collected in a community-level survey that accompanied the voter-level exit polls, implying that there are many fewer observations for this analysis than for B1. Survey questions here do not distinguish gifts from different party representatives, and instead reference any party official or candidate for office, where the offices include President, MP, Local Councillor, and Local Council Chair, all of which were contested during the single General Election. Hypothesis C1 includes some additional community-level outcomes, like hosting political rallies, that are not included in B1.

3.2. Individual-level Experiment

The series of treatment arms randomized at the individual level unpacks voter responsiveness to the distinct types of information delivered by the debates (domain D). We use the same hypotheses and outcomes regarding voter behavior (A1-A5 above), but now test the absolute effect of each of the three treatment arms (debate, get to know you, and radio report) compared to the control group, as well as the relative effect of each treatment arm compared to the other treatments.

Analysis of effects in domain D takes the form:

$$;_{ihçã} = Ú_4 + Ü6_{ûçâ} + {}_{,}\bar{Û}_â{}^{°} + \dagger\bar{â}_â £ + \bar{Ü}_{ûçâ}{}_{,} + ?_â + \acute{Y}_{ûûçâ} \tag{2}$$

where outcome Y (i.e. vote choice) is measured for individual i living in household h assigned to treatment arm t registered in polling center p located in Parliamentary constituency c; T is a dummy variable indicating assignment to treatment arm t; X is a vector of indicator variables that denote the stratification bin from which the household was drawn (based on the gender composition of registered voters); and Z, W, c and θ remain as defined in (1). For each treatment arm, the

13

coefficient of interest is $\ddot{U}$the average treatment effect for treatment t compared to the control group. The control group includes respondents in the surveyed and "pure" control arms as well as participants in the lab-in-the-field arm (who received no political information). We further test a series of hypotheses about the relative effects of the different treatment arms that take the form $\ddot{U} M\ddot{U}$. Tests of average treatment effects are one-sided in the direction of the hypothesis statement, and tests of relative effects are two-sided. We lodged a second PAP to cover this analysis (Appendix B).

3.3. Constituency-level Experiment

Moving from the pre- to post-election time period, the constituency-level experiment explores medium term accountability effects of the debate treatment on the candidates who subsequently won the seat (domain E). This analysis operates at the highest level of aggregation, where recall that we randomly allocated 14 of 28 constituencies to host debates. There are four hypotheses:

E1. Accountability pressure of constituent exposure to debates is expected to increase the activity and engagement level of elected MPs

E2. The publicity of the debates helps solve the candidate commitment problem and makes their post-election behavior in Parliament more consistent with their pre-Election promises

E3. Accountability pressure of constituent exposure to debates is expected to increase post-election engagement with constituents

E4. Accountability pressure of constituent exposure to debates is expected to increase development expenditure under the CFF

The econometric specification for domain E is:

$$;_{\ddot{U}}=_{\ddot{O}}\dot{J}_4 + \ddot{U}\zeta_{\ddot{O}} + {}_{,}\overset{\tilde{n}\circ}{\dot{U}} + \tilde{a}_{\ddot{O}} + \acute{Y}_{\dot{\ddot{O}}} \tag{3}$$

where Y is outcome for MP candidate i who won the seat for constituency c, T is an indicator signaling that the constituency was assigned to the debates participation treatment, X is a vector of MP-level controls {gender, public office experience} selected by their contribution to increasing the R^2 in analysis of the control group data, and $\tilde{a}$ are fixed effects for the randomization strata used in the constituency-level assignment (three bins of ethnic-party bias). Tests are one-sided in the direction of better performance. Given the small sample at this level, standard error estimators that are robust to heteroskedasticity are likely downward biased. To reduce this bias, we present standard errors that are the maximum value of conventional ordinary least squares and bias corrected HC estimators in MacKinnon and White (1985), following discussion in Angrist and

14

Pischke (2009). We do not have power to adjust for multiple inference in this domain.

Outcome data for these hypotheses draw on several sources. We surveyed all candidates in treated and control constituencies pre-election, and surveyed the 28 winning MPs shortly after the election. The 14 treated winners were also given a video of the debate they participated in, edited to include only their own statements, and told how many voters had seen their debate. Performance outcomes for the winners were drawn from Parliamentary administrative records, MP self-reports, and extensive fieldwork in their home constituencies. Note that collection of performance data did not commence until after the MPs had completed their first year in office. We registered a third PAP to cover these last hypotheses (Appendix C).

Overall, we lodged three PAPs and two updates to accommodate data collection efforts spanning two years. The important thing to note about this iterative process is that the hypotheses and outcome measures for domains A, B, C and D were all established with the first plan in November 2012 before the exit poll fieldwork was completed; and those for domain E were lodged in June 2014 before the fieldwork tracking MP activity was completed. There are three substantive revisions to the first PAP worth noting here. First, we "demoted" the hypothesis about turnout from primary to secondary after official election results were published revealing very high (87%) turnout rates, implying that we would have limited power to detect treatment effects. Second, we combined two hypotheses in the initial plan—policy alignment and policy persuasion—into one single hypothesis, as they capture different mechanisms leading to the same observable outcomes.[10]   Third, we added analysis of survey priming. Its earlier omission was a simple oversight as the research design explicitly includes surveyed and pure controls in order to capture these effects. Throughout the rest of the paper, we flag analyses that were not pre-specified and should thus be considered exploratory rather than confirmatory in nature.

4. Estimated Treatment Effects by Research Domain

4.1. Effects of Debate Group Screenings on Voters (Domain A)

Overall, we find that exposure to debate screenings increased political knowledge, moved voters into better policy alignment with their selected candidate, increased vote shares for candidates who performed the best during the debates, and enhanced voter openness to participating candidates.

---

[10] Having two hypotheses covering the exact same set of outcomes in the original PAP was clearly a mistake.

Treatment effect estimates for these four hypothesis-level indices are significant at above the 95% confidence level when considered on their own, and remain above 90% confidence under conservative FWER corrections (Table 1). We find that while voters moved back and forth across ethnic-party lines to support strong debate performers, this had no net effect on the overall incidence of ethnicity-based voting (i.e. hypothesis A4 is not confirmed).

More specifically, watching debates increased the mean effect on political knowledge by 0.281 standard deviation units (standard error 0.028) across the 20 individual outcomes included. To give a sense of magnitude and substantive content, Table 2 unpacks this index into a selection of its component measures (see Appendix Table A.3 for all component estimates). Under general political knowledge, the percentage of voters who could correctly state the amount in the constituency facilitation fund (CFF), allowing for a generous range around the true figure of 43.8 M Leones, rose from 3.4% in control centers to 17.4% in treatment areas, a fivefold increase. Voters also learned about public entitlements: the proportion who knew who was eligible to receive free healthcare increased by 5.6 percentage points (s.e. 3.3) on a base of 70.6%. And they gained a better sense of what elected officials were meant to do in office: the number of correctly reported MP roles and responsibilities increased significantly. The statistical strength of these results is largely unchanged when we adjust *p*-values to control the false discovery rate (FDR) across all 38 outcomes within domain A (column 5).

We find that voters exposed to debates had substantially more knowledge of candidate characteristics: treatment effect estimates for 6 of 7 outcomes measured are significantly positive at 95% confidence. For example, the proportion of voters who could infer which candidate was better educated rose from 24.3% to 40.2%, and the proportion who knew which candidate (if any) had been an MP in the past increased from 49.0% to 60.1%, both significant at 99% confidence. Voter knowledge of the candidates' public office experience and ability to name candidates from all three parties also increased significantly.

Debates further increased voter knowledge of candidate policy positions. For each of (up to) three participating candidates, on each of three national policy issues,[11] voter ability to correctly place the candidate on the specific policy spectrum increased significantly (at 99% confidence) for 8 of 9 estimates. For example, the proportion of voters who could correctly identify the SLPP candidate's first priority for government spending doubled, from 14.2 to 29.1%; the proportion

---

[11] To keep the exit poll survey short, we included only 3 of the 5 national policy issues discussed during the debates.

who knew the APC candidate's view on whether free healthcare (FHC) was being well implemented or needed to be significantly reformed rose from 25.2 to 44.9%; and the proportion who knew whether the PMDC candidate would vote in favor of the gender equity bill (GEB) rose from 24.4 to 45.3%.

Together, these results suggest that watching debates substantially increased voters' political knowledge. Recall that respondents experienced a one to five week lag between exposure to debates and the exit polls, indicating that the gains in knowledge were relatively persistent. The next natural question is whether these gains translated into changes in vote choice on Election Day.

