DEBATES

The Impact of Voter Knowledge Initiatives in Sierra Leone

Kelly Bidwell Katherine Casey* Rachel Glennerster

Jameel Poverty Action Lab Stanford GSB Jameel Poverty Action Lab

December 15, 2014

Abstract

Debates between candidates for public office have a rich historical tradition and remain an integral part of campaign strategy in contemporary politics. There is, however, no definitive evidence of whether debates affect actual voting behavior. Limited media penetration implies that the effects of publicizing debates could be more pronounced, persistent and more directly linked to electoral outcomes in the developing world. This paper experimentally manipulates citizen exposure to debates between Parliamentary candidates in Sierra Leone to measure their impacts on, and the interconnections between, voter behavior, campaign spending, and the performance of elected politicians. We find evidence of strong positive impacts on voter's political knowledge, policy alignment, and importantly, their votes cast on Election Day. We then document an endogenous spending response by participating candidates, who increased their campaign effort in communities where videotapes of the debates were screened in large public gatherings. A complementary series of individual treatment arms unpacks the different types of information delivered by the debates, and finds evidence that voters respond to both candidate charisma and "hard facts" about policy stances and professional qualification. Ongoing work tests whether the publicity of debate screenings helps solve candidate commitment problems and thereby enhances the consistency between pre-election promises and post-election behavior.

1   Introduction

Debates between candidates for public office have a rich historical tradition and remain an integral part of campaign strategy in contemporary politics. In the United States, famous and influential candidate debates include the Lincoln-Douglas Senatorial debates of 1858 and the first televised Presidential debate between Kennedy and Nixon in 1960. More recently, the first Obama-Romney debate of 2012 generated widespread criticism of the President's performance and much speculation about its implications for his re-election prospects (e.g. Shear 2012). Such debates constitute significant campaign events: a large share of the voting public tunes in to view 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. There is, however, no definitive evidence of whether debates have any real impact on subsequent vote shares. Moving to the developing world, the markedly lower penetration of mass media suggests that the effects of publicizing debates could be more pronounced, persistent and more directly linked to electoral outcomes. Given the paucity of reliable political information in these markets, allowing candidates to stand on equal footing and express their views on key policy issues could further facilitate the election of more competent candidates and enhance the subsequent accountability of elected officials. This paper evaluates these claims in the context of a recent election in Sierra Leone.

We worked in partnership with the civil society organization Search for Common Ground to host and film structured, inter-party debates between Parliamentary (MP) candidates in fourteen competitive constituencies in the lead up to the November 2012 General Election. We then experimentally manipulated citizen exposure to the debates to measure their impacts on, and the interconnections between, voter behavior, campaign spending, and the performance of elected politicians. The paper makes four contributions. First, we randomly allocated a "road show" across polling centers that screened videotapes of the debates in large public gatherings to capture effects on voters. We document strong positive impacts of watching debates on political knowledge, policy alignment and importantly, actual votes cast on Election Day. Second, we document an endogenous spending response by participating candidates who increased their campaign effort in villages where these screenings were held. Documenting this indirect result makes a small advance from the typical partial equilibrium orientation of field experiments towards capturing a more general equilibrium effect. Third, we used a series of individual-level treatment arms to unpack the causal mechanisms driving the observed changes in voter behavior, isolating the role of more superficial attributes such as candidate appearance and charisma from that of "hard facts" about policy stance and professional qualifications. Fourth, ongoing work evaluates whether the publicity of the debate

screenings can help solve candidate commitment problems and thereby enhance consistency between pre-election promises and post-election behavior in office. Two additional aspects of the research are worth noting: the experiments were in part designed to test specific theoretical predictions from our earlier work; and the empirical analysis is governed by publicly registered pre-analysis plans.

Large groups of voters were exposed to films of the MP debates via a mobile cinema that visited 112 of 224 randomly selected polling centers in the six weeks before the election. We find strong positive impacts of watching the debates on voter knowledge—of politics in general, and of specific candidate attributes 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, these gains in voter knowledge translate into actual changes in votes cast, where we document a five percentage point average increase in vote shares for the candidates who performed best during the debates. This effect is positive and significant in both our exit poll data and in the National Electoral Commission's official voting returns. In the context of strong historical ties between ethnic groups and political parties, we find no net impact of debates on voting across ethnic-party lines, however document heterogeneous effects along the dimension of ethnic loyalty that are consistent with theoretical predictions from Casey (2014).

As one might expect of any intervention that alters voter behavior on a nontrivial scale, we next document an endogenous campaign response by candidates who participated in the debates. While the candidates were not informed of which polling centers were assigned to treatment, the debate screenings were large public gatherings in areas with limited entertainment options, making their locations fairly easy to track down. We find evidence that candidates increased their campaign effort, as measured by gift giving, the monetary value of gifts, and the number of visits to communities, in villages where the screenings were held. The increase in expenditure is consistent with a "swing" voter investment model if the debates made exposed areas appear more competitive, by making expected vote margins narrower and/or more uncertain.[1] Tracking the endogenous response of MP candidates to the road show treatment aims to move one small step from a partial towards a more general equilibrium approach. While not similar in design, the goal of documenting the indirect effects of the experiment in its market is similar in spirit to Crépon et al (2013).

Ongoing work evaluates whether the publicity of the debate screenings can help solve candidate commitment problems and thereby enhance consistency between pre-election promises and the post-election behavior of winning MPs. The motivating idea is that it is more dif-

---

[1]See Lindbeck and Weibull 1987, Dixit and Londregan 1996, 1998, and Bardhan and Mookherjee 2010; and Casey 2014 for application to ethnic politics.

2

ficult to renege on campaign promises when constituents have a very public and enduring record of the candidate's own statements. To evaluate these links to electoral accountability, the 14 constituencies that participated in the debate were randomly selected from what we anticipated would be the 28 most competitive races. We are currently tracking the performance of all 28 winning MPs in terms of attendance and participation in Parliamentary sittings, allocation of the constituency facilitation fund earmarked for the development of and transport to their jurisdictions, and engagement with their constituents.

The final set of research questions aims to isolate which of the different types of information revealed by debates—from superficial attributes such as appearance and charisma, to more "hard facts" about policy stance and professional qualifications—drives the observed changes in voter behavior. A series of treatment arms administered at the individual level unpacks these channels of effect. Specifically, some voters were exposed to brief "get to know you" videos of the candidates speaking informally about themselves and their hobbies, which capture persona but not professional or policy information. Others listened to a radio report or journalistic summary of the debate, which articulated all the "hard facts" about policy stances and professional qualifications that arose during the debates, but delivered no information about persona. Still others watched the full debate on a tablet device. Control respondents were surveyed at the time of treatment and the election, and another group was surveyed only at the election. We find evidence that voters update their views of candidates and policy in response to information regarding persona and hard facts, but that only the combination of the two delivered by the debates moves them into better policy alignment and triggers changes in actual voting choices. Comparing the two sets of controls allows us to further isolate the role of priming (Zwane et al. 2011), where we find that the experience of being surveyed accounts for roughly one third of the overall effect on general political knowledge. Reassuringly, survey priming does not contribute to knowledge of specific candidate attributes or policy stances, nor does it move voters to update their policy preferences or vote choice.

There is a large literature concerning the impact of debates in American politics (see for example, Jamieson and Birdsell 1990). Much of this work is limited to panel opinion polls, with the familiar identification challenges. Two more recent studies of interest use randomized experiments. Mullainathan, Washington and Azari (2010) telephoned New York City voters to request that they watch the final 2005 mayoral debate or a placebo public broadcasting program, and find no evidence of meaningful changes in political opinion. Druckman (2003) revisits the Kennedy-Nixon debate with a lab experiment that exposed contemporary students to either a televised or audio only recording of the debate. He finds that television primes people to focus more on personality in evaluating candidates and enhances learning.

3

Our individual-level treatments delivered via tablet devices contribute to this line of research by testing for the impacts of debates in an information poor political environment; unpacking voter responses to multiple different slices of information delivered by debates; and documenting effects of debate exposure on actual voting returns. We view the scale and intensity of the polling center level public screenings as a new contribution. Interestingly, we find an effect size of these group screenings on vote shares that is markedly similar to what Gerber et al. (2011) document in a Texan gubernatorial race for the most intense "dose" of televised campaign advertising on voting intentions. The effects in the US, however, dissipate very rapidly, reduced to zero in a matter of days, whereas the impact of our debates persisted over a range of one to six weeks after treatment exposure and affected actual vote choices on Election Day. We are not aware of any other study that links debates to endogenous campaign responses by candidates, nor to accountability effects over elected officials.