Voters used their enhanced policy knowledge to move into better policy alignment with their chosen candidate. Alignment is measured as a match between the voter's reported policy position in the exit poll and the position the candidate they voted for expressed during the debate. Estimates suggest that debate exposure increased policy alignment by 0.106 standard deviation units (s.e. 0.035) on average across three national policy issues discussed during the debates (Table 1, row 2). This effect is significant at 99% confidence for both per comparison and FWER controlled p-values. To provide a sense of magnitude, consider the results in Panel A2 of Table 2. The empirical match between the voter's first priority issue and the view articulated by their chosen candidate during the debate increased by 9.0 percentage points (s.e. 3.1) on a base of 42.5%. We find similar effects for free healthcare, where alignment increased by 9.2 points (s.e. 3.5) on a base of 39.4%. We see no effect for the gender equity bill, although note that there was little divergence in views expressed during the debates (only two candidates voiced strong objection to the bill).

What drives this improvement in policy alignment: choosing candidates based on previously determined policy preferences, or changing policy positions based on comments from the candidates? Using priority sector as an example, alignment improves if: i) voters who prefer education select a candidate who also supports education; and/or ii) voters update their view that education is the most important sector after observing their preferred candidate advocate for education. The former is what one would expect from canonical proximity voting models, originating with Hotelling. By contrast, Abramowitz (1978) suggests that the latter was at work in the Carter-Ford Presidential debates of 1976, where voters adopted their previously preferred candidate's view on unemployment policy after watching the two candidates debate the issue. Lenz (2009) further argues that these effects are concentrated among voters who learned the candidates' positions from the debate. We find evidence that both channels are likely at work.

We have no pre-treatment data on policy preferences for the control group, so to identify these channels we split our sample into party stalwarts, who are most likely to change their policy view to match their preferred candidate's position, and flexible voters, who have crossed party lines in the past and may be more likely to choose candidates based on policy. We define stalwarts as those who voted for their ethnically aligned party in all of the other elections for which we collected vote choice: 2007 Parliamentary, 2012 Presidential and 2012 Local Council.[12] For these voters, alignment increases by 12.2 points (s.e. 3.5) for priority issue and by 11.2 (s.e. 4.1) for healthcare, suggestive of position updating. Flexible voters are those who voted across ethnic-party lines in at least one of these three other elections. We find very similar results: among voters who have demonstrated willingness to vote for the rival party, alignment improves with treatment by 9.8 percentage points (s.e. 4.6) for first priority issue and by 11.2 (s.e. 4.0) for free healthcare, suggestive of selecting candidates based on policy. Note however that the party stalwarts represent a much larger share of the study sample (72% compared to 15%)[13] suggesting that position updating may be the more empirically substantive channel.

The main outcome of interest is whether debates changed how people voted. We find significant positive impacts on votes cast for the candidates who performed best during the debates. Estimates for the mean effect index in Table 1 suggest an increase of 0.086 standard deviation units (s.e. 0.043), significant at 97% confidence on a per comparison basis and 92% confidence under FWER adjustment. This index compiles two measures of debate performance: one determined by the audience and another by our expert panel. Audience judgments were recorded in a survey that immediately followed the implementation of the group-level screening. The expert panel consists of twenty five members of government and civil society who watched the debate videos and scored candidate responses to each debate question. These two sets of evaluations coincide on who performed best in 10 of the 14 debates. Where they diverge, the expert panel was more likely to pick a less popular candidate, including one from the PMDC, the smallest party that was not very competitive in this election (they won no seats nationwide).

Table 2 Panel A3 reports treatment effects for these two measures in our exit poll data (primary test) and in the National Electoral Commission's (NEC) official polling-center level returns

---

[12] We omit current MP 2012 choices as these may have been affected by the debate treatment.
[13] The remaining 12% represents voters from ethnic groups not strongly affiliated with either party. These voters do not appear to move into policy alignment via either channel.

(secondary test).[14] The correlation between party vote shares measured across the two datasets is 0.93 for the APC and 0.92 for the SLPP, suggesting that misreporting of vote choice in the exit polls is not a major concern. All four treatment effect estimates for votes for the debate winner are positive, and three are significant at 95% confidence. The estimate that is largest in magnitude is for votes for the candidate that audience members judged to have performed best, measured in the exit polls, where we see a 4.9 percentage point (s.e. 2.1) increase in votes for the debate winner. As a benchmark, this is comparable to the incumbency advantage of American state legislators (Ansolabehere and Snyder [2002]). The corresponding estimate using the official NEC returns is 3.5 percentage points (s.e. 1.7). While the two are not statistically distinguishable from each other, it makes sense that the point estimate in the NEC data is smaller, since the returns include votes from peripheral villages not exposed to treatment.[15] Note that vote shares for these candidates were already high (71% in the NEC returns for control areas), indicating that in this set of constituencies, the candidate who was locally popular tended to also perform best during the debates.

We find no evidence that these shifts in vote shares translate into any net impact on the prevalence of voting across ethnic-party loyalties. In Table 1, the coefficient for the mean effects index for hypothesis A4 is small in magnitude and not statistically distinguishable from zero, as are all estimates for its individual component measures. How can we reconcile a five percentage point shift in votes toward the debate winner, with no commensurate change in voting along ethnic lines? First, note that a move toward the debate winner only crosses party lines if the voter is from a rival ethnic group. Voters traditionally loyal to the debate winner should neither change their vote nor cross ethnic lines after exposure. This is what we see in the data. For historically aligned voters, there is no treatment effect (1.6 percentage points, s.e. 1.4) of watching the debate on their vote choice, as presumably they were already planning to vote for that candidate. These voters constitute 81% of the study sample and had baseline rates of 90% voting for the aligned candidate (i.e. debate winner) in the control group. By contrast, voters from ethnic groups historically

---

[14] The NEC sample excludes one constituency where the SLPP candidate was disqualified immediately before the Election but his name remained on the ballot. A full 48% of ballots cast were deemed invalid (many of which were likely SLPP votes). The winner was eventually determined via the courts. Treatment effect estimates are similar with its inclusion (0.032*, s.e. 0.016 for audience pick and 0.032*, s.e. 0.015 for expert panel, $N = 224$).

[15] The PAP commits to showing estimates when including an additional 29 "pure" control polling centers that were randomized out of our study sample. As we defined the randomization strata after their exclusion, which was a mistake, we must alter the main specification somewhat to include these extras. Treatment effect estimates remain similar with their inclusion: 2.8 percentage points for both votes for audience and expert best, with one-sided p-values of 0.077 and 0.073, respectively.

affiliated with the rival party (i.e. the candidate running against the debate winner), represent only 7% of the sample and had a much larger treatment effect estimate of 10.6 percentage points (s.e. 7.5), which is significant at 92% confidence in a one-sided test.[16]

Second, we should expect more votes to move toward the debate winner where the rival party candidate strongly outperforms the local favorite. Consistent with this, for the subsample where the audience deemed that the "outsider" candidate (who received only 26% of votes in the control group) won the debate, the treatment effect on votes for the winner is four times larger than in the full sample (19.1 percentage points, s.e. 11.0, $N = 381$) and significant at 94% confidence in a one-sided test. Thus the effects on switching one's vote to the debate winner are concentrated in "upset" contests and among voters historically affiliated with the rival party, consistent with the model in Casey (2015). Both of these subsamples, however, are small. This reflects the failure of our pre-election estimation to select (*ex post*) the most competitive races, creating a sample of races more advantaged toward the locally popular candidate than we anticipated. As a result, tests for these subgroups are underpowered and not pre-specified, so should be considered exploratory in nature.

Estimates for the fifth and final hypothesis suggest that exposure to the debates enhanced voter openness to candidates from all participating parties. In Table 1, we see that the treatment effect for the mean effect index is 0.091 standard deviation units (s.e. 0.048), significant at 97% confidence (or 92% confidence under FWER adjustments). This index compiles information from 10 point likeability scales, where all five treatment effect estimates in the individual outcomes are positive and one is statistically significant at conventional levels. While clearly voters updated more positively for some candidates than others, the fact that their opinions rose across the board is important for securing candidate participation in future debates.

Results in domain A are fairly robust to excluding control variables and conducting two-sided tests. Only three of the 23 individual outcome estimates that are significant at 95% confidence under our preferred specification fall to the 90% confidence level in the "raw" results of Appendix Table A.3. These are for knowledge of free healthcare eligibility and the two (secondary) measures of votes for the debate winner in the NEC data. Two of the mean effect indices in Table 1—votes for best and voter openness—similarly fall from 95 to 90% confidence when the control variables

---

[16] The remaining 12% of the sample are voters from ethnic groups that do not have strong historical ties to either party, so are excluded from the crossing party lines estimate. About half (57%) of these voters chose the debate winner in the control sample. The point estimate on the treatment effect for this group is also large, at 10.1 percentage points (s.e. 8.4), but not statistically significant (one-sided *p*-value of 0.115).

are dropped and two-sided tests are used.

4.2. Endogenous Response by Candidates and Other Party Officials (Domains B and C)
Domain B explores whether participating MP candidates altered their campaign strategy in response to the debates road show, given its strong effects on voters' political knowledge and opinions. Table 3 presents evidence that candidate campaign spending serves as a complement to the publicity of the polling center screenings. The treatment effect for the mean effects index is 0.103 standard deviation units (s.e. 0.039), significant at 99% confidence using a two-sided test.

Unpacking the index, treatment effect estimates for all nine component measures—covering MP candidates from each of three parties and each of three campaign outcomes—are positive in sign. These reflect increases in voter reports of having received a gift from the particular candidate, the monetary value of the gift (expressed in logs), and the number of times the candidate visited the community, all with reference to the weeks leading up to the election. The response by candidates from the two major parties, the APC and the SLPP, is similar when measured as the percentage increase on their base level of spending in control communities. Third party candidates, who generally had less of a chance of winning, appear to have responded more strongly to the road show: estimates for each of the three PMDC campaign measures are statistically significant. All estimates are robust to excluding control variables (see Appendix Table A.4).

What drives this reallocation of campaign effort? One explanation is that by equipping voters with greater political knowledge and changing their voting choices, debate screenings made these areas more competitive. This would be consistent with a standard swing voter model (Lindbeck and Weibull [1987]). Extending the exploratory analysis above (and again noting that it was not pre-specified), the treatment effect on the campaign index is five times larger in the constituency where the "outsider" candidate won the debate (at 0.41 standard deviation units, s.e. 0.16) compared to the other constituencies in the sample, which is precisely where the debates had the largest impact on the competitiveness of the race. The coefficient on this difference (0.33, s.e. 0.16) is significant at 95% confidence. Note, however, that the coefficient for the remaining constituencies, where the screenings *de facto* made the races less competitive as the locally popular candidates performed better in the debates, remains positive and statistically significant at 95% confidence (0.08, s.e. 0.04). This can be reconciled with the idea of greater competition if the debates made vote shares in screening communities more uncertain, as recall that the actual impact of the debates on voting was not revealed until Election Day. This is consistent with the extended

model in Casey (2015), where information increases voter responsiveness to individual candidate attributes, thereby making it harder to infer vote shares from the ethnic composition of a locality, and thus widening the set of potentially competitive areas.

To assess whether this spending worked to reinforce or unwind the impact of the screenings, we explore whether the intensity of the campaign response covaries with candidate performance during the debate. Appendix Figure A.2 reveals an inverted U-shaped relationship between the size of treatment effect on campaign expenditure and the share of audience members who said that candidate won the debate (Panel A).[17] This suggests that the campaign response to the road show was strongest where the debates themselves were most closely contested. There is also some asymmetry in the tails, where the treatment effect estimate for candidates who received the fewest audience votes is negative (at left) while estimates for those who received many votes (at right) are positive although noisily estimated. The relatively stronger campaign response by those who performed well versus poorly would work to reinforce the impact of debates. Panel B presents the same estimates for third party candidates and shows that they responded most strongly where they had performed well during the debate. This again suggests that the spending response was strongest where the debates worked to increase the competitiveness of the race.

We find little evidence that centralized party bosses and candidates for President, Local Councillor and Local Council Chair altered their campaign strategy in response to dissemination of the MP candidate debates (domain C, Appendix Table A.5). While the treatment effect for the mean effects index is positive in sign (0.08 standard deviation units), it is not significant at conventional levels (s.e. 0.05 and p-value 0.11). Similarly, while the majority (16 of 21) of treatment effect estimates for individual outcomes are positive, none are significant at conventional levels. This can be viewed as a pseudo placebo test: candidates for offices not involved in the debates should not alter their campaign strategy in response to the MP debate road show. This would make sense if the parties did not strongly coordinate campaigns across candidates for different offices, or if the road show was not a salient enough event to justify reallocating campaign support from other party members to support the participating candidates. While this seems plausible, we do not place too much weight on this interpretation for two reasons. The sample for this community-level survey is small ($N = 224$), so power to reject the null is limited. And, the community survey questions bundled together the campaign efforts of all party officials and

---

[17] Estimates control for the underlying ethnic-party loyalty of the constituency.

candidates for all offices, which includes Parliament, so they do not clearly exclude the MP candidates as one would do for a true placebo.

4.3. Unpacking Voter Response via the Individual Treatment Arms (Domain D)

What is it about debates—the revelation of policy positions or the showcasing of candidate personalities—that voters respond to? Results from the suite of individual-level treatments suggest that voters update their views of candidates based on both policy and charisma, but that only access to the combination of both types of information motivates them to change the way they vote.

Table 4 compares voter response to the debates, the get to know you videos, and the radio reports (all delivered privately via tablet). Estimates in row 1 show that all three arms were effective in transmitting political information, where the treatment effect on the mean effect index is positive and significant at 99% confidence for each (columns 1, 3 and 5). The coefficients for debates (0.109, s.e. 0.021) and the radio reports (0.095, s.e. 0.018) are more than twice as large in magnitude as that for the get to know you videos (0.041, s.e. 0.016), differences that are statistically significant under naïve p-values and FDR corrected q-values (columns 8 and 12).[18] While the coefficient for debates is slightly larger than that for the radio reports, the difference is not statistically distinguishable from zero (column 9).

The pattern of treatment effects for questions about general political knowledge mirrors that for the hypothesis overall: all three arms yield strong positive effects; and the debate and radio report estimates are larger in magnitude than those for get to know you videos. Similarly, estimates for knowledge of candidate characteristics are positive and significant for all three arms, although now magnitudes are comparable across treatments. Interestingly, this implies that voters were equally as able to infer things like which candidate was better educated and which one had more public office experience by watching the five minute get to know you video as they were after watching 45 minutes of debate. These topics were generally not asked directly, but could plausibly be inferred from the candidate's manner of speech, physical carriage, or confidence. For policy knowledge, only the debates and radio reports enhanced voter ability to correctly locate candidates on the three policy spectra. The estimate for the get to know you videos, which focused solely on candidate persona and delivered no policy information, is near zero and thus reassuring for the soundness of the research design.

---

[18] The FDR q-values adjust across all 24 (two-sided) comparative tests run.

Notably, only debates moved voters into better policy alignment with the candidates they selected. The treatment effect for debates (0.081, s.e. 0.029) is positive and significantly larger than that for the other two arms, which are both indistinguishable from zero (row 5). For the get to know you videos, this is intuitive and consistent with the null result on policy knowledge. For the radio reports, however, it implies that the acquired knowledge of policy positions did not translate into better policy alignment as it did for the debates. Similarly, only the debates arm had an impact on votes for the debate winner (0.058, s.e. 0.040), which is statistically larger than the result for the radio reports. The fact that radio was equally as effective in building knowledge, but only debates impacted policy preferences and voting choices, suggests a key role for personality in persuading voters to change their behavior. None of the three treatment arms impacted crossing party lines, consistent with what we saw earlier for the public screenings, and none of them affected candidate likeability scores.

Overall, while the debate, radio report and get to know you treatments all affected political knowledge, only debates moved voters to change whom they voted for and update their policy views. We can thus evaluate whether the treatment effect for debates is larger than the sum of the effects of the radio plus get to know you arms (not pre-specified). For policy alignment, the treatment effect for debates is larger than the sum of the other two by 0.11 standard deviation units (s.e. 0.04), and for votes for the debate winner, it is larger by 0.10 (s.e. 0.07). Under one-sided tests, we can reject the null at 99 and 92 percent confidence, respectively, or at 99 and 85 percent under two-sided tests. This pattern of results is consistent with the idea that debates are additive in both charisma and policy or professional content, and that the combination is more powerful than either in isolation.