In the field of development economics, our approach of working with political candidates in the course of their actual campaigns follows in the tradition Wantchekon (2003) and Fujiwara and Wantchekon (forthcoming) who convinced candidates to randomly alter their campaign platform and the allocation of town hall meetings, respectively. The latter experiment finds that public deliberation between a single candidate's representative and his constituents decreased the prevalence of clientelism in recent Presidential elections in Benin. We instead focus on the interaction between rival candidates from different parties, where the head-to-head debates were designed to reveal information about the relative quality and policy differences between candidates. Testing the efficacy of debates further contributes to the literature exploring the impacts of information of voting. Ferraz and Finan (2008) and Banerjee et al (2011), among others, show that providing specific information about incumbent performance and candidate qualifications can have large effects on voting. Debates are distinctive in that they provide more general and comprehensive information about candidates, including information about persuasion and charisma, which can be considered productive attributes of an effective legislator. Moreover, if information about an arguably sufficient competence statistic—like corruption—is not available, the generality of debates could further be important for two 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 et al 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 public statements.

The rest of this paper is structured as follows. Section 2 explains the institutional context,

research design and econometric specifications. Section 3 discusses evidence for treatment effects on voters, candidates and elected officials. Section 4 explores survey priming, treatment effect heterogeneity and social mobilization. Section 5 concludes with policy and cost-benefit considerations.

2 Context and Research Design

2.1 Institutions and Treatments

Sierra Leone has 112 Parliamentary constituencies, which are single member jurisdictions elected by first-past-the-post plurality. The winning Member of Parliament (MP) represents the local area in the national legislature. In these elections, the ethnic composition of the 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 tribes in the South, most prominently the Mende (see Kandeh 1992). As an example of the contemporary strength of these ties, in the control group of this study, 89 percent of citizens reported voting for the MP from the party that is historically associated with their ethnic group.

Casey (2014) suggests that one reason for the strength and persistence of these ethnic-party ties is that voters have little alternative information on which to base their vote. If true, then providing citizens with better information about the candidates themselves could broaden their decision calculus, potentially making voting more responsive to things like professional qualifications, past performance or persuasiveness. Yet how to deliver such information in a way that is credible, engaging, and accessible to voters who are predominantly poor, illiterate and geographically remote remains an open question. It is out of this challenge that our interest in publicizing debates between candidates from competing parties initially arose, although this study quickly expanded into a broad exploration of debates and their effects on political markets. This paper thus explores whether and how debates affect voting behavior, how such changes might in turn affect the allocation of campaign expenditure by candidates, and over the longer run, whether debates can enhance the electoral accountability of elected officials.

To answer these questions, we worked in partnership with Search for Common Ground (SFCG) to host, film and disseminate debates between Parliamentary candidates in the lead up to the November 2012 elections in Sierra Leone. Before the candidates were officially announced, we randomly selected 14 constituencies from what we estimated would be the 28

most competitive races, stratifying on the degree of ethnic-party bias favoring one party over the other. 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)—that were contesting a given seat to participate in the debate. No other parties won seats in the previous election, and these three parties respectively held 59, 39 and 9 percent of the existing seats.

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, where each candidate was allowed two to three minutes to respond. The first policy question concerned the candidate's top priority for additional government spending. The second covered plans for how to spend the constituency facilitation fund (CFF), which is an untied 43.8 million Leones (approximately US$ 11K) grant given annually to each MP for transport to and development of his or her constituency. The third issue asked for the candidate's strategy to uplift the youth, where "youth" is defined by the government as 18 to 35 year old adults. 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 or against 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, which varied by constituency and were tailored to prominent issues in the local area. All debates were conducted in Krio, Sierra Leone's lingua franca.

Within the fourteen constituencies selected for participation in the debates, we first allocated polling centers to the group screening treatment and control group. All citizens had to register anew for this election, and these centers—typically a primary school or community center—are where they went to register and then later to vote. 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 polling center (2.4 miles on average). SFCG took videotapes of the debates on a "road show" to 112 of these polling centers, selected randomly. In constituencies where there were a sufficient number of polling centers left over, we randomly allocated 40 of the remaining larger and closer together polling centers into the individual-level treatment arms group. Note that the individual treatments were

thus administered in a completely separate set of villages from the public screenings. A few months before administering any intervention or survey, we conducted a household listing of registered voters in all 264 polling centers to develop the sampling frame for individual 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 protocol for these screenings was 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; the video was played a second time with or without translation; and a secondary screening was held in the largest accessible satellite village earlier in the day, in most cases without translation. A total of 112 primary and 85 secondary screenings were held, and we estimate that roughly 19,000 people were thereby exposed to one of the debates.

At the time of screening in treated polling centers, some of the 25 respondents who received attendance incentives were also surveyed, where specifically: i) 12 completed surveys both before and after the screening; ii) 4 completed only after screening surveys; and iii) the 9 remaining were not surveyed but were contacted only to deliver the incentive. We later conducted exit polls on Election Day and the days immediately after in all 224 treatment 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 thus be estimating intention to treat effects, where 82% of exit poll respondents indicated that they had attended a debate screening, as did 4% of those in the control group.

Within each of the polling centers assigned to individual-level treatments, households were divided into those with only female registered voters, only male, and both male and female registered voters (based on the earlier household listing exercise). We randomly assigned the five arms to households within each of these bins, and randomly selected respondents within each household to receive the individual-level treatments and/or survey(s). The five treatment arms at the individual level were as follows: (i) debate treatment, where individuals were shown the exact same debate screened in polling centers on a personal handheld device; (ii) "getting to know you" treatment, where individuals were shown a short video of the same two candidates speaking informally about their hobbies and interests; (iii) "radio

7

report" treatment, where individuals listened to a recording of an independent moderator or journalist summarizing 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 treatment arm was administered that involved a lab-in-the-field experiment that exposed voters to photos and 20 second video clips of candidates (analyzed in our related work). This treatment did not relay any political information and the media snippets did not cover any candidates from the voter's own constituency, and is thus grouped with the control subsamples.

There were 400 individuals assigned per arm, save 600 in the surveyed control group. Unlike for the polling center level intervention, the exact same respondents who participated in the individual treatment arms were relocated in the exit polls. As we had perfect compliance and near zero attrition, average treatment effect estimates for the individual treatment arms are comparable to treatment on the treated effects.

2.2   Hypotheses and Econometric Framework

We registered the first and main pre-analysis plan (PAP) governing this analysis with the Abdul Latif Jameel Poverty Action Lab on November 20, 2012 before fieldwork for the exit poll, which is the primary source of data for this analysis, was completed. We later migrated the PAP when the American Economic Association's randomized control trial registry opened, where our entry can be found here: https://www.socialscienceregistry.org/trials/26. To establish an iterative process that used earlier data analysis to inform subsequent analysis, and in light of data collection efforts spanning 18 months, we later lodged additional plans and revisions with time stamps to the registry (explained further below). Overall, the plans established five research domains with hypotheses under each domain; grouped outcomes under these hypotheses; and specified the econometric framework including subgroup analysis, dimensions of heterogeneous effects, and which tests would be one-sided and in which direction.[2]

The PAP commits to reporting treatment effects for all individual outcomes as well as mean effects indices by hypothesis, and then adjusting for multiple inference across hypotheses within a domain and across outcomes within a given hypothesis. Compilation of the mean effects index follows Kling, Liebman and Katz (2007) to first orient each individual

---

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

outcome so that larger numbers imply "better" outcomes, translate each into standard deviation units with reference to the mean and standard error of the control group, and then compute the equally weighted average of all transformed outcomes under a given hypothesis.[3] Following Anderson (2008), we then apply family wise error rate (FWER) adjustments 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. Note that we do not adjust across research domains (e.g. across voters in the polling-center versus individual-level treatment arms), as each domain concerns a distinct sample: covering different actors, datasets and/or randomizations.[4]

The PAP lists the following hypotheses for the first research domain (A), which concerns the effects of the polling center-level debate screenings on voters:

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

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

A5. Exposure to debates enhances voter openness to other parties

Secondary hypotheses: (i) Exposure to debates mobilizes the public and leads to greater turnout; (ii) Exposure to debates increases the perceived legitimacy of elections; and (iii) Exposure to debates increases interest in politics

Analysis of treatment effects for domain A takes the form:

$$Y_{ipc} = \beta_0 + \delta T_{pc} + \mathbf{X}'_{ipc}\mathbf{\Pi} + \mathbf{Z}'_{pc}\mathbf{\Gamma} + \mathbf{W}'_{ipc}\mathbf{\Psi} + c_p + \varepsilon_{ipc} \tag{1}$$

where outcome Y (i.e. 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; $\mathbf{X}$ is a vector of indicator variables

---

[3]Missing values for individual index component outcome measures imputed at the random assignment group mean (see KLK 2007 footnote 11).

[4]Note that the word "domain" is often used to refer to different groups of outcomes tested on the same dataset. Our "domains" are thus quite distinct from that usage and imply a much stronger degree of separation between tests.

that denote the stratification bin from which exit poll respondents were drawn (where the bins were constructed by age and gender); $\mathbf{Z}$ is a vector of indicator variables that denote the stratification bin from which the polling center was drawn (where the bins were constructed by number of registered voters and distance to nearest neighboring center); $\mathbf{W}$ is a set of individual controls determined by a pre-specified algorithm that uses control group data to select the subset of {gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership}[5] that predicts the mean effects index for a given hypothesis at 95% confidence; c is a set of constituency-specific fixed effects (the level of debate and candidates); and ε is an idiosyncratic error term clustered at the polling center level. The coefficient of interest is δ and recall that we are estimating intention to treat effects. Unless otherwise stated, all tests are one-sided in the direction indicated in the statement of the hypothesis. The PAP further specifies the following dimensions of potential heterogeneous effects: (i) competitiveness of constituency; (ii) candidate performance; (iii) lesser known candidates (secondary); and (iv) subgroup analysis by gender, age and fluency in Krio; which are discussed in Section 4.

For the second research domain (B), the PAP lists only one hypothesis concerning the effects of polling center-level debate screenings on candidates:

> B1. Candidate allocation of campaign effort and expenditure is responsive to debate publicity

Here we are interested in whether campaign investment complements or substitutes for treatment allocation, and thus conduct two-sided tests. In terms of theoretical motivation, if debates influence vote shares in a way that makes the races appear more competitive, then they could attract greater resources to areas where the debates were screened. This would be consistent with a "swing voter" model of political redistribution (Lindbeck and Weibull 1987), which Casey (2014) finds evidence for in Sierra Leone. Recall that while we did not inform the candidates of which polling centers were assigned to treatment or control, the screenings were large public events whose locations would not have been difficult to track down after they occurred. This is thus an endogenous response of candidates to the polling center-level treatment assignment. Treatment effects in domain A concerning voters thus capture the combination of exposure to debate and the campaign response. The econometric specification is exactly the same as presented in Equation (1), save the outcomes are linked to individual candidates: e.g., an outcome Y (i.e. 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.

---

[5]Interest in politics was originally part of this set and listed as a dimension of heterogeneous effects, but is now omitted as it appears as an outcome under one of the secondary hypotheses.

The PAP repeats the hypothesis above, only now applied to political parties more generally as opposed to individual candidates, to establish the third research domain (C). Data for this domain were collected in a community-level survey that accompanied the voter-level exit polls, implying that there are many fewer (by an order of magnitude) observations for this analysis than for domain B. 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 under study. The hypothesis covers additional outcome measures, like political rallies and number of posters displayed, that apply only at the community-level.

For the fourth research domain (D), we registered a separate PAP to govern the analysis of the individual treatment arms. The hypotheses and outcomes are the same as those specified for domain A above, but we are now interested in the absolute treatment 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 net or relative effect of each treatment arm compared to the other treatments. Analysis of individual-level treatment arms takes the form:

$$Y_{ihtpc} = \beta_0 + \delta T_{htpc} + \mathbf{X}'_{hpc}\mathbf{\Pi} + \mathbf{Z}'_{pc}\mathbf{\Gamma} + \mathbf{W}'_{ihtpc}\mathbf{\Psi} + c_p + \varepsilon_{ihtpc} \qquad (2)$$

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; $\mathbf{X}$ is a vector of indicator variables that denote the stratification bin from which the household was drawn (where the bins were determined by the gender composition of registered voters); and $\mathbf{Z}$, $\mathbf{W}$, c and ε remain as defined in (1). For each treatment arm, the coefficient of interest is δ_t, the average treatment effect for treatment t compared to the control group. The control group is defined as respondents in both the surveyed and "pure" control arms as well as participants in the sixth lab-in-the-field arm (who received no information about these MP races). We further test a series of hypotheses about the relative effects of the different treatment arms that take the form $\delta_t \neq \delta_{\text{-}t}$. Tests of average treatment effects are one-sided in the direction of the hypothesis statement, and tests of relative effects are two-sided.

The fifth and final research domain (E), explores medium term accountability effects of the debate treatment on the candidates who won the seat. This analysis operates at the highest level of aggregation, where we randomly allocated 14 of the 28 constituencies into debate participation. All candidates in the 28 constituencies were surveyed pre-election, and the 28 winning MPs were then tracked over their first 18 months in office. Their performance

outcomes were drawn from Parliamentary administrative records, MP self-reports, and a set of surveys fielded in their home constituencies. There are four hypotheses in this domain:

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

These tests will be one sided in the direction of better performance. Analysis remains ongoing.

As referenced above, we in total lodged four PAPs in an iterative process that tracked the sequential analysis of the many datasets we collected. The important thing to note 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 collection of the primary data, the exit poll, was completed; and those for domain E were lodged in June 2014 before the constituency-level fieldwork tracking the activity of winning MPs was completed. In between these two, one additional plan sets out the mechanisms related to the individual treatment arms analysis, and a second transparently records updates and revisions to the first PAP as analysis of earlier data collection efforts informed subsequent analysis. All revisions are carefully documented in the appendix and AEA registry. There were three substantive revisions to the first PAP worth noting here. First, we "demoted" the hypothesis about effects on turnout from primary to secondary after official election results were published revealing very high (87.3%) 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. Third, we added analysis of survey priming, however its earlier omission was a simple oversight as the original research design explicitly includes surveyed and pure controls to capture these effects. Throughout the rest of the paper, we clearly indicate a small number of analyses that were not pre-specified and should be thus considered exploratory rather than confirmatory in nature.

3 Estimated Treatment Effects by Research Domain

3.1 Effects of Debate Group Screenings on Voters (Domain A)

Table 1 presents summary results at the hypothesis level regarding effects of group exposure to the debates via the mobile cinema road show. We find evidence for strong positive treatment effects for four out of the five primary hypotheses concerning impacts on voter behavior. For the first hypothesis concerning political knowledge, the treatment coefficient for the mean effects index in row 1 suggests that watching debates increases political knowledge on average by 0.278 standard deviation units (standard error 0.027) across the 20 individual outcome measures included. The second column displays the per comparison p-values based on one-sided tests in the direction pre-specified by the hypothesis statement in the PAP. This per comparison, or "naïve," p-value is appropriate for any researcher with an a priori interest in the specific hypothesis presented (see discussion in Kling, Liebman and Katz 2007). For political knowledge, the p-value indicates significance at greater than 99% confidence. The final column adjusts p-values to account for the fact that we are simultaneously testing five distinct hypotheses on the same dataset. These family wise error rate (FWER) adjustments are quite conservative, as may be appropriate for assessing the overall effectiveness of the program and making policy decisions about whether additional funds should be allocated towards scaling up implementation (see discussion in Anderson 2008). This adjustment does not affect the significance of the effect on political knowledge (FWER p-value of 0.000).

To understand what drives the observed increase in political knowledge, Panel A1 of Table 2 unpacks the mean effects index into its individual components, showing results in units natural to the specific outcome. We break this hypothesis into three knowledge subsets covering general political information, candidate attributes and candidate policy stances. The first four outcomes, preceded by (i), capture effects on general political knowledge. In the first row, the first column shows that the proportion of voters in control polling centers who could correctly state the amount in the constituency facilitation fund (CFF) was only 0.034 or 3.4%. Note that answers were coded to correct for a generous range around the actual 43.8 M Leone figure (roughly US$ 11 K). Column 2 presents a treatment effect estimate of 0.140 (s.e. 0.018) or 14 percentage points, indicating that the proportion of voters who knew the amount in the CFF increased fivefold with exposure to treatment. This effect is highly significant, with a per comparison or "naïve" p-value of 0.000. Similarly, in row 2 the proportion of voters 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%. Row 3 shows that the treatment effect estimate for correctly stating that the Gender Equity Bill would reserve between 2 and 4 out of 10 seats for women is positive, but not statistically significant. Lastly, in

row 4, the number of correctly reported roles and responsibilities of an MP also increased significantly. Note that the statistical strength of these results is largely unchanged when we adjust p-values to control the false discovery rate (FDR) across all 43 primary plus 5 secondary outcomes within domain A (column 5).[6] While the FDR adjustments generally inflate p-values throughout the table, for knowledge of the GEB quota and some other larger p-values they actually adjust downwards, which can be expected when there are many true rejections in the test set.

The next set of individual outcomes under hypothesis A1 concerns voter knowledge of specific candidate attributes, demarcated by (ii) in Table 2. Here we again find strong positive treatment effects, which are significant at 99% confidence for 5 of the 7 outcomes measured. As some examples, 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 (s.e. 4.4 and 3.2, respectively). Voter ability to name candidates also increased significantly for all three parties.