4.4. Effects of Debate Participation on Elected Members of Parliament (Domain E)

Moving from the election to the behavior of the winning candidates once in office, Table 5 presents results for the four longer term accountability hypotheses. Overall, eight of the eleven individual treatment effect estimates are positive and five are at least marginally significant. The positive effects are concentrated in the hypotheses about constituency engagement and public spending. Here we find increases in verifiable development expenditures, which is the outcome that most directly enhances constituency welfare. While these findings are substantively important, they are based on a limited sample and are thus more speculative than results presented for other domains.

Discussing the hypotheses in order, we find little evidence for treatment effects on the activity

24

level of elected MPs during sittings of Parliament. Outcomes cover the period from when MPs were inaugurated in December 2012 through the end of 2013, or 57 sittings in total. Specifically, there is no difference in the number of sittings attended, committees joined or public statements made in Parliament (note the low control mean of four statements).

There is no evidence for effects on enhancing consistency between the candidate's first priority sector articulated during the campaign, and their subsequent effort promoting that sector. We defined the priority sector for each MP based on their pre-election response to the question, "If you had to prioritize one issue in Sierra Leone to receive additional funding in the national budget, what issue would you prioritize?" The modal response was education (44 percent), followed by roads, health and agriculture (each with 15 percent). Treated MPs were no more likely to make public statements during a Parliamentary agenda item concerning their preferred sector, although note that only one MP in the sample did so. They similarly were no more likely to join committees dedicated to that sector, and their constituents are no more likely to report that they focus on that sector. We were not able to evaluate consistency in voting in line with pre-stated positions on key national policy issues of interest, as relevant bills have either not yet been introduced (including the gender equity bill) or were passed unanimously (including a freedom of information act).

We do, by contrast, find positive and significant effects of debate participation on subsequent constituency engagement. Participating MPs made on average 1.3 (s.e. 0.6) additional community visits, on a base of 2.9, and held 1.1 (s.e. 0.6) more public meetings, on a base of 1.0. These represent increases of 145 and 210 percent, respectively. Their constituents on average named more sectors in which they viewed the MP as doing "a good job in promoting" that sector, and medical staff in clinics were more likely to report that the MP was doing a good job in promoting health. The mean effects index covering all four outcomes is positive and highly significant (0.8 standard deviation units, s.e. 0.3).

Most importantly, we find significantly higher spending on development projects by MPs who participated in a debate. Recall that the constituency facilitation fund (CFF) is an annual allotment of 43.8 M Leones (approximately US$ 11,000) intended to support the development of, and the MP's own transport to, their constituency. MPs are fairly unconstrained in how they spend this money and are not subject to monitoring or reporting requirements. During the debates, each candidate was asked to articulate their plans for spending the CFF. All candidates, save one, promised to spend some, if not all, of the funds on development projects. To compile data on how

the CFF was actually spent, we first surveyed each elected MP to generate a detailed itemized list of expenditures and project locations for the first CFF allotment. Our research teams then conducted exhaustive field work to verify these expenditures in the MP's home constituency, which involved in-person visits and physical examination of all purported projects, and multiple interviews with community leaders, clinic staff, teachers and residents of villages where money was reported to have been spent. We did not attempt to verify the MP's own transport expenses, so unaccounted for funds represent either legitimate travel costs or leakage. Note, however, that substantially larger travel expenses in the control group is not consistent with the evidence above that control MPs held fewer meetings with their constituents.[19]

For the control group, Table 5 shows that only 36 percent of the $11,000 allotment could be verified as spent on the development of the constituency. The treatment effect estimate of 54.7 (s.e. 31.7) suggests that MPs who participated in the debates spent 2.5 times as much on verifiable development expenditures. The effect is significant at 95% confidence and the point estimate corresponds to average gains of roughly six thousand dollars per constituency. Appendix Figure A.3 transparently plots the distribution of this outcome by treatment assignment. Comparing the two subplots shows that the positive treatment effect estimate is driven by differences in both tails: there are more low values among control MPs and more high values among treated MPs. Estimates are robust to dropping the top outlier (treatment effect of 46.5, s.e. 29.1 and one-sided p-value 0.06) and to using a binary outcome measure equal to one for any nonzero amount of verified development spending (treatment effect of 0.29, s.e. 0.18 and one-sided p-value 0.06).

Appendix Table A.4 reruns all specifications in this section without the control variables and performs two-sided tests. Of the six treatment effect estimates that are significant at 95% confidence, two fall to 90% confidence and one falls just below the 90% threshold with these two adjustments. The latter is the estimate for CFF spending, where the point estimate falls to 49.5 (s.e. 29.3), with a p-value of 0.105 under a two-sided test.

The small sample for this domain calls for additional robustness checks (not pre-specified). The final row of Table 5 reports estimates for a mean effects index that covers all 11 outcomes. It suggests that participation in a debate enhanced the post-election performance of MPs by an average of 0.36 standard deviation units (s.e. 0.17), which is significant at 95% confidence under

---

[19] It also cannot be explained by differential distance to the capital or availability of major roads as both of these characteristics are well balanced across treatment assignment (see Appendix Table A.1).

both one- and two-sided tests, with and without controls. Gelman and Carlin (2014) recommend reporting the Type S (for "sign") error rate when working with noisy estimates. A Type S error is the probability, for a given true effect size, that a hypothetical replication yields an estimate with the incorrect sign, conditional on it being statistically significant. If the true effect on MP accountability is half as large as our estimate, the Type S error rate would be less than one percent. If the true effect equals what we found for candidates' campaign response (roughly a third of the accountability estimate), it would be five percent. These are reassuringly low probability estimates. It is only when we scale down the true effect size by a large amount that we begin to see nontrivial Type S error rates: for example, if the true effect size is only one tenth of our estimate, the Type S error rate would be 27 percent.

What drives these expenditure effects? One possible interpretation is accountability pressure: many more voters now know how much money the MP has at his disposal, know what he promised to spend it on, and the MP knows that they know. Our results show that debate exposure made voters more informed, and these voters could potentially take political action. At the same time, the watchfulness of the media might be more salient to treated MPs, as they interacted with SFCG during the debate and were given a video record of their commitments.[20] Standard media models suggest that both are needed: it is the combination of an engaged public and an observant media that generates accountability effects.[21]

An interesting but unlikely alternative explanation is candidate selection. As we gave the central party bosses a list of constituencies where we would host debates, they could have strategically responded by allocating different candidates to those races. If the attributes the parties thought were associated with favorable debate performance also correlated with performance in office, then the treatment effect would be operating through a change in the candidate pool instead of the accountability and commitment channel. While this would constitute an exciting general equilibrium response worth exploring in future, it is unlikely to hold in this experiment for several reasons. First, there was little time between our notification to parties and the close of candidate registration. Given the relative newness of debates in Sierra Leone, it further seems unlikely that parties would respond strongly to an unproven concept. Moreover, Appendix Table A.1 presents

---

[20] By contrast, the salience of the research would have been equivalent for treated and control MPs, who were asked the same questions pre- and post-Election, and were not contacted about their CFF expenditures until a year later.
[21] In another context, this might suggest a weaker response by term limited politicians, however there are no term limits for MPs in Sierra Leone.

little evidence that candidate characteristics vary systematically across constituencies assigned to debates participation and controls: while candidates in treated constituencies had somewhat less political experience; measures of age, gender, years of schooling, managerial experience, ethnicity and pre-election quiz scores are all comparable across the two groups.

Relatedly, if debates made voting more responsive to competence, these effects could be explained by selection via the electoral process. Yet recall that while public screenings were held in one quarter of the polling centers, only a small fraction of the total registered voting population in a given constituency attended. This means that the road show did not change the outcome of who won any of the fourteen races covered. If the program were scaled up, however, there could be potential impacts on MP selection.

4.5. Secondary Outcomes

There were several outcomes we thought were interesting but less directly related to the debates intervention, so segregated them in the PAP to a more speculative, exploratory category (Appendix Table A.6). The results for voter turnout are mixed and power to reject the null is low. For group screenings, estimates reflect negative yet insignificant treatment effects on very high baseline levels: turnout was 98 percent in the control areas of our exit poll sample, which is drawn from households in the immediate vicinity of the polling center itself; and 83 percent in the NEC returns, which cover voters from the entire catchment area of the polling center. In the individual treatment sample where turnout was slightly lower (96 percent), we find positive and significant effects for the debate and get to know you arms, and no effect for the radio report. Magnitudes are small: for example, the estimate for the debates arm in percentage point terms would be 1.4 (s.e. 0.69). We conclude that debates exposure had little to no effect on turnout.

We find some suggestive evidence that exposure to debates spurred voter interest in politics more generally. Voter ability to name the two Presidential candidates and frequency of discussing politics increased (although the latter is not statistically significant). Consistent with this, we also find positive and significant treatment effects on voter ability to name local council candidates from all three parties (outcomes not pre-specified). There is no evidence that the debates increased electoral misconduct, which was reasonably low according to community reports; or that it increased voter confidence that the elections were free and fair, which was high (92 percent).

Self-reported behavior by the MPs themselves tracks what we saw for the objective performance measures in Table 5. We find no divergence in reported rates of activity in Parliament

or consistency in policy positions over time. By contrast, treated MPs say that they spent more days in their home constituency and held more meetings with constituents. These debate participants further claimed to have spent more money under the CFF and allocated a larger portion towards development projects.

5. Additional Analysis

5.1. Survey Priming

How much of the effects on voters can be attributed to the content of the treatment itself as compared to the experience of being surveyed in depth about one's political views? This distinction is important in light of findings that the act of surveying has nontrivial impacts on behavior (Zwane *et al.* [2011]). Using two separate estimation techniques, we find significant priming effects on general political knowledge. Reassuringly, all results hold net of these effects.

First, in Panel A of Appendix Table A.7, we compare surveyed to "pure" controls in the individual-level experiment. At the time of treatment implementation, surveyed controls were given the same survey that accompanied the debates treatment, which may have primed respondents to seek out information on outcome variables of interest or increased their salience in the weeks leading up to the Election. By contrast, "pure" controls were asked only basic demographic questions, and were not asked any political questions until the exit polls. Estimates in column 4 suggest that the survey experience on its own led to a 0.099 standard deviation unit (s.e. 0.035) increase in general political knowledge (row 2). Column 1 compares those in the debate arm to surveyed controls and estimates a 0.211 standard deviation unit (s.e. 0.042) increase in general political knowledge, which can be attributed to the content of treatment, above and beyond the survey experience. Together, these estimates suggest that survey priming accounts for one third of the total treatment effect on general political knowledge. There is only one other marginally significant priming effect, however it does not replicate in the larger sample of Panel B.

Our second approach uses the group screening sample to capture a survey reinforcing effect by comparing treatment effect estimates for those assigned to treatment with survey versus "pure" treatment (all controls are "pure" controls). This is the converse of the above, where we now measure whether being surveyed at the time of treatment facilitates greater absorption of the political information conveyed by the debates. Estimates in Column 1 of Panel B suggest that the "pure" treatment effect of watching the debate without being surveyed is a 0.233 standard

deviation unit (s.e. 0.055) increase in general political knowledge. In column 4, there is evidence for an additional 0.099 standard deviation unit (s.e. 0.037) effect of being surveyed alongside treatment, suggesting that the survey reinforcing effect similarly accounts for roughly a third of the total effect on general knowledge. For all other knowledge, policy alignment and voting outcomes the "pure" treatment effect remains positive and highly significant, and there is no evidence of survey reinforcing effects.

5.2. Treatment Effect Heterogeneity

We find little evidence for systematic heterogeneity in treatment effects on voters. Appendix Table A.8 estimates heterogeneous effects by sub-groups of gender, age and lack of fluency in Krio (the language of the debates). Specifications use the hypothesis level mean effects index and include all subgroup terms and their interaction with treatment status in a single regression. Across the fifteen estimates of interest, only the negative coefficient on political knowledge for women (-0.08 standard deviation units, s.e. 0.02) is significant at conventional levels. This suggests that women acquired only 75% as much political knowledge from the debates when compared to men. We find little evidence that voter responsiveness varied with the expected competitiveness of the race, based on 2007 vote margins, or with candidate performance in the debate, based on expert panel scores (results not shown). Our results also do not appear to be driven by large effects in any particular constituency. As an example, the treatment effect estimate on voting for the debate winner is robust to excluding each constituency one by one.

Considering dissipation of effects over time, we find suggestive evidence for an immediate drop in political knowledge gains in the days after treatment, but no evidence for additional decay between treatment exposure and the election. Confining attention to the treatment group, voter knowledge doubled from the before- to after-screening surveys: voters on average correctly answered 24 percent of political knowledge questions at baseline, which jumped to 46 percent immediately after watching the group screening. By the time of the exit poll, this percentage had fallen to 40, implying that a third of the initial gains had dissipated. Similar estimates obtain for those who watched the debate privately via tablet. Bringing in the control group, we estimate whether this attenuation covaries with the time lag between the screening and the exit poll, which ranges from 6 to 35 days. The time variation is not random, so estimates rely on the assumption that factors determining field deployment (e.g. remoteness) are orthogonal to voter responsiveness to treatment. Here we find no evidence for heterogeneity over time: effects for those treated far

from the election, e.g. 30 days earlier, are similar to estimates for those treated close to the election, e.g. within 10 days of the exit poll. Our interpretation is that some knowledge gains dissipate quickly after exposure, while the remaining gains persist for several additional weeks.

5.3. Debate Delivery: Individual versus Group Exposure

We next consider why the treatment effects observed for the group screenings of debates are generally larger in magnitude than those found for individual private viewing. Since many aspects of the experience differ across these two modes of debate delivery, we will not be able to pin down exact mechanisms, but can speculate as to how salient differences might drive the divergence.

The first pronounced difference is that the screenings involved large public gatherings of a couple hundred people, while the individual treatment had respondents watch the debate alone on a tablet.[22] Consistent with a substantive role for social mobilization, lab experiments show that exposure to the reactions of audience members—either real or fabricated—affects evaluations of debate performance and candidate attributes (Fein, Goethals and Kugler [2007], Davis, Bowers and Memon [2011]). The public nature of group screenings may also generate common knowledge that eases coordination problems and reinforces the messages conveyed (Chwe [2001]). Note that the papers by Wantchekon and co-authors cited earlier all involve public treatments, where groups of voters come together in town hall meetings. The second salient difference is that it would have been very difficult for candidates to track the locations of the individual experiments and respond with greater campaign expenditure. Assuming that voters value the additional candidate visits and gifts, the uptick in campaign effort could contribute to a larger total effect for the group screenings.

Table 6 presents the cleanest comparison of the two delivery mechanisms by limiting the group screening estimates to the 8 constituencies where the individual treatments were also implemented, and restricting the individual estimates to comparisons between the debates and pure control arms. First, note that the qualitative pattern of effects for the two delivery modes is the same: strong positive treatment effects on political knowledge, policy alignment, and votes for the debate winner; and no evidence of effects on crossing party lines or voter openness. Second, note that the treatment effect for the group screening is larger in magnitude than that of the individual viewing

---

[22] The content of the debate films was exactly the same under the two conditions. Other differences in delivery are that individual treatments were administered in larger polling centers (as measured by total registered voters); and the implementation procedures varied, where group screenings played music before the debates, played the debates twice, and had simultaneous translation into the relevant local language.

everywhere save on votes for the best performer, where it is equal. This difference is more pronounced when we scale up the intention-to-treat effects for the group screening to estimate average treatment effects on compliers (column 2), which is more directly comparable to the individual treatments where compliance was near perfect. Notice that the difference in magnitude is largest for knowledge of candidate characteristics (ten times larger), knowledge of candidate policy positions (three times larger), and moving into policy alignment (twice as large). These differences are consistent with the idea that watching the films in a group setting facilitated discussion among voters that clarified and reinforced the information about candidates and policy conveyed by the debates.[23] The fact that point estimates for votes for the debate winner are the same across modes suggests that any impact of additional campaign effort did not translate into differences in vote choices, perhaps because the candidates who responded most strongly were from the relatively uncompetitive third party.

6. Conclusion

These experiments suggest that voters acquire significant political knowledge from watching candidate debates, knowledge that persists over a number of weeks, and importantly, influences their vote choice on Election Day. By equipping voters with knowledge that changes their voting behavior, debate screenings further attracted greater campaign investment by participating candidates. This spending response is consistent with debate exposure making vote margins appear narrower or more uncertain *ex ante*, even in areas where it was revealed *ex post* that debates favored the more popular candidate. Debates convey comprehensive information about candidates—including charisma, professional qualifications, and policy stances—and the combination of factors appears more powerful than each in isolation. Over the longer run, participation in debates enhanced the accountability pressure on elected officials, increasing their subsequent engagement with constituents and expenditure on development projects. The finding that debates seem to strengthen accountability, even in relatively uncompetitive areas where direct electoral pressure is limited, is important.