For the third and final area of political knowledge, we find evidence that voter knowledge of candidate policy positions also increased markedly. For each of (up to) three participating candidates, on each of three national policy issues, voter ability to correctly place the candidate on the specific policy spectrum increased significantly (at 99% confidence) for 8 of 9 estimates. As some examples, the proportion of voters who could correctly identify the SLPP candidate's first priority for government spending doubled, from 14.2 to 29.1% (s.e. 2.8); the proportion who knew the APC candidate's view on whether free healthcare was being well implemented or needed to be significantly reformed rose from 25.2 to 44.9% (s.e. 3.5); 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% (s.e. 5.3).

Together, these results suggest that exposure to debates led to substantial improvements in voter knowledge. Recall that respondents experienced a one to six week lag between exposure to debates and the exit polls, indicating that these gains in knowledge were relatively persistent. The next natural question is thus whether these knowledge gains translated into changes in voting choices on Election Day.

Returning to Table 1, estimates in the second row suggest that voters acted on the gains in policy knowledge to move into better policy alignment with their preferred candidate. Alignment is measured as a match between the voter's reported policy position in the exit poll and that of the candidate they voted for as expressed by the candidate during the debate.

---

[6]This adjustment across all outcomes within Domain A is actually more conservative than what we specified in the PAP, which was to adjust only across outcomes within each of the five hypotheses.

The treatment effect coefficient suggests that debate exposure increased policy alignment by 0.106 standard deviation units (s.e. 0.035) on average across the three major national policy issues discussed during the debates. This effect is significant at 99% confidence for both per comparison and FWER controlled p-values. To provide a sense of magnitude, consider estimates in Panel A2 of Table 2. The empirical match between the voter's first priority issue and the view articulated by their preferred candidate during the debate increased by 9.0 percentage points (s.e. 3.1) on a base of 42.5%. We find similar effects on alignment with respect to free healthcare, where alignment increased by 9.2 percentage points (s.e. 3.5) on a base of 39.4%. We find no effect for the gender equity bill, although note that alignment was already markedly higher in the control group (61.3%) and there was little divergence in views expressed by candidates during the debates, where only one candidate expressly strong objection to the bill.

What drives this improvement in policy alignment: choosing aligned candidates or persuasion and opinion updating? Using first priority issue as an example, alignment improves if: i) voters who prefer education select a candidate who also supports education; and/or ii) voters are persuaded to adopt the policy views of their preferred candidate and update their position that education is the most important sector if that is what their favorite candidate advocates. The former is what one would expect from canonical minimum policy distance voting models (e.g. Downs 1957). By contrast, Abramowitz (1978) suggests that the latter was at work in the Carter-Ford Presidential debates of 1976, where voters adopted their preferred candidate's view on unemployment policy after watching the two candidates debate the issue. One way to distinguish these channels in our data is to compare changes in alignment for voters who selected the same party in the previous Parliamentary elections (in 2007), suggestive of persuasion; to those for voters who changed their party choice for this election, suggestive of selecting a more aligned candidate. For all three issues, the coefficient on the interaction between voting for the same party in both elections and treatment is not statistically distinguishable from zero. Estimating effects on alignment for the two types of voters separately, the coefficient on treatment is larger in magnitude for the persuasion subsample for first priority issue but smaller for views on free healthcare (all four estimates remain positive and three statistically significant in the subsamples). This suggests that improvements in alignment are roughly comparable for both types of voter, or that both channels are in fact at work. Note, however, that the population size of the group who supported the same party over time is much larger (encompassing 83% of the sample), suggesting that persuasion is the more empirically substantive channel.

The treatment effect of ultimate interest is on the third hypothesis, where we find significant positive impacts on votes cast for the candidate who performed best during the debates.

Specifically, estimates for the mean effect index in row 3 of Table 1 suggest an increase of 0.086 standard deviation units (s.e. 0.043), significant at 97.5% confidence on a per comparison basis and 91.9% confidence under FWER adjustment. Table 2 breaks this index into its two component measures, where the debate winner is determined by the audience and our expert panel, respectively. Audience votes were recorded in a survey that immediately followed the implementation of the group-level screening. The expert panel consists of over thirty members of government and civil society who watched the debate videos and scored each candidate's response to each debate question. These two sets of evaluations coincided on who performed best in 10 out of the 14 debates. Where they diverged, 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 zero seats nationwide). Table 2 reports treatment effects for these two measures in our exit poll data in Panel A3 and in the National Electoral Commission's (NEC) official polling-center level returns at the bottom of the table.[7] Note that the correlation between party vote shares measured across the two datasets is .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 the best, measured in the exit poll data, where we see a 4.9 percentage point (s.e. 2.1) increase in votes for the debate winner. Note that vote shares for these candidates were already high, at 71% in the NEC returns for the control group polling centers, indicating that in this set of constituencies, the candidate who was locally popular tended to also perform best during the debates.

For hypothesis A4, we find no evidence for treatment effects on voter willingness to vote across ethnic-party lines. In Table 1, the treatment effect coefficient for the mean effects index is small in magnitude and not statistically distinguishable from zero, as are all three estimates for the associated individual outcome measures in Table 2. How can we reconcile a 5 percentage point shift in votes toward the winner of the debate, with no commensurate change in voting across ethnic-party lines? Here exploration of heterogeneous treatment effects along two separate dimensions are informative, although note that neither were included in the PAP so their analysis should be considered exploratory and not confirmatory in nature.

---

[7]The NEC sample excludes constituency 15 where: the SLPP candidate was disqualified immediately before the Election, but his name remained on the ballot; 48% of ballots cast were deemed invalid (many of which were likely SLPP votes); and the winner was determined via the courts. Treatment effect estimates remain largely unchanged when this constituency is included (0.032, s.e. 0.016* for audience best and 0.032, s.e. 0.015* for expert best, N=224).

First, the treatment effect on votes for the debate winner are inversely proportional to both the strength of the voters' ethnic loyalty to that candidate's party and their population share of the study sample. Specifically, for voters historically aligned with the party of the debate winner, 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, roughly 12% of the sample are composed of voters from ethnic groups that do not have strong historical ties to either party. About half (57%) of these voters chose the debate winner in the control sample (as one would expect if they were truly unaffiliated). The point estimate on the treatment effect for this group is substantially larger, at 10.1 percentage points (s.e. 8.4), but not statistically significant (one sided p-value of 0.115). Note that these unaffiliated voters are excluded from the sample for the crossing ethnic-party lines estimate, as they do not have an ethnic choice to move away from. And finally, voters from ethnic groups historically affiliated with the rival party (i.e. the candidate running against the debate winner), represent only 7% of the sample and had a treatment effect estimate of 10.6 percentage points (s.e. 7.5), which is significant at 92% confidence in a one sided test. Thus, while the net effect of voting across party lines is zero, heterogeneous effects on switching one's vote to the debate winner are concentrated in the small group of voters historically affiliated with the rival tribe, who are indeed moved to cross party lines. For the second test, we focus on the single constituency where the audience deemed that the "outsider" candidate (who received only 26% of the votes in the control group) was the debate winner. Here 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) and significant at 94.8% confidence in a one-sided test. These related pieces of evidence are consistent with the theoretical model in Casey (2014), where better information about candidates increases voter willingness to vote across ethnic-party lines, but only in the event that the rival party runs a sufficiently superior quality candidate.

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 97.1% confidence based on unadjusted p-values and 92.1% confidence based on FWER adjusted p-values. This index compiles information from 10 point likeability scales, where all five treatment effect estimates in the individual outcomes are positive (in Table 2) 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 for all candidates is an important feature for securing candidate participation in future debates.

3.2 Endogenous Response by Candidates and Other Party Officials (Domains B and C)

Domain B explores whether candidates altered their campaign strategy in response to the debates road show, given its strong effects on voters' political knowledge and preferences. Table 3 presents evidence that candidate campaign spending serves as a complement to the publicity of the polling center debates screenings. In the first row, the treatment effect for the mean effects index is 0.103 standard deviation units (s.e. 0.039), which is significant at 99% confidence using a two-sided test. The next nine rows unpack this index into its component measures and present treatment effect estimates for each individual outcome, per comparison or naïve p values, and false discovery rate (FDR) controlled q values that adjust for multiple inference. Three things are worth noting. First, all nine treatment effect estimates – covering candidates from each of three parties and each of three campaign outcomes considered – 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 was reported to have visited the village, all with reference to the weeks leading up to the Election. Second, the response by candidates from the two major parties, the APC and the SLPP, is roughly proportional when measured as the percentage increase on their base level of spending in control communities. Third, the candidates who appeared to have responded most strongly with complementary spending are those from the PMDC, the third party that generally had less of a chance of winning seats.