From a policy perspective, this project demonstrates that interparty debates are logistically

---

[23] The fact that the impact on general political knowledge is more comparable across the two modes suggests that basic differences in comprehension (attributable to waning attention or the lack of local language translation in individual delivery) cannot fully explain the divergence in magnitude of effect.

feasible to host and disseminate, and could be replicated on a larger scale. In considering the costs and benefits of scaling up, fixed video production costs for the debates themselves were modest in this setting: roughly five thousand dollars per constituency. The point estimate on increased development expenditure associated with debate participation is large enough to fully cover this cost. In terms of marginal dissemination costs, the mobile cinema in rural areas was a relatively resource intensive way to publicize the debates. Mobile cinemas in urban areas could reach substantial numbers at lower cost. In settings where mass media penetration is higher, dissemination via television or radio broadcast are obvious alternatives. While the individual treatments suggest that video is more effective than audio alone, the radio report we tested was rather dry. A livelier program that captures a real time debate between candidates in the recording studio might come closer to the impacts of the film screening, and could reach large voting audiences at negligible marginal cost.

One could imagine multiple equilibria that might arise if debates were taken to scale. At the pessimistic end, politicians could learn to game the debates and unravel any benefit to voters. Candidates could, for example, coordinate on making only vague statements so that debates do not reveal their relative policy positions and the public record contains no concrete promises for voters to later follow up on. The novelty value of debates might also fade over time, making each subsequent debate less interesting to voters and less impactful for electoral and policy outcomes. More optimistically, the knowledge that debates provide information to voters could drive candidate effort and policy more in line with the interests of citizens. Incumbent awareness that debate videos exist and could be used to hold them to account could further motivate better performance in office. And, by making voting more responsive to candidate quality, debates could strengthen incentives for political parties to invest in recruiting more competent candidates. We leave these questions of effects at scale and persistence over repeated events to future research.

References

Abramowitz, Alan I, "Impact of a Presidential Debate on Voter Rationality," *American Journal of Political Science*, 22:3(1978), 680-90.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103:484(2008),1481-95.

Angrist, Joshua D. and Jorn-Steffen Pischke, *Mostly Harmless Econometrics* (Princeton: Princeton University Press, 2009).

Ansolabehere, Stephen and James M. Snyder Jr., "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000," *Election Law Journal*, 1:3(2002), 315-38.

Banerjee, Abhijit V., Selvan Kumar, Rohini Pande and Felix Su, "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India," manuscript (2011).

Bardhan, Pranab and Dilip Mookherjee, "Determinants of Redistributive Politics: An Empirical Analysis of Land Reforms in West Bengal, India," *American Economic Review*, 100:4(2010), 1572-1600.

Benjamini, Y., Krieger, A., and Yekutieli, D., "Adaptive Linear Ste*p*-Up Procedures That Control the False Discovery Rate," *Biometrika*, 93(2006), 491-507.

Besley, Timothy and Robin Burgess, "The Political Economy of Government Responsiveness: Theory and Evidence from India," *Quarterly Journal of Economics*, 117:4(2002): 1415-51.

Callander, Steven and Catherine H. Wilson, "Context-dependent Voting," *Quarterly Journal of Political Science*, 1(2006), 227-54.

Casey, Katherine, "Crossing Party Lines: The Effects of Information on Redistributive Politics," *American Economic Review*, 105:8(2015), 2410-48.

Casey, Katherine, Rachel Glennerster and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *Quarterly Journal of Economics*, 127:4(2012), 1755-1812.

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

Chwe, Michael Suk-Young, *Rational Ritual: Culture, Coordination, and Common Knowledge* (Princeton University Press, 2001).

Cranes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts, "Leadership and Pandering: A Theory of Executive Policymaking," *American Journal of Political Science*, 45:3(2001), 532-50.

Cruz, Cesi, Philip Keefer and Julien Labonne, "Incumbent Advantage, Voter Information and Vote Buying." manuscript (2015).

Davis, Colin J., Jeffrey S. Bowers and Amina Memon, "Social Influence in Televised Election Debates: A Potential Distortion of Democracy," *PLOS One*, 6:3(2011), e18154.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 122:3(2007), 1187-1234.

Dixit, Avinash, and John Londregan, "The Determinants of Success of Special Interests in Redistributive Politics," *Journal of Politics*, 58(1996), 1132-55.

Dixit, Avinash, and John Londregan, "Ideology, Tactics, and Efficiency in Redistributive Politics," *Quarterly Journal of Economics*, (1998), 497-529.

Druckman, James N., "The Power of Television Images: The First Kennedy-Nixon Debate Revisited," *Journal of Politics*, 65:2(2003), 559-71.

Eifert, Benn, Edward Miguel and Daniel N. Posner, "Political Competition and Ethnic Identification in Africa," *American Journal of Political Science*, 54:2(2010), 494-510.

Enikolopov, Ruben, Maria Petrova and Ekaterina V. Zhuravskaya, "Media and Political Persuasion: Evidence from Russia," *American Economic Review*, 101:7(2009), 3253-85.

Fein, Steven, George Goethals, Matthew Kugler, "Social Influence on Political Judgments: The Case of Presidential Debates," *Political Psychology*, 28:2(2007), 165-92.

Ferraz, Claudio and Frederico Finan, "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes," *Quarterly Journal of Economics*, 123:2(2008), 703-45.

Fridkin, Kim L., Patrick J. Kenney, Sarah Allen Gershon, Karen Shafer and Gian Serignese Woodall, "Capturing the Power of a Campaign Event: The 2004 Presidential Debate in Tempe," *Journal of Politics*, 69:3(2007), 770-85.

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil," *Econometrica*, 83:2(2015), 423-64.

Fujiwara, Thomas, and Leonard Wantchekon, "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin," *American Economic Journal: Applied Economics*, 5:4(2013), 241-55.

Gelman, Andrew and John Carlin, "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors," Perspectives on Psychological Science, 9:6(2014), 641-51.

Gentzkow, Matthew, "Television and Voter Turnout," *Quarterly Journal of Economics*, 121:3(2006), 931-72.

Gentzkow, Matthew, Jesse M. Shapiro and Michael Sinkinson, "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review*, 101:7(2011), 2980-3018.

Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw, "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment," *American Political Science Review*, 105:1(2011), 135-50.

Hellweg, Susan A., Michael Pfau and Steven R. Brydon, *Televised Presidential Debates: Advocacy in Contemporary America* (Praeger, 1992).

Hotelling, Harold, "Stability in Competition," *Economic Journal*, 39(1929), 41-57.

Humphreys, Macartan and Jeremy M. Weinstein, "Policing Politicians: Citizen Empowerment and Political Accountability in Uganda Preliminary Analysis," manuscript (2012).

Kandeh, Jimmy D., "Politicization of Ethnic Identities in Sierra Leone," *African Studies Review*, 35(1992): 81-99.

Kendall, Chad, Tommaso Nannicini and Franceso Trebbi, "How Do Voters Respond to Information? Evidence from a Randomized Campaign," *American Economic Review*, 105:1(2015), 322-53.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75:1(2007), 83-119.

Lenz, Gabriel S., "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis," *American Journal of Political Science*, 53:4(2009), 821-37.

Liessem, Verena, and Hans Gersbach, "Incentive Contracts and Elections for Politicians with Multi-Task Problems," SSRN 243518 (2003).

Lindbeck, Assar and Jorgen W. Weibull, "Balanced-budget redistribution as the outcome of political competition," *Public Choice*, 52(1987): 273-97.

MacKinnon, James G. and Halbert White, "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29(1985), 305-25.

McKinnon, Lori Melton, John C. Tedesco and Lynda Lee Kaid, "The Third 1992 Presidential Debate: Channel and Commentary Effects," *Argumentation and Advocacy*, 30:2(1993), 106-18.

Mullainathan, Sendhil, Ebonya Washington and Julia R. Azari, "The impact of electoral debate on public opinions: an experimental investigation of the 2005 New York City mayoral election," in *Political Representation*, Ian Shapiro, Susan C. Stokes, Elisabeth Jean Wood, and Alexander S. Kirshner, eds. (Cambridge University Press, 2010).

Olken, Benjamin A., "Promises and Perils of Pre-Analysis Plans," *Journal of Economic Perspectives*, 29:3(2015), 69-80.

Paluck, Elizabeth Levy and Donald P. Green, "Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda," *American Political Science Review*, 103:4(2009), 622-44.

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

Prat, Andrea, "The Wrong Kind of Transparency," *American Economic Review*, 95:3(2005), 862-77.

Prior, Markus, "Who watches presidential debates? Measurement problems in campaign effects," *Public Opinion Quarterly*, 76:2(2012), 350-63.

Reinikka, Ritva, and Jakob Svensson, "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda," *Journal of the European Economic Association*, 3(2005): 259-67.

Shear, Michael D., "After Debate, a Torrent of Criticism for Obama," *New York Times*, October 5, 2012.

Snyder, James M., and David Strömberg, "Press Coverage and Political Accountability," *Journal of Political Economy*, 118:2(2010), 355-408.

Strömberg, David, "Radios Impact on Public Spending," *Quarterly Journal of Economics*, 119:1(2004), 189-221.

Strömberg, David, "Media Coverage and Political Accountability: Theory and Evidence," in *Handbook of Media Economics*, Simon Anderson, David Strömberg and Joel Waldfogel, eds. (North Holland, 2015).

Wald, Kenneth D. and Michael B. Lupfer, "The Presidential Debate as a Civics Lesson," *Public Opinion Quarterly*, 42:3(1978), 342-53.

Wantchekon, Leonard, "Clientelism and Voting Behavior: Evidence from A Field Experiment In Benin," *World Politics*, 55(2003), 399-422.

Wantchekon, Leonard, Gabriel Lopez-Moctezuma, Thomas Fujiwara, Cecilia Pe Lero and Daniel Rubenson, "Policy Deliberation and Voting Behavior: A Campaign Experiment in the Philippines," manuscript (2014).

Westfall, P., and Young, S., *Resampling-Based Multiple Testing*, (Wiley, 1993).

Zwane, A., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Giné, E. Duflo, F. Devoto, B. Crepon and A. Banerjee, "Being surveyed can change later behavior and related parameter estimates," *Proceedings of the National Academy of Sciences*, 108:5(2011), 1821-6.

Table 1: Domain A - Treatment Effects of Polling Center Debate Screenings on Voters

Mean effects index by hypothesis	Treatment effect (standard error)	Per comparison p-value (1 sided)	FWER adjusted p-value (1 sided)
	(1)	(2)	(3)
A1. Exposure to debates increases political knowledge	0.281	0.000**	0.000**
(20 outcomes)	(0.028)		
A2. Exposure to debates increases policy alignment	0.106	0.002**	0.010**
(3 outcomes)	(0.035)		
	0.086	0.023*	0.078+
	(0.043)		
	-0.018	0.718	0.705
	(0.032)		
	0.091	0.028*	0.078+
	(0.048)		
Observations	5,247		

Note: This table presents treatment effect estimates for the public debate screenings on voter behavior, summarized at the hypothesis level. In this analysis: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01 based on one-sided tests in the direction of the hypothesis statement in the pre-analysis plan (PAP); ii) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) robust standard errors clustered by polling center; iv) all specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; v) specifications further include the full set of respondent control variables (gender, age, years of schooling, polygamous marriage, farming occupation and radio ownership); vi) adjustments to control the familywise error rate (FWER) computed following Westfall and Young (1993) and Anderson (2008); and vii) data source is the exit poll survey.

Table 2: Domain A - Treatment Effects for Selected Individual Outcomes

Outcome measures by hypothesis	Control mean (1)	Treatment effect (2)	Standard error (3)	*p*-value (1 sided) (4)	FDR *q*-value (5)	*N* (6)
Panel A1. Political knowledge increases						
i. Knows amount of the constituency facilitation fund (CFF)	0.034	0.140	0.018	0.000**	0.001**	5,400
i. Knows who is entitled to free healthcare (FHC)	0.706	0.056	0.033	0.044*	0.038*	5,399
i. Knows MP job responsibilities (out of 3)	0.555	0.217	0.070	0.001**	0.003**	5,400
ii. Knows which candidate was most educated	0.243	0.159	0.044	0.000**	0.001**	3,097
ii. Knows which candidates had been an MP before	0.490	0.111	0.032	0.000**	0.002**	5,400
ii. Knows candidate with most public office experience	0.319	0.076	0.037	0.021*	0.021*	2,972
iii. Knows SLPP candidate's first priority issue	0.142	0.149	0.028	0.000**	0.001**	5,398
iii. Knows APC candidate's view of FHC	0.252	0.197	0.035	0.000**	0.001**	4,579
iii. Knows PMDC candidate's position on GEB	0.244	0.209	0.053	0.000**	0.001**	3,291
Panel A2. Policy alignment increases						
Voter's priority issue matches that of chosen candidate	0.425	0.090	0.031	0.002**	0.005**	5,160
Voter's view on FHC matches that of chosen candidate	0.394	0.092	0.035	0.004**	0.008**	4,727
Voter's view on GEB matches that of chosen candidate	0.613	-0.024	0.024	0.843	0.432	5,160
Panel A3. Votes for best performing candidate in the debate increase						
Voted for debate winner, as judged by audience	0.803	0.049	0.021	0.012*	0.015*	5,219
Voted for debate winner, as judged by expert panel	0.712	0.011	0.022	0.312	0.218	5,219
Vote share of debate winner, audience pick (NEC returns)	0.711	0.035	0.017	0.033*		206
Vote share of debate winner, expert panel (NEC returns)	0.617	0.035	0.016	0.027*		206
Panel A4. Voting across ethnic-party lines increases						
Voted across ethnic-party lines	0.107	-0.012	0.013	0.811	0.432	4,569
Voted for a different party for MP in 2012 than in 2007	0.163	0.004	0.019	0.425	0.259	4,405
Panel A5. Voter openness to candidates increases						
Voter likeability rank for APC candidate (10 point scale)	6.523	0.161	0.261	0.269	0.199	5,073
Voter likeability rank for PMDC candidate	2.369	0.579	0.273	0.018*	0.020*	3,299
Voter likeability rank for SLPP candidate	4.952	0.229	0.230	0.160	0.115	5,414

Note: This table presents treatment effect estimates for a selection of individual outcome measures under the five hypotheses concerning voter response to public debate screenings (see Appendix Table A.3 for the complete set of domain A estimates). In this analysis: i) significance levels $+ p < 0.10$, $* p < 0.05$, $** p < 0.01$ based on one-sided tests in the direction pre-specified in the PAP; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; iv) additional controls vary by hypothesis from the set (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership) as determined by a pre-specified algorithm that selects the subset with the greatest predictive power in the control group data; v) false discovery rate (FDR) adjustments computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 33 primary and 5 secondary outcomes in domain A; vi) data source is the exit poll survey, except for the last two estimates in panel A3 which use the National Electoral Commission polling center-level returns; and vii) the NEC returns exclude one constituency where the SLPP candidate was disqualified immediately before the election (see footnote 14).

Table 3: Domain B - Treatment Effects of Polling Center Screenings on Candidates

Mean effects index and outcome measures	Control mean	Treatment effect	Standard error	Naïve p-value (2 sided)	FDR q-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
Hypothesis B1. Mean effects index	*0.000*	*0.103*	*0.039*	*0.008***		*5,399*
Received any gift from the APC candidate	0.160	0.011	0.027	0.686	0.666	5,055
Received any gift from the PMDC candidate	0.007	0.013	0.006	0.027*	0.089+	3,220
Received any gift from the SLPP candidate	0.089	0.007	0.020	0.725	0.666	5,397
Value of gift received from APC candidate (in log(value+1))	0.412	0.121	0.098	0.217	0.364	4,989
Value of gift received from PMDC candidate (in log(value+1))	0.014	0.034	0.014	0.016*	0.089+	3,213
Value of gift received from SLPP candidate (in log(value+1))	0.210	0.077	0.063	0.222	0.364	5,347
Voter report of number of APC candidate visits to village	1.292	0.147	0.137	0.285	0.398	5,056
Voter report of number of PMDC candidate visits to village	0.353	0.219	0.093	0.019*	0.089+	3,291
Voter report of number of SLPP candidate visits to village	1.273	0.070	0.186	0.709	0.666	5,399

Note: This table estimates the campaign response by MP candidates who participated in a debate to the allocation of group screenings across polling centers. In this analysis: i) significance levels $+ p < 0.10$, $* p < 0.05$, $** p < 0.01$ based on two-sided tests; ii) the mean effects index in row 1 is constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) the individual outcomes in all other rows are expressed in units natural to the measure; iv) robust standard errors clustered by polling center; v) specifications include stratification bins for the polling center (number of registered voters and distance to next nearest), respondent (youth status and gender) and constituency fixed effects; vi) additional controls determined by analysis of control group data and include gender, age, years of schooling, and radio ownership; vii) adjustments to control the false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); and viii) data source is the exit poll survey.