What drives this reallocation of campaign effort? One potential 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 "swing voter" campaign investment model (see for example Lindbeck and Weibull 1987). Extending the exploratory analysis above (and again noting that this 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 (0.41 compared to 0.08 standard deviation units), 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 rest of the sample, 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 the vote shares in screening villages more uncertain, as recall that the actual impact of the debates on voting was not revealed until Election Day. This is consistent with the model in Casey (2014) that by increasing voter responsiveness to individual candidate attributes, information makes it harder for candidates to infer vote shares from the ethnic composition of a locality, which widens the set of potentially competitive areas.

Table 4 presents results for Domain C, about whether other party officials beyond the candidates directly involved in the debates responded to the publicity of the road show. We find little evidence that other party officials, including 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. While the treatment effect for the mean effects index is positive in sign (0.082 standard deviation units), it is not significant at conventional levels (s.e. 0.052 and p-value 0.113). Similarly, while the majority (16 of 21) of treatment effect estimates for the individual outcomes are positive in sign, none are significant at conventional levels. One interpretation of these results is as a pseudo placebo test, where candidates for offices not involved in the debates would not be expected to alter their campaign strategy in response to the 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 MP candidates. While this seems plausible, we do not want to place too much weight on this interpretation, for two reasons. The sample size for this community-level survey is small (224 communities), so power to reject the null is limited. And, the community survey questions bundled together the campaign efforts of all party officials and candidates for all offices, which includes Parliament, so they do not clearly exclude the MP candidates as one would do for a true placebo.

3.3 Unpacking Causal Mechanisms vis Individual Treatment Arms (Domain D)

Table 5 turns to the series of treatment arms administered to individual voters to explore the relative effects of different types of information conveyed by the debates. Each hypothesis is summarized by its mean effects index, where columns 1 and 2 present the treatment effect estimates and associated per comparison, one-sided p-values for the debate arm, columns 3 and 4 for the "get to know you" video arm, and columns 5 and 6 for the radio report, all with respect to the full set of controls. We show results for political knowledge at both the hypothesis and sub-hypothesis level. Columns 7 through 12 estimate the relative size

of treatment effects across the 3 arms, using two sided tests for each two-way combination. In these columns we present both per comparison p-values and FDR q values that adjust across all 24 tests run (three treatment arm combinations by five hypotheses and three sub-hypotheses).

The first row of columns 1 through 6 test for treatment effects on hypothesis A1 concerning political knowledge and suggest that all three arms were effective in transmitting political information. The treatment effect on the mean effect index is positive and significant at 99% confidence for all three arms. Making comparisons across arms, the coefficient for debates and the radio report are more than twice as large in magnitude as that for the get to know you video, differences that are both statistically significant under per comparison p and FDR adjusted q values in Columns 7 and 11. Specifically, the treatment effect for debates is 0.109 standard deviation units (s.e. 0.021) compared to 0.041 (s.e. 0.016) for get to know you, yielding a difference of 0.068 standard deviation units (s.e. 0.022) which is significant at greater than 98% confidence under both a single two-sided test and the FDR adjustment for the fact that it is one of 24 comparative tests run. While the coefficient on debates is slightly larger than that for the radio report, the difference is not statistically distinguishable from zero.

The next three rows separate the knowledge questions into the three sub-hypothesis indices specified in the PAP. The pattern of treatment effects for general political knowledge mirrors that of the hypothesis overall, where all three arms yielded strong positive treatment effects, and the debates and the radio report estimates are larger in magnitude than those for get to know you video. Estimates in row three cover the 7 outcomes concerning knowledge of candidate characteristics, where again, all three estimates are positive and significant, although now the magnitudes are comparable across arms. 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 5 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 and physical carriage or confidence.

Moving to policy, row 4 shows that both debates and radio reports enhanced voter knowledge of the candidates' policy positions. These estimates are much larger in magnitude than the estimate for the get to know you video, which is near zero, and the differences are significant at 99% confidence. The fact that voters learned nothing about policy from the get to know you video, which focused solely on candidate persona, is reassuring for the soundness of the basic research design. Consistent with these estimates, row 5 concerning hypothesis A2 suggests that debates moved voters into better policy alignment with the candidates

they selected, while the get to know you video had no impact, a difference that is highly significant (0.074, s.e. 0.033). By contrast to what we see for debates, knowledge of policy positions acquired through the radio report did not translate into better policy alignment. 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 arm. 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 had an impact on 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 video all affected political knowledge, it is only debates that moved voters to change their voting choices and update their policy views. This pattern of results is consistent with the idea that debates are additive in both charisma and policy/professional information, and that the combination is more powerful than either in isolation.

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

Analysis in this research domain remains ongoing.

3.5 Secondary Hypotheses

There were a handful of outcomes that 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 1 presents estimates for these outcomes grouped under the three secondary hypotheses of voter turnout, perceived legitimacy of the electoral process, and interest in politics more broadly construed. We find little evidence to support treatment effects for any of these hypotheses.

Specifically, the results for turnout are mixed across voter samples. For the group screening intervention, estimates in the first three rows of Appendix Table 1 reflect negative yet insignificant treatment effects in our exit polls and in the National Electoral Commission (NEC)'s official polling center-level returns. Note that baseline turnout was very high in the control areas, measured at 98.4 percent in our exit poll sample, which is drawn from households in the immediate vicinity of the polling center itself; and 79.1 percent in the NEC returns, which cover voters from the entire catchment area of the polling center. The next three rows concern the individual treatment arms, where we find positive and significant

effects for the debate and "get to know you" treatments, and no effect for the radio report. For direct comparability with the estimates in Table 5, recall that these effects are expressed in standard deviation units. To get a better sense of magnitude, note that the treatment effect estimate for the debates treatment is 1.4 percentage points (s.e. 0.69) on a base of 96.1 percent for controls. Since these results do not replicate in the larger polling center level sample, we do not place much weight on them, and conclude that debates exposure if anything had small positive impacts on turnout.

There is no evidence that the debates increased voter confidence that the elections were free and fair, although again baseline confidence was extremely high (91.9 percent for controls). In the bottom of the table, we find some suggestive evidence that exposure to debates spurred voter interest in politics more generally. The mean effects index across the three outcomes of interest is positive and marginally significant. This is driven primarily by a positive effect on voter ability to name the two Presidential candidates, and on voters reporting that they were more likely to discuss politics (although the latter is not statistically significant).

4   Additional Analysis

4.1   Survey Priming

Table 6 turns to the question of how much of these effects 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 is important in light of findings that the act of surveying has nontrivial impacts on behavior (see Zwane et al 2011). The design of our experiment tackles this issue in two ways. First, for the individual-level treatments, we can compare surveyed controls to "pure" controls to capture a survey priming effect. Surveyed controls were given the same detailed survey that accompanied the delivery of the debates and other treatment arms, and then surveyed again in the exit polls. The earlier survey may have primed respondents to seek out information on the outcome variables of interest or increased their salience in the weeks leading up to the Election. By contrast, at the time of treatment implementation the "pure" controls were asked only basic demographic questions, and were not asked about the political outcomes of interest until the exit polls. They thus experienced no prime that would have enhanced their interest in or attention to political information in the weeks leading up to the Election. Analysis in Panel A of Table 6 estimates treatment effects for the surveyed versus pure controls subsample (to capture the survey priming effect), and then separately for the debates treatment versus surveyed controls subsample (to capture the additional

treatment effect above and beyond the survey priming effect).

Our second approach uses the group screening sample to capture a survey reinforcing effect by tracking those assigned to treatment with survey versus "pure" treatment across treated and control polling centers. This is thus the converse of the above approach, where we now aim to measure whether being surveyed at the time of treatment facilitates greater comprehension or absorption of the political information conveyed by the debates. Respondents in the treatment plus survey group were given an incentive to attend and also surveyed at the debate screening. Members of the "pure" treatment group were given the attendance incentive but not surveyed at the time of screening. All respondents in the control polling centers are "pure" controls. Analysis in Panel B of Table 6 uses the pooled sample to estimate the coefficient on assignment to a treatment polling center (the pure treatment effect) as well as the interaction term between treatment polling center and assignment to the surveyed group (the additional survey reinforcing effect).

Starting from the right hand side of Table 6 Panel A, estimates in the first row of column 5 suggest that the experience of being surveyed had a small positive effect on political knowledge in the absence of any individual-level treatment. This overall effect is driven by impacts on general political knowledge, where the second row of column 5 suggests that being surveyed accounts for a 0.099 standard deviation unit (s.e. 0.035) increase in general political knowledge compared to pure controls, significant at 99% confidence. In the left hand side of the same row, column 1 compares those in the debate arm to surveyed controls to reveal 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 two estimates suggest that the survey priming effect accounts for one third of the total treatment effect on general political knowledge. Treatment effects on the other two indices of political knowledge (candidate characteristics and policy stances) appear to be driven wholly by the debate content, where treatment effect estimates in column 1 are positive and significant while those in column 5 are not distinguishable from zero. There is weak evidence for a survey priming effect on votes for the debate winner in the sixth row of column 5, however it is noisily estimated and does not hold up in the larger group screening sample of Panel B.