Table 4: Domain D - Mechanisms Explored via the Individual-level Treatments

Mean effects index	Debate		Get to Know You		Radio Report		Debate vs. GTKY		Debate vs. Radio		Radio vs. GTKY	
	Treatment effect (std error)	Naïve p (1 sided)	Treatment effect (std error)	Naïve p (1 sided)	Treatment effect (std error)	Naïve p (1 sided)	Treatment effect (std error)	Naïve p FDR q (2 sided)	Treatment effect (std error)	Naïve p FDR q (2 sided)	Treatment effect (std error)	Naïve p FDR q (2 sided)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A1. Political knowledge	0.109** (0.021)	0.000	0.041** (0.016)	0.006	0.095** (0.018)	0.000	0.068** (0.022)	0.002 0.011	0.014 (0.018)	0.425 0.436	0.053* (0.022)	0.016 0.068
i. General knowledge	0.175** (0.040)	0.000	0.095** (0.035)	0.005	0.160** (0.045)	0.000	0.079+ (0.043)	0.066 0.174	0.014 (0.034)	0.674 0.585	0.065 (0.050)	0.192 0.275
ii. Candidate characteristics	0.049** (0.019)	0.006	0.068** (0.025)	0.005	0.042* (0.020)	0.021	-0.019 (0.026)	0.455 0.436	0.007 (0.026)	0.793 0.585	-0.026 (0.032)	0.411 0.436
iii. Policy stances	0.127** (0.031)	0.000	-0.003 (0.017)	0.575	0.106** (0.023)	0.000	0.130** (0.028)	0.000 0.001	0.020 (0.026)	0.434 0.436	0.110** (0.026)	0.000 0.001
A2. Policy alignment	0.081** (0.029)	0.004	0.007 (0.027)	0.395	-0.040 (0.024)	0.945	0.074* (0.033)	0.025 0.087	0.121** (0.032)	0.000 0.002	-0.047+ (0.027)	0.083 0.176
A3. Vote for best	0.058+ (0.040)	0.077	0.006 (0.037)	0.440	-0.046 (0.043)	0.851	0.052 (0.045)	0.241 0.291	0.104* (0.052)	0.046 0.136	-0.051 (0.040)	0.203 0.275
A4. Cross party lines	-0.030 (0.035)	0.802	0.004 (0.031)	0.453	0.058 (0.045)	0.103	-0.033 (0.044)	0.447 0.436	-0.088+ (0.050)	0.076 0.176	0.055 (0.042)	0.195 0.275
A5. Openness	0.006 (0.023)	0.395	-0.022 (0.025)	0.812	0.014 (0.030)	0.322	0.029 (0.034)	0.403 0.436	-0.008 (0.033)	0.818 0.585	0.036 (0.029)	0.215 0.275
Observations	1,698		1,695		1,695							

Note: This table considers the suite of treatment arms randomized at the individual-level. It presents results for absolute treatment effects compared to control group data (in Columns 1 to 6) as well as relative effects of the three treatments compared to one another (in Columns 7 to 12). In this analysis: i) significance levels + p <0.10, * p <0.05, ** p <0.01 based on per comparison one-sided tests in the direction prespecified in the PAP in Columns 1 to 6 and two-sided tests in Columns 7 to 12; ii) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) robust standard errors clustered by polling center; iv) specifications include stratification bins for the household (gender and age composition), polling center (number of registered voters and distance to next nearest) and constituency fixed effects; v) additional controls vary by hypothesis from the set (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership) as determined by analysis of the control group data; vi) adjustments to control the false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 24 comparative tests run; and vii) data source is the exit poll survey for the individual treatment sample.

Table 5: Domain E - Treatment Effects of Debate Participation on Accountability

Mean effects index and outcome measures by hypothesis	Control mean	Treatment effect	Std. error max of (OLS, HC$_2$)	Naïve p-value (1 sided)	N
	(1)	(2)	(3)	(4)	(5)
Hypothesis E1. Activity in Parliament, index	*0.000*	*0.129*	*0.320*	*0.345*	*28*
Percent of 2012-13 sittings attended (out of 57 sittings in total)	76.692	3.371	3.003	0.137	28
Total number of public comments in Parliamentary sittings 2012-13	4.286	-1.569	2.224	0.780	28
Committee membership (total number)	3.929	0.524	0.625	0.186	28
Hypothesis E2. Consistency with pre-election promises, index	*0.000*	*-0.219*	*0.226*	*0.829*	*28*
Total public comments in priority sector agenda items	0.154	-0.170	0.166	0.842	27
Membership in priority sector committee	0.231	0.201	0.187	0.147	27
Constituent assessment of focus on priority sector	0.571	-0.343	0.150	0.984	27
Hypothesis E3. Constituency engagement, index	*0.000*	*0.779*	*0.307*	*0.009***	*28*
Total number of constituent visits	2.915	1.316	0.619	0.022*	28
Total number of public meetings held with constituents	1.018	1.089	0.606	0.043*	28
Total number of sectors that constituents assess good performance	1.417	0.882	0.476	0.039*	28
Health clinic staff reported good performance in health	0.202	0.187	0.135	0.089+	28
Hypothesis E4. CFF spending, index	*0.000*	*1.097*	*0.619*	*0.045**	*28*
Development spending verified in the field (as % of 2012 CFF allotment)	35.560	54.738	31.707	0.049*	27
Domain E. All 11 outcomes, index	*0.000*	*0.359*	*0.166*	*0.021**	*28*

Note: This table leverages the constituency-level randomization to estimate the effects of participating in a debate as a candidate on the subsequent performance of the elected MP in office. In this analysis: i) significance levels + p <0.10, * p <0.05, ** p <0.01 based on one-sided tests in the direction prespecified in the PAP; ii) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; iii) estimates for individual outcomes are expressed in units natural to the measure; iv) the standard error presented is the maximum value of conventional OLS and bias corrected HC$_2$ estimators in MacKinnon and White (1985), following discussion in Angrist and Pischke (2009); v) specifications include stratification bins for the constituency (3 bins of ethnic-party bias), MP gender and an indicator for whether the MP held an elected position in the past; and vi) missing values for hypothesis E2 concern one control MP who did not provide a pre-election priority sector and for hypothesis E4 concern one treated MP who did not take office until December 2013 (one year after the election) and thus did not receive the 2012 CFF.

Table 6: Debate Delivery - Group Screening versus Individual Private Viewing

Mean effects index	Group Screening			Individual Viewing	
	Treatment effect (std error)	ATE on compliers (std error)	*N*	Treatment effect (std error)	*N*
	(1)	(2)	(3)	(4)	(5)
A1. Political knowledge	0.324**	0.392**	3,507	0.140**	708
	(0.034)	(0.040)		(0.023)	
i. General political knowledge	0.371**	0.448**	3,507	0.304**	708
	(0.034)	(0.068)		(0.053)	
ii. Candidate characteristics	0.268**	0.324**	3,507	0.031+	708
	(0.039)	(0.047)		(0.023)	
iii. Policy stances	0.348**	0.420**	3,507	0.152**	708
	(0.049)	(0.058)		(0.037)	
A2. Policy alignment	0.129**	0.156**	3,514	0.087*	748
	(0.047)	(0.057)		(0.039)	
A3. Votes for best performer in the debate	0.102+	0.123+	3,510	0.102*	748
	(0.068)	(0.082)		(0.051)	
A4. Cross party lines	-0.022	-0.027	3,511	-0.052	748
	(0.042)	(0.050)		(0.050)	
A5. Openness	0.059	0.071	3,514	-0.002	748
	(0.065)	(0.078)		(0.031)	

Note: This table estimates voter response to debates when delivered via public group screenings (Columns 1 to 3) as compared to individual private viewing (Columns 4 and 5). To make the two distinct sets of estimates as comparable as possible, the group screening sample is limited to the 8 constituencies where the individual-level treatments were also administered, and the individual level estimates are limited to the debates treatment arm versus the "pure" control groups. In this analysis: i) significance levels + p <0.10, * p <0.05, ** p <0.01 based on one-sided per comparison tests; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the randomization procedure and constituency fixed effects; iv) additional controls vary by dataset and hypothesis from the set (gender, age, years of schooling, polygamous marriage, farming occupation and radio ownership); v) hypothesis-level mean effects indices are constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units, with missing values for component measures imputed at random assignment group means; and vi) Column 2 reports 2SLS estimates of the average treatment effect (ATE) on compliers instrumenting group screening exposure with treatment assignment.