Panel B takes the second approach, where column 1 estimates a "pure" treatment effect from watching the debate without being surveyed. Similar to the above, in the second row there is a 0.238 standard deviation units (s.e. 0.055) treatment effect of watching the debates compared to respondents in control polling centers. In column 5, there is evidence for an additional 0.103 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 political knowledge. For the other two sub-indices of political knowledge, and for policy alignment and voting for the best performer in the debate, the "pure" treatment effects remain positive and highly significant, and there is no evidence for an additional survey reinforcing effect. There are two marginally significant reinforcing effects on crossing party lines and openness that lack a strong counterpart in the "pure" treatment effect, which would suggest that impacts only arise when debate exposure is followed up with a survey experience. However, these are noisily estimated and do not replicate in the individual-level sample of Panel A.

Overall, these two sets of results suggest that the experience of being surveyed about politics accounts for a third of the total treatment effect on general political knowledge. There is no evidence, however, that survey priming or reinforcing explain much of the other treatment effects described in Section 3. It seems sensible that being surveyed can affect respondent acquisition of straightforward and readily available information, like the eligibility criteria for free healthcare, but not more nuanced information like candidate views on specific policies or which candidate has more education. Given these null results for priming effects on knowledge of candidate attributes and policy stances, it makes sense that we find no evidence for subsequent priming impacts on policy alignment or votes for the candidate who performed best during the debate.

4.2   Treatment Effect Heterogeneity

Overall, we find little evidence for systematic heterogeneity in treatment effects. Appendix Table 2 estimates heterogenous effects by respondent sub-groups of gender, age and lack of fluency in Krio (the *lingua franca* of Sierra Leone and language of the debates). These 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 (-0.076 standard deviation units, s.e. 0.021) is significant at conventional levels (under two sided tests). In terms of magnitude, this suggests that women acquired only 75% as much political knowledge from the debates when compared to men.

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 best candidates is robust to excluding each constituency one by one. Some additional heterogeneity analyses remain to be compiled.

4.3 Social Mobilization: Individual versus Group Exposure

This section considers how the delivery of the debate content—via group screening versus individual private viewing— affects the impact it has on voter behavior. Since many aspects of the experience differ across the two delivery modes, we will not be able to pin down exact mechanisms, but can speculate as to how salient differences might drive divergence in treatment effect intensity. First note that the content of the debate films was exactly the same under the two conditions. The most pronounced difference in delivery is that the screenings involved large public gatherings of a couple hundred people, while the individual treatment arm involved respondents watching the debate alone on a tablet device. Thus in our PAP we specify that the comparison of the two delivery modes is a test of social mobilization effects. Other aspects, however, differ across the two modes: individual treatments were administered in larger polling centers (as measured by total registered voters); and the group screenings played music before the debates, played the debates twice, had simultaneous translation into the relevant local language, and occurred later in the day.

Table 7 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 a comparison of the debates treatment arm to the pure control group (dropping respondents from the surveyed control and lab experiment control arms). First note that the qualitative pattern of effects for the two delivery modes on these comparable subsamples is the same: strong positive treatment effects on political knowledge, including effects for all three sub-indices, on policy alignment, and on votes for the best performer; and no evidence of treatment 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 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 level screening to estimate treatment on the treated effects (Column 3), which are more directly comparable to the individual level treatments where compliance with treatment assignment was nearly perfect. Notice that the difference in magnitude is most pronounced for knowledge of candidate characteristics, knowledge of candidate policy positions, and moving into policy alignment. These differences are consistent with the idea that watching the films in a group setting facilitated greater discussion amongst voters that clarified and reinforced the information about candidates and policy conveyed by the debates. The fact that the impact on general political knowledge is comparable across the two modes suggests that basic differences in comprehension (attributable to the lack of local language translation in the individual-level arms) cannot fully explain the divergence in magnitude of effect.

25

5   Conclusions to Date

Taken together, these results suggest that voters acquired significant political knowledge from watching the debates, knowledge that persisted over a number of weeks, and importantly, influenced their voting choices on Election Day. By equipping voters with knowledge that changed their voting behavior, debates further attracted greater campaign investment by participating candidates. 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.

From a policy perspective, this project demonstrates that interparty debates between candidates are logistically feasible to host and disseminate, and can create large, persistent gains in voters' political knowledge. Commissioners of the National Electoral Commission of Sierra Leone, who have been consults on this project from its inception, responded positively to our early dissemination events and expressed interest in taking debates to scale in the next elections. In considering the costs and benefits, the mobile cinema was a relatively resource intensive way to publicize the debates. Fixed video production costs for the debates themselves were modest in this setting: roughly five thousand dollars per constituency to assemble the candidates, host, film and edit the debates. In settings where mass media penetration is higher, televising the debates would be an obvious alternative dissemination mechanism with lower marginal costs than our road show. Where television broadcast or ownership is limited, radio becomes the obvious choice. While the individual treatment arms suggest that video is more effective than audio alone, the radio report we tested was rather dry. One could imagine a livelier radio program that captures a real time debate between candidates in the recording studio that might come closer to the impacts of the film screening. This kind of radio counterpart could reach large voting audiences at negligible marginal cost.

Discussion so far has focused on the question of whether debates can facilitate more informed voting, influence electoral support for better performing candidates, and affect the allocation of campaign spending. The final outstanding question is the extent to which debates can facilitate greater accountability over elected officials. Data collection and analysis for this research area remain ongoing as we track the winning MPs from treated and control constituencies over their first 18 months in office.

References

Abramowitz, Alan I. "Impact of a Presidential Debate on Voter Rationality," American Journal of Political Science, 22:3(1978): 680-690.

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.

Benjamin, Daniel J. and Jesse M. Shapiro. "Thin-slice Forecasts of Gubernatorial Elections." Review of Economics and Statistics, 91:3(2009): 523–536.

Casey, Katherine. "Crossing Party Lines: The Effects of Information on Redistributive Politics." Stanford University Working Paper No. 2099R (2014).

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.

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):532-550.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot and Philippe Zamora. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." The Quarterly Journal of Economics (2013), 531–580

Dixit, Avinash, and John Londregan, "The Determinants of Success of Special Interests in Redistributive Politics," The Journal of Politics, 58(1996), 1132-1155.

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

Downs, Anthony (1957). An Economic Theory of Democracy. New York: Harper.

Druckman, James N. "The Power of Television Images: The First Kennedy-Nixon Debate Revisited." Journal of Politics, 65:2(2003): 559–571.

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): 703-745.

Fujiwara and Wantchekon. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." American Economic Journal: Applied Economics, forthcoming (2013).

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): 135-150.

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

Jamieson, Kathleen Hall and David S. Birdsell. Presidential Debates: The Challenge of Creating an Informed Electorate. Oxford University Press, 1990.

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

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

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 Shapiro, Ian, Susan C. Stokes, Elisabeth Jean Wood, and Alexander S. Kirshner, eds. Political representation. Cambridge University Press, 2010.

Prat, Andrea. "The Wrong Kind of Transparency." American Economic Review, 95:3(2005): 862-877.

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

Todorov, Alexander, Anesu N. Mandisodza, Amir Goren, and Crystal C. Hall. "Inferences of competence from faces predict election outcomes." Science, 308:5728(2005): 1623-1626.

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

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer et al. "Being surveyed can change later behavior and related parameter estimates." Proceedings of the National Academy of Sciences 108:5(2011): 1821-1826.

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

Mean Effects Indices by Hypothesis	Treatment effect (standard error) (1)	Per comparison p-value (one sided) (2)	FWER adjusted p-value (one sided) (3)
A1. Exposure to debates increases political knowledge (20 outcomes)	0.278 (0.027)	0.000**	0.000**
A2. Exposure to debates increases policy alignment (3 outcomes)	0.106 (0.035)	0.002**	0.009**
A3. Exposure to debates increases vote shares for the candidate that performed the best in the debates (2 outcomes)	0.086 (0.043)	0.023*	0.079+
A4. Exposure to debates increases the willingness to vote across party lines (3 outcomes)	-0.018 (0.032)	0.718	0.710
A5. Exposure to debates enhances voter openness to other parties (5 outcomes)	0.091 (0.048)	0.028*	0.079+

Notes: i) significance levels indicated by $+ p < 0.10$, $* p < 0.05$, $** p < 0.01$; ii) robust standard errors clustered by polling center; iii) p-values are based on one-sided tests in the direction of the hypothesis statement as pre-specified in our pre-analysis plan (PAP); 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 control variables (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership); vi) treatment effects are on the hypothesis-level mean effects index constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units; vii) adjustments to control familywise error rate (FWER) computed following Westfall and Young (1993) and Anderson (2008); and viii) data source is the exit poll survey.

Table 2: Domain A Treatment Effects on All Individual Outcomes

Individual outcomes by hypothesis	Control mean (1)	Treatment effect (2)	Standard error (3)	Naïve p-value (4)	FDR q-value (5)	N (6)
Panel A1. Political knowledge increases						
i. Knows amount of the constituency facilitation fund	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*	0.040*	5,399
i. Knows the gender equity bill (GEB) is 30%	0.352	0.012	0.030	0.346	0.237	5,398
i. Knows MP job responsibilities (out of 3)	0.555	0.217	0.070	0.001**	0.003**	5,400
ii. Candidate choice driven by individual characteristic	0.209	-0.010	0.028	0.638	0.350	5,229
ii. Knows which candidates had been an MP before	0.490	0.111	0.032	0.000**	0.002**	5,400
ii. Knows which candidate was most educated	0.243	0.159	0.044	0.000**	0.001**	3,097
ii. Knows candidate with most public office experience	0.336	0.049	0.038	0.101	0.087+	2,576
ii. Knows APC candidate's name	0.442	0.181	0.034	0.000**	0.001**	5,058
ii. Knows PMDC candidate's name	0.115	0.105	0.031	0.000**	0.002**	3,291
ii. Knows SLPP candidate's name	0.395	0.168	0.031	0.000**	0.001**	5,400
iii. Knows APC candidate's first priority issue	0.190	0.088	0.030	0.002**	0.005**	5,057
iii. Knows PMDC candidate's first priority issue	0.099	0.064	0.026	0.008**	0.012*	3,288
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 view of FHC	0.119	0.007	0.037	0.421	0.259	2,812
iii. Knows SLPP candidate's view of FHC	0.123	0.072	0.029	0.007**	0.011*	4,921
iii. Knows APC candidate's position on GEB	0.285	0.096	0.035	0.003**	0.007**	5,058
iii. Knows PMDC candidate's position on GEB	0.244	0.209	0.053	0.000**	0.001**	3,291
iii. Knows SLPP candidate's position on GEB	0.331	0.155	0.038	0.000**	0.001**	5,400
Panel A2. Policy alignment increases						
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
Voter's priority issue matches that of chosen candidate	0.425	0.090	0.031	0.002**	0.005**	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
Panel A4. Votes across ethnic-party lines increase						
Voted across ethnic-party lines	0.107	-0.012	0.013	0.811	0.432	4,569
Voted for a different party for MP than did in 2007	0.163	0.004	0.019	0.425	0.259	4,405
Split ticket across parties for MP vs President	0.058	-0.009	0.010	0.833	0.432	5,212
Panel A5. Voter openness to candidates increases						
Voter likeability rank for APC candidate (10 point scale)	6.523	0.161	0.261	0.269	0.212	5,073
Voter likeability rank for own party's candidate	7.971	0.245	0.237	0.151	0.120	5,160
Voter likeability rank for PMDC candidate	2.369	0.579	0.273	0.018*	0.020*	3,299
Voter likeability rank for rival party's candidate	3.395	0.117	0.226	0.303	0.218	4,906
Voter likeability rank for SLPP candidate	4.952	0.229	0.230	0.160	0.123	5,414
Panel A3 Alternate: NEC Official Returns						
Vote share of debate winner, as judged by audience	0.711	0.035	0.017	0.033*		206
Vote share of debate winner, as judged by expert panel	0.617	0.035	0.016	0.027*		206

Notes: i) significance levels + p <0.10, * p <0.05, ** p <0.01 based on one-sided tests in direction pre-specified in PAP; ii) robust standard errors clustered by polling center; iv) 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) additional controls vary by hypothesis from the set (gender, age, years of schooling, polygamous marital status, farming occupation and radio ownership); vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 48 primary and secondary individual outcomes in domain A; viii) data source is the exit poll survey in panels A1-A5 and the National Electoral Commission polling center-level returns in A3 Alt: and ix) NEC returns exclude constituency 15 where the SLPP candidate was disqualified immediately before the election.

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

Outcome	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 (all 9 outcomes)	*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

Notes: i) significance levels + p <0.10, * p <0.05, ** p <0.01; 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 determined by analysis of control group data and include gender, age, years of schooling, and radio ownership; v) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); and vii) data source is the exit poll survey.

Table 4: Domain B - Treatment Effects of Polling Center Screenings on Party Officials

Outcome	Control mean	Treatment effect	Standard error	Naïve p-value (2 sided)	FDR q-value	N
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Effects Index	*0.000*	*0.082*	*0.052*	*0.113*		*224*
Any visits by party officials, APC	0.819	0.004	0.060	0.941	1	210
Number of visits by party officials, APC	1.857	0.427	0.383	0.266	1	210
Any political rallies, APC	0.248	0.045	0.055	0.412	1	207
Number of political rallies, APC	0.467	0.046	0.147	0.753	1	209
Any party officials distributed gifts? APC	0.481	0.080	0.063	0.206	1	205
Number of community members receiving gifts, APC	0.295	0.069	0.044	0.114	1	208
Number of posters displayed in community, APC	0.699	-0.022	0.042	0.601	1	207
Any visits by party officials, PMDC	0.368	0.008	0.069	0.912	1	133
Number of visits by party officials, PMDC	0.515	0.386	0.342	0.260	1	134
Any political rallies, PMDC	0.044	-0.028	0.022	0.198	1	134
Number of political rallies, PMDC	0.044	0.004	0.043	0.934	1	134
Any party officials distributed gifts? PMDC	0.045	-0.003	0.032	0.928	1	133
Number of community members receiving gifts, PMDC	0.025	-0.002	0.017	0.924	1	133
Number of posters displayed in community, PMDC	0.235	0.048	0.052	0.353	1	132
Any visits by party officials, SLPP	0.739	0.051	0.062	0.409	1	222
Number of visits by party officials, SLPP	1.679	0.066	0.230	0.773	1	224
Any political rallies, SLPP	0.159	0.011	0.044	0.809	1	213
Number of political rallies, SLPP	0.315	-0.079	0.104	0.448	1	219
Any party officials distributed gifts? SLPP	0.368	0.078	0.072	0.278	1	213
Number of community members receiving gifts, SLPP	0.226	0.038	0.040	0.339	1	214
Number of posters displayed in community, SLPP	0.555	0.047	0.035	0.187	1	221

Notes: i) significance levels + p <0.10, * p <0.05, ** p <0.01; 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) and constituency fixed effects; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; v) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008); and vi) data source is the community level exit poll survey.

Table 5: Domain D - Causal Mechanisms Explored through Relative Treatment Effects Across Individual Treatment Arms

Hypothesis 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 value 1 sided	Treatment effect (std error)	Naïve p value 1 sided	Treatment effect (std error)	Naïve p value 1 sided	Treatment effect (std error)	2 sided Naïve p FDR q	Treatment effect (std error)	2 sided Naïve p FDR q	Treatment effect (std error)	2 sided Naïve p FDR q
	(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 0.012	0.014 (0.018)	0.425 0.521	0.053* (0.022)	0.016 0.077
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 0.197	0.014 (0.034)	0.674 0.736	0.065 (0.050)	0.192 0.370
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 0.521	0.007 (0.026)	0.793 0.819	-0.026 (0.032)	0.411 0.521
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 0.521	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 0.101	0.121** (0.032)	0.000 0.002	-0.047+ (0.027)	0.083 0.199
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 0.386	0.104* (0.052)	0.046 0.159	-0.051 (0.040)	0.203 0.370
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 0.521	-0.088+ (0.050)	0.076 0.199	0.055 (0.042)	0.195 0.370
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 0.521	-0.008 (0.033)	0.818 0.819	0.036 (0.029)	0.215 0.370
Number of observations	1,698		1,695		1,695							

Notes: i) significance levels, corresponding to one sided per comparison p values, + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors clustered by polling center; iii) 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; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and is expressed in standard deviation units; v) Columns 1 to 6 run one-sided tests in the (prespecified) direction of the hypothesis statement, while Columns 7 to 12 run two-sided tests; vi) adjustments to control false discovery rate (FDR) computed following Benjamini, Krieger and Yekutieli (2006) and Anderson (2008) across all 24 tests run; and vii) data source is the individual treatment sample exit poll survey.

Table 6: Survey Priming and Reinforcing Effects

Panel A: Survey priming effects across individual treatment arms

Mean Effects Index by Hypothesis	Treatment effect beyond survey: Debates vs Surveyed controls				Survey priming effect: Surveyed controls vs. Pure controls			
	Treatment effect	Standard error	Naïve p-value	N	Treatment effect	Standard error	Naïve p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A1. Political knowledge	0.118**	0.023	0.000	875	0.026+	0.015	0.051	875
i. General Knowledge	0.211**	0.042	0.000	875	0.099**	0.035	0.004	875
ii. Candidate Characteristics	0.057**	0.023	0.008	875	-0.021	0.019	0.858	875
iii. Policy Stances	0.125**	0.033	0.000	875	0.030	0.024	0.114	875
A2. Policy Alignment	0.079**	0.032	0.010	933	0.012	0.038	0.372	935
A3. Vote for best	0.025	0.053	0.317	933	0.079+	0.053	0.070	935
A4. Cross party lines	-0.010	0.039	0.605	933	-0.045	0.046	0.834	935
A5. Openness	-0.003	0.025	0.546	933	0.006	0.028	0.411	935

Panel B: Survey reinforcing effects across group screening arms

Mean Effects Index by Hypothesis	"Pure" treatment effect: Debates without survey vs. controls				Survey reinforcing effect: Interaction term for debates with survey			
	Treatment effect	Standard error	Naïve p-value	N	Treatment effect	Standard error	Naïve p-value	N
	(1)	(2)	(3)	(8)	(5)	(6)	(7)	(8)
A1. Political knowledge	0.263**	0.031	0.000	5,415	0.028+	0.020	0.079	5,415
i. General Knowledge	0.238**	0.055	0.000	5,415	0.103**	0.037	0.003	5,415
ii. Candidate Characteristics	0.247**	0.036	0.000	5,415	0.010	0.025	0.340	5,415
iii. Policy Stances	0.286**	0.039	0.000	5,415	0.009	0.022	0.344	5,415
A2. Policy Alignment	0.105**	0.041	0.005	5,415	-0.001	0.030	0.510	5,415
A3. Vote for best	0.088*	0.047	0.033	5,415	-0.015	0.036	0.664	5,415
A4. Cross party lines	-0.054	0.037	0.925	5,415	0.056+	0.034	0.052	5,415
A5. Openness	0.064	0.050	0.103	5,415	0.042+	0.026	0.052	5,415

Notes: i) significance levels, corresponding to one sided per comparison p values, + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors clustered by polling center; iii) specifications include stratification bins for the relevant randomization procedure and constituency fixed effects; iv) mean effects index constructed following Kling, Liebman and Katz 2007 and expressed in standard deviation units; and v) mean effects indices are standardized with respect to the pure control group for all of panel A.

Table 7: Social Mobilization Effect - Group vs Individual Delivery of Debates

Mean Effects Index by Hypothesis	Group Screening			Individual Viewing	
	Treatment effect (ITT) (std error)	TOT scaling	N	Treatment effect (≈TOT) (std error)	N
	(1)	(2)	(3)	(4)	(5)
A1. Political knowledge	0.324** (0.034)	0.416	3,507	0.140** (0.023)	708
i. General political knowledge	0.371** (0.034)	0.475	3,507	0.304** (0.053)	708
ii. Candidate characteristics	0.268** (0.039)	0.344	3,507	0.031+ (0.023)	708
iii. Policy stances	0.348** (0.049)	0.446	3,507	0.152** (0.037)	708
A2. Policy alignment	0.129** (0.047)	0.165	3,514	0.087* (0.039)	748
A3. Votes for best performer in the debate	0.102+ (0.068)	0.130	3,510	0.102* (0.051)	748
A4. Cross party lines	-0.022 (0.042)	-0.028	3,511	-0.052 (0.050)	748
A5. Openness	0.059 (0.065)	0.075	3,514	-0.002 (0.031)	748

Notes: i) significance levels, corresponding to one sided per comparison p values, + p <0.10, * p <0.05, ** p <0.01; 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 marital status, farming occupation and radio ownership); v) mean effects index constructed following Kling, Liebman and Katz 2007 and expressed in standard deviation units; vi) the group screening sample is limited to the 8 constituencies where the individual-level treatments were also administered; vii) the individual level estimates are limited to the debates treatment arm versus the "pure" control group.

Appendix Table 1: Treatment Effect Estimates for Secondary Hypotheses

Outcomes and datasets by hypothesis	Control mean (1)	Treatment effect (2)	Standard error (3)	Naïve p-value (4)	FDR q-value (5)	N (6)
A6.i. Exposure to debates mobilizes the public and leads to greater turnout (1 outcome)						
Mean effects index, group screening exit polls	0.000	-0.210	0.117	0.963		5,247
Individual turnout verified by voter ID card, group screening exit polls	0.984	-0.026	0.015	0.961	0.490	5,346
Polling center-level turnout, National Electoral Commission official returns	0.791	-0.004	0.008	0.675		224
Mean effects index, individual debate treatment	0.000	0.072	0.036	0.026*		1,595
Mean effects index, individual "get to know you" treatment	0.000	0.110	0.061	0.040*		1,590
Mean effects index, individual "radio report" treatment	0.000	0.024	0.044	0.291		1,592
A6.ii. Exposure to debates increases the perceived legitimacy of elections (1 outcome)						
Mean effects index, group screening exit polls	0.000	-0.092	0.083	0.866		5,247
Individual confidence that election was free and fair, group screening exit polls	0.919	-0.025	0.023	0.858	0.432	5,415
Mean effects index, individual debate treatment	0.000	-0.008	0.050	0.563		1,698
Mean effects index, individual "get to know you" treatment	0.000	0.020	0.042	0.320		1,695
Mean effects index, individual "radio report" treatment	0.000	-0.083	0.055	0.930		1,695
A6.iii. Exposure to debates increases interest in politics (3 outcomes)						
Mean effects index, group screening exit polls	0.000	0.054	0.036	0.071+		5,247
Individual frequency of discussing politics, group screening exit polls	0.243	0.031	0.027	0.124	0.102	5,411
Individual names MP candidate not in the debate, group screening exit polls	0.135	-0.019	0.032	0.718	0.395	1,645
Individual knows Presidential candidates' names (of 2), group screening exit polls	1.360	0.097	0.041	0.009**	0.013*	5,413
Mean effects index, individual debate treatment	0.000	-0.030	0.028	0.853		1,595
Mean effects index, individual "get to know you" treatment	0.000	0.015	0.033	0.332		1,590
Mean effects index, individual "radio report" treatment	0.000	0.037	0.031	0.122		1,592

Notes: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) all mean effects indices are expressed in standard deviation units; iii) estimates for mean effects indices in group screening exit polls specified as in Table 1; iv) estimates for individual outcomes in group screening exit polls specified as in Table 2, where false discovery rate (FDR) adjustments apply across all 48 primary and secondary outcomes; and v) estimates for mean effects indices on individual treatments specified as in Table 5.

Appendix Table 2: Treatment Effect Heterogenity, Subgroup Analysis

Dependent variable = Mean effect index by hypothesis	A1. Political Knowledge	A2. Policy Alignment	A3. Votes for Debate Winner	A4. Votes Across Party Lines	A5. Voter Openness
	(1)	(2)	(3)	(4)	(5)
Female * Treatment	-0.076**	0.037	-0.034	0.036	-0.003
	(0.021)	(0.031)	(0.039)	(0.040)	(0.030)
Youth * Treatment	0.021	-0.033	0.070	-0.069+	0.011
	(0.019)	(0.031)	(0.045)	(0.041)	(0.032)
No Krio * Treatment	-0.023	0.038	0.107+	-0.100+	-0.045
	(0.042)	(0.061)	(0.061)	(0.053)	(0.058)
Treatment	0.311**	0.095*	0.050	0.015	0.096+
	(0.030)	(0.042)	(0.050)	(0.043)	(0.050)
Female	-0.062**	0.024	0.060+	-0.078*	-0.027
	(0.016)	(0.027)	(0.035)	(0.037)	(0.029)
Youth	-0.008	0.011	-0.027	0.030	0.002
	(0.014)	(0.022)	(0.033)	(0.034)	(0.023)
No Krio	-0.099**	-0.022	0.064+	-0.040	-0.004
	(0.025)	(0.037)	(0.036)	(0.040)	(0.037)
Observations	5,246				

Notes: i) significance levels indicated by + p <0.10, * p <0.05, ** p <0.01; ii) robust standard errors clustered by polling center; iii) all specifications include stratification bins for the polling center (number of registered voters and distance to next nearest) and constituency fixed effects; iv) specifications further include additional control variables years of schooling, polygamous marital status, farming occupation and radio ownership; v) treatment effects are on the hypothesis-level mean effects index constructed following Kling, Liebman and Katz (2007) and expressed in standard deviation units; and vi) data source is the exit poll survey from the group screening sample